

72974_Auto_EditedC.docx

Name of Journal: *World Journal of Gastroenterology*

Manuscript NO: 72974

Manuscript Type: ORIGINAL ARTICLE

Basic Study

Gut mucosal microbiota profiles linked to colorectal cancer recurrence

Huo RX *et al.* Colorectal cancer and gut mucosal microbiota

Rui-Xue Huo, Yi-Jia Wang, Shao-Bin Hou, Wei Wang, Chun-Ze Zhang, Xue-Hua Wan

Abstract

BACKGROUND

Emerging evidences link gut microbiota to various human diseases including colorectal cancer (CRC) initiation and development. However, gut microbiota profiles associated with CRC recurrence and patient prognosis are not completely understood yet, especially in a Chinese cohort.

AIM

To investigate the relationship between gut mucosal microbiota profiles and CRC recurrence and patient prognosis.

METHODS

We obtained the composition and structure of gut microbiota collected from 75 patients diagnosed with CRC and 26 healthy controls. The patients were followed up by regular examination to determine whether tumors recurred. ¹ Triplet-paired samples from on-tumor, adjacent-tumor and off-tumor sites of patients diagnosed with/without CRC recurrence were analyzed to assess spatial-specific patterns of gut mucosal microbiota by 16S ribosomal RNA sequencing. Next, we carried out bioinformatic analyses,

Kaplan-Meier survival analyses and Cox regression analyses to determine the relationship between gut mucosal microbiota profiles and CRC recurrence and patient prognosis.

RESULTS

We observed spatial-specific patterns of gut mucosal microbiota profiles linked to CRC recurrence and patient prognosis. A total of 17 bacterial genera/families were identified as potential biomarkers for CRC recurrence and patient prognosis, including *Anaerotruncus*, *Bacteroidales*, *Coriobacteriaceae*, *Dialister*, *Eubacterium*, *Fusobacterium*, *Filifactor*, *Gemella*, *Haemophilus*, *Mogibacteriaceae*, *Pyramidobacter*, *Parvimonas*, *Porphyromonadaceae*, *Slackia*, *Schwartzia*, TG5 and *Treponema*.

CONCLUSION

In this work, we have identified critical bacterial taxa that are associated with prognosis and survival of CRC patients. Our work suggests that intestinal microbiota can serve as biomarkers to predict the risk of CRC recurrence and patient death. Further functional analyses of these bacterial physiological roles in CRC recurrence will pave a way for the development of novel strategies for CRC treatment.

Key Words: Gut microbiota; Colorectal cancer; Prognosis

Huo RX, Wang YJ, Hou SB, Wang W, Zhang CZ, Wan XH. Gut mucosal microbiota profiles linked to colorectal cancer recurrence. *World J Gastroenterol* 2022; In press

Core Tip: Emerging evidences indicate that besides genetic and epigenetic factors, gut microbiota is capable of driving colorectal cancer (CRC) progression. Here, we analyzed the gut mucosal microbiota of 75 triplet-paired samples collected from on-tumor, adjacent-tumor and off-tumor sites of patients diagnosed with/without CRC recurrence and 26 healthy controls. After a long-term follow-up, we identified spatial-specific

bacterial taxa whose abundances are associated with overall survival and disease-free survival. Our data reveal the profiles of gut mucosal microbiota that risk CRC recurrence and affect patient prognosis, which may serve as potential new biomarkers for CRC diagnosis.

INTRODUCTION

19

Colorectal cancer (CRC) is a major cause of cancer-related deaths with the third highest mortality and the fourth highest incidence worldwide, according to GLOBOCAN 2020 (global cancer statistics)^[1], which brings a public health issue. Although many new treatment options of CRC have doubled overall survival times for advanced-stage patients to 3 years, survival prognosis for early-stage patients with non-metastasised disease is still better than that for advanced-stage patients^[2]. Currently, treatment of CRC mainly depends upon tumor-node-metastasis (TNM) stage of disease, and patients' physical fitness and intent. However, previous studies showed that disease recurrence was observed in 40% of treated patients with stage I-II CRC (lymph node negative postoperative) and 70% of those with stage III CRC^[3], which is proposed to be related to rat sarcoma viral oncogene and microsatellite statuses^[4]. Thus, treatment strategy merely depending on CRC TNM stage may cause over-treatment or under-treatment, leading to CRC recurrence. Moreover, the molecular mechanisms behind CRC recurrence are not yet completely understood. Therefore, we still need to explore more suitable biomarkers for assessing prognosis of CRC patients, in order to achieve optimal personalized treatment.

In addition to hereditary and lifestyle factors, human gut microbiota is considered as an important risk factor for CRC initiation as well as development. ⁴ With the continuous development of high-throughput sequencing, the compositional structure of human gut microbiota is revealed to be closely involved in CRC initiation, development and treatment^[5]. Moreover, pathogenic bacteria have been identified from human gut microbiota and their pro-carcinogenic properties have been demonstrated to play a role in causing gut microbial dysbiosis in the complex environment of human intestinal tract. Certain bacterial species, including *Bacteroides fragilis*, *Clostridium septicum*, *Enterococcus faecalis*, *Escherichia coli*, *Fusobacterium* spp., *Helicobacter pylori* and *Streptococcus bovis* have been identified to play a role in driving colorectal carcinogenesis^[6]. The mechanisms behind these pathogenic bacteria involve bacterial-

derived genotoxicity and other virulence factors that regulate host defenses systems, metabolisms, oxidative stress and anti-oxidative defense modulation^[6].

Evidences have linked gut microbiota to prognosis of CRC. A pilot study has shown that high abundances of *Fusobacterium nucleatum* (Fn) and *Bacteroides fragilis* were independent indicators for poor patient's survival, while high abundance of *Faecalibacterium prausnitzii* predicted improved survival rate^[7]. Besides, most studies supported that increased relative abundances of *Bacteroides fragilis* and Fn were associated with short-term survival and late stage of CRC^[8]. Enterotoxigenic *Bacteroides fragilis* (ETBF) and *Fusobacterium* spp. levels were significantly higher in late stage (III/IV) of CRC than those in early stage (I/II) of CRC, and high abundance of *Fusobacterium* was reported to be associated with high microsatellite instability (MSI-H)^[9,10]. Compared to Fn-negative cases, multivariable hazard ratio (HR) for CRC-specific mortality in Fn-high cases was 1.58^[9]. After neoadjuvant chemoradiotherapy, Fn persistence was associated with high relapse rates in locally advanced CRC^[11]. CRC patients with low levels of Fn had significantly longer overall survival (OS) and disease-free survival (DFS) than patients with moderate and high levels of Fn that were obvious in late-stage CRC patients^[12-15]. However, extensive bacterial taxa associated with prognosis of CRC are unclear, especially in a Chinese cohort.

To investigate the profiles of gut mucosal microbiota associated with CRC recurrence and survival of CRC patients, we collected gut mucosal microbiota from CRC patients when they received radical resection or palliative surgery in Tianjin Union Medical Center, China. For each patient, triplet-paired CRC samples were collected from on-tumor (T), adjacent-tumor (P) and off-tumor (N) sites. Additional samples were collected from 26 healthy controls. We performed 16S ribosomal RNA (rRNA) gene sequencing and analyses on these gut mucosal microbiota. Next, we carried out Kaplan-Meier survival curve analyses for OS and DFS. Our data suggest that a number of bacterial genera/families including *Anaerotruncus*, *Bacteroidales*, *Coriobacteriaceae*, *Dialister*, *Eubacterium*, *Fusobacterium*, *Filifactor*, *Gemella*, *Haemophilus*, *Mogibacteriaceae*, *Pyramidobacter*, *Parvimonas*, *Porphyromonadaceae*, *Slackia*, *Schwartzia*, TG5 and *Treponema*,

are associated with survival of CRC patients and CRC recurrence. We further performed the univariate and multivariate Cox regression analyses. Our data reveal that high abundances of *Anaerotruncus*, *Bacteroidales*, *Coriobacteriaceae*, *Dialister*, *Eubacterium*, *Filifactor*, *Haemophilus*, *Mogibacteriaceae*, *Pyramidobacter*, *Slackia*, *Treponema* and TG5 are associated with worse OS or DFS.

² **MATERIALS AND METHODS**

Patients and sample collection procedures

This study analyzed 69 pathologically confirmed CRC cases, who received radical or palliative surgeries at Tianjin Union Medical Center, Tianjin, China, from December 2016 to September 2017. These patients were followed up by regular examination or telephone survey. DFS was defined as the time period from the date of surgery to the time of tumor recurrence, and OS was defined as the time period from the date of surgery to the time of death. Our study was conducted in accordance with the Declaration of Helsinki. Every patient provided written informed consent for the collection of samples and subsequent analysis when required, and the study was approved by the Ethics Committee of Tianjin Union Medical Center. The TNM staging was determined according to the American Joint Committee on Cancer staging handbook (8th edition). Whether every patient received chemotherapy or not depends on his/her TNM stage, physical state score and intention. Imaging examination was arranged for every 2-6 mo or when patients' conditions changed, to determine whether the disease recurrence occurred.

DNA library preparation and 16S rRNA sequencing

²² ZR Fungal/Bacterial DNA kit (Zymo Research, Irvine, CA, United States) was used to isolate bacterial DNA from intestinal microbiota samples according to the manufacturer's instructions. Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher, Sunnyvale, CA, United States) was used to quantify the DNA amounts. The hypervariable regions of 16S rRNA gene amplicon libraries targeting the V3-V4 regions

were prepared according to the Illumina manufactory manual. The amplification primers were used according to the Illumina manufactory manual. The forward and reverse primers were used as follows:
 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
 and
 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAAT
 CC. The products of polymerase chain reaction were examined using agarose gel electrophoresis (1% concentration of agarose gel; 100 V; 40 min electrophoresis time). The AMPure XP beads (Beckman Coulter, Fullerton, CA, United States) was used to purify the amplified DNA libraries and Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher, Sunnyvale, CA, United States) was used to quantify the DNA library amounts. All the DNA samples were stored at -20°C until used.

Quality control, operational taxonomic unit picking and diversity analyses of 16S rRNA amplicons

The 16S rRNA amplicon libraries were sequenced for 2 × 300 bp on Illumina MiSeq platform. The reads information was listed in Supplementary Table 1. FastQC program was used to perform quality control and filtering of raw paired-end reads (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, accessed on 5 June 2019). Next, PandaSeq v2.10 was used to assemble paired-end reads with default parameters^[16]. The QIIME pipeline v1.9.1 (http://qiime.org/home_static/dataFiles.html, accessed on June 5, 2019) and Greengenes database v13.8 were used for de novo operational taxonomic unit (OTU) picking and taxonomic assignment^[17]. Briefly, we generated the mapping file that contains all the information about the sequencing samples including sample ID, barcode sequences and sample types. The python command, pick_otus.py, was used to pick OTUs based on sequence similarity (threshold of 97%) of the assembled reads, which is commonly used to define bacterial species. The python command, pick_rep_set.py, was used to pick the representative sequences for each OTU.

USEARCH 6.1 was used for chimera detection and filtering. The python command, assign_taxonomy.py, was used to assign taxonomy to OTU representative sequences. The python command, make_rarefaction_plots.py and online MicrobiomeAnalyst (https://dev.microbiomeanalyst.ca/MicrobiomeAnalyst/home.xhtml, accessed on 28 August 2021) were used to calculate alpha diversities within the samples¹³. The 16S rRNA paired-end reads have been submitted to the Sequence Read Archive database at the National Center for Biotechnology Information website, under accession number PRJNA606879.

Bioinformatic analyses of gut mucosal microbiota for CRC patients

The beta diversity of gut mucosal microbiota among the samples collected from CRC patients with and without recurrence was calculated using principal component analysis (PCA) on MicrobiomeAnalyst website^{18,19}. The linear discriminant analysis (LDA) effect size (LEfSe) was analyzed using online MicrobiomeAnalyst^{18,19}.

¹ *Statistical analysis*

Statistical analyses were performed using statistical analysis in social science program (SPSS) Version 23. The statistical significances of multiple sample comparisons were calculated using one-way ANOVA with Kruskal-Wallis test. ^a $P < 0.05$; ^b $P < 0.01$; ^c $P < 0.001$; ^d $P < 0.0001$. The Kaplan-Meier survival curves for OS and DFS and the optimal cut-off value of RiskScore were calculated using R software package maxstat¹⁸ (maximally selected rank statistics with several P value approximations version: 0.7-25)^[20]. We set the minimum number of samples in the grouping to be greater than 25%, and the maximum number of samples in the grouping to be less than 75%, and finally obtained the optimal cut-off value of each bacterial taxa. Based on this cut-off threshold, the patients were divided into groups with high and low abundances of the bacterial taxa. Furthermore, the survfit function of R software was used to analyze the prognostic differences between the two groups, and the significance of prognostic differences between the two groups was evaluated by logrank test. According to the

cut-off value of each bacterial taxon, we divided the patients into groups with high and low abundances of the bacterial taxa and then conducted Cox regression analyses. The Cox regression analyses based on the Proportional hazards model were carried out using SPSS v. 23 program.

RESULTS

Clinicopathological features of CRC patients

We summarized the patients' clinicopathological data in Table 1. ⁴ The median age of 75 patients was 63.4 years and 60% of them were males. The pathological type of all patients was categorized as adenocarcinoma, and most of the tumor cells were poorly differentiated. Most patients were diagnosed as microsatellite stable (MSS) based on immunohistochemical evaluation of components of the mismatch repair machinery, including MLH1, MSH2, MSH6 and PMS2. All the patients did not receive chemotherapy or other antitumor therapy before surgery. After surgery, 44 of 75 patients received chemotherapy based on FOLFOX (CapeOX) or FOLFIRI regimen, 4 of them received radiotherapy, and 6 of them received targeted therapy. ² The median follow-up duration for all cases after surgery was 51.2 mo (2.90-57.03 mo). There were 6 patients failed to be followed up. Until August 23, 2021, a total of 17 patients diagnosed with I-III stages had relapsed and 20 patients had dead.

Altered alpha-diversity of gut mucosal microbiota at adjacent-tumor site for CRC recurrence

Although tumor tissues can be removed by surgery, the microbiota residing in the tumor surrounding tissues, *e.g.* those at adjacent-tumor site, as well as those in the remaining intestinal tissues where tumors are removed, may retain pathogenic bacteria that have the capabilities to drive CRC, leading to CRC recurrence. Therefore, we hypothesize that gut mucosal microbiota profiles at on-tumor or adjacent-tumor sites may be linked to CRC recurrence. To examine this, ¹ we assessed microbial alpha-diversities of biopsy samples collected from on-tumor (T), adjacent-tumor (P) and off-

tumor (N) sites of patients with and without CRC recurrence. The 16S rRNA gene hypervariable V3-V4 regions were sequenced and analyzed for five α -diversity indices including Chao1, Fisher, Observed OTU, Shannon and Simpson (Figure 1). Analyses of species variations based on these five metrics consistently indicated that species diversities at on-tumor (T) or off-tumor (N) sites of patients with and without CRC recurrence showed no significant differences (Student *t*-test, $P < 0.05$) (Figure 1). Analyses of species variations based on Chao1 (Student *t*-test, $P = 0.0092$), Fisher (Student *t*-test, $P = 0.0092$) and observed OTU (Student *t*-test, $P = 0.0092$), but not Shannon (Student *t*-test, $P = 0.1314$) and Simpson (Student *t*-test, $P = 0.2513$) showed significant differences at adjacent-tumor (P) sites of patients with and without CRC recurrence (Figure 1). Although Shannon and Simpson metrics didn't show statistical significance, analyses of the α -diversities based on all the five metrics suggested that at adjacent-tumor (P) sites, those from patients with CRC recurrence were higher than those from patients without recurrence (Figure 1). These data suggest that there is no significant difference of compositions of gut mucosal species at on-tumor (T) and off-tumor (N) sites between patients with and without CRC recurrence. However, compositions of internal species showed differences at adjacent-tumor (P) sites between patients with and without CRC recurrence.

Microbiota profiles linked to CRC recurrence

To assess whether the structure diversities (beta-diversity) of gut microbiota show difference between patients with and without CRC recurrence, we performed PCA for abundances of genera identified at on-tumor (T), adjacent tumor (P) and off-tumor (N) sites of CRC patients diagnosed with and without disease recurrence. The microbiota structure diversities at on-tumor (T), adjacent tumor (P) and off-tumor (N) sites of CRC patients diagnosed with and without disease recurrence showed differences (Figure 2), suggesting that the structures of gut microbiota were different at these sites in patients with and without CRC recurrence.

We next compared the differential relative abundances of bacterial taxa at phylum level between patients with and without CRC recurrence. Among all the samples, *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, and *Fusobacteria* showed the top relative abundances (Figure 3). *Firmicutes*, *Bacteroidetes*, and *Fusobacteria* showed higher relative abundances at on-tumor (T) sites of patients diagnosed with CRC recurrence than those at on-tumor (T) sites of patients diagnosed without CRC recurrence (Figure 3). At adjacent-tumor (P) sites, the relative abundances of *Fusobacteria* in patients with CRC recurrence were dramatically higher than those in patients without CRC recurrence (Figure 3). By contrast, the relative abundances of *Fusobacteria* at off-tumor (N) sites showed no differences between patients with and without CRC (Figure 3). The relative abundances of *Firmicutes* and *Bacteroidetes* at off-tumor (N) sites of patients with CRC recurrence were lower than those of patients without CRC recurrence (Figure 3). Collectively, these data suggest that species from *Fusobacteria*, *Bacteroidetes* and *Firmicutes* may play roles in worsening or improving the patient prognosis.

Altered gut mucosal microbiota signatures in CRC recurrence

We next carried out the LEfSe to predict site-specific biomarkers that are associated with CRC recurrence. By setting LDA score > 3.0 (FDR adjusted *P* value < 0.1), a total of 100 genera were identified to show significant differences of relative abundances between patients with and without CRC recurrence (Figure 4, Supplementary Table 2). The top 10 genera/families with the highest LDA scores contained *Fusobacterium* (LDAScore = 5.56, FDR adjusted *P* = 0.0026), *Faecalibacterium* (LDAScore = 5.29, FDR adjusted *P* = 0.0176), *Peptostreptococcus* (LDAScore = 5.13, FDR adjusted *P* = 0.0026), *Streptococcus* (LDAScore = 5.08, FDR adjusted *P* = 0.0639), *Parvimonas* (LDAScore = 5.04, FDR adjusted *P* = 0.0072), *Burkholderiales* (LDAScore = 5.03, FDR adjusted *P* = 0.0018), *Pseudomonas* (LDAScore = 4.82, FDR adjusted *P* = 0.0301), *Caulobacteraceae* (LDAScore = 4.77, FDR adjusted *P* = 0.0043), *Mitsuokelia* (LDAScore = 4.77, FDR adjusted *P* = 0.0011), and *Pseudomonadales* (LDAScore = 4.76, FDR adjusted *P* = 0.0106) (Supplementary Table 2). Seven genera/families, including *Fusobacterium*, *Pyramidobacter*, *Mogibacteriaceae*,

¹ *Coriobacteriaceae*, *Anaerotruncus*, *Slackia* and *Bacteroidales*, showed higher abundances at on-tumor (T) and adjacent-tumor (P) sites [but not off-tumor (N) sites] of patients with CRC recurrence than those without CRC recurrence and healthy controls (Supplementary Table 2). Four genera, including *Dialister*, *Selenomonas*, TG5 and *Schwartzia*, showed higher abundances at ¹ on-tumor (T), adjacent-tumor (P) and off-tumor (N) sites of patients with CRC recurrence than those without CRC recurrence and healthy controls (Supplementary Table 2). By contrast, some well-recognized CRC drivers, e.g. *Peptostreptococcus* and *Streptococcus*, showed lower abundances at on-tumor (T) site of patients with CRC recurrence than those without CRC recurrence. These CRC-recurrence-associated genera may play roles in interacting with host cells and worsening patient prognosis.

Gut mucosal microbiota profiles associated with prognosis and survival of CRC patients

Next, we selected those genera/families that were significantly associated with CRC recurrence based on LEfSe analysis, and used their abundances to evaluate their effects on prognosis and patient survival. The Kaplan-Meier survival analyses curves showed that seven bacterial genera at on-tumor (T) site and nine bacterial genera at adjacent-tumor (P) site were capable of significantly predicting the OS of patients ($P < 0.05$) (Figure 5). High abundances of *Anaerotruncus*, *Bacteroidales*, *Fusobacterium*, *Pyramidobacter*, *Pseudoramibacter_Eubacterium* and TG5 at on-tumor (T) site predicted shorter OS, whereas high abundance of *Parvimonas* was associated with longer OS (Figure 5A). At adjacent-tumor (P) site, high abundances of *Anaerotruncus*, *Coriobacteriaceae*, *Dialister*, *Filifactor*, *Mogibacteriaceae*, *Pyramidobacter* and *Treponema* predicted shorter OS, whereas high abundances of *Haemophilus* and *Bacteroidales* indicated longer OS (Figure 5B). The cut-off values were shown in supplementary materials Supplementary Table 3.

We next performed Kaplan-Meier survival analyses to analyze DFS between patients with high and low abundances of specific bacterial taxa. A total of eleven genera at on-

tumor (T) site and eight genera at adjacent-tumor (P) site were capable of predicting DFS obviously ($P < 0.05$) (Figure 6). High abundances of *Anaerotruncus*, *Bacteroidales*, *Erysipelotrichaceae_Eubacterium*, *Filifactor*, *Mogibacteriaceae*, *Pyramidobacter*, *Pseudoramibacter_Eubacterium*, *Porphyromonadaceae*, *Slackia* and TG5 at on-tumor (T) site were associated with shorter DFS, whereas high abundance of *Gemella* was associated with longer DFS (Figure 6A). At adjacent-tumor (P) site, high abundances of *Anaerotruncus*, *Coriobacteriaceae*, *Dialister*, *Filifactor*, *Pyramidobacter* and *Schwartzia* showed shorter DFS, while high abundances of *Haemophilus* and *Bacteroidales* were associated with longer DFS (Figure 6B).

According to the abundance cut-off value of each genus calculated by Kaplan-Meier survival analyses (Supplementary Table 3), we divided the patients into two groups with high or low abundance of each genus. Then, we investigated the survival risk associated with the abundance of each genus/family and other clinicopathological features via ¹⁷ univariate and multivariate COX regression analyses. According to the univariate Cox regression analysis results, we chose bacterium or clinicopathological feature with significance ($P < 0.1$) and additional bacterium with significance in K-M analysis ($P < 0.05$) to further perform multivariate Cox regression analysis. Interestingly, although gender, age, differentiation, smoking and drinking histories showed no significance ($P > 0.05$) in univariate DFS or OS Cox regression analyses (Supplementary Tables 4 and 5), the TNM stage and location of tumor in intestine showed significant influences on patient's survival based on multivariate OS Cox regression analysis ($P < 0.05$) (Supplementary Figures 1 and 2). CRC III/IV stage and CRC located in the right colon revealed shorter OS than CRC I/II stage and CRC located in the left colon or rectum, respectively. Moreover, only CRC III/IV stage showed shorter DFS than CRC I/II stage (Supplementary Figure 2). Besides, microsatellite instability was associated with improved DFS rates in all patients based on multivariate Cox regression analysis (Supplementary Figure 3). Univariate Cox regression OS analysis indicated patients with high abundances of *Anaerotruncus*, *Bacteroidales*, *Pyramidobacter*, *Pseudoramibacter_Eubacterium* and TG5 at on-tumor (T) site

were associated with shorter OS rates (Supplementary Figure 1). However, no significant differences were found between these bacteria and OS in multivariate analysis (Supplementary Figure 1). At adjacent-tumor (P) site, univariate analysis showed high abundances of *Anaerotruncus*, *Coriobacteriaceae*, *Dialister*, *Filifactor*, *Mogibacteriaceae*, *Pyramidobacter* and *Treponema* and low abundance of *Bacteroidales* were associated with shorter OS rates. In multivariate analysis, high abundance of *Anaerotruncus* and low abundance of *Haemophilus* were associated with shorter OS rates (Supplementary Figure 2).

High abundances of *Anaerotruncus*, *Bacteroidales*, *Erysipelotrichaceae_Eubacterium*, *Mogibacteriaceae*, *Pyramidobacter* and *Pseudoramibacter_Eubacterium* at on-tumor (T) site predicted shorter DFS rates in univariate analysis, whereas high abundances of *Mogibacteriaceae* and *Slackia* were associated with shorter DFS rates in multivariate analysis (Supplementary Figure 3). In univariate analysis of bacterial taxa at adjacent-tumor (P) site, high abundances of *Anaerotruncus*, *Coriobacteriaceae*, *Dialister*, *Filifactor*, *Pyramidobacter* and *Schwartzia* indicated shorter DFS rates (Supplementary Figure 4). Meanwhile, low abundance of *Bacteroidales* was associated with shorter DFS rates (Supplementary Figure 4). In multivariate analysis, there was no correlation between these bacteria and DFS rates (Supplementary Figure 4).

DISCUSSION

Emerging evidences indicate that gut microbiota plays pivotal roles in CRC incidence and progression^[21]. Most of previous studies have focused on identifying gut microbiota profiles linked to CRC carcinogenesis and revealing the physiological roles of specific species, e.g. *Fn*, *Peptostreptococcus anaerobius*, *Bacteroides fragilis* and *Eubacterium rectale*, in CRC tumorigenesis and development^[22-25]. However, few reports have fully screened the gut microbiota profiles linked to prognosis and survival of CRC patients^[7]. Up to date, only *F. nucleatum* and *B. fragilis* were evaluated for their impacts on patient prognosis^[7,8,13,14]. Therefore, our understanding of bacterial taxa associated with clinical outcome of CRC is incomplete. Profiling these bacterial taxa will pave a

way for further understanding their functional roles in impairing clinical outcome of CRC and development of novel strategies for prevention of CRC recurrence. In this study, we explored large-scale screening on ¹gut mucosal microbiota of triplet-paired biopsy samples collected from on-tumor (T), adjacent-tumor (P) and off-tumor (N) sites of CRC patients, and identified critical bacterial taxa that are linked to prognosis and survival of CRC patients.

Our data revealed that a number of bacterial genera/families at on-tumor (T) and adjacent-tumor (P) sites are capable of influencing DFS and OS rates. High abundances of *Anaerotruncus* and *Pyramidobacter* indicated shorter DFS and OS rates in K-M survival analyses, and increased the risk of CRC recurrence and patient death according to our Cox regression analyses ($P < 0.05$). High abundances of *Coriobacteriaceae*, *Dialister* and *Filifactor* at adjacent-tumor (P) site and high abundances of *Bacteroidales*, *Pseudoramibacter_Eubacterium* and TG5 at on-tumor (T) site indicated shorter DFS and OS rates, and increased the risk of CRC recurrence and patient death ($P < 0.05$). Conversely, high abundance of *Parvimonas* at on-tumor (T) site showed longer OS rates. Meanwhile, high abundances of *Bacteroidales* and *Haemophilus* at adjacent-tumor (P) site indicated longer DFS and OS rates, but only *Haemophilus* decreased the risk of death in multivariate Cox regression analysis ($P < 0.05$). High abundances of *Fusobacterium* at on-tumor (T) site and *Treponema* at adjacent-tumor (P) site indicated shorter OS rates, but only *Treponema* increased the risk of death ($P < 0.05$). High abundances of *Erysipelotrichaceae_Eubacterium*, *Gemella*, *Porphyromonadaceae* and *Slackia* at on-tumor (T) site and *Coriobacteriaceae* at adjacent-tumor (P) site indicated shorter DFS rates, and *Erysipelotrichaceae_Eubacterium*, *Gemella*, *Slackia* and *Schwartzia* increased the risk of recurrence ($P < 0.05$). In multivariate regression analysis, *Haemophilus* showed more convinced protective effect and *Anaerotruncus* showed risky effect when referred to death ($P < 0.05$), and *Mogibacteriaceae* and *Slackia* showed obvious risky effects when referred to recurrence ($P < 0.05$).

As well-recognized oral pathogens, *Fn*, *Parvimonas micra* and *Gemella morbillorum* were significantly enriched in both right- and left-sided CRC tumors^[26]. *F. nucleatum* in

tumor samples was reported to be associated with worse outcomes in terms of OS, DFS, or cancer-specific survival, with hazard ratios ranging from 1.58 to 19.96^[27]. However, our data showed that the abundance of *F. nucleatum* only affected OS but not DFS rates. This observation is consistent with two previous studies. Wei *et al*^[7] reported that high abundance of *F. nucleatum* or *B. fragilis* was associated with poor OS rates after surgery. A meta-analysis found that enrichment of *F. nucleatum* in tumor tissue was associated with worse OS among CRC patients ($n = 5$ studies, HR = 1.87; 95%CI: 1.12–3.11; $I^2 = 60.6\%$), but was not associated with DFS ($n = 3$ studies, HR = 1.48; 95%CI: 0.84–2.59; $I^2 = 88.5\%$)^[28]. *F. nucleatum* may contribute to CRC progression in an established tumor microenvironment (TME), due to its highly adherent, invasive and pro-inflammatory nature that can take advantage of a compromised colonic epithelial cell layer. Moreover, *F. nucleatum* is an asaccharolytic bacterium that will not compete for glucose, a preferred substrate in tumor metabolism, and *F. nucleatum* can tolerate the hypoxic tumor environment^[26,29]. In addition, *F. nucleatum* is capable of functioning as a bridge-forming bacterium to interact with other bacterial colonizers, leading to most complex biofilm formation in the human body^[30,31]. The increases of *F. nucleatum* abundance were observed in both biopsy site and saliva samples of CRC^[32]. Flynn *et al*^[33] have proposed a polymicrobial synergy model that certain oral pathogens may cooperate to fight off host immune system for survival and be able to establish a niche containing mixed species in the gut. The intestinal mucosa and epithelium are falling off and replaced constantly, which provide available nutrients and binding sites for adhesive bacteria. Intestinal and oral environments share similar pH, which is conducive for bacteria to form biofilm and persist in host. Initiation or progression of tumorigenesis benefit these bacteria for proliferation, because local inflammatory responses during tumorigenesis increase available nutrients in the niche^[33]. Polymicrobial colonisation at tumor site by bacterial species that are phylogenetically related to those classified as oral pathogens (*e.g. Fusobacterium, Anaerococcus* and *Parvimonas*), could promote tumorigenesis by modifying the TME and eliciting an elevated response of T helper 17 cells, which is linked to a poor prognosis of CRC patient^[34].

The order *Bacteroidales* contains more than 35 species, including the best studied genus *Bacteroides*, which is highly abundant in gut microbiota of healthy human^[35-39]. Due to its high density, it is proposed that the species in this order may form mutualistic relationships with the host gut and play a role in stabilizing the compositional structure of gut microbiota. However, certain species from this order is pathogenic and considered as a potential driver bacterium for CRC. For example, ETBF secretes a zinc metalloprotease toxin, known as *B. fragilis* toxin, which is associated with inflammatory bowel disease and CRC^[40]. ETBF can induce productions of reactive oxygen species in host cells that cause oxidative DNA damage, induce inflammation, and disrupt the integrity of the epithelial barrier. In addition, ETBF is able to activate β -catenin nuclear signaling cascade and induce proliferation of host cells^[41]. Although *B. fragilis* may enhance the efficiency of immune checkpoint inhibitor therapy^[42], ETBF has been reported to decrease OS and DFS of CRC patients^[27,42]. Unexpectedly, our data showed that *Bacteroides* at on-tumor (T) and adjacent-tumor (P) sites displayed distinguishable effects on OS and DFS (Figures 5 and 6). High abundance of *Bacteroides* at on-tumor (T) site manifested worse OS and DFS, whereas those at adjacent-tumor (P) site showed better OS and DFS. It was reported that high levels of defined chemokines (e.g. CCL5 and CCL17) in CRC tissues may attract beneficial T cells [cytotoxic T lymphocytes, T-helper (Th)1 cells (Th1), interleukin (IL)-17-producing Th cells (Th17) and regulatory T cells] and lead to improved patient survival, and titers of *Bacteroidales* were positively correlated with expression levels of individual chemokines and extent of T cell infiltration^[43]. Loading *Bacteroidales* to tumor xenografts recruited T cells, indicating that *Bacteroidales* is capable of controlling extent of tumor infiltration by beneficial immune cells^[43]. However, CRC tissue was infiltrated by more “not effector” T cells (Th2/Th0/Tregs/Tnull) with regulatory or anergic properties, which are unable to kill CRC cells and may contribute to the CRC promotion^[44]. Therefore, *Bacteroidales* at off-tumor (P) site here may play a protective role in recruiting beneficial T cells (e.g. Th1, Th17, etc.), leading to improved prognosis of CRC patients. On the other hand,

Bacteroidales at on-tumor (T) site may contribute to CRC recurrence as pathogens or thrive as passenger bacteria.

In our K-M survival analyses, high abundance of *Anaerotruncus* at either on-tumor site or adjacent-tumor site was associated with shorter OS and DFS, suggesting that the genus increases the risk of CRC recurrence and patient death. The genus *Anaerotruncus* contains only one validly published species, namely *Anaerotruncus colihominis*. *A. colihominis* was first identified as a Gram-positive, anaerobic bacillus that was isolated from the stool specimens of two children^[45]. Moreover, this species was isolated from the blood culture of human with nosocomial bacteremia^[45]. Significant enrichment of *Anaerotruncus* was found in the endometrium of patients with endometrial cancer, suggesting that this bacterium promotes inflammation and tumorigenesis^[46]. High fat diets (HFDs) are generally considered as a high-risk factor for CRC. Abundance of *Anaerotruncus* is considered to be linked to consumption of saturated fatty acids in both men and women^[47]. In mouse model experiments, HFDs and high-sucrose diets lead to conditional pathogenic bacterial growth, such as *Anaerotruncus* and *Bacteroides*. These bacteria played a pro-inflammatory role in disrupting the integrity of epithelial barrier function^[48-50].

Several genera, including *Pyramidobacter* and *Mogibacteriaceae*, are sulfidogenic and associated with CRC^[51-55]. *Pyramidobacter* is mainly isolated from human oral cavity, upper gastrointestinal tract and bile^[56,57]. The genus *Pyramidobacter* contains anaerobic, Gram-negative bacilli that produce acetic, isovaleric acids and many other trace chemicals as metabolism products^[51]. Abundance of *Pyramidobacter* was found to be higher in older adults, and was positively correlated with pro-inflammatory cytokine IL-6 that promotes CRC development^[58]. In addition, the family *Mogibacteriaceae* was reported to be associated with CRC and observed to co-occur with *Fn*^[53,59]. Our previous study also reported that *Mogibacterium* from the family *Mogibacteriaceae* was associated with *Peptostreptococcus* in gut microbiota isolated from CRC patients^[60].

Several genera identified in this study were reported to be associated with CRC previously. The role of *Eubacterium* in CRC initiation was underestimated for long time.

Recently, we identified *Eubacterium* as potential driver bacteria contributing to CRC and experimentally showed that the lipopolysaccharide of *Eubacterium rectale* activates the transcription factor NF- κ B, which regulates innate and adaptive immune responses in normal colon epithelial cells^[22]. The genus *Dialister* was detected in blood of patients with oral infections that cause bacteremia^[61]. A meta-analysis based on 26 studies that use next generation sequencing to analyze microbiota showed that the relative abundance of *Dialister* was significantly higher in cancer patients than those in control samples^[62]. The metabolism end products of *Dialister* include acetate, lactate and propionate that may cause carcinogenesis^[62]. *Treponema denticola* is an oral pathogen and associated with an increased risk of CRC^[63]. The genus *Schwartzia* is linked to CRC carcinogenesis-related gene methylation^[64]. The family *Coriobacteriaceae* is considered as commensal or probiotic bacteria residing in human gut^[65]. Their roles in driving CRC recurrence are unknown yet.

Unlike other critical genera analyzed in this study, low abundance of *Haemophilus* at adjacent-tumor (P) site indicated longer OS and DFS than those with high abundance in our K-M analyses. *Haemophilus* is a commensal microorganism, which belongs to the phylum *Proteobacteria*. It is an opportunistic pathogen that may lead to infections such as endocarditis and pneumonia. *Haemophilus* in stool samples showed significantly higher proportions in CRC group than in control group^[66]. The proportions of *Haemophilus* decreased after tumor removal *via* surgery, indicating it is a carcinogenesis pathogen indirectly^[67]. However, decreased abundance of *Haemophilus* was observed in stage I/II CRC compared to stage 0 (earliest) CRC, which may be due to overgrowth of other harmful bacteria that acted as competitors during the transition from pre-cancerous lesions to late-stage CRC^[68].

According to our COX regression analysis, location (right/left colon or rectum) is another risk factor for patient death. Compared to left colon or rectum, right colon obviously increased risk of death. Consistent with our results, a study showed that the 10-year OS of patients with CRC at right-sided colon was shorter than that of patients with CRC at left-sided colon^[68]. There is more abundant blood supply in the right-sided

colon than other parts of colon, which benefits tumor growth and results in common clinical symptoms including anaemia, emaciation, fever, and dyscrasia. The right-sided CRC exhibits more mucinous and advanced TNM stage^[69]. Obviously different patterns of microbiota structures and abundances were found between left-sided and right-sided colons of CRC patients^[69]. Moreover, invasive bacterial biofilms were found in 89% of right-sided CRC cases but in only 12% of left-sided CRC cases^[69]. Therefore, right-sided CRC specific bacterial species with concomitant procarcinogenic epithelial responses may contribute to the development of right-sided CRC^[70].

We further identified that TNM stage showed an obvious impact on survival of CRC patients. Patients with III/IV stage CRC had higher risk of death and disease recurrence, which is consistent with previously reported results^[71]. At the same time, microsatellite status was another factor that influenced patient survival in our study. MSI may decrease the risk of CRC recurrence compared to MSS in the overall patients. Moreover, it was reported that MSI shows stage-specific impacts on prognostic of CRC patients^[70]. In stages II and III CRC, MSI-H tumors had superior prognosis compared with MSS-high tumors. In stage IV CRC, although merely 4% of tumors were identified as MSI-H tumors, these tumors were recognized to be associated with inferior survival^[70]. Sequencing paired colon tumor and normal-adjacent tissue and mucosa samples revealed significant enrichment of *Bacteroides fragilis* and *Fn* in deficient mismatch repair (MMR) CRC, but not in proficient MMR CRC^[72].

CONCLUSION

In this work, we have identified critical bacterial taxa that are associated with prognosis and survival of CRC patients. Our work suggests that intestinal microbiota can serve as biomarkers to predict the risk of CRC recurrence and patient death. Unexpectedly, most of these identified genera have not been investigated for their physiological roles in interacting with host intestinal cells. On the other hand, some well-recognized CRC drivers, *e.g.* *Peptostreptococcus* and *Streptococcus*, are not associated with CRC recurrence according to our observations. Thus, the mechanism behind bacteria-driving CRC

recurrence may be different from those proposed for bacteria-driving CRC development. The activities of complex immune cells, *e.g.* various types of T cells (*e.g.* Th1, Th2 and Th17) and macrophages, in response to these bacterial activities may need to be considered. Further functional analyses of physiological roles of these bacteria in patient prognosis and CRC recurrence will shed light on developing novel strategies for CRC treatment and prevention.

14

ARTICLE HIGHLIGHTS

Research background

Colorectal cancer (CRC) is one of the most common malignant tumors. Gut mucosal microbiota is considered to be one of the key factors promoting CRC. There are evidences that certain gut bacteria are linked to the prognosis (recurrence, overall survival and disease-free survival) of CRC, but there is a lack of research on the relationship between large-scale intestinal microbiota profiles and CRC recurrence/patient prognosis.

Research motivation

Our study focused on the relationship between the abundances of intestinal microbiota at different position and CRC recurrence/patient prognosis. This study provides novel potential biomarkers for patient prognosis in the future.

20

Research objectives

The main objective of this study was to evaluate whether the abundances of intestinal microbiota at on-tumor or adjacent-tumor sites can predict CRC recurrence and patient prognosis. Our study has preliminarily suggested that some gut bacteria may have predictive values for CRC recurrence and patient prognosis. These results can provide new biomarkers for prediction of CRC recurrence in the future.

Research methods

We collected intestinal bacteria from different locations of intestinal mucosa of patients and healthy controls. The bacterial taxa and abundances were determined by high-throughput 16S ribosomal RNA sequencing. The relationship between gut mucosal microbiota profiles and CRC recurrence and patient prognosis was explored by bioinformatics analysis, Kaplan Meier survival analysis and Cox regression analysis. These methods have been well established in the field.

Research results

Through analysis, gut mucosal microbiota profiles are associated with CRC recurrence and patient prognosis. Abundances of some bacterial genera/families, *e.g. Anaerotruncus, Bacteroidales* and *Fusobacterium*, may have prognostic values for CRC recurrence and patient prognosis. The mechanism studies exploring the roles of gut mucosal microbiota in CRC recurrence and patient prognosis need to be carried out in the future.

Research conclusions

This study provides new potential biomarkers identified from gut mucosal microbiota for CRC recurrence and patient prognosis.

Research perspectives

In the future, it is necessary to explore the mechanism of how gut mucosal bacteria affect the CRC recurrence and patient prognosis.

12%

SIMILARITY INDEX

PRIMARY SOURCES

- 1 www.mdpi.com 352 words — 5%
Internet
- 2 Kensuke Yamamura, Daisuke Izumi, Raju Kandimalla, Fuminori Sonohara et al. " Intratumoral Levels Predict Therapeutic Response to Neoadjuvant Chemotherapy in Esophageal Squamous Cell Carcinoma ", Clinical Cancer Research, 2019 66 words — 1%
Crossref
- 3 gutpathogens.biomedcentral.com 41 words — 1%
Internet
- 4 v3r.esp.org 32 words — < 1%
Internet
- 5 Zhiliang Wei, Shougen Cao, Shanglong Liu, Zengwu Yao, Teng Sun, Yi Li, Jianteng Li, Dongfeng Zhang, Yanbing Zhou. "Could gut microbiota serve as prognostic biomarker associated with colorectal cancer patients' survival? A pilot study on relevant mechanism", Oncotarget, 2016 27 words — < 1%
Crossref
- 6 Julia L. Drewes, James R. White, Christine M. Dejea, Payam Fathi et al. "High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm 25 words — < 1%

status reveal common colorectal cancer consortia", npj Biofilms and Microbiomes, 2017

Crossref

-
- | | | |
|--|---|-----------------|
| <div style="background-color: #8B4513; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">7</div> | pureadmin.qub.ac.uk
<small>Internet</small> | 25 words — < 1% |
|--|---|-----------------|
-
- | | | |
|--|---|-----------------|
| <div style="background-color: #005596; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">8</div> | wjso.biomedcentral.com
<small>Internet</small> | 22 words — < 1% |
|--|---|-----------------|
-
- | | | |
|--|---|-----------------|
| <div style="background-color: #800080; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">9</div> | research.vu.nl
<small>Internet</small> | 20 words — < 1% |
|--|---|-----------------|
-
- | | | |
|---|---|-----------------|
| <div style="background-color: #6B8E23; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">10</div> | www.ncbi.nlm.nih.gov
<small>Internet</small> | 18 words — < 1% |
|---|---|-----------------|
-
- | | | |
|---|--|-----------------|
| <div style="background-color: #000080; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">11</div> | Andreas Koulouris, Christos Tsagkaris, Ippokratis Messaritakis, Nikolaos Gouvas et al. "Resectable Colorectal Cancer: Current Perceptions on the Correlation of Recurrence Risk, Microbiota and Detection of Genetic Mutations in Liquid Biopsies", Cancers, 2021
<small>Crossref</small> | 17 words — < 1% |
|---|--|-----------------|
-
- | | | |
|---|---|-----------------|
| <div style="background-color: #000080; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">12</div> | Mangili, G., J. Ottolina, G. Cormio, Vera Loizzi, P. De Iaco, D.A. Pellegrini, M. Candiani, G. Giorda, G. Scarfone, S.C. Cecere, L. Frigerio, A. Gadducci, C. Marchetti, and G. Ferrandina. "Adjuvant chemotherapy does not improve disease-free survival in FIGO stage IC ovarian granulosa cell tumors: The MITO-9 study", Gynecologic Oncology, 2016.
<small>Crossref</small> | 16 words — < 1% |
|---|---|-----------------|
-
- | | | |
|---|---|-----------------|
| <div style="background-color: #FF0000; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">13</div> | academic.oup.com
<small>Internet</small> | 16 words — < 1% |
|---|---|-----------------|
-
- | | | |
|---|---|-----------------|
| <div style="background-color: #FF00FF; color: white; display: inline-block; width: 30px; height: 30px; text-align: center; line-height: 30px;">14</div> | www.wjgnet.com
<small>Internet</small> | 15 words — < 1% |
|---|---|-----------------|

15	annonc.oxfordjournals.org Internet	14 words — < 1%
16	assets.researchsquare.com Internet	14 words — < 1%
17	Zhuo Lin, Zhuoyan Chen, Dan Miao, Qian Xu, Liuwei Zeng, Luying Zhao, Jun Xu, Xian Song, Fujun Yu. "Pyroptosis-related genes IRAK1, NOD2, POP1 and YWHAB impact the prognosis through immune functional pathways in hepatocellular carcinoma", Research Square Platform LLC, 2021 Crossref Posted Content	13 words — < 1%
18	cran.r-project.org Internet	13 words — < 1%
19	core.ac.uk Internet	12 words — < 1%
20	bsdwebstorage.blob.core.windows.net Internet	11 words — < 1%
21	estia.hua.gr Internet	11 words — < 1%
22	www.frontiersin.org Internet	11 words — < 1%
23	Sofia C Tortora, Vimal M Bodiwala, Andrew Quinn, Laura A Martello, Shivakumar Vignesh. "Microbiome and colorectal carcinogenesis: Linked mechanisms and racial differences", World Journal of Gastrointestinal Oncology, 2022 Crossref	10 words — < 1%

EXCLUDE QUOTES OFF
EXCLUDE BIBLIOGRAPHY OFF

EXCLUDE MATCHES OFF