

Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer

Qiaoyi Liang¹, Jonathan Chiu¹, Yingxuan Chen², Yanqin Huang³, Akira Higashimori^{1,4}, Jingyuan Fang², Hassan Brim⁵, Hassan Ashktorab⁵, Siew Chien Ng¹, Simon Siu Man Ng⁶, Shu Zheng², Francis Ka Leung Chan¹, Joseph Jao Yiu Sung¹, and Jun Yu¹

Abstract

Purpose: Gut microbiota have been implicated in the development of colorectal cancer. We evaluated the utility of fecal bacterial marker candidates identified by our metagenome sequencing analysis for colorectal cancer diagnosis.

Experimental Design: Subjects (total 439; 203 colorectal cancer and 236 healthy subjects) from two independent Asian cohorts were included. Probe-based duplex quantitative PCR (qPCR) assays were established for the quantification of bacterial marker candidates.

Results: Candidates identified by metagenome sequencing, including *Fusobacterium nucleatum* (Fn), *Bacteroides clarus* (Bc), *Roseburia intestinalis* (Ri), *Clostridium hathewayi* (Ch), and one undefined species (labeled as m7), were examined in fecal samples of 203 colorectal cancer patients and 236 healthy controls by duplex-qPCR. Strong positive correlations were demonstrated between the quantification of each candidate by our qPCR assays and metagenomics approach ($r = 0.801\text{--}0.934$, all $P < 0.0001$). Fn

was significantly more abundant in colorectal cancer than controls ($P < 0.0001$), with AUROC of 0.868 ($P < 0.0001$). At the best cut-off value maximizing sum of sensitivity and specificity, Fn discriminated colorectal cancer from controls with a sensitivity of 77.7%, and specificity of 79.5% in cohort I. A simple linear combination of four bacteria (Fn + Ch + m7-Bc) showed an improved diagnostic ability compared with Fn alone (AUROC = 0.886, $P < 0.0001$) in cohort I. These findings were further confirmed in an independent cohort II. In particular, improved diagnostic performances of Fn alone (sensitivity 92.8%, specificity 79.8%) and four bacteria (sensitivity 92.8%, specificity 81.5%) were achieved in combination with fecal immunochemical testing for the detection of colorectal cancer.

Conclusions: Stool-based colorectal cancer-associated bacteria can serve as novel noninvasive diagnostic biomarkers for colorectal cancer. *Clin Cancer Res*; 23(8); 2061–70. ©2016 AACR.

Introduction

Colorectal cancer is one of the most common malignancies worldwide. Many Asian countries including China have experienced a 2- to 4-fold increase in colorectal cancer incidence during the past decade (1). Abnormality in the composition of the gut

microbiota has been implicated as a potentially important etiologic factor in the initiation and progression of colorectal cancer (2). With the widespread application of metagenome sequencing and pyrosequencing in the investigation of intestinal microbiota, an increasing number of bacteria have been identified to be positively associated with the incidence of colorectal cancer (3–7). Recent studies have shown that *Fusobacterium*, especially *Fusobacterium nucleatum* (Fn), is associated with colorectal cancer. Fn is enriched in both the feces and colonic mucosa of colorectal cancer patients (3, 5, 8) and plays important roles in colorectal carcinogenesis (9, 10). In our recent study, using 16S rRNA sequencing to catalog the microbial communities in human gut mucosa at different stages of colorectal tumorigenesis, *Fusobacterium* was also found to be enriched in colorectal tumors (11). Then by using metagenomics analysis to compare the fecal microbiome of 74 colorectal cancer patients and 54 healthy subjects, we have identified bacterial candidates that may serve as noninvasive biomarkers for colorectal cancer (12), including Fn, *Bacteroides clarus* (Bc), *Roseburia intestinalis* (Ri), *Clostridium hathewayi* (Ch), one undefined species (labeled as m7). Unlike Fn, the other bacteria have not yet been associated with colorectal cancer. Moreover, the translational application of these bacterial candidates into diagnostic biomarkers needs further investigation using simple, cost-effective, and targeted methods such as quantitative PCR (qPCR).

In this study, we validated the stool-based bacterial candidate markers in a large cohort of 203 colorectal cancer patients and 236

¹Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, CUHK Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong, China. ²Division of Gastroenterology & Hepatology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China. ³Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, Key Laboratory of Molecular Biology in Medical Sciences, Cancer Institute, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ⁴Department of Gastroenterology, Osaka City University Graduate School of Medicine, Abeno-ku, Osaka, Japan. ⁵Cancer Center and Pathology Department, Howard University College of Medicine, Washington, DC. ⁶Department of Surgery, The Chinese University of Hong Kong, Hong Kong, China.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Authors: Jun Yu, The Chinese University of Hong Kong, Rm707A, Li Ka Shing Medical Sciences Building, Prince of Wales Hospital, Shatin, N.T., Hong Kong. Phone: 852-3763-6099; Fax: 852-2144-5330; E-mail: junyu@cuhk.edu.hk; and Qiaoyi Liang, JessieQY@cuhk.edu.hk

doi: 10.1158/1078-0432.CCR-16-1599

©2016 American Association for Cancer Research.

Translational Relevance

Changes in gut microbiota have been associated with colorectal cancer. In this study, diagnostic utility of bacterial marker candidates identified by metagenome sequencing, including *Fusobacterium nucleatum* (*Fn*), *Bacteroides clarus* (*Bc*), *Roseburia intestinalis* (*Ri*), *Clostridium hathewayi* (*Ch*), and one undefined species (*m7*), were evaluated in two independent Asian cohorts of samples. Duplex-qPCR assays were established for translational application, which quantitated candidate markers consistently with metagenome sequencing and validated the significantly elevated *Fn*, *Ch*, and *m7* and decreased *Bc* and *Ri* in colorectal cancer patients compared with controls. *Fn* showed good performance in discriminating colorectal cancer from controls, and combination of four markers (*Fn* + *Ch* + *m7*-*Bc*) further improved the diagnostic ability of *Fn*. An increased performance of bacterial markers was achieved in combination with the fecal immunochemical test (FIT) for colorectal cancer detection (sensitivity 92.8%, specificity 81.5%). Bacterial markers can be used alone or in combination with existing methods, such as FIT, for noninvasive diagnosis of colorectal cancer.

control subjects to identify a panel of markers with good sensitivity and specificity as a novel diagnostic tool for colorectal cancer. We established probe-based duplex qPCR assays for the quantification of the bacteria; the technique involved is easy and less costly to perform compared with the currently available tests.

Materials and Methods

Human fecal sample collection

Fecal samples ($n = 439$) were collected from the two independent cohorts, including cohort I-Hong Kong: 370 subjects, consisting of 170 patients with colorectal cancer (mean age, 67.2 ± 11.6 years; 100 males and 70 females) and 200 normal controls (59.3 ± 5.8 years; 77 males and 123 females), at the Prince of Wales Hospital, the Chinese University of Hong Kong between 2009 and 2013 (Supplementary Table S1), and cohort II-Shanghai: 69 subjects, consisting of 33 patients with colorectal cancer (mean age, 63.4 ± 9.6 years; 17 males and 16 females) and 36 normal controls (53.2 ± 12.2 years; 10 males and 26 females), at Renji Hospital, Shanghai Jiaotong University between 2014 and 2015 (Supplementary Table S1). Fecal samples from 97 Hong Kong patients with adenoma (60.5 ± 4.7 years; 50 males and 47 females) were further included in this study. Subjects recruited for fecal sample collection included individuals presenting symptoms such as change of bowel habit, rectal bleeding, abdominal pain or anemia, and asymptomatic individuals aged 50 or above undergoing screening colonoscopy as in our previous metagenomics study (12). Samples were collected before or one month after colonoscopy, when gut microbiome should have recovered to baseline (13). The exclusion criteria were: (i) use of antibiotics within the past 3 months; (ii) on a vegetarian diet; (iii) had an invasive medical intervention within the past 3 months; (iv) had a past history of any cancer, or inflammatory disease of the intestine. Subjects were asked to collect stool samples in standardized containers at home, and store the samples in their home -20°C freezer immediately. Frozen samples were then delivered to the

hospitals in insulating polystyrene foam containers and stored at -80°C immediately until further analysis. Patients were diagnosed by colonoscopic examination and histopathologic review of any biopsies taken. Informed consents were obtained from all subjects. The study was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong and the Ethics Committee of Renji Hospital, Shanghai Jiaotong University.

DNA extraction

Fecal samples were thawed on ice and DNA extraction was performed using the QIAamp DNA Stool Mini Kit according to manufacturer's instructions (Qiagen). Extracts were then treated with DNase-free RNase to eliminate RNA contamination. DNA quality and quantity were determined using a NanoDrop2000 spectrophotometer (Thermo Fisher Scientific).

Design of primers and probes

Primer and probe sequences for the internal control were designed manually on the basis of the conservative fragments in bacterial 16S rRNA genes (14), and then they were tested using the tool PrimerExpress v3.0 (Applied Biosystems) for determination of T_m , GC content, and possible secondary structures. We included degenerate sites in the primers and probes to increase target coverage; degenerate sites were not close to 3' ends of primers and 5' end of the probes. Amplicon target was nt 1,063–1,193 of the corresponding *E. coli* genome.

Five bacterial marker candidates identified by previous metagenome sequencing were selected for qPCR quantification, including *Fn*, *Bc*, *Ri*, *Ch*, one undefined species (labeled as *m7*) (Supplementary Table S2). These candidates were identified by eliminating confounding effects of colonoscopy using blocked independent Wilcoxon rank-sum tests with colonoscopy as a stratifying factor in our previous metagenome study (12). *Fn* has also been identified to be enriched in colorectal cancer patients by others (3, 5, 8), while the other four have not associated with colorectal cancer by other researchers. Primer and probe sequences targeting the *nusG* gene of *Fn* (Accession# GMHS-1916) and gene markers identified by our previous metagenome sequencing study, including *Bc* (ID 370640), *Ch* (ID 2736705), *Ri* (ID 181682), and *m7* (ID 3246804; ref. 12), were designed using PrimerExpress. The primer-probe sets specifically detect our targets and not any other known sequences, as confirmed by Blast search. Each probe carried a 5' reporter dye FAM (6-carboxy fluorescein) or VIC (4,7,2'-trichloro-7'-phenyl-6-carboxyfluorescein) and a 3' quencher dye TAMRA (6-carboxytetramethyl-rhodamine). Primers and hydrolysis probes were synthesized by Invitrogen. Nucleotide sequences of the primers and probes are listed in Supplementary Table S3. PCR amplification specificity was confirmed by direct Sanger sequencing of the PCR products or by sequencing randomly picked TA clones.

Quantitative PCR

Quantitative PCR (qPCR) amplifications were performed in a 20- μL reaction system of TaqMan Universal Master Mix II (Applied Biosystems) containing 0.3 $\mu\text{mol/L}$ of each primer and 0.2 $\mu\text{mol/L}$ of each probe in MicroAmp fast optical 96-well reaction plates (Applied Biosystems) with adhesive sealing. Thermal cycler parameters of an ABI PRISM 7900HT sequence detection system were 95°C 10 minutes and (95°C 15 seconds, 60°C 1 minute) \times 45 cycles. A positive/reference control and a negative

control (H₂O as template) were included within every experiment. Measurements were performed in triplicates for each sample. qPCR data was analyzed using the Sequence Detection Software (Applied Biosystems) with manual settings of threshold = 0.05 and baseline from 3–15 cycles for all clinical samples. Experiments were disqualified if their negative control Cq value was <42. Data analysis was carried out according to the ΔCq method, with $\Delta Cq = Cq_{\text{target}} - Cq_{\text{control}}$ and relative abundances = $2^{-\Delta Cq}$.

Fecal immunochemical test

The HemoSure immunogold labeling FIT dipsticks (WHPM Co. Ltd), which are certified by the State Food and Drug Administration of China, were used as described previously (15).

Statistical analysis

Values were all expressed as mean \pm SD or median [interquartile range (IQR)] as appropriate. The differences in specific bacterial abundance were determined by Wilcoxon signed-rank test or Mann–Whitney *U* test. Continuous clinical and pathologic variables were compared by *t* test, while categorical variables were compared by χ^2 test. Spearman correlation coefficient was used to estimate the association of the bacterial abundances and several factors of interest. Factors independently associated with colorectal cancer diagnosis were estimated using univariate and multivariate logistic regression. Receiver operating characteristic (ROC) curve was used to evaluate the diagnostic value of bacterial candidates in distinguishing colorectal cancer and controls. The best cut-off values were determined by ROC analyses that maximized the Youden index ($J = \text{Sensitivity} + \text{Specificity} - 1$; ref. 16). Pairwise comparison of areas under ROC (AUROC) for each method/marker was performed using a nonparametric approach (17). Logistic regression model was applied to obtain probability plot values for estimating the incidence of colorectal cancer among all subjects. ROC curves were then constructed for the logistic regression models. All tests were done by GraphPad Prism 5.0 (GraphPad Software Inc.) or SPSS software v17.0 (SPSS). $P < 0.05$ was taken as statistical significance.

Results

Duplex qPCR assays for convenient and reliable quantification of bacterial abundances

To make the quantification of bacterial content convenient, we designed a degenerate primer–probe (VIC-labeled) set with an amplicon size suitable for qPCR quantification to target a 131-bp conserved region of the 16S rRNA genes. The primer and probe sequences cover >90% of the eubacterial population within the Ribosomal Database Project Release version 10.8 (14). Tests using different fecal DNA samples indicated that this internal control assay was capable of evaluating the total bacteria with DNA templates of <10 ng/ μ L in the final reaction systems (Fig. 1A). Higher template concentrations inhibited PCR amplification probably due to the general impurities within DNA isolated from feces, as no inhibition was observed for pure total DNA extracted from cultured *E. coli* up to at least 25 ng/ μ L. Using templates with concentration <10 ng/ μ L, Cq values correlated well with Log₂ DNA quantities ($R = 0.804$) (Fig. 1B). Then duplex qPCR assays were developed using the VIC-labeled internal control and FAM-labeled primer–probe

sets to specifically target our bacterial candidates. The relative abundance of target bacterium in individual samples could be quantitated consistently with templates of <10 ng/ μ L (Fig. 1C), but template concentration should be >0.1 ng/ μ L to avoid false-negative results in samples with low abundance of the target bacterium (Fig. 1D). Quantification of bacterial abundances using our qPCR assays can be well repeated (Supplementary Fig. S1) and was not interfered by human DNA contamination (Supplementary Fig. S2). Our platform and well-defined experimental conditions may guarantee reliable and convenient quantification of bacterial targets using duplex qPCR assays.

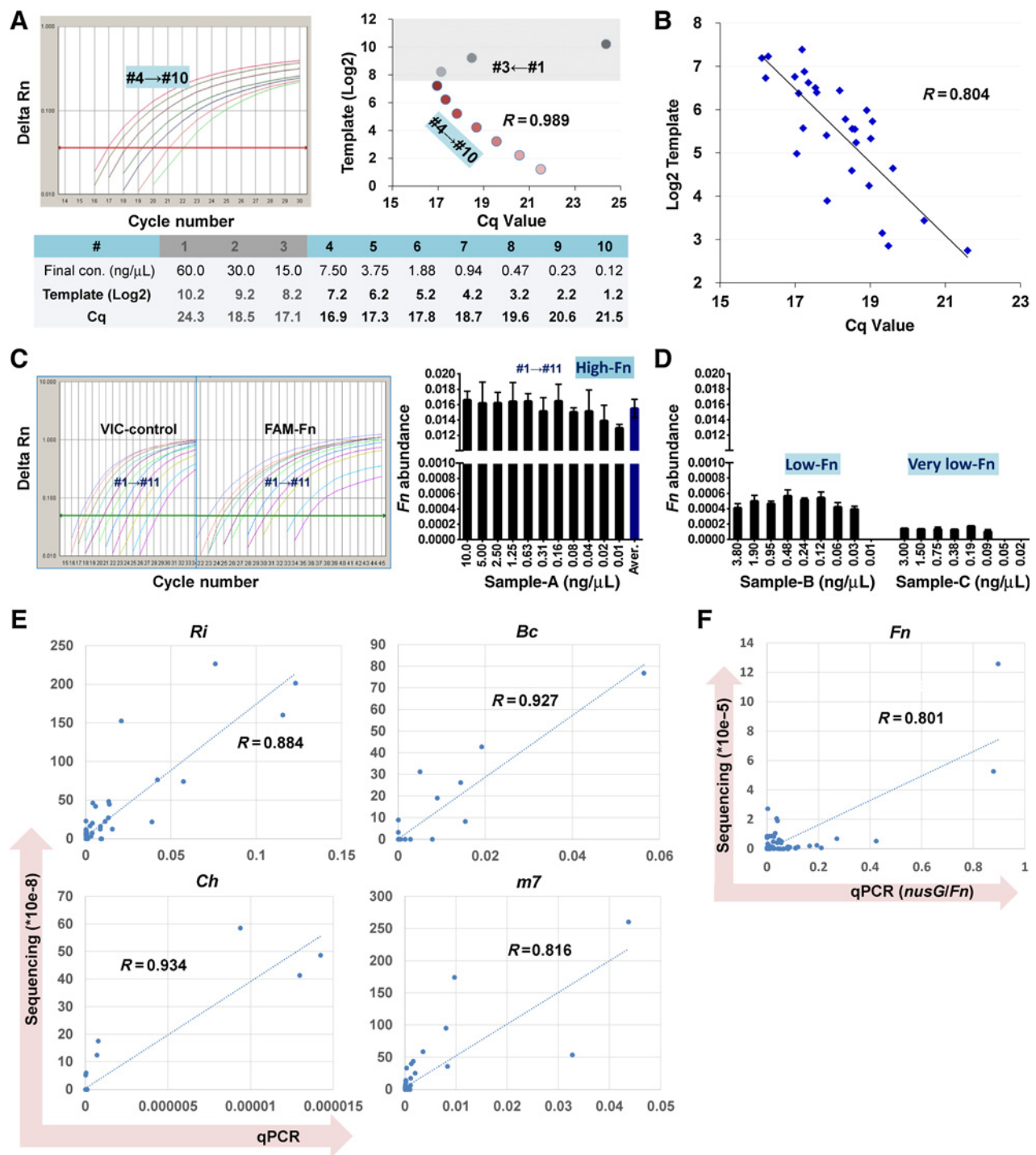
The quantification of each bacterial candidate by metagenomics is correlated with qPCR assays

To verify whether the relative abundances of candidate markers measured by qPCR assays are comparable with metagenomics sequencing, the relative abundances of four bacterial candidates (*Bc*, *Ch*, *Ri*, and *m7*) in a subset of subjects (51 colorectal cancer and 45 controls) by qPCR were compared with metagenomic sequencing. Quantification of each of these bacteria showed strong correlations by qPCR assays compared with metagenomic sequencing (Spearman $r = 0.816$ – 0.934 ; Fig. 1E). The gene marker, butyryl-CoA dehydrogenase from *Fn* (m1704941; 99.13% identity), showed an occurrence of only 52.7% (39/74) in colorectal cancer patients, while at the species level, *Fn* showed an occurrence of 83.8% (62/74) in colorectal cancer patients (Supplementary Table S4). Therefore, *Fn* at species level showing a higher occurrence in colorectal cancer is better than gene marker m1704941, which may represent a specific strain of *Fn*, for the diagnosis of colorectal cancer. Therefore, we established a duplex-qPCR assay targeting the *nusG* gene of *Fn*, which was reported to be transcriptionally more active in colorectal tumors than in matched normal samples (5), to assess the diagnostic value of *Fn* for colorectal cancer. This qPCR assay showed good correlation with *Fn* at species level by metagenome sequencing (Fig. 1F), suggesting qPCR targeting *nusG* may cover more strains of *Fn* and could be more sensitive in detecting colorectal cancer.

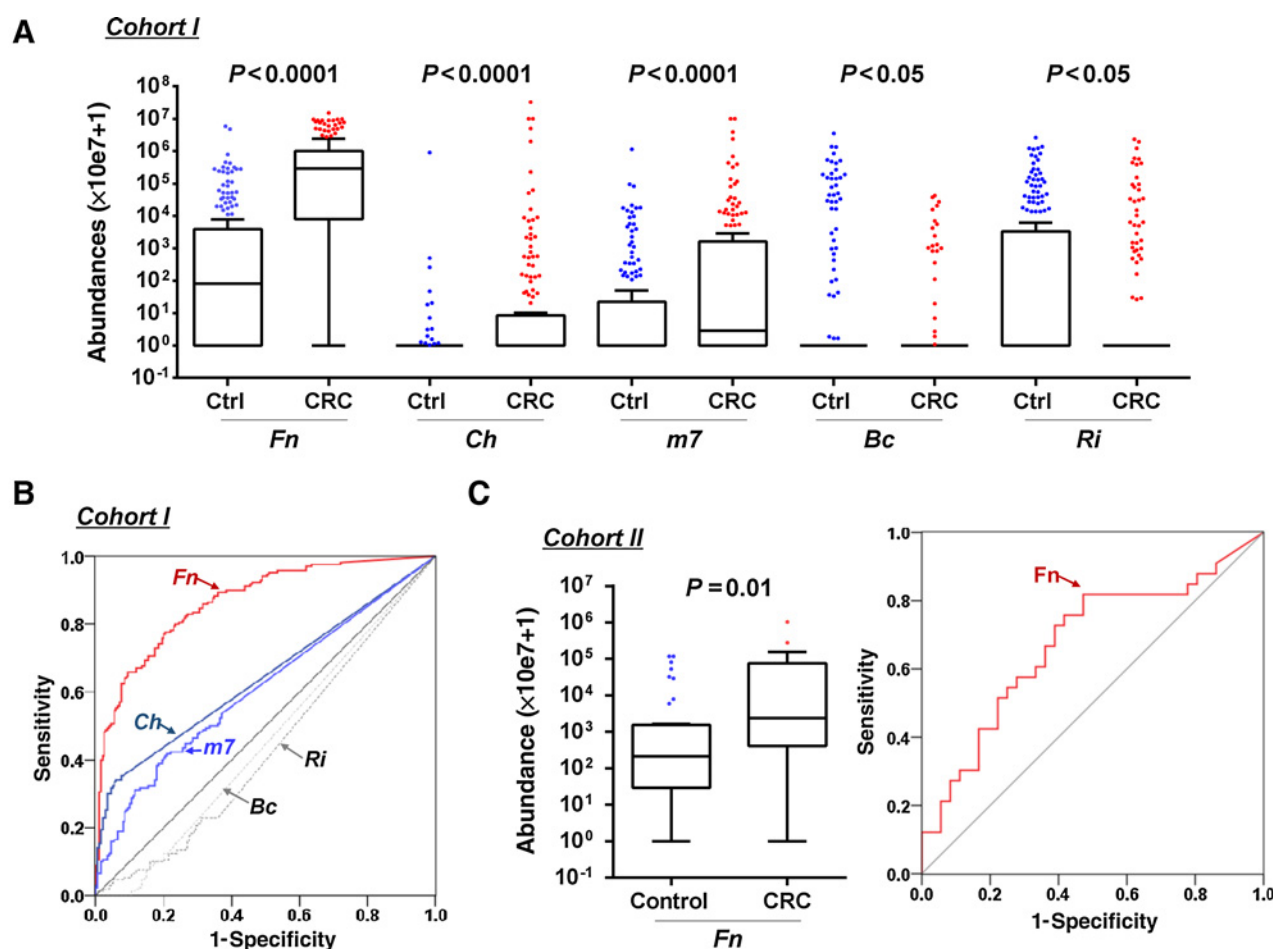
Significantly elevated abundances of *Fn*, *Ch*, and *m7* and decreased abundances of *Bc* and *Ri* in colorectal cancer patients compared with healthy controls

We found that the relative abundance of fecal *Fn* was predominantly higher in colorectal cancer patients ($n = 170$) as compared with healthy controls ($n = 200$; $P < 0.0001$, Fig. 2A; Supplementary Table S5) by qPCR quantification. In addition, we also demonstrated the significantly elevated abundances of *Ch* ($P < 0.0001$) and *m7* ($P < 0.0001$), and decreased abundances of *Bc* ($P < 0.05$) and *Ri* ($P < 0.05$) in colorectal cancer patients compared with control subjects. Bivariate correlation test showed that relative abundances of all the five bacteria were significantly associated with colorectal cancer, but not with tumor–node–metastasis (TNM) staging or tumor location (Supplementary Table S6). The occurrence rates of these five bacteria differed significantly between colorectal cancer patients and healthy control subjects (Supplementary Table S7). These results collectively confirmed the potential of these bacterial marker candidates in discriminating colorectal cancer patients from healthy subjects.

Liang et al.

**Figure 1.**

Assessment of the qPCR assays. **A**, Correlation between template quantity and Cq values of the 16S rDNA control. A representative example of qPCR evaluation on samples #1–10 serially diluted from a mixture of 10 randomly selected fecal samples. qPCR results correlated well with template quantity when final DNA concentrations were <10 ng/μL (#4–#10), whereas DNA of >10 ng/μL inhibited PCR amplification (#1–#3). **B**, Correlation between Cq values of the internal control and DNA quantities ($n = 29$). **C**, The new duplex qPCR assay can stably assess relative target abundance with an appropriate DNA template concentration from fecal samples. An example showing the relative abundance of *Fusobacterium nucleatum* (Fn) was stably assessed in one randomly selected fecal sample with several final DNA concentrations <10 ng/μL. **D**, The relative abundance of Fn was stably assessed in samples, known to have low and very low Fn abundance, with final DNA concentrations <10 ng/μL, but extremely low DNA concentrations may cause false-negative detection in samples of low Fn abundance. **E**, Correlation for the quantification of *Roseburia intestinalis* (Ri), *Roseburia clarus* (Bc), *Clostridium hathewayi* (Ch), and one undefined species (labeled as m7), by metagenomics approach (gene level) and qPCR assay. **F**, Correlation for the quantification of Fn by metagenomics approach (species level) and qPCR assay. Abundances assessed by qPCR in **C–F** are relative values over the internal control of 16S rDNA and thus are shown in arbitrary units.

**Figure 2.**

Quantitative detection of fecal bacterial markers in the diagnosis of colorectal cancer (CRC) patients. **A**, Relative abundances of *Fusobacterium nucleatum* (*Fn*), *Bacteroides clarus* (*Bc*), *Roseburia intestinalis* (*Ri*), *Clostridium hathewayi* (*Ch*), and one undefined species (label as *m7*) in fecal samples differed significantly between healthy control subjects ($n = 200$) and colorectal cancer patients ($n = 170$) of the first cohort. **B**, Receiver operating characteristic (ROC) curves for markers *Fn*, *Ch*, *m7*, *Bc*, and *Ri* in discriminating colorectal cancer patients from healthy control subjects of cohort I. **C**, Relative abundance of *Fn* in fecal samples of 33 colorectal cancer patients and 36 healthy subjects from an independent cohort II and the corresponding ROC curve for *Fn* in discriminating colorectal cancer patients from healthy control subjects in this cohort. Medians with interquartile ranges are shown in the box and whisker plots by Tukey method. Abundances in **A** and **C** are plotted as "relative abundances $\times 10^{7+1}$ " (zero abundance plotted as 1).

Fn is a potential noninvasive fecal biomarker for diagnosing colorectal cancer patients

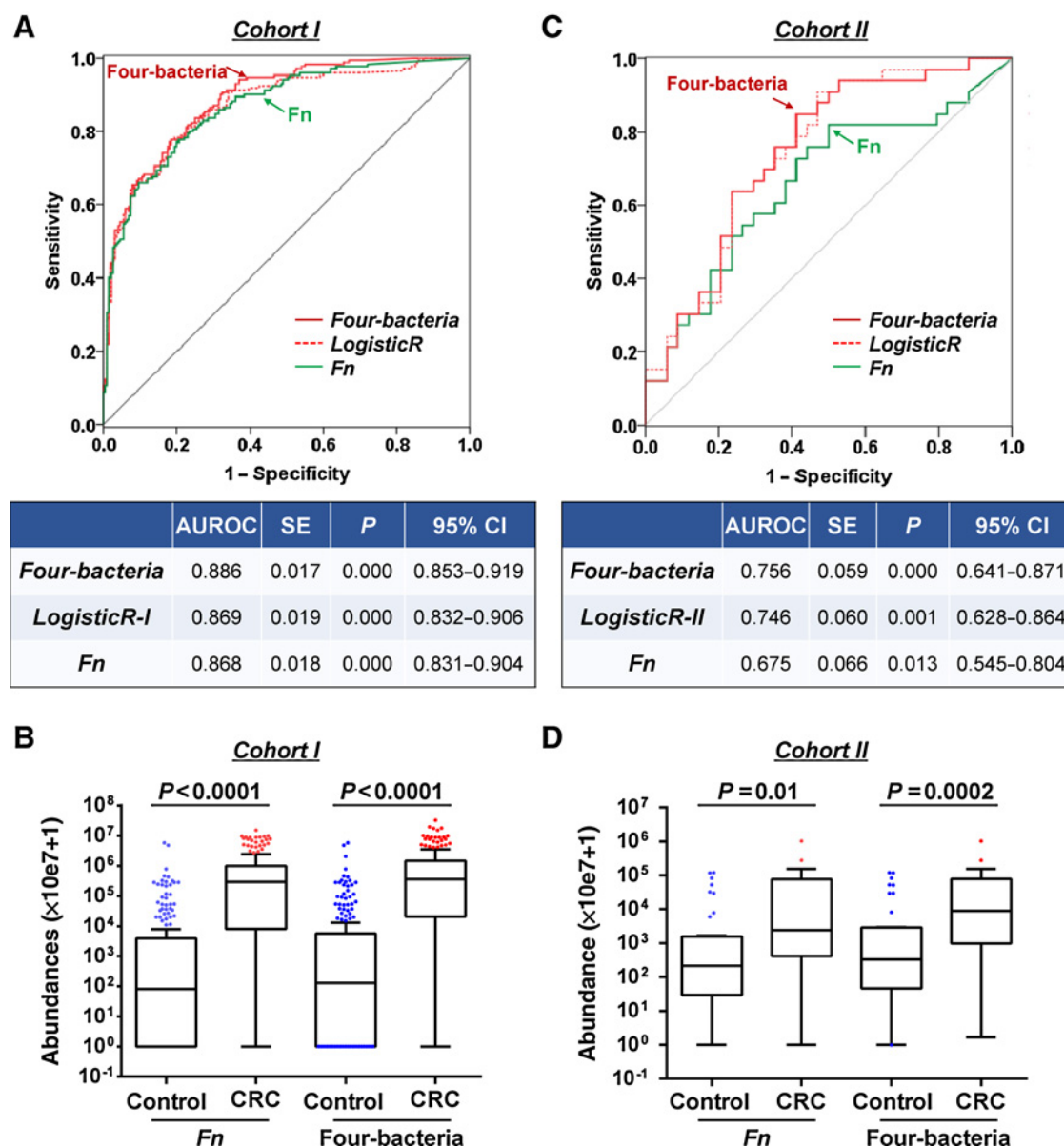
Among all the five bacteria, *Fn* showed the best performance in discriminating colorectal cancer from healthy controls, giving an area under receiver operating curve (AUROC) of 0.868 (95% confidence interval, 0.831–0.904; $P < 0.0001$; Fig. 2B). At the best cut-off value that maximized the sum of sensitivity and specificity by ROC analysis, *Fn* could discriminate colorectal cancer from controls with a sensitivity of 77.7%, specificity of 79.5%, negative predictive value (NPV) of 80.7%, and positive predictive value (PPV) of 76.3% in the first cohort of 170 colorectal cancer patients and 200 healthy subjects. This was further verified in a second independent cohort of 33 colorectal cancer patients and 36 healthy controls. The relative abundance of *Fn* was significantly higher in colorectal cancer patients as compared with healthy controls ($P = 0.012$; Fig. 2C). As a single factor in discriminating between colorectal cancer patients and control subjects, fecal *Fn*

had an AUROC of 0.675 (0.545–0.804; $P = 0.013$). The best cut-off value of *Fn* could discriminate colorectal cancer from controls with a sensitivity of 81.8%, specificity of 52.8%, NPV of 76.0%, and PPV of 61.4% in this second cohort.

The combination of *Fn*, *m7*, *Bc*, and *Ch* improves the diagnostic ability of *Fn* alone for colorectal cancer patients

According to metagenome sequencing data, combination of the tested bacterial markers showed improved diagnostic performance as compared with *Fn* alone, with AUROC increased from 0.748 to 0.843 (Supplementary Fig. S3). Therefore, we evaluated the abilities of combining the other bacterial markers with *Fn* for the diagnosis of colorectal cancer by qPCR assays. We found that a simple linear combination of *Ch*, *m7*, and *Bc* with *Fn* (four-bacteria: *Fn+Ch+m7-Bc*) gave an increased AUROC (0.886) as compared with other combinations (2–5 markers; all ≤ 0.877) and *Fn* only (0.868), as well as the logistic regression model with

Liang et al.

**Figure 3.**

Combination of four markers showed improved diagnostic ability for colorectal cancer (CRC). **A**, ROC curves for simple linear combination of four selected bacterial marker candidates including *Fusobacterium nucleatum* (*Fn*), *Bacteroides clarus* (*Bc*), *Clostridium hathewayi* (*Ch*), and one undefined species (labeled as *m7*), *Fn* only, and probability plot values of logistic regression model in the first cohort. **B**, Relative fecal abundances of *Fn* and four-bacteria (*Fn*, *Bc*, *Ch*, and *m7*) in colorectal cancer patients compared with healthy control subjects of the first cohort. **C**, ROC curves for four bacteria (*Fn*, *Bc*, *Ch*, and *m7*), *Fn* only, and probability plot values of logistic regression model in an independent second cohort. **D**, Relative fecal abundances of *Fn* and four bacteria in colorectal cancer patients and healthy control subjects of cohort II. Medians with interquartile ranges are shown in the box-and-whisker plots by Tukey method. Abundances shown in **B** and **D** are plotted as "relative abundances $\times 10e7+1$ " (zero abundance plotted as 1).

inclusion of the four bacteria (*Fn*, *Ch*, *m7*, and *Bc*; 0.869) in the first cohort (Fig. 3A). The combined relative abundance of four-bacteria was significantly higher in colorectal cancer patients as compared with healthy controls ($P < 0.0001$; Fig. 3B). At the best cut-off value, this panel of four-bacteria (*Fn*, *m7*, *Bc* and *Ch*) could discriminate colorectal cancer patients from healthy controls with a sensitivity of 77.7%, specificity of 81.5%, NPV of 81.1%, and PPV of 78.1%, showing a better diagnostic performance than *Fn*

only (Table 1). Pairwise comparison of AUROCs showed that the four-bacteria panel was superior to *Fn* alone for colorectal cancer diagnosis ($P = 0.05$).

The improved performance of four-bacteria was further validated in the second independent cohort. The combination of the four-bacteria also demonstrated an increased AUROC (0.756) as compared with *Fn* only (0.675) or the logistic regression model (0.746; Fig. 3C). The combined relative abundance of the

Table 1. Performance of *Fn* alone and in combination with other bacteria for colorectal cancer diagnosis

Variable	<i>Fn</i>	Combination of <i>Fn</i> , <i>Bc</i> , <i>Ch</i> , and <i>m7</i>
AUROC	0.868	0.886
Cut-off ^a	0.0007072	0.001774
Sensitivity	77.7%	77.7%
Specificity	79.5%	81.5%
PPV	76.3%	78.1%
NPV	80.7%	81.1%

Abbreviations: AUROC, area under receiver operating characteristics curve; NPV, negative predictive value; PPV, positive predictive value.

^aThe best cut-off values that maximize sensitivity and specificity were used. *n* = 370 (170 colorectal cancer and 200 healthy controls).

four-bacteria was significantly higher in colorectal cancer patients than in healthy controls ($P = 0.0002$; Fig. 3D). At the best cut-off value, this panel of bacteria could discriminate colorectal cancer from controls with a sensitivity of 84.9%, specificity of 61.1%, NPV of 81.5%, and PPV of 66.7%, which also shows a better diagnostic performance than *Fn* only. Therefore, the four-bacteria panel of *Fn*, *m7*, *Bc*, and *Ch* could improve the diagnostic ability of *Fn* alone in discriminating colorectal cancer from healthy controls.

The combination of bacterial markers with FIT improves the diagnostic ability of bacteria alone for colorectal cancer patients

FIT was performed on the stool samples of 111 colorectal cancer patients and 119 control subjects. We found that 70.3% (78/111) fecal samples of colorectal cancer patients showed FIT positive. The detection rate of FIT was less than the quantification of *Fn* alone (82.0%) or the four-bacteria panel (83.8%; both

$P < 0.05$ by χ^2) in this subcohort of colorectal cancer patients. Pairwise comparison of the ROC curves showed that the four-bacteria panel was significantly superior to *Fn* alone or FIT for colorectal cancer diagnosis (both $P < 0.05$; Supplementary Table S8). FIT was marginally associated with TNM staging ($P = 0.084$), while the relative abundances of the four-bacteria or *Fn* alone showed no correlation with TNM staging (Supplementary Table S9). Comparative results for the detection of cancer, according to TNM stage subsets, demonstrated that the quantification of bacterial markers showed significantly higher sensitivities compared with FIT for stage I cancer (Fig. 4). Elevated detection rates of stages II and III cancers were also observed by the bacteria than by FIT but not late stage IV. These results demonstrated that the quantification of bacterial markers was significantly more sensitive than FIT for the detection of colorectal cancer, especially for nonmetastatic colorectal cancer.

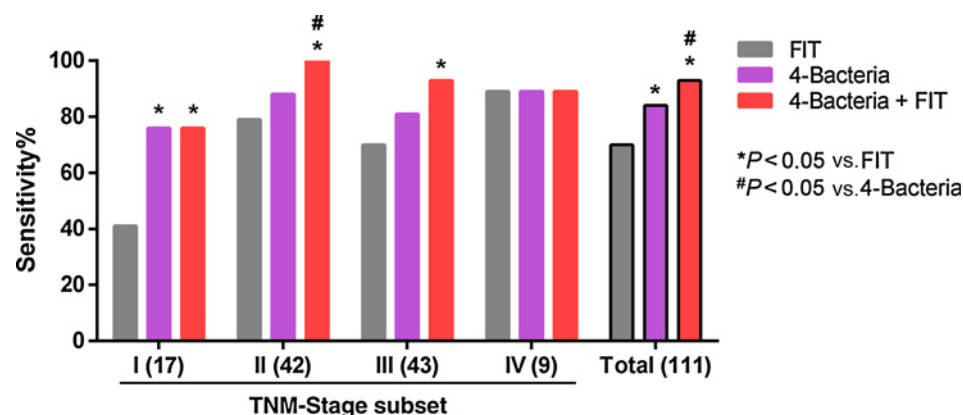
The combination of bacterial markers with FIT significantly increased the sensitivity of *Fn* from 82.0% to 92.8% and the four-bacteria from 83.8% to 92.8%, along with improved PPV and NPV and almost unchanged specificity (Table 2). According to TNM staging, combination of bacterial markers with FIT showed significantly higher sensitivities than using FIT only for stages I, II, and III cancers (Fig. 4). These results suggested that the combination of bacterial markers and FIT had the highest sensitivity and specificity for the noninvasive diagnostic value of patients with colorectal cancer.

Discussion

According to the most updated Asia Pacific consensus recommendations on colorectal cancer screening, FIT is applied to select high-risk patients for colonoscopy (18). FIT has also been widely

Figure 4.

Sensitivity of the commercial fecal immunochemical test (FIT) and bacterial markers according to tumor-node-metastasis (TNM) stage subsets. Shown are the sensitivities of FIT, 4-bacteria, and their combination for the detection of colorectal cancer according to tumor stage. The numbers in parentheses are the number of participants in each category.



TNM stage	Sensitivity		
	FIT	4-Bacteria	4-Bacteria + FIT
I	41.2%	76.5%	76.5%
II	78.6%	88.1%	100.0%
III	69.8%	81.4%	93.0%
IV	88.9%	88.9%	88.9%
Total	70.3%	83.8%	92.8%

Table 2. Performance of FIT or *Fn* alone and in combination with other bacteria for colorectal cancer diagnosis

Variables ^{a,b}	FIT	<i>Fn</i>	<i>Fn</i> + FIT	4-Bac	4-Bac + FIT
Sensitivity	70.3%	82.0%	92.8%	83.8%	92.8%
Specificity	98.3%	80.7%	79.8%	83.2%	81.5%
PPV	97.5%	79.8%	81.1%	82.3%	82.4%
NPV	78.0%	82.8%	92.2%	84.6%	92.4%

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

^a230 Subjects (111 colorectal cancer and 119 healthy controls) with FIT results in Hong Kong cohort were included.^b4-Bac includes *Fn*, *Bc*, *Ch*, and *m7*.

used in other regions of the world (19). However, the sensitivity of FIT shows limitations for colorectal cancer (0.79; 95% CI, 0.69–0.86) and differed greatly among various studies, according to a recent systematic review and meta-analysis by Lee and colleagues (19). Nevertheless, the wide application of FIT makes fecal samples easily obtainable. Detection of molecular biomarkers in fecal samples for the noninvasive diagnosis of colorectal cancer may be a more promising alternative than blood/plasma biomarkers to be implemented in present clinical settings. With the widespread application of pyrosequencing and metagenome sequencing in the field of microbiota, an increasing number of colorectal cancer-associated bacteria have been identified, including those identified by us (12). There is an urgent need to validate these candidate markers and to evaluate their clinical application values by targeted quantification methods.

To develop a convenient and reliable method for the targeted quantification of bacterial candidates on their validity and potential clinical implementation, we established a qPCR platform for the quantification in fecal samples. Different internal controls have been reported for qPCR-based quantification of bacterial abundances, including the *Bacteroides* genus (20), absolute DNA quantities (8), and 16S rDNAs (21–23). Considering that the proportion of *Bacteroides* varies among subjects of different enterotypes (24), and that the absolute DNA contained both bacterial DNAs and host DNAs, 16S rDNA could serve as a suitable internal control with conserved sequences uniformly distributed in most species. We thus designed a degenerate primer–probe set to guarantee sufficient coverage of 16S rDNAs by targeting the conservative fragments in bacterial 16S rDNA sequences (14). This internal control was proven to represent the bacterial DNA content in different samples. Then the probe-based duplex-qPCR assay allows the detection of both internal control and target in the same reaction for each sample, saving both reagents and samples, and producing more reliable data. Target marker abundance is calculated relative to total bacterial content by the ΔC_p method. We defined for the first time that DNA template concentration should be limited (<10 ng/μL) to avoid inhibitory effects caused by fecal DNA and >0.1 ng/μL to avoid false-negative assessments of the targets using our duplex qPCR assays. We further showed a good correlation in the quantification of bacterial candidates by metagenomics approach and qPCR assays. Therefore, our duplex-qPCR assays are reliable, convenient, and of great clinical application value in the quantitative detection of target bacteria.

Using this platform, we examined the clinical application values of 5 selected marker candidates (*Fn*, *Bc*, *Ri*, *Ch*, and *m7*); these marker candidates could be well quantitated by TaqMan probe-based duplex qPCR assays (Fig. 1), and their combination could improve the diagnostic performance of *Fn* according to metagenome sequencing data (Supplementary Fig. S3). We corroborated the potential value of *Fn* as a biomarker for the stool-

based diagnosis of colorectal cancer. The relative abundance of fecal *Fn* was significantly higher in colorectal cancer patients than in healthy control subjects. As a single factor in discriminating colorectal cancer patients from healthy subjects, *Fn* had a sensitivity of 77.7% and specificity of 79.5% in the first cohort of 170 colorectal cancer patients and 200 healthy control subjects. We also showed the significantly increased or decreased relative fecal abundances of *Bc*, *Ri*, *Ch* and *m7* in colorectal cancer patients than in control subjects, as consistent with metagenomics findings. Although the ability of these individual bacteria to discriminate colorectal cancer patients from healthy subjects was limited due to the limited occurrence rates in colorectal cancer patients or control subjects, we found that combining the relative abundances of *Bc*, *Ch*, and *m7* with that of *Fn* could improve the diagnostic ability of *Fn* for colorectal cancer. The relative abundance of *Ri* did not improve the diagnostic ability of *Fn* for colorectal cancer, showing a decreased AUROC of 0.846 by *Ri* in combination with *Fn* as compared with *Fn* alone (0.868), and was thus excluded in the further analyses. At the best cut-off value that maximized the sum of sensitivity and specificity, the combined four-bacteria panel had a sensitivity of 77.7% and specificity of 81.5% in the first cohort of 370 subjects. Importantly, *Fn* and the combination of four-bacteria markers (*Fn*, *Bc*, *Ch*, and *m7*) for the diagnosis of colorectal cancer was also verified in a second independent cohort of fecal samples of colorectal cancer patients and healthy controls. In particular, after adjustment for potential confounding factors including age and gender by multivariate logistic regression analyses, *Fn* and four-bacteria were found to be independent risk factors for colorectal cancer (Supplementary Table S10).

We have further examined the four-bacteria biomarker in 97 fecal samples from adenoma patients (60.5 ± 4.7 years; 50 males and 47 females). A significantly increased relative abundance of *Fn* was detected in adenoma patients as compared with healthy controls [7.1×10^{-5} (4.8×10^{-6} –0.0007) vs. 8.1×10^{-6} (0–0.0004), median (IQR); $P < 0.05$], which is consistent with the previous finding by Suehiro and colleagues in Japanese population (25). *Fn* alone gave an AUROC of 0.609 (0.545–0.674; $P < 0.05$) in discriminating adenoma patients from control subjects. Adding the other three bacteria did not significantly improve the predictive power of *Fn* for adenoma diagnosis, suggesting that the three newly identified bacterial markers (*Bc*, *Ch*, and *m7*) are specific for colorectal cancer. Host genetics and diets have been associated with the shaping of or changes in gut microbiome (26, 27). Gut microbes have been identified to be associated with obesity (28). As genetics, dietary habits and BMI vary among different populations, whether the bacterial markers verified in the two Asian cohorts in this study could be applied in other populations needs further investigation.

Compared with FIT, the bacterial markers were found to be superior in sensitivity for colorectal cancer diagnosis, especially for nonmetastatic colorectal cancer. It is intriguing that 16 and

15 samples in stage II and III respectively, showed positive in either bacterial markers or FIT (Supplementary Table S11), summing up to 36.5% of stage II and III cases. Together with the 60% cases (II: 26/42 and III: 25/43) showing positive in both bacterial markers and FIT, the combination of bacteria with FIT detected 96.5% of stages II and III colorectal cancer. It has been shown that metagenomic analysis combined with the standard fecal occult blood test (FOBT) improved colorectal cancer detection sensitivity (29). It is thus anticipated that the inclusion of the bacterial marker quantification assays, in the noninvasive diagnosis of colorectal cancer, with the widely applied FIT may improve diagnosis sensitivity.

Bc is a gram-negative, obligately anaerobic, non-spore-forming, rod-shaped bacterium species that was isolated from human feces in 2010 (30). *Ch* is a strictly anoxic, gram-positive, spore-forming, rod-shaped bacterium that participates in glucose metabolism using carbohydrates as fermentable substrates to produce acetate, ethanol, carbon dioxide, and hydrogen (31). Unlike the well-characterized *Fn*, which is known to promote colorectal cancer tumorigenesis, whether the altered abundances of *Ri*, *Bc*, or *m7* play a causative role in colorectal cancer development or serve as a consequence of colorectal cancer development needs further investigation.

In conclusion, the quantification of *Fn* alone can serve as a noninvasive diagnostic method for colorectal cancer with a moderate sensitivity and specificity. The combination of four bacterial markers (*Fn*, *Bc*, *Ch*, and *m7*) improved the diagnostic ability of *Fn* alone for colorectal cancer. Moreover, the combination of the bacterial markers and FIT showed the highest sensitivity and specificity for the diagnosis of colorectal cancer, especially for nonmetastatic colorectal cancer. Thus, stool-based detection of bacterial markers can serve as a novel noninvasive diagnostic method for patients with colorectal cancer.

References

- Sung JJ, Lau JY, Goh KL, Leung WK. Increasing incidence of colorectal cancer in Asia: implications for screening. *Lancet Oncol* 2005;6:871–6.
- Irrazabal T, Belcheva A, Girardin SE, Martin A, Philpott DJ. The multifaceted role of the intestinal microbiota in colon cancer. *Mol Cell* 2014;54:309–20.
- Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst* 2013;105:1907–11.
- Kassinen A, Krogius-Kurikka L, Makivuokko H, Rinttilä T, Paulin L, Corander J, et al. The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology* 2007;133:24–33.
- Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res* 2012;22:299–306.
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 2012;490:55–60.
- Shin NR, Lee JC, Lee HY, Kim MS, Whon TW, Lee MS, et al. An increase in the *Akkermansia* spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. *Gut* 2014;63:727–35.
- McCoy AN, Araujo-Perez F, Azcarate-Peril A, Yeh JJ, Sandler RS, Keku TO. *Fusobacterium* is associated with colorectal adenomas. *PLoS One* 2013;8:e53653.
- Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 2013;14:207–15.
- Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe* 2013;14:195–206.
- Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WK, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* 2015;6:8727.
- Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 2015 Sep 25. [Epub ahead of print].
- Jalanka J, Salonen A, Salojärvi J, Ritari J, Immonen O, Marćani L, et al. Effects of bowel cleansing on the intestinal microbiota. *Gut* 2015;64:1562–8.
- Wang Y, Qian PY. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* 2009;4:e7401.
- Huang Y, Li Q, Ge W, Cai S, Zhang S, Zheng S. Predictive power of quantitative and qualitative fecal immunochemical tests for hemoglobin in population screening for colorectal neoplasm. *Eur J Cancer Prev* 2014;23:27–34.
- Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
- Sung JJ, Ng SC, Chan FK, Chiu HM, Kim HS, Matsuda T, et al. An updated Asia Pacific Consensus Recommendations on colorectal cancer screening. *Gut* 2015;64:121–32.
- Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Ann Intern Med* 2014;160:171.

Disclosure of Potential Conflicts of Interest

F.K.L. Chan reports receiving speakers bureau honoraria from AstraZeneca, Eisai, Pfizer, and Takeda. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: Q. Liang, F.K.L. Chan, J.J.Y. Sung, J. Yu

Development of methodology: Q. Liang, J.J.Y. Sung, J. Yu

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Q. Liang, J. Chiu, Y. Chen, Y. Huang, A. Higashimori, J. Fang, H. Brim, H. Ashktorab, S.S.M. Ng, J. Yu

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Q. Liang, A. Higashimori, S.S.M. Ng, J.J.Y. Sung, J. Yu

Writing, review, and/or revision of the manuscript: Q. Liang, J. Chiu, J. Fang, H. Brim, H. Ashktorab, S.C. Ng, S.S.M. Ng, F.K.L. Chan, J. Yu

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Q. Liang, H. Ashktorab, S.C. Ng, S. Zheng, J. Yu

Study supervision: Q. Liang, S. Zheng, J. Yu

Grant Support

This project was supported by China 863 program fund (2012AA02A506), China 973 program fund (2013CB531401), National Nature and Science Foundation of China (81272304), The National Key Technology R&D Program (2014BAI09B05), Shenzhen Municipal Science and Technology R&D funding (JCYJ20130401151108652), and Shenzhen Virtual University Park Support Scheme to CUHK Shenzhen Research Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 24, 2016; revised September 1, 2016; accepted September 14, 2016; published OnlineFirst October 3, 2016.

Liang et al.

20. Flanagan L, Schmid J, Ebert M, Soucek P, Kunicka T, Liska V, et al. *Fusobacterium nucleatum* associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. *Eur J Clin Microbiol Infect Dis* 2014;33:1381–90.
21. Suzuki MT, Taylor LT, DeLong EF. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. *Appl Environ Microbiol* 2000;66:4605–14.
22. Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* 2002;148:257–66.
23. Ott SJ, Musfeldt M, Ullmann U, Hampe J, Schreiber S. Quantification of intestinal bacterial populations by real-time PCR with a universal primer set and minor groove binder probes: a global approach to the enteric flora. *J Clin Microbiol* 2004;42:2566–72.
24. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature* 2011;473:174–80.
25. Suehiro Y, Sakai K, Nishioka M, Hashimoto S, Takami T, Higaki S, et al. Highly sensitive stool DNA testing of *Fusobacterium nucleatum* as a marker for detection of colorectal tumours in a Japanese population. *Ann Clin Biochem*. 2016 Apr 28. [Epub ahead of print].
26. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. *Cell* 2014;159:789–99.
27. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 2011;334:105–8.
28. Walters WA, Xu Z, Knight R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett* 2014;588:4223–33.
29. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 2014;10:766.
30. Watanabe Y, Nagai F, Morotomi M, Sakon H, Tanaka R. *Bacteroides clarus* sp. nov., *Bacteroides fluxus* sp. nov. and *Bacteroides oleiciplenus* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2010;60:1864–9.
31. Steer T, Collins MD, Gibson GR, Hippe H, Lawson PA. *Clostridium hathewayi* sp. nov., from human faeces. *Syst Appl Microbiol* 2001; 24:353–7.

Clinical Cancer Research

Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer

Qiaoyi Liang, Jonathan Chiu, Yingxuan Chen, et al.

Clin Cancer Res 2017;23:2061-2070. Published OnlineFirst October 3, 2016.

Updated version	Access the most recent version of this article at: doi: 10.1158/1078-0432.CCR-16-1599
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2016/10/01/1078-0432.CCR-16-1599.DC1

Cited articles	This article cites 29 articles, 8 of which you can access for free at: http://clincancerres.aacrjournals.org/content/23/8/2061.full#ref-list-1
Citing articles	This article has been cited by 12 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/23/8/2061.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/23/8/2061 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.