Leveraging Fecal Microbial Markers to Improve the Diagnostic Accuracy of the Fecal Immunochemical Test for Advanced Colorectal Adenoma

Yuhan Zhang, MBBS1, Ming Lu, MBBS1, Bin Lu, MBBS1, Chengcheng Liu, MBBS1, Yiming Ma, PhD2, Li Liu, PhD3, Xiaoping Miao, PhD3, Junjie Qin, PhD4, Hongda Chen, PhD1 and Min Dai, PhD1

INTRODUCTION: Fecal immunochemical tests (FITs) detect colorectal adenoma inefficiently. The gut microbiota participates in colorectal cancer development. We aimed to explore fecal microbial signatures for advanced adenomas and evaluate their diagnostic value and complementary capacity to FIT.

METHODS: Using 16S rRNA sequencing, we studied gut microbiota in feces from 1,546 subjects in a screening setting, including 268 patients with advanced adenomas, 490 patients with nonadvanced adenomas, and 788 healthy subjects. Feature selections were performed using linear discriminant analysis effect size, multivariate association with linear models, and least absolute shrinkage and selection operator. The diagnostic performance of microbial signatures and their auxiliary role to FITs and the added value of the Asia-Pacific Colorectal Screening score were evaluated. We applied 0.632+ bootstrapping to adjust the potential overfitting.

RESULTS: We identified 13 microbial signatures to show the joint diagnostic value for advanced adenoma, with genus *Tyzzerella* 4 demonstrating the highest adjusted area under the curve (AUC) of 0.545 (95% confidence interval [CI], 0.520–0.610). The 13-bacteria increased the adjusted AUC to 0.607 (95% CI, 0.548–0.660). Compared with individual FIT (adjusted AUC = 0.527; 95% CI, 0.519–0.571), 13-bacteria and FITs collectively reached an adjusted AUC of 0.641 (95% CI, 0.579–0.691). At cutoff values yielding specificities of 90% and 80%, the adjusted sensitivities were 28.4% (95% CI, 19.3–36.8) and 41.1% (95% CI, 29.9–49.4), respectively. The Asia-Pacific Colorectal Screening score further boosted the adjusted AUC to 0.706 (95% CI, 0.648–0.750).

DISCUSSION: In this study using fecal samples from a screening setting, the identified microbial signatures could complement FITs for detecting advanced adenomas. Gut microbiota can act as a promising tool to optimize the current colorectal cancer screening modalities.

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/CTG/A654.

Clinical and Translational Gastroenterology 2021;12:e00389. https://doi.org/10.14309/ctg.0000000000000389

INTRODUCTION

According to the GLOBOCAN 2018 (1), colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second leading cause of cancer deaths worldwide. The development of most sporadic CRC cases follows the adenoma-carcinoma sequence, and the progression from colorectal adenoma to invasive cancer typically takes 5–10 years, leaving a large enough time window for early detection and treatment (2). Because of the continuously increasing coverage of CRC screening, a reduction in the incidence and mortality of CRC has been observed in some countries, including the United States (3–5). However, the current screening modalities, including colonoscopy and the fecal immunochemical test (FIT), have several limitations, such as a poor compliance rate and high cost or limited diagnostic

1Office of Cancer Screening, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; 2State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China; 3Department of Epidemiology and Biostatistics, and the Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; 4Promegene Institute, Shenzhen, China. Correspondence: Min Dai, PhD. E-mail: daimin2002@hotmail.com or Hongda Chen, PhD. E-mail: hongda.chen@cicams.ac.cn.

Received January 22, 2021; accepted July 8, 2021; published online August 19, 2021

© 2021 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of The American College of Gastroenterology
performance in detecting advanced adenomas (6,7). Therefore, exploration of novel noninvasive tests for early detection of CRC, especially preneoplastic lesions, is urgently required.

Previous in vivo and in vitro experiments as well as epidemiological studies have demonstrated that alterations in the gut microbial environment are closely associated with colorectal tumorigenesis (8,9). Furthermore, a series of fecal microbial markers have shown significant potential for the early detection of CRC (10–12). For instance, in a hospital–based case-control study by Liang et al. (10) that included 203 CRC cases and 236 healthy controls, it was demonstrated that Fusobacterium nucleatum (Fn) had good discriminative power for detecting CRC with an area under the curve (AUC) of 0.868, which could be further enhanced by combining with 3 other bacteria, yielding an AUC of 0.886. However, fecal microbial biomarkers for detection of colorectal adenomas have been less explored. In addition, in terms of translational application, such significant findings should be prospectively validated in samples collected from asymptomatic individuals, who represent the target screening population. However, this was not common in previous studies (13–15). Moreover, further improvement of the diagnostic performance by combination of fecal microbial markers and the FIT, together with the established risk stratification score of CRC, was highly anticipated.

In this study, we conducted 16S rRNA sequencing of 1,546 fecal samples (including 268 advanced adenoma, 490 nonadvanced adenomas, and 788 controls) obtained from an ongoing multicenter population-based CRC screening trial. By conducting a multistep selection, we aimed to explore and construct a fecal microbial signature panel that is associated with colorectal adenoma and further explore its auxiliary role in complementing FIT for improving the diagnostic performance of CRC screening for detecting advanced adenomas.

MATERIALS AND METHODS

Subject inclusion

Subjects in this study were selected from the baseline phase of the TARGET-C study, an ongoing randomized controlled trial comparing the effectiveness of colonoscopy, FIT, and risk-adapted screening strategies for CRC in China. Additional details of the study protocol have been reported previously (16,17). In brief, residents aged 50–74 years in the study areas were recruited, with the exclusion of the ineligible subjects according to the study protocol. After signing the informed consent, eligible participants were randomly assigned into 3 groups to undergo 1-time colonoscopy, annual FIT, and risk-adapted screening every year. All abnormal findings during colonoscopy were checked under standard clinical procedures and confirmed by pathological evaluations. Final clinical diagnoses were determined according to the most advanced finding. High-grade dysplasia, adenomas with villous or tubular-villous histologic features, or adenomas greater than 1 cm were classified as advanced adenomas. Healthy controls were defined as subjects with findings of no adenomas, including the detection of hyperplastic polyps or no polyps. This study was approved by the Ethics Committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College (18-013/1615).

In total, 29 CRCs, 283 advanced adenomas, and 740 nonadvanced adenomas were detected during the baseline screening of the TARGET-C study between May 1, 2018, and April 30, 2019. In accordance with the study design, participants undergoing colonoscopy were required to collect stool samples within 24 hours before bowel preparation for colonoscopy. After excluding subjects with no available specimens, fecal samples from all 29 CRC cases, all 270 patients with advanced adenoma, randomly selected 494 nonadvanced adenoma carriers, and 789 healthy controls were subjected to 16S rRNA sequencing. During the quality control process of sequencing data, 7 samples with unqualified data were excluded. A detailed flow diagram of the sample selection process was shown in Figure 1.

Fecal sample collection and processing

Participants undergoing colonoscopy were instructed to collect 2 stool samples at home before bowel preparation for colonoscopy within 24 hours. One of the collected raw stool samples was kept in a stool-filled container, then packaged in an insulated box with ice packs, and brought to the clinical site on the day of the colonoscopy. On receipt, the fecal samples were frozen at −80 °C and then transported by a cold chain to the central biobanks for further research. The other sample was collected using the FIT sample collection device for microbiome analysis. Existing evidence suggests that feces collected by the fecal occult blood test devices are stable at room temperature and are suitable for microbiota studies (18). After placed in a storage box, the stool-filled containers were delivered to a central laboratory and immediately frozen at −20 °C until DNA extraction. For this study, we used fecal samples collected in the FIT tubes for 16S rRNA sequencing.

DNA extraction and 16S rRNA gene sequencing

According to the manufacturer’s instructions, DNA was extracted using the QIAamp Fast DNA Stool Mini Kit (Qiagen, Hilden, Germany). The V4 region of the microbial 16S rRNA gene was amplified and sequenced on the Illumina MiSeq sequencing platform (Illumina, San Diego, CA). Afterward, 16S rRNA amplicon sequences were processed based on QIIME2 (19). To avoid end-read sequencing errors, all reads were truncated at the 150th base and the median Q score > 20. Noisy sequences, chimeric sequences, and singletons were removed, and then, amplicon sequence variants (ASVs) were inferred from the sequencing data using the DADA2 pipeline (20). Based on the Greengenes 13.8 database, the classifications for ASVs were identified using the classify-sklern classification methods through the q2-feature-classifier plugin (https://data.qiime2.org/2018.11/common/gg-13-8-99-515-806-nb-classifier.qza). To quantify the taxonomic composition, all sequences were rarefied to an even sampling depth of 10,000. Only the taxa and taxa present in at least 5% of the samples and with an average relative abundance greater than 0.01% were included in the downstream analyses. The results of relative abundances of the microbial markers were characterized as continuous variables in subsequent analyses.

Fecal immunochemical test

Frozen stool aliquots were used for the FIT, and hemoglobin concentrations were tested using FIT (OC-Sensor; Eiken Chemical, Tokyo, Japan) following a standard operating procedure. First, the fecal samples were thawed at room temperature (20 °C–25 °C) for 5 minutes. Then, 1-g stool in total from each sample was separated from 3 detached sites and transferred into an empty tube. Defrosted fecal samples were mixed and dipped using a collection stick. Subsequently, the stool-filled stick was inserted back into the OC-Sensor test tube, which was manually shaken 10
times and left overnight at 4 °C. All stool samples were tested simultaneously following the manufacturer's instructions the next day. The laboratory staff were blinded to the colonoscopy results. Quantitative FIT results were characterized as a dichotomous variable with a threshold of 20-µg Hb/g feces in subsequent analyses.

Asia-Pacific Colorectal Screening Score
The Asia-Pacific Colorectal Screening (APCS) score is an established risk stratification score for selecting high-risk populations suitable for CRC screening and comprises age, sex, family history of CRC, and smoking (21). A modified APCS risk score with body mass index (BMI) incorporated into the predictors has been validated to improve the risk prediction of advanced neoplasia (22).

In this study, the modified APCS risk score was calculated, which ranged from 0 to 6 and included the following risk factors: age (50–54 years, 0; 55–64 years, 1; and 65–74 years, 2), sex (female, 0; male, 1), family history of CRC among first-degree relatives (absent, 0; present, 1), smoking (never smokers, 0; current or past smokers, 1), and BMI (<23 kg/m², 0; ≥23 kg/m², 1). Detailed information has been reported previously (17). The APCS score was characterized as a continuous variable in subsequent analyses.

Microbial diversity and composition analysis
Microbial diversity was calculated using the sequencing data at the ASV level and presented by multiple alpha-diversity indices calculated using the R package vegan, including richness, Shannon index, Simpson index, Chao1 index, pielou index, and faith_pd index. Principal coordinates analysis was conducted to display the microbiome space between samples by adopting the Bray-Curtis distance matrix, also known as β-diversity. Compositional comparisons were performed using permutational multivariate analysis of variance (the adonis function in the vegan R package).

Microbial feature selection and statistical analysis
Based on the normalized relative abundance matrix, microbial features with significant differences between the advanced adenoma group and the healthy control group were selected in the following steps: (i) generating microbial feature candidates identified by the linear discriminant analysis (LDA) effect size (LEfSe) method (23) (http://huttenhower.sph.harvard.edu/galaxy/), which was embedded with the Kruskal-Wallis rank-sum test to detect features with significantly distinct abundance in regard to the status of interest ($P < 0.05$), and LDA score to assess the effect size of each feature (LDA score ≥ 2 as the cutoff value); (ii) generating microbial feature candidates identified using the multivariate association with linear models (MaAsLin) method with the purpose of adjusting for covariates, including region, sex, age, BMI, smoking, alcohol consumption, and physical activities, and subjected to false discovery rate correction (Q-value < 0.25 as the cutoff value); and (iii) for the union set of the microbial features selected by the 2 before-mentioned methods, applying the least absolute shrinkage and selection operator (LASSO) binary regression analysis to identify features with significant changes in abundance between the advanced adenoma group and the healthy control group.
logistic regression procedure to select variables. MaAsLin algorithm was performed in R software using the Maaslin2 package (v.1.0.0), and the LASSO regression procedure was implemented in R software using the glmnet package (v. 4.0-2).

Statistical analyses were performed using R software (v.3.6.3). The differences in the distribution of categorical factors among study groups were evaluated using the χ² test, while the Kruskal-Wallis H test and Wilcoxon test were used to examine the differences of continuous variables among groups. For the identified microbial markers and panel, and their combination with FIT and APCS scores, logistic regression models were used to construct the prediction models. The diagnostic performance of each model was evaluated using the receiver operating characteristic curve, area under the receiver operating characteristic, and sensitivity at cutoffs yielding 80% and 90% specificity. The corresponding 95% confidence intervals (CIs) were calculated based on 1,000 bootstrap samples. To adjust for the potential overestimation of the directly calculated estimators, the 0.632+ bootstrap method with 1,000 replicates was applied (24). The differences in apparent diagnostic indicators between models were compared using the Delong test (25). All apparent diagnostic indicator calculations were performed using the pROC package (v.1.16.2), while the 0.632+ bootstrap adjusted values were calculated based on the Daim package (v.1.1.0).

RESULTS

Study population characteristics
Because of the limited number of CRC cases (n = 29) and our main interest in adenomas, we finally included 16S rRNA sequencing data and metadata from 1,546 samples in the primary analyses, including 268 patients with advanced adenoma (mean age, 61.7 ± 6.2 years; 197 men and 71 women), 490 nonadvanced adenoma carriers (mean age, 61.4 ± 6.3 years; 317 men and 173 women), and 788 healthy controls (mean age, 59.9 ± 6.2 years; 378 men and 410 women). More details of sociodemographic characteristics in each group were shown in Supplementary Table S1 (Supplementary Digital Content 1, http://links.lww.com/CTG/A654).

Gut microbial diversity alterations
No significant differences in the alpha-diversity of microbiota were observed among the 3 groups (advanced adenomas, non-advanced adenomas, and healthy controls) using richness, Chao 1, Shannon, Simpson, pieou, and faith_pd indices. Likewise, no remarkable differences were detected in pairwise comparisons (Figure 2a). At the ASV level, principal coordinates analysis was performed at the Bray-Curtis distance, focusing on the compositional comparison of microbial communities. No significant bacterial compositional differences between the groups were identified by permutational multivariate analysis of variance. Nevertheless, the first 2 principal coordinates among the patients with advanced adenoma, nonadvanced adenoma carriers, and healthy controls differed significantly (Figure 2b, P < 0.05).

Identification of fecal microbial markers for detecting advanced adenomas
We mainly focused on advanced adenomas with respect to the exploration of potential microbial markers to facilitate early detection of precancerous lesions. The LEfSe analysis result (LDA score > 2) provided a picture of the microbiota alteration between patients with advanced adenomas and healthy controls, characterized by higher relative abundances of several taxa in patients with advanced adenomas (Figure 3a), including genus Fusobacterium, an unnamed species under it and its related taxa at higher levels; genus Tyzzerella 4 and its unnamed species; genus Phascolarctobacterium and related taxa at the higher level; an unnamed species under genus Clostridium sensu stricto 1; an unnamed species under genus Streptococcus; genus Gemella, an unnamed species under it and its related taxa at the higher level; genus Actinomyces, an unnamed species under it and its related taxa at higher levels; and genus Terrisporobacter, an unnamed species under it and its related taxa at the higher level. However, some taxa were more abundant in the healthy control group (Figure 3a), including genus Faecalibacterium and an unnamed species under it; species Blautia masiliensis; an unnamed species under genus Lachnospira; species Bacteroides thetaiotaomicron; Ruminococcaceae UCG 002 and an unnamed species under it; genus Lachnospiraceae UCG 004 and an unnamed species under it; an unnamed genus under family Rumminococcaceae; genus Odoribacter and its related species and family; species Bifidobacterium bifidum; and an unnamed species under genus Ruminococcaceae NK4A214 group.

Taking confounders influences into consideration, MaAsLin analysis showed that genus Fusobacterium and its related taxa; genus Gemella and its related taxa; an unnamed species under genus Streptococcus; and an unnamed species under genus Clostridium sensu stricto 1 were still abundant in the advanced adenoma group. In addition, an unnamed species under genus Rothia, genus Streptococcus and the higher family, genus Clostridium sensu stricto 1 and another unnamed species under it, and another unnamed species under genus Fusobacterium were also identified by MaAsLin analysis (Figure 3b).

We then performed LASSO regression for further selection of the most useful predictive features from the union set consisting of feature candidates identified using LEfSe and MaAsLin analyses to avoid missing out on valuable microbial markers. Of feature candidates, a total of 49 microbial features (see Supplementary Table S2, Supplementary Digital Content 1, http://links.lww.com/CTG/A654) were reduced to 13 predictors on the basis of 268 patients with advanced adenoma and 788 healthy controls (see Supplementary Figure S1, Supplementary Digital Content 1, http://links.lww.com/CTG/A654), including genus Tyzzerella 4, genus Gemella, an unnamed species under genus Faecalibacterium, genus Lachnospiraceae UCG 004, an unnamed species under family Ruminococcaceae, genus Clostridium sensu stricto 1, genus Streptococcus, an unnamed species under genus Clostridium sensu stricto 1, an unnamed species under genus Lachnospira, an unnamed species under genus Fusobacterium, an unnamed species under genus Rothia, species Bifidobacterium bifidum, and an unnamed species under genus Fusobacterium.

Diagnostic performance of microbial markers
The apparent AUCs of 13 candidate microbial biomarkers in detecting advanced adenoma ranged from 0.520 (95% CI, 0.494–0.547) to 0.578 (95% CI, 0.542–0.614) with genus Tyzzerella 4 performing best (AUC = 0.578; 95% CI, 0.542–0.614; see Supplementary Table S3, Supplementary Digital Content 1, http://links.lww.com/CTG/A654). Through the 0.632+ bootstrap method, the adjusted AUCs ranged from 0.503 (95% CI, 0.483–0.548) to 0.545 (95% CI, 0.52–0.61), also with genus Tyzzerella 4 performing best (AUC = 0.545; 95% CI, 0.52–0.61; Supplementary Table S3, Supplementary Digital Content 1, http://links.lww.com/CTG/A654 and Figure 4a). A logistic regression model with inclusion of 13 candidate biomarkers
demonstrated a remarkable improvement in the diagnostic performance (apparent AUC = 0.641; 95% CI, 0.601–0.681; adjusted AUC = 0.607; 95% CI, 0.548–0.660; Figures 4b and 4c).

To determine whether microbial biomarkers could be used to improve the diagnostic performance of FIT for advanced adenoma, we calculated the AUC of microbiota-FIT–based and only FIT-based logistic regression models, respectively, and compared their diagnostic performance. Owing to the unavailable FIT results of 76 subjects, we included 248 patients with advanced adenoma and 732 healthy controls in the subsequent comparisons. Compared with individual FIT (apparent AUC = 0.544; 95% CI, 0.525–0.564; adjusted AUC = 0.527; 95% CI, 0.519–0.571; Figures 4b and 4c), a significant improvement in the AUC was observed for the combination of 13 microbial markers and FIT (apparent AUC = 0.671; 95% CI, 0.631–0.711; adjusted AUC = 0.641; 95% CI, 0.579–0.691; P < 0.001; Figures 4b and 4c). At the recommended cutoff value suggested by the manufacturer, the sensitivity and specificity of FIT for detecting advanced adenoma were 10.5% and 98.4%, respectively. For the combination of FIT and 13 microbial markers, when setting specificity of 80%, the apparent and adjusted sensitivities were 45.2% (95% CI, 38.3–52.0) and 41.1% (95% CI, 29.9–49.4), respectively.

In this study, we further used the APCS score, a risk assessment tool integrating 5 commonly recognized and easily collected CRC-related risk factors, for additional improvement of the diagnostic performance. Results showed that the APCS score could further improve the AUC of the bacteria-FIT union for advanced adenomas from 0.671 (95% CI, 0.631–0.711) to 0.727 (95% CI, 0.691–0.763; adjusted AUC = 0.706; 95% CI, 0.648–0.750; P < 0.001; Figures 4b and 4c). At cutoff values yielding 90% and 80% specificities, the higher sensitivity was observed for the tripartite combination of FIT, 13 microbial markers, and APCS score, but was not of statistical significance (Figure 4c). Similarly, additional microbial features panel could further promote the diagnostic performance of the combined FIT and the APCS score. The microbial features panel improved the AUC for advanced adenoma from 0.674 (95% CI, 0.637–0.712) to 0.727 (95% CI, 0.691–0.763; P < 0.001), with adjusted AUCs from 0.661 (95% CI, 0.621–0.728) to 0.706 (95% CI, 0.648–0.750; Figures 4b and 4c). Significant promotions of sensitivities at cutoff values yielding 90% and 80% were observed.

Figure 2. The shift of gut microbiota in patients with advanced adenoma, nonadvanced adenoma, and healthy controls at the amplicon sequence variant level. (a) Alpha (α)-diversity in patients with advanced adenoma, nonadvanced adenoma, and healthy controls. No significant diversity difference was found using 6 diversity indexes (richness, chao 1, shannon, simpson, pielou, and faith pd). Boxplots and dot plots were both presented. The horizontal lines in the boxplots represent median values; upper and lower ranges of the box represent the 75% and 25% quartiles. (b) PCoA of the microbiota based on the Bray-Curtis distance. ANOSIM, R² = 0.0014, P = 0.051. AA, patients with advanced adenoma; bc_dist, Bray-Curtis distance; HC, healthy controls; NAA, patients with nonadvanced adenoma; PCoA, principal coordinates analysis.
DISCUSSION

In this study, we performed the microbial feature selection using prospectively collected fecal samples of 268 patients with advanced adenoma and 788 healthy controls from a population-based CRC screening program. In total, we identified 13 fecal microbial markers at the genus and species level with good potential for detecting advanced adenoma. Combined with microbial markers, the diagnostic performance improved notably than FIT solely. In addition, commonly recognized CRC risk factors could further boost the diagnostic value for advanced adenoma diagnosis. These results confirmed the complementary role of gut microbiota in differentiating healthy individuals from those with colonic lesions. Findings of our study filled the current research gap owing to study population limitation and demonstrated the feasibility of using fecal microbiota markers as a complementary tool to the established CRC screening modalities, such as FITs.

Figure 3. LEfSe and MaAsLin analyses revealed differences in taxonomic composition of patients with advanced adenoma compared with healthy controls. (a) Differentially abundant taxa identified using LEfSe analysis with LDA scores showing significant bacterial differences between patients with advanced adenoma (red) and healthy controls (green). (b) Differentially abundant taxa identified using MaAsLin analysis adjusting sex, age, region, body mass index, smoking, and alcohol consumption, with Q-value lower than 0.25 presented. AA, patients with advanced adenoma; HC, healthy controls; LDA, linear discriminant analysis; LEfSe, linear discriminant analysis effect size; MaAsLin, multivariate association with linear models.

Of the selected microbial features, Tyzzerella 4 exhibited the highest diagnostic value for detecting advanced adenomas and showed its promising diagnostic potential for early detection of advanced adenoma for the first time. Evidence from quantitative gut microbiota data sets of Homo sapiens suggested its higher abundance in patients diagnosed with Crohn’s disease (26,27). Considering that gut microbiota may be dominantly shaped by geographical locations (28) and feature selection results originating from LEfSe without adjusting covariates were of risk of obtaining false positives, we further applied separated LEfSe analysis among the included participants stratified by the 5 residential areas. We systematically reviewed remarkably altered microbial markers at the genus and species levels in patients with advanced adenoma residing in more than 2 regions (see Supplementary Table S4, Supplementary Digital Content 1, http://links.lww.com/CTG/A654). It is noteworthy that genus Tyzzerella 4 and its members were identified among populations of 3 different regions, covering 228 patients with advanced adenoma and 669 healthy controls, thus achieving a cross-regional validation of its potential association with advanced adenoma in populations from East China and Central China. Also, MaAsLin algorithm identified the potential association between genus Tyzzerella 4 and advanced adenoma after adjusting for region and other confounders.

Apart from genus Tyzzerella 4, previous evidence has offered support for other selected microbial markers’ participation in the initiation and progression of colorectal mucosa tumorigenesis, as well. Gemella, recognized as 1 of the predominant genera in the upper gastrointestinal tract, was almost absent in the lower gastrointestinal tract of healthy individuals (29). Researchers have reported the enrichment of Gemella morbillorum, a member of the genus Gemella, in CRC microbiomes based on stool sample analysis and its potential role in colorectal carcinogenesis promotion (30,31). Clostridium could promote colorectal carcinogenesis by increasing aberrant crypt foci induced by 1,2-dimethylhydrazine (8). The cluster I in the 16S rRNA tree was regarded as the true representatives of the genus Clostridium (i.e., Clostridium sensu stricto) (32). The risk of CRC was observed increasing in patients with bacteremia from the species Clostridium perfringens under genus Clostridium sensu stricto (31). As for genera Streptococcus and Rothia, several studies have reported the associations between CRC and these oral originating microbiota (33). Fusobacterium adhesin A from species of Fusobacterium may contribute to the activation of procarcinogenic signaling pathways and ultimately lead to molecular changes and colonic mucosa carcinogenesis (8,34).
Probiosics were observed in association with CRC (35,36), which was prevalent in the gut microbiota of healthy adults (37,38). Faecalibacterium prausnitzii was the only known species under genus Faecalibacterium (39), and findings from experiments in vitro and animal models reported its ability to produce anti-inflammatory metabolites (40–42), and its diagnostic value was confirmed as well (43). Lachnospiraceae UCG-004 is an unclassified genus under the highly polyphyletic family Lachnospiraceae, which possesses the predominant production of butyrate, a short-chain fatty acid with anti-inflammatory and antitumorigenic properties (44,45). Members of the Lachnospiraceae family have been shown to be specific to the fecal microbial communities of patients with preneoplastic lesions (46). Researchers also observed depleted abundance of members from family Ruminococcaceae among gut microbiome of conventional adenoma cases (47). Results from the mouse model suggested that Bifidobacterium bifidum had a beneficial effect on CRC (48,49).

Results in this study, together with the existing evidence, suggested the promising potential of fecal microbial signatures as a novel noninvasive tool for CRC screening. However, multiple necessary steps are required for their translation into screening products, including robust sequencing techniques having high reproducibility and easy-to-apply platform, prospective validation in samples from a true screening setting. In our study, although genus Tyzzerella 4 performed best individually among the 13 identified microbial features, its underlying biological mechanism and reproducibility among diverse populations have yet to be investigated in depth. In agreement with previous studies (8), the fact herein that model performance could be promoted by panels of taxa compared with individual ones could be partly explained by the mutual complement of diagnostic performance to avoid missing cases (10,15). Moreover, meta-analyses from geographically diverse metagenomes suggested that poly-microbial classifiers were globally applicable and robust against technical and geographical differences (50).

Currently, stool-based products consisting of multiple targets are the focus of most attention. In addition to the integration of multiple microbial features, the combination of diverse types of stool-based tools exemplifies the advantages of a multitarget CRC test in which individual items can complement each other to reduce missed cases. The addition of microbial signatures to FIT can further strengthen its diagnostic performance than FIT alone (8,10,13,51), especially for colorectal adenoma detection (13–15). Similarly, in view of the fact that the multitarget stool DNA test could detect CRCs with a sensitivity of 92.3%, but was limited by a suboptimal sensitivity of 42.4% for advanced adenomas (52), the auxiliary role of microbial signatures in the feces for colorectal neoplasms deserved further investigation. Moreover, our study showed that the overall diagnostic performance of the combination of FIT and microbial signature could be potentially enhanced if combining an established questionnaire-based risk prediction model (the APCS score) according to the elevated AUCs. Furthermore, at the cutoff values yielding 70% specificity, for instance, the sensitivity increased from 52.0% (95% CI, 45.6–58.9) for the combination of FIT and microbial markers to 62.9% (95% CI, 55.2–70.2; P = 0.003) for the combination of FIT, microbial markers, and the APCS score in our study (data not shown).
shown in the results). However, in true screening setting, the determination of the optimal cutoff should be considered depending on the targeted sensitivity, specificity, and positivity rates.

Our study has several strengths. In contrast to hospital-based case-control studies, participants in this study were recruited from a practical screening program targeting asymptomatic individuals with average risk. Stool samples in this study were collected in a manner compatible with the population screening practice. The large sample size of our cases and controls assures high reliability of our results. In addition, we successfully identified a novel microbial marker of considerable potential in the Chinese population for ensuring more effective detection of advanced adenomas. Despite our best efforts, this study has some limitations. The underlying molecular mechanisms and possible signaling pathways involved in the association between Tyzzerella 4 and advanced adenoma require further exploration. Simultaneous investigation of mucosal samples may contribute to a better understanding of roles of the gut microbiota in CRC carcinogenesis. In addition, with the inherent properties of 16S rRNA sequencing in annotation resolution, systematic investigation of the key bacterial species or strains by metagenomic sequencing based on the current findings may further improve the diagnostic value of the microbiota markers for advanced adenoma. Furthermore, targeted quantification studies of the candidate microbial markers, such as quantitative polymerase chain reaction, and prospective validation in larger populations can accelerate their introduction to population-based screening practices as novel noninvasive tools for the early diagnosis of colorectal precancerous lesions.

In summary, we identified a novel genus of gut microbiota (Tyzzerella 4) potentially associated with the colorectal mucosa carcinogenesis process. Both individual Tyzzerella 4 and 13-bacteria panel exhibited good discriminative ability for the early detection of advanced adenoma. Combining the microbial panel with FIT and an established risk stratification score (APCS) may further enhance the diagnostic performance. Additional evidence from mechanistic studies and microbiome quantitative analyses is required. Altogether, it can be foreseen that the identified fecal microbial biomarkers in our study may contribute to the development of novel, effective, and noninvasive CRC screening tools in the near future.

CONFLICTS OF INTEREST
Guarantor of the article: Yuhan Zhang, MBBS, Hongda Chen, PhD, and Min Dai, PhD.
Specific author contributions: H.C. and M.D. conceptualized and designed the study. Y.Z., M.L., B.L., Y.M., L.L., X.M., J.Q., H.C., and M.D. participated in acquisition of data and analysis and interpretation; manuscript preparation, review, or approval; and the decision to submit the manuscript for publication. Potential competing interests: None to report.

Study Highlights

**WHAT IS KNOWN**

- Colorectal cancer (CRC) is a leading cause of cancer incidence and mortality globally. Early detection and treatment lead to the decreased incidence and increased survival.
- The widely used fecal immunochemical tests (FITs) were less efficient to detect colorectal adenomas.
- Gut microbiota participates in CRC development and has a promising potential for colorectal neoplasm detection.

**WHAT IS NEW HERE**

- We identified 13 microbial signatures to show the joint diagnostic value for advanced adenomas, with genus *Tyzzerella 4* demonstrating the highest adjusted area under the curve.
- The 13-bacteria panel improved the diagnostic performance of FIT for advanced adenoma detection.
- Commonly recognized CRC risk factors could further facilitate the detection of advanced adenoma on the basis of FITs and microbial markers.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394–424.
2. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet 2014;383(9927):1490–502.
3. Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: Systematic review and meta-analysis of randomised controlled trials and observational studies. BMJ 2014;348:g2467.
4. Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Screening for colorectal cancer: US Preventive Services Task Force Recommendation Statement. JAMA 2016;315(23):2564–75.
5. Wolf AMD, Fontham ETH, Church TR, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin 2018;68(4):250–81.
6. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. Ann Intern Med 2009;150(3):162–9.
7. Lee JK, Liles EG, Bent S, et al. Accuracy of fecal immunochemical tests for colorectal cancer: Systematic review and meta-analysis. Ann Intern Med 2014;160(3):171.
8. Wong SH, Yu J. Gut microbiota in colorectal cancer: Mechanisms of action and clinical applications. Nat Rev Gastroenterol Hepatol 2019;16(11):690–704.
9. Song M, Chan AT, Sun J. Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. Gastroenterology 2020;158(2):322–40.
10. Liang Q, Chiu J, Chen Y, et al. Fecal bacteria act as novel biomarkers for noninvasive diagnosis of colorectal cancer. Clin Cancer Res 2017;23(8):2061–70.
11. Shah MS, DeSantis TZ, Weinmaier T, et al. Leveraging sequence-based fecal microbial community survey data to identify a composite biomarker for colorectal cancer. Gut 2018;67(5):882–91.
12. Dai Z, Coker OO, Nakatsu G, et al. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. Microbiome 2018;6(1):70.
13. Baxter NT, Ruffin MT, Rogers MAM, et al. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med 2016;8(1):37.
14. Wong SH, Kwong TNY, Chow TC, et al. Quantification of faecal Fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. Gut 2017;66(8):1441–8.

15. Liang JQ, Li T, Nakatsu G, et al. A novel faecal Lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut 2020;69(7):1248–57.

16. Chen H, Li N, Shi J, et al. Comparative evaluation of novel screening strategies for colorectal cancer screening in China (TARGET-C): A study protocol for a multicentre randomised controlled trial. BMJ open 2019;9(4):e025935.

17. Chen H, Lu M, Liu C, et al. Comparative evaluation of participation and diagnostic yield of colonoscopy vs faecal immunochromatographic test vs risk-adapted screening in colorectal cancer screening: Interim analysis of a multicenter randomized controlled trial (TARGET-C). Am J Gastroenterol 2020;115(8):1264–74.

18. Vogtmann E, Chen J, Kibriya MG, et al. Comparison of fecal collection methods for microbiota studies in Bangladesh. Appl Environ Microbiol 2017;83(10):e00361–17.

19. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37(8):852–7.

20. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13(7):581–3.

21. Yeoh KG, Ho KY, Chiu HM, et al. The Asia-Pacific Colorectal Screening score: A validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects. Gut 2011;60(9):1236–41.

22. Sung JYY, Wong MCS, Lam YTY, et al. A modified colorectal screening score for prediction of advanced neoplasia: A prospective study of 5744 subjects. J Gastroenterol Hepatol 2018;33(1):187–94.

23. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;12(6):R60.

24. Efron B, Tibshirani R. Improvements on cross-validation: The .632+ bootstrap method. J Am Stat Assoc 1997;92(438):580–60.

25. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under 2 or more correlated receiver operating characteristic curves: A nonparametric approach. Biomometrics 1988;44(3):837–45.

26. Olaisen M, Flatberg A, Granlund AVB, et al. Bacterial mucosa-associated microbiome in inflamed and proximal noninflammatory ileum of patients with Crohn's disease. Inflamm Bowel Dis 2021;27(1):12.

27. Zhang Q, Yu K, Li S, et al. gutMEGA: A database of the human gut microbiome studies of human disease. Nature 2020;587(7834):448–54.

28. Vasapoli R, Schütte K, Schulz C, et al. Analysis of transcriptionally active bacteria throughout the gastrointestinal tract of healthy individuals. Gut microbiota studies of human disease. Nature 2020;587(7834):448–54.

29. Vasapoli R, Schütte K, Schulz C, et al. Analysis of transcriptionally active bacteria throughout the gastrointestinal tract of healthy individuals. Gut microbiota studies of human disease. Nature 2020;587(7834):448–54.

30. Yu J, Feng Q, Wong SH, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 2017;66(1):70–8.

31. Kwong TNY, Wang X, Nakatsu G, et al. Association between bacteria from specific microbes and subsequent diagnosis of colorectal cancer. Gastroenterology 2018;155(2):383–90.e8.

32. Gupta RS, Gao B. Phylogenetic analyses of clostridia and identification of novel protein signatures that are specific to the genus Clostridium sensu stricto (cluster D). Int J Syst Evol Microbiol 2009;59(Pt 2):285–94.

33. Flamer B, Warren RD, Barrett MP, et al. The oral microbiota in colorectal cancer is distinctive and predictive. Gut 2018;67(11):1454–63.

34. Rubinstein MR, Wang X, Liu W, et al. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/catenin signaling via its FadA adhesin. Cell Host Microbe 2013;14(2):195–206.

35. Fernández-Halder CV, Faria AVS, Andrade SS. Action and function of Faecalibacterium prausnitzii in health and disease. Best Pract Res Clin Gastroenterol 2017;31(6):643–8.

36. López-Siles M, Duncan SH, Garcia-Gil LJ, et al. Faecalibacterium prausnitzii: From microbiology to diagnostics and prognostics. ISME J 2017;11(4):841–52.

37. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012;486(7402):207–14.

38. Falony G, Joossens M, Vieira-Silva S, et al. Population-level analysis of gut microbiome variation. Science 2016;352(6285):560–4.

39. De Filippis F, Pasolli E, ERCOLINI D. Newly explored Faecalibacterium diversity is connected to age, lifestyle, geography, and disease. Curr Biol 2020;30(24):4932–4943.e4.

40. Quérévan E, Maubert MA, Michon C, et al. Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii, a commensal bacterium deficient in Crohn’s disease. Gut 2016;65(3):415–25.

41. Martin R, Bermúdez-Humárán L, Langella P. Searching for the bacterial effector: The example of the multi-skilled commensal bacterium Faecalibacterium prausnitzii. Front Microbiol 2018;9:346.

42. Zhou L, Zhang M, Wang Y, et al. Faecalibacterium prausnitzii produces butyrate to maintain Th17/Treg balance and to ameliorate colorectal colitis by inhibiting histone deacetylase 1. Inflamm Bowel Dis 2018;24(9):1926–40.

43. Guo S, Li L, Xu B, et al. A simple and novel fecal biomarker for colorectal cancer: Ratio of Fusobacterium nucleatum to probiotics populations, based on their antagonistic effect. Clin Chem 2018;64(9):1327–37.

44. Manichanh C, Rigottier-Goïs L, Bonnau E, et al. Reduced diversity of faecal microbiota in Crohn’s disease revealed by a metagenomic approach. Gut 2006;55(2):205–11.

45. Frank DN, St Amand AL, Feldman RA, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci USA 2007;104(34):13780–5.

46. Mori G, Rampelli S, Orena BS, et al. Shifts of faecal microbiota during sporadic colorectal carcinogenesis. Sci Rep 2018;8(1):10329.

47. Peters BA, Dominianni C, Shapiro JA, et al. The gut microbiota in conventional and serrated precursors of colorectal cancer. Microbiome 2016;4(1):69.

48. Heydari Z, Rahaei M, Alizadeh AM, et al. Effects of lactobacillus acidophilus and Bifidobacterium bifidum probiotics on the expression of microRNAs 135b, 26b, 18a and 155, and their involving genes in mice colon cancer. Probiotics Antimicrob Proteins 2019;11(4):148–62.

49. Wang W, Wang K, Wu W, et al. Administration of Bifidobacterium bifidum CGMCC 15068 modulates gut microbiota and metabolome in azoxymethane (AOM)/dextran sulphate sodium (DSS)-induced colitis-associated colon cancer (CAC) in mice. Appl Microbiol Biotechnol 2020;104(13):5915–28.

50. Wang Q, Wang K, Wu W, et al. Administration of Bifidobacterium bifidum CGMCC 15068 modulates gut microbiota and metabolome in azoxymethane (AOM)/dextran sulphate sodium (DSS)-induced colitis-associated colon cancer (CAC) in mice. Appl Microbiol Biotechnol 2020;104(13):5915–28.

51. Zeller G, Tap J, Voigt AY, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 2014;10(11):766.

52. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014;370(14):1287–97.

Open Access This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.