



Original Article

Gut Bacterial Characteristics and Noninvasive Detection of Colorectal Lesions



Bangzhou Zhang^{1,2#}, Qiongyun Chen^{1,2#}, Yanyun Fan^{1#}, Man Cao³, Yiqun Zhao¹, Changsheng Yan¹, Xiaoning Yang¹, Jingjing Liu¹, Hongzhi Xu^{1,2,4,5*} and Jianlin Ren^{1,2,4,5*}

¹Department of Gastroenterology, The National Key Clinical Specialty, Zhongshan Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, China; ²Xiamen Key Laboratory of Intestinal Microbiome and Human Health, Zhongshan Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, China; ³Department of Mathematics and Numerical Simulation and High-Performance Computing Laboratory, School of Sciences, Nanchang University, Nanchang, China; ⁴Institute for Microbial Ecology, School of Medicine, Xiamen University, Xiamen, China; ⁵Department of Digestive Disease, School of Medicine, Xiamen University, Xiamen, China

Received: August 15, 2022 | Revised: February 16, 2023 | Accepted: February 21, 2023 | Published online: March 24, 2023

Abstract

Background and objectives: The gut microbiota are frequently reported to be associated with colorectal cancer, while less attention has been paid to precancerous tumors. This study aimed to characterize the intestinal bacteria in patients with colorectal lesions and to assess the potential of bacteria as noninvasive biomarkers of colorectal tumors

Methods: We prospectively collected and sequenced 463 fecal samples from Zhongshan Hospital, Xiamen University, by targeting 16S rRNA V3_V4 on a Hiseq instrument with PE250 reagents. We analyzed the gut bacterial communities, determined the bacterial characteristics, and constructed models to classify colorectal tumors after feature selection, especially for precancerous lesions.

Results: There was a significant difference in fecal bacterial communities among the controls with normal colons (healthy subjects; HS) and the four stages of colorectal tumors. The fecal bacterial diversity increased in colorectal tumors. The phylum Firmicutes was significantly decreased, while Bacteroidetes was increased in colorectal tumors vs. HS. Correspondingly, a total of 81 genera, 589 operational taxonomic units, and 157 predicted pathways were remarkably different in relative abundances among the five groups. Relatively weak differences were observed among colorectal hyperplastic or inflammatory polyps (CRP), small adenomas (CRA), and advanced adenomas (Adv_CRA). Based on feature selection from genera, operational taxonomic units, pathways, and age, the models achieved an area under the receiver operating characteristic curve of 0.92 for classifying colorectal tumors vs. HS, 0.91 for the precancerous tumors vs. CRC, 0.80 for Adv_CRA vs. CRP, and 0.70 for CRA vs. CRP.

Conclusions: Alterations in the bacterial diversity, composition, and predicted pathways were identified across multistep colorectal tumorigenesis. The selected bacterial features represent potential noninvasive predictors of colorectal tumors, especially in discriminating benign polyps and adenomas.

Keywords: Gut bacteria; Colorectal tumors; 16S rRNA gene; Feature selection; Random forest.

Abbreviations: Adv_CRA, colorectal adenomas equal to or larger than 10 mm; AUC, area under the receiver operating characteristic curve; CRA, colorectal adenomas smaller than 10 mm; CRC, colorectal cancer; CRP, colorectal hyperplastic or inflammatory polyp; HS, normal colonoscopy; OTU, operational taxonomic unit; ZOTU, zero-radius operational taxonomic unit.

*Correspondence to: Jianlin Ren and Hongzhi Xu, Department of Gastroenterology, Zhongshan Hospital, Xiamen University, 201 Hubin South Road, Xiamen 361004, China. ORCID: <https://orcid.org/0000-0001-7901-594X> (JR) and <https://orcid.org/0000-0002-2725-5359> (HX). Tel: +86 0592-0592-2292232 (JR) and +86 0592-0592-2292232 (HX), E-mail: renjianl@xmu.edu.cn (JR) and xuhongzhi@xmu.edu.cn (HX)

#These authors contributed equally to this work.

How to cite this article: Zhang B, Chen Q, Fan Y, Cao M, Zhao Y, Yan C, et al. Gut Bacterial Characteristics and Noninvasive Detection of Colorectal Lesions. *Cancer Screen Prev* 2023;2(1):2–13. doi: 10.14218/CSP.2022.00017.

Introduction

According to a report of the International Agency for Research on Cancer, colorectal cancer (CRC) is the third most frequently diagnosed cancer and the second leading cause of cancer deaths worldwide, with more than 1.9 million new cases and 935,000 deaths estimated in 2020.¹ CRC is a heterogeneous disorder arising through different precursor lesions, different molecular pathways, and different end-stage carcinomas.^{2,3} Colorectal adenomas or adenomatous polyps are the most common precursors for CRC,^{4,5} which is the well-known “adenoma-carcinoma sequence.”⁶ It is estimated that over 50% of the screening-age population have one

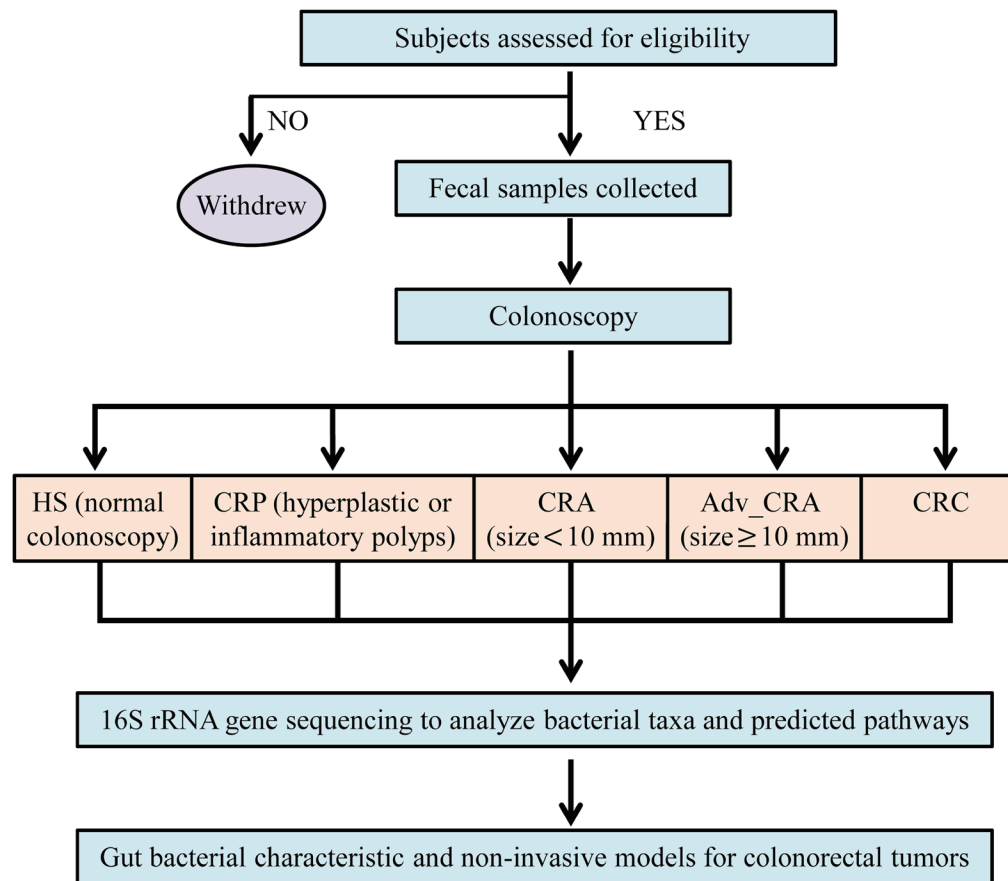


Fig. 1. The workflow chart of this study. Adv_CRA, colorectal adenomas equal to or larger than 10 mm; CRA, colorectal adenomas smaller than 10 mm; CRC, colorectal cancer; CRP, colorectal hyperplastic or inflammatory polyp; HS, normal colonoscopy.

or more precancerous adenomas or polyps.⁷ Furthermore, since the size of the adenoma is considered one of the important markers for the potential risk of cancerization,⁸ adenomatous polyps are considered advanced adenomas (Adv_CRA) when the size is equal to or larger than 10 mm in diameter.⁹

The 5-year survival rate is around 13% when CRC is detected at the advanced metastatic stage, but it exceeds 90% if the tumor is detected and treated at an early, localized stage.¹⁰ The early detection of colorectal tumors, especially adenomas, can significantly facilitate successful treatment and is important for decreasing CRC morbidity, mortality, and economic burden.¹¹ Colonoscopy is recognized as the golden standard of CRC screening. However, this test is poorly adhered to due to the invasiveness, frequency, and expensive price. For example, only 14% of high-risk people evaluated by a scoring system finally undertook colonoscopy screening in a recent survey in China.¹² Other widely used noninvasive tests, including the fecal immunochemical test and the fecal occult blood test, show unsatisfying sensitivities for CRC and have low sensitivity for colorectal adenomas or precancers.^{13,14} Relatively new tests based on multi-target stool DNA, such as ColoGuard, are still low in sensitivity for nonadvanced adenomas and are too expensive for large-scale screening.¹⁵ These shortcomings highlight the urgent need for the development of noninvasive and sensitive tests for CRC and precancerous lesions to improve the screening rate.

Acting as environmental factors of the human body, the gut

microbiota are frequently reported to play important roles in the initiation and progression of CRC^{16–19} and have been extensively studied to identify noninvasive biomarkers reflecting the disease,^{10,20–23} including *Fusobacterium nucleatum*,^{24,25} *Peptostreptococcus* sp., *Porphyromonas*, *Campylobacter jejuni*,²⁶ and some other specific genes.²⁷ Recently, microbe-derived metabolites also have been reported to serve as biomarkers of CRC.²⁸ However, unifying microbial signatures have not been identified for CRC across studies. Furthermore, it is not clear whether these individual biomarkers of CRC can effectively predict/classify adenomas, which appear at the early stage of CRC. In fact, the current knowledge on associations between the microbiota and adenomas is limited,^{11,23} since only a few studies have investigated the microbial alterations in adenomas.^{29–31} Moreover, few studies have explored the shifts of the gut microbiota of subjects with colorectal hyperplastic or inflammatory polyps (CRP),^{30,32} which are usually benign types of polyps, nor have they focused on the differences between CRP and adenomas.

In this study, we collected fecal samples across colorectal carcinogenesis and analyzed the fecal microbiota of participants with CRP, adenomas smaller than 10 mm (CRA), Adv_CRA, CRC, or a normal colonoscopy (healthy subjects; HS) by 16S rRNA gene sequencing (Fig. 1). The aims of this study were as follows: 1) to elucidate the shifts and characteristics of gut bacterial communities across the adenoma-carcinoma sequence with comprehensive stages, 2) to determine whether gut bacterial features can be used

to classify colorectal tumors, and 3) to explore the differences between CRP vs. Adv_CRA and CRP vs. CRA, and to evaluate the performances of the bacterial models in classifying them.

Methods

Subject enrollment and sample collection

All participants were voluntarily enrolled in this study before the colonoscopy. Exclusion criteria were as follows: the detection of bloodstream or gastrointestinal infections; use of antibiotics or probiotics one month before enrollment; prior colorectal resection; preparation for pregnancy; a history of other diseases affecting the gut microbiota, such as metabolic syndromes and autoimmunity; and contraindication to colonoscopy. Fecal samples were prospectively collected by the participants before bowel preparation. Briefly, fecal samples were mixed and collected in sterile tubes after defecation, and they were immediately stored at -80°C until DNA extraction. Lesion assessments included the location, size, number, and architecture during colonoscopy. Lesions were removed from the colon mucosa under the guidance of colonoscopy and were submitted for histological classification. According to the feedback of pathologists who had an average of five years of experience in the field, samples were grouped into CRP, CRA (size <10 mm, including adenomas with a tubular, tubulovillous, villous, or serrated growth pattern), Adv_CRA (size ≥ 10 mm, including adenomas with a tubular, tubulovillous, villous, or serrated growth pattern), CRC, or HS. Small sample sizes for CRC and HS were set up, since the gut microbiota of these two groups have been widely studied.

DNA extraction and amplicon sequencing

The fecal samples were thawed and homogenized, followed by DNA extraction using a Powerfecal Kit (QIAGEN, Hilden, Germany), and quality checked as previously described.³³ Extracted DNA samples were amplified by polymerase chain reaction with the forward primer 5'-CCTACGGGNBGCASCAG-3' and the reverse primer 5'-GGACTACNVGGGTWTCTAAT-3', which targets the 16S rRNA gene V3 and V4 region. The products were purified and checked with Qubit 3.0 (Thermo Fisher Scientific, Waltham, MA, USA) and then sequenced on a HiSeq 2500 platform (Illumina, San Diego, CA, USA) using a 250-bp paired-end sequencing protocol at Xiamen Treatgut Biotechnology Co.

Bioinformatic analyses

The raw paired-end reads were assembled using FLASH with default parameters except for parameters $-M = 200$ and $-x = 0.15$ ³⁴ and were further filtered using Usearch with the parameters $-fastq_maxee\ 0.5$.³⁵ High-quality reads were denoised into zero-radius operational taxonomic units (ZOTUs) with UNOISE3.³⁶ All analyses performed on the ZOTU table were rarefied to the sequencing depth of 13,793 reads per sample for downstream analyses. Taxonomic assignment of ZOTUs was performed in QIIME 1.9.1³⁷ using the SILVA132 database.³⁸ The microbial function was predicted by PICRUST2.³⁹

Developing machine learning models

To train multivariable statistical models for the prediction of different stages (HC, CRP, CRA, Adv_CRA, and CRC), three levels of bacterial features (genus, OTU, or pathway) and age were permuted and combined to develop prediction models separately. Data were randomly split into training and testing sets in a 5 \times -repeated

5-fold cross-validation, followed by the generation of random forest models using the randomForest R package v4.6-14. Finally, all predictions were used to calculate the area under the receiver operating characteristics curve (AUC) using the pROC R package v1.17.01. To optimize the performance, a feature selection step was developed for each model. Briefly, the importance ranking of each potential feature was first obtained based on the random forest importance parameters, mean decrease accuracy, or mean decrease in Gini values. Features were filtered within the cross-validation (that is, for each training set) by first calculating the AUC of the top-ranked feature and then removing features when the AUC dropped after adding the next feature, thereby keeping features informative in the model.

Statistical analyses

Alpha diversity indexes, including observed ZOTUs (Obs), Chao1, Shannon, and Pielou's evenness were computed based on the ZOTU table using the vegan package.⁴⁰ The differences in the diversity indexes and individual taxa were determined using the nonparametric Wilcoxon rank-sum test for two groups or the Kruskal–Wallis rank-sum test with Benjamini–Hochberg corrections for multiple groups using the agricolae package.⁴¹ The beta diversity of the overall bacterial communities was measured and visualized by distance-based redundancy analysis using the Euclidean distance, and the significance was determined with PERMANOVA with 9999 permutations using the vegan package. Visualization was mainly based on ggplot2⁴² or Venn Diagram.⁴³ All of these analyses were in R language.⁴⁴

Results

Demographic and clinical information

In total, fecal samples from 490 participants were prospectively collected, and 463 samples were included and subjected to 16S rRNA gene sequencing after a strict pathological diagnosis and exclusion process. Briefly, 45 HS, 120 CRP, 150 CRA, 113 Adv_CRA, and 35 CRC patients were included and randomly divided into the discovery phase (training set, 371 samples) and the validation phase (testing set, 92 samples) in this study. The ages of the patients were matched and were not significantly different among the five groups. The male percentages of CRC (60%), Adv_CRA (63%), CRA (73%), and CRP (67%) likely reflect the male preponderance of colorectal tumors (Table 1).

Shifts in gut microbial diversity

A total of 58,185,919 high-quality reads were obtained from 463 samples (mean = 125,672). We subsampled 13,793 reads for each participant according to the sample with the lowest sequence number. Compared with the HS group, the fecal bacterial richness (Observed and Chao1) was significantly ($p < 0.05$) increased in patients with colorectal tumors (Fig. 2a). A marginal significance ($p = 0.07$) was obtained for the test of difference in Shannon diversity among the five groups. Among the four disease groups, bacterial richness was significantly decreased in CRA ($n = 120$) vs. CRP ($n = 95$) and was significantly increased in CRC ($n = 30$) vs. CRA or Adv_CRA ($n = 90$). The fecal bacterial Shannon diversity and evenness were not significantly different among the five groups. Moreover, a Venn diagram showed that 1,289 of 4,689 OTUs were shared among the five groups, while 51 (2.80%), 321 (10.25%), 445 (13.29%), 330 (10.84%), and 238 (10.00%) OTUs were unique for HS, CRP, CRA, Adv_CRA, and CRC, respectively

Table 1. Clinical characteristics of the enrolled participants in training set

Clinical group	Mean age, years (±SD)	n	Male (%)	Female (%)
Normal colonoscopy (HS)	54 (±9)	36	21 (58%)	15 (42%)
Hyperplastic or inflammatory polyps (CRP)	57 (±11)	95	64 (67%)	33 (33%)
Adenomas (CRA, diameter < 10 mm)	57 (±12)	120	87 (73%)	33 (27%)
Advanced adenomas (Adv_CRA, diameter ≥ 10 mm)	56 (±11)	90	57 (63%)	33 (37%)
Colorectal cancer (CRC)	57 (±11)	30	18 (60%)	12 (40%)

(Fig. 2b). Beta). The beta diversity was visualized by db_RDA and indicated distinct clustering of samples from different groups, in which HS was associated with a greater abundance of *Faecalibacterium*, *Roseburia*, and *Ruminococcus_2* in the top 10 genera, while

Escherichia_Shigella was greater in CRC (Fig. 2c). Permutation analysis showed significant differences (PERMANOVA, $F = 1.60$, $p < 0.001$) in overall bacterial community differences among samples from the five groups.

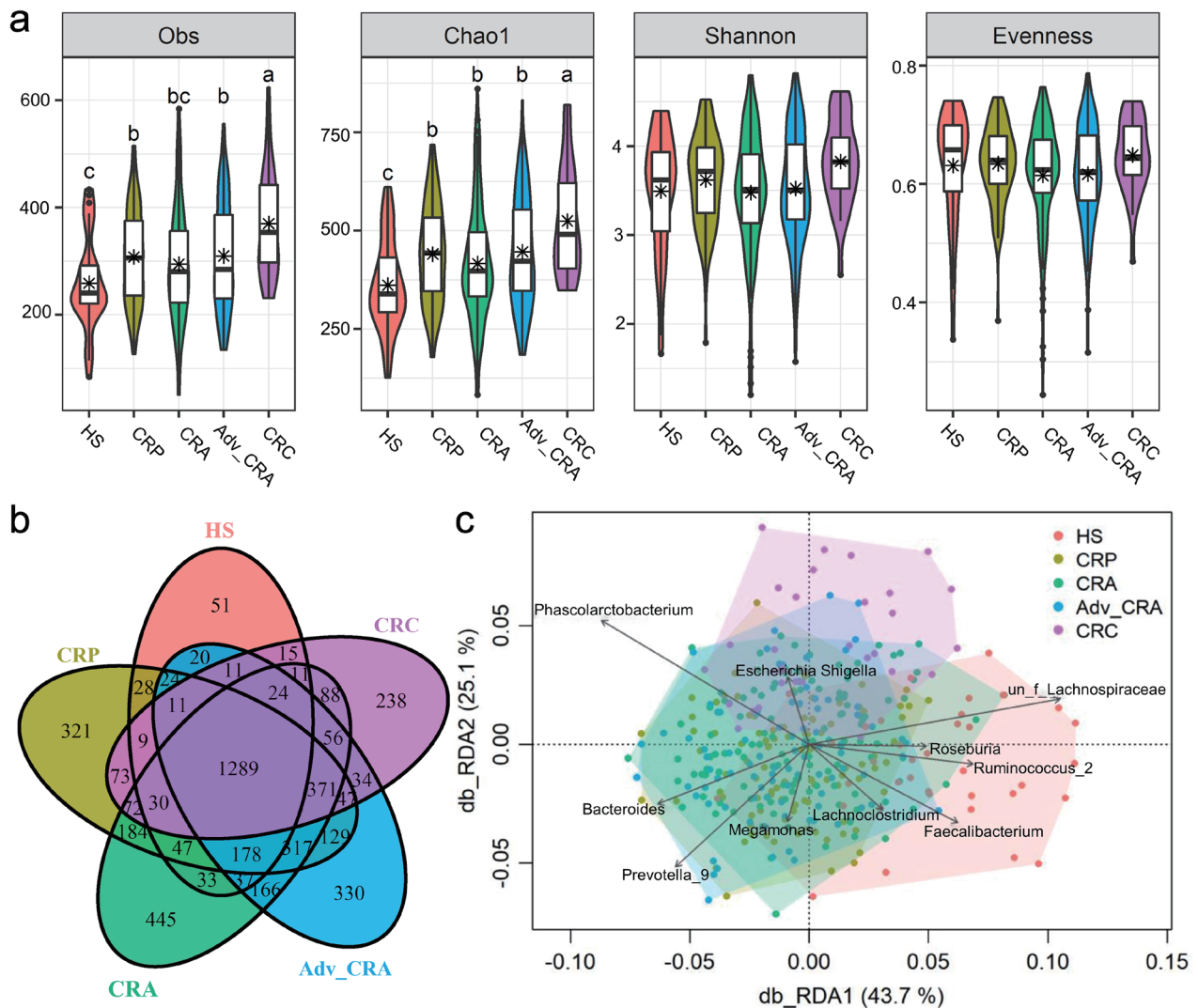


Fig. 2. Fecal bacterial diversity in patients with colorectal tumors at different stages and healthy subjects. (a) Alpha-diversity indexes including richness (observed ZOTUs and estimated Chao1), Shannon, and evenness were compared. The lower-case letters indicate the different groupings with significant differences. (b) The Venn diagram displaying the number of unique ZOTUs for each group and overlaps among the five groups. (c) The beta diversity was analyzed using db-RDA and indicates a distinct clustering of samples from different groups. Adv_CRA, colorectal adenomas equal to or larger than 10 mm; CRA, colorectal adenomas smaller than 10 mm; CRC, colorectal cancer; CRP, colorectal hyperplastic or inflammatory polyp; db-RDA, distance-based redundancy analysis; HS, normal colonoscopy; ZOTU, zero-radius operational taxonomic unit.

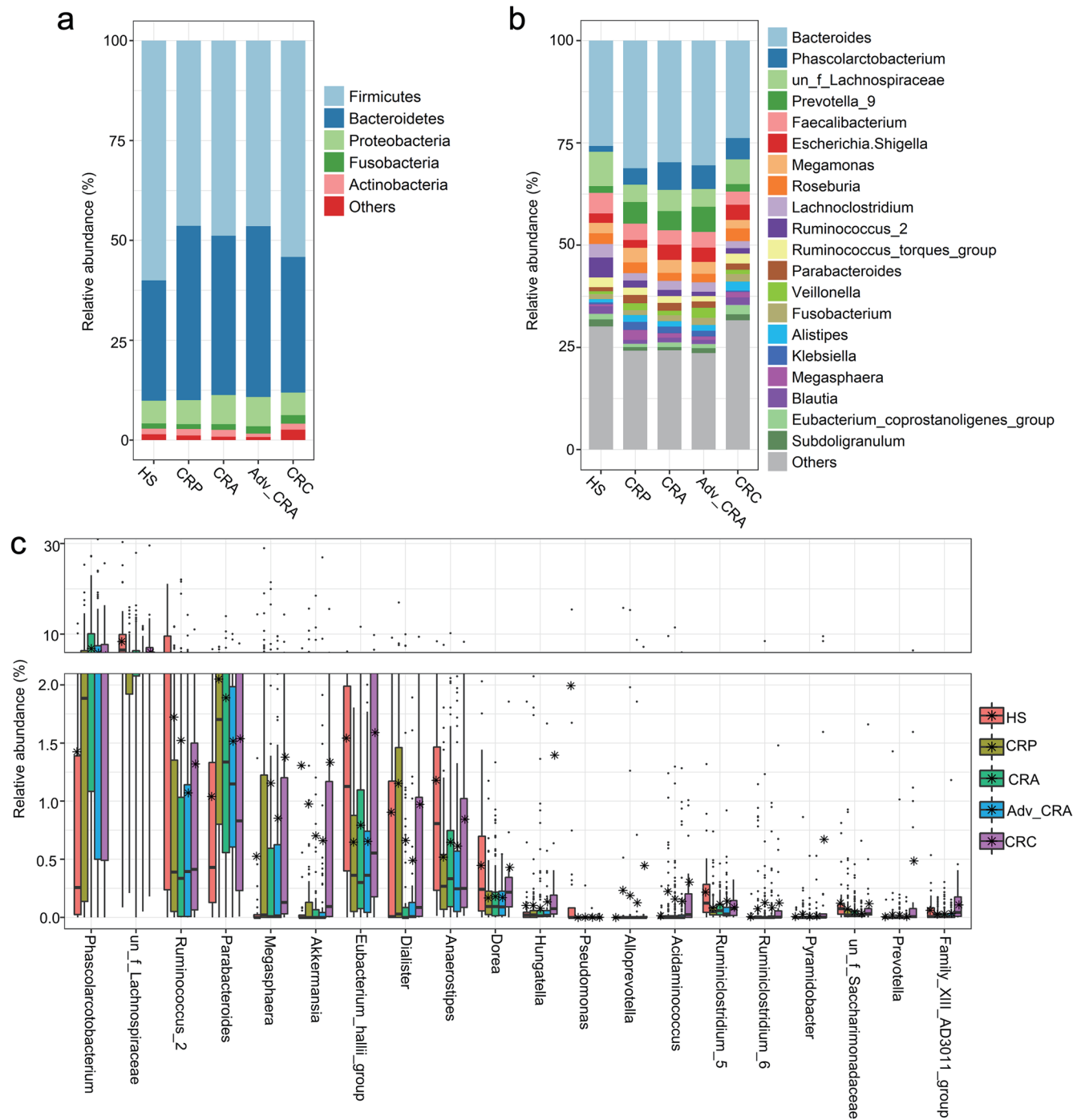


Fig. 3. Fecal bacterial profiles among patients with colorectal tumors at different stages and healthy subjects. Composition of fecal bacteria at the (a) phylum level and (b) genus level among the five groups. (c) Relative abundances of the top 20 genera that are significantly different among the five groups. Adv_CRA, colorectal adenomas equal to or larger than 10 mm; CRA, colorectal adenomas smaller than 10 mm; CRC, colorectal cancer; CRP, colorectal hyperplastic or inflammatory polyp; HS, normal colonoscopy.

Phylogenetic profiles of fecal microbial communities

The gut bacterial profiles were dominated by Bacteroidetes, Firmicutes, and Proteobacteria at the phylum level, together accounting for more than 90% of sequences (Fig. 3a). On average, *Bacteroides*, *Phascolarctobacterium*, *un_f_Lachnospiraceae*, *Prevotella_9*, and

Faecalibacterium were the top five genera (Fig. 3b). Firmicutes was significantly decreased in the colorectal tumor groups, while Bacteroidetes and Verrucomicrobia were significantly increased (Supplementary Fig. 1a). A total of 81 genera were detected to be significantly different among the five groups (Supplementary

Table 1). *Phascolarctobacterium*, *Megasphaera*, and *Desulfovibrio* displayed increasing trends (enriched) along with the development of the disease, while *un_f_Lachnospiraceae*, *Anaerostipes*, *Butyricimonas*, and *Dorea* were significantly decreased in the disease groups (Fig. 3c). Among the four disease groups, *Parabacteroides* decreased along with the progression of disease. At the finer amplicon sequence variant (ZOTU) level, 589 of 4,689 amplicon sequence variants were significantly different among the five groups (Supplementary Table 2). Moreover, a total of 409 microbial functional pathways were predicted, 157 of which were detected to be significantly different among the five groups (Supplementary Table 3). Pyruvate fermentation to isobutanol (PWY-7111), pyruvate fermentation to acetate and lactate II (PWY-5100), and galactose degradation I (PWY-6317) were depleted in the disease groups (Supplementary Fig. 1d and Supplementary Table 3). Additionally, the bacterial difference at the class and family levels among the five groups were compared and are shown in the online supplementary figure (Supplementary Fig. 1b and c).

Classification of colorectal tumors

To illustrate the diagnostic value of fecal bacteria for colonic tumors, we constructed a random forest classifier model that could specifically identify patients with colorectal lesions (non_HS) from the HS group as well as the four individual stages from HS. The combination of bacterial features showed non-HS prediction accuracy with an AUC of 0.922 (95% CI: 0.901–0.944) for the training set and 0.882 (95% CI: 0.780–0.983) for the testing set (Fig. 4a). This performance resulted from feature selection based on genera, OTU, pathways, or age with mean decrease accuracy or mean decrease Gini measures (Fig. 4c). Finally, a total of 35 features were identified, including 21 OTUs, 13 pathways, and age (Fig. 4d). Specifically, 12 OTUs from *Ruminococcus_2*, *Lachnoclostridium*, *Akkermansia*, etc. were depleted in the non-HS groups, while 9 OTUs from *Desulfovibrio*, *Phascolarctobacterium*, etc. were enriched in the non-HS groups (Supplementary Fig. 2). Eight pathways, including galactose degradation I (Leloir pathway) (PWY-6317), L-lysine biosynthesis I (DAPLYSINESYN-PWY), etc. had a lower abundance in the non-HS groups, while the superpathway of pyrimidine deoxyribonucleotides *de novo* biosynthesis (*E. coli*) (PWY0-166), pyrimidine deoxyribonucleotides *de novo* biosynthesis I (PWY-7184), etc. were upregulated in the non-HS groups (Supplementary Fig. 2). High performance of the random forest models was obtained for classifying CRC (AUC: 0.952, 95% CI: 0.931–0.972), Adv_CRA (AUC: 0.902, 95% CI: 0.877–0.927), CRA (AUC: 0.924, 95% CI: 0.903–0.945), and CRP (AUC: 0.959, 95% CI: 0.945–0.973) from HS (Fig. 4a, Supplementary Table 4). Next, we trained random forest models for differentiating CRC from individual precancerous stages. With a similar strategy, 31 OTU features, 27 OTU features, and 21 OTU features (Supplementary Table 5) were finally selected and achieved high performance in classifying Adv_CRA (AUC: 0.942, 95% CI: 0.919–0.966), CRA (AUC: 0.94, 95% CI: 0.917–0.964), and CRP (AUC: 0.91, 95% CI: 0.885–0.935) from CRC (Fig. 4b).

Bacterial differences between CRP, CRA, and Adv_CRA

We further explored the alterations of the gut microbial composition from benign colorectal polyps to adenomas and advanced adenomas. Analysis of beta diversity via principal component analysis revealed no significant differences in bacterial communities among the CRP, CRA, and Adv_CRA groups (PERMANOVA, $F = 1.034$, $p = 0.357$; Fig. 5a). No significant difference was observed in the alpha-diversity indexes (Fig. 2a). A total of 4 families and 12 gen-

era were significantly different, with no differences at the phylum or class level (Supplementary Fig. 3). These results indicated relatively weak differences between polyps and adenomas or advanced adenomas. As expected, these more biologically similar outcomes were more difficult to differentiate but might still be accessible via some bacterial features. Thus, we went on to identify specific taxa at the finer OTU level that were significantly enriched/depleted between CRP and Adv_CRA or CRA. There were 117 OTUs and 91 OTUs that were significantly different in relative abundances in CRP vs. CRA and CRP vs. Adv_CRA (Fig. 5b), of which 18 OTUs were shared and assigned as *Phascolarctobacterium*, *Lachnoclostridium*, *Fusobacterium*, *Butyricimonas*, *Subdoligranulum*, etc. (Supplementary Table 6). The relative abundances and fold changes of the top 20 OTUs that were different between CRP and Adv_CRA are displayed in Figure 5c. Based on these altered OTUs, we performed feature selection using the mean decrease in Gini value ranking to build microbial models for the classifications of CRP and Adv_CRA or CRA (Fig. 5c–d), including feature engineering by the combination of OTUs enriched (C1) or depleted (C2) in Adv_CRA into new features. To classify Adv_CRA from CRP, 19 features were finally selected as markers with OTUs from *Lachnoclostridium*, *Bacteroides*, *Ruminoclostridium_5*, etc., and AUC values of 0.802 (95% CI: 0.774–0.830) and 0.762 (95% CI: 0.612–0.902) for the training and testing sets, respectively, were achieved (Fig. 5e). Similarly, 14 genera including *Butyricimonas*, *Porphyromonas*, *Akkermansia*, etc. (Fig. 5f) were identified to discriminate CRP from CRA, with an AUC of 0.697 (95% CI: 0.666–0.728) for the training set and 0.706 (95% CI: 0.569–0.843) for the testing set (Fig. 5d). Both models reflected the potential of bacterial characteristics to distinguish advanced adenomas or adenomas from polyps.

Discussion

In this study, we profiled and analyzed the fecal bacterial communities and predicted the metabolic pathways of participants across five different stages of colorectal tumorigenesis, with a particular focus on the differences between benign polyps (hyperplastic or inflammatory) and precancerous adenomatous polyps. The overall bacterial communities were significantly different among the healthy controls and patients with colorectal tumors, and the patients with CRC had a greater fecal bacterial richness than the healthy controls and patients with polyps or adenomas. A total of 81 genera, 589 ZOTUs, and 157 predicted pathways were significantly different in relative abundances among the five groups. Importantly, the combination of bacterial genera, ZOTUs, pathways, or clinical information showed a promising potential for the non-invasive diagnosis of lesions. Based on feature selection, the bacterial models could achieve an average AUC of 0.92 for classifying colorectal tumors vs. HS, 0.91 for precancerous tumors vs. CRC among colorectal tumors, 0.80 for Adv_CRA vs. CRP, and 0.70 for CRA vs. CRP. Our findings suggest that alterations in the bacterial structures and pathways are associated with the occurrence and development of colorectal tumors and that the selected bacterial features may be a potential noninvasive predictor of colorectal lesions, especially in discriminating benign polyps (CRP) and precancerous adenomatous polyps (CRA or Adv_CRA).

Accumulating evidence has revealed that variations in the gut microbiota are associated with colorectal tumors. We did observe significant differences in the overall bacterial communities among the five groups at various stages of colorectal tumorigenesis. Although the gut microbial characteristics in patients with hyperplas-

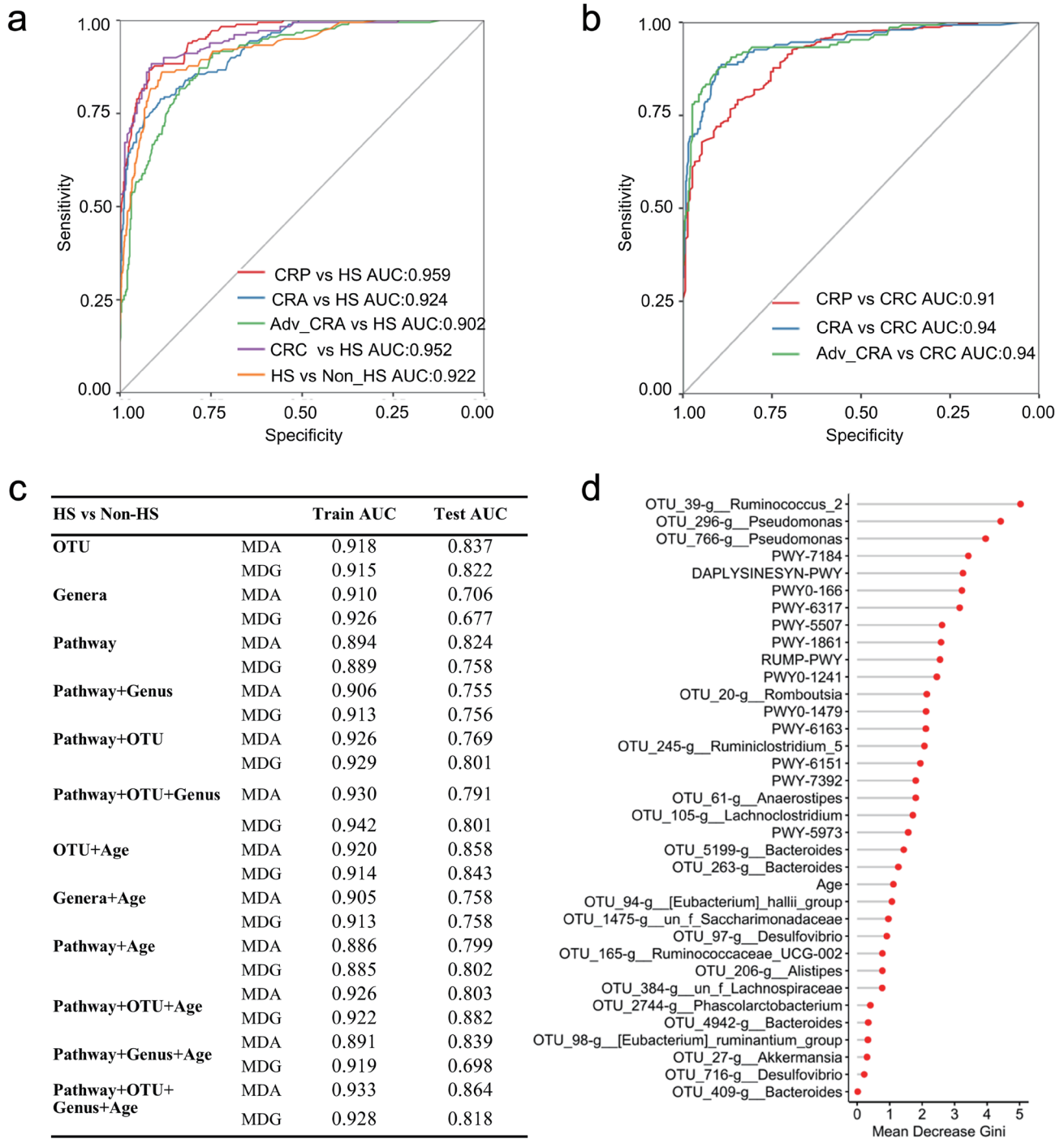


Fig. 4. Bacterial features demonstrate the potential to classify colorectal tumors. Receiver operating characteristic curves of models distinguishing (a) healthy subjects from those with CRP, CRA, Adv_CRA, and CRC as well as (b) CRC patients from those with CRP, CRA, and Adv_CRA (B). (c) Model performance and (d) feature importance (mean decrease in Gini value) of final features for distinguishing healthy patients from those with colorectal tumors after feature selection based on genera, OTUs, pathways, and age. Adv_CRA, colorectal adenomas equal to or larger than 10 mm; AUC, the area under the receiver operating characteristic curve; CRP, colorectal hyperplastic or inflammatory polyp; HS, normal colonoscopy; OTU, operational taxonomic unit.

tic or inflammatory polyps were merely illustrated, individuals with CRC and adenomas have been extensively reported to have different taxonomic compositions of fecal microbiota compared to healthy controls,^{10,18,21,28} which is referred to as “dysbiosis.”⁴⁵

More functionally, the gavage of fecal samples from patients with CRC promotes intestinal tumorigenesis, including the number of polyps, levels of intestinal dysplasia, and proliferation in mice.⁴⁶ In terms of bacterial diversity, our finding of increased species

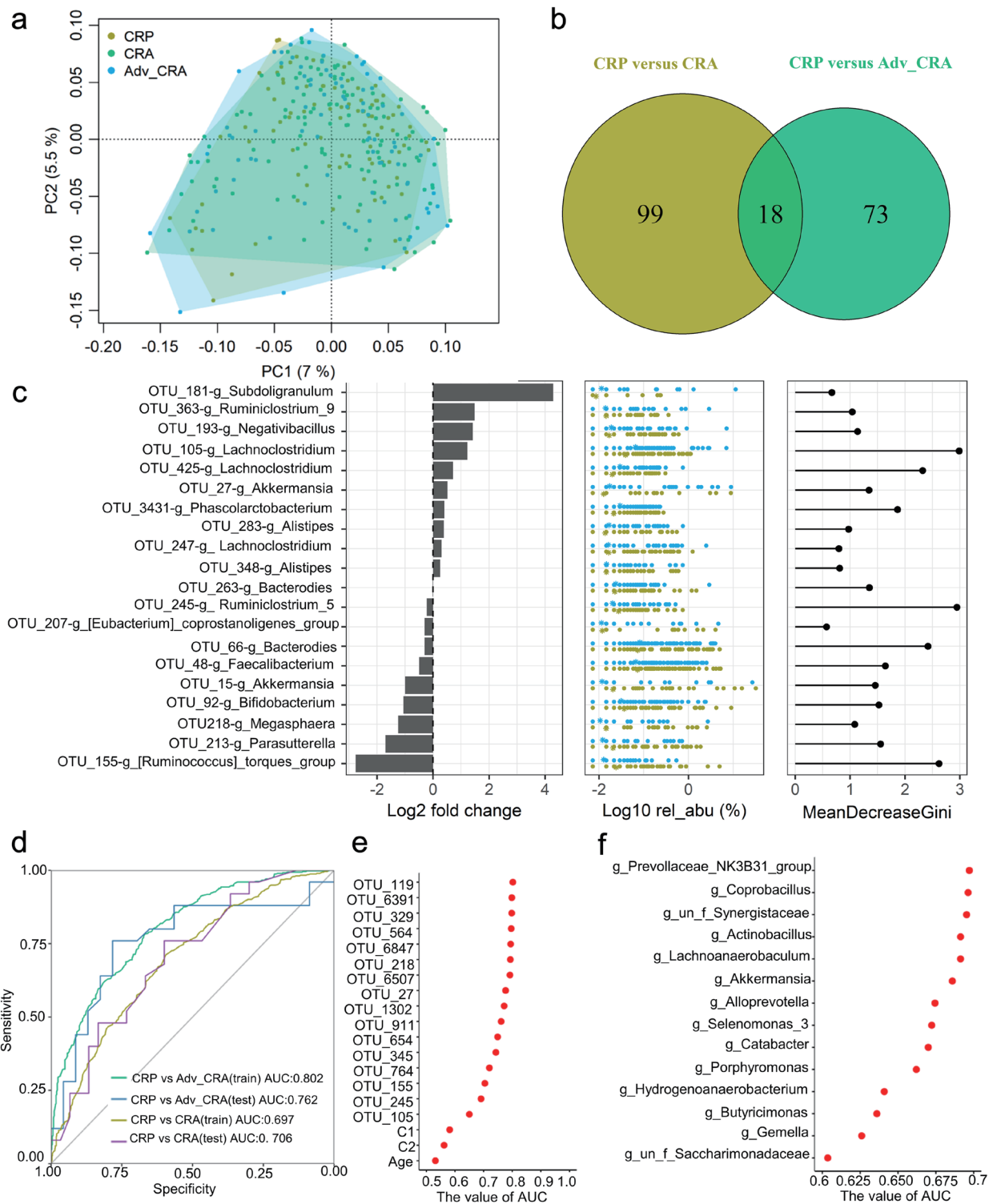


Fig. 5. Bacterial community differences and performance of models classifying CRP, CRA and Adv_CRA. (a) The bacterial beta diversity of the three groups was analyzed using principal component analysis. (b) The Venn diagram displaying the number of unique and shared differential OTUs between CRP vs. Adv_CRA and CRP vs. CRA. (c) Fold changes, relative abundances, and mean decrease in Gini values of the top 20 OTUs for distinguishing CRP from Adv_CRA. (d) Receiver operating characteristic curves of the training and testing sets for CRP vs. Adv_CRA and CRP vs. CRA. Feature importance (cumulative AUCs) of selected features for distinguishing (e) CRP vs. Adv_CRA and (f) CRP vs. CRA. C1 and C2 are combinations of OTUs enriched and depleted in Adv_CRA compared with CRP, respectively. Adv_CRA, colorectal adenomas equal to or larger than 10 mm; AUC, the area under the receiver operating characteristics curve; CRA, colorectal adenomas smaller than 10 mm; CRP, colorectal hyperplastic or inflammatory polyp; OTU, operational taxonomic unit.

richness in adenoma, particularly in CRC vs. HS, is the opposite to some previous reports.^{47,48} However, Nina *et al.* have reported an increased diversity in adenoma than controls, which is consistent with the current study,⁴⁹ followed by two studies reporting that the gut microbial richness is greater in CRC than adenoma.^{50,51} Similarly, bacterial diversity has been reported to be significantly increased in early hepatocellular carcinoma compared to that in liver cirrhosis.⁵² The increased richness and diversity at the severe stage of disease may be due to the recruitment and overgrowth of various pathogenic or harmful bacteria;⁵² this finding is supported by the high proportion (more than 10%) of ZOTUs that were unique to each tumor group in this study.

We detected plenty of bacterial characteristics at the genus, ZOTU, and pathway levels that were significantly different across the stages. Individual taxa with abnormal abundances have been extensively reported to be associated with CRC and even with adenoma.^{20,21} The genera of *Anaerostipes* and *Butyricimonas* decreased along with the tumor stage in this study; these taxa are well known to produce short-chain fatty acids,⁵³ which are essential to maintaining human health by providing energy to the intestinal epithelium, modulating the immune system, and affecting diverse metabolic routes. In fact, the microbial pathways that produce short-chain fatty acids, such as pyruvate fermentation to acetate and lactate II (PWY-5100), were depleted in patients with tumors. *Desulfovibrio*, including the I_97 aIOTU_716 selected as model features, can produce hydrogen sulfide,⁵⁴ a genotoxic insult to the colonic epithelium,⁵⁵ representing a potential pathogen that directly increases the risk of the development of colorectal tumors. Some genera were reported in this study. Similarly, *Phascolarctobacterium* has been reported to abundantly colonize the human gastrointestinal tract⁵⁶ and has been positively associated with autism spectrum disorder⁵⁷ and Alzheimer's disease.⁵⁸ Although less investigated, *Megasphaera* has been reported to increase in abundance after appendectomy in both children and adults,^{33,59} while it seems to be beneficial for those with diarrheal cryptosporidiosis.⁶⁰ Interestingly, we found that the galactose degradation I pathway (PWY-6317) was depleted in disease stages; therefore, it was selected as an important feature for the model classifying tumors. Galactose from fruits and vegetables can prevent CRC by binding and inhibiting lectins that can stimulate colon epithelial proliferation.⁶¹ Recently, β -galactosidase, which hydrolyzes lactose into galactose, has been reported to prevent tumor formation by inhibiting cell proliferation, promoting apoptosis of CRC cells, and retarding the growth of CRC xenografts.⁶² Certainly, more studies are needed to illustrate the mechanisms of the individual taxa and pathways acting on tumorigenesis.

As expected, good performances were obtained for the random forest models based on a combination of genera, OTUs, pathways, and/or age after feature selection when classifying individual stage of lesions (AUCs: 0.84–0.96) or overall colorectal tumors (AUC = 0.88) vs. healthy controls, as well as further discriminating CRC vs. other precancerous lesions (AUCs: 0.74–0.88). Good performance of microbiome-based models for classifying CRC vs. healthy controls has been published previously, with AUCs greater than 0.8 based on the meta-analysis of metagenomic^{20,21} and 16S rRNA gene sequencing datasets.^{10,23} Unfortunately, models for adenoma have been less investigated and usually provide a lower performance.^{11,23} Recently, Young *et al.* have reported 16S rRNA sequencing-based models distinguishing neoplasm (CRC or adenoma) vs. blood-negative guaiac fecal occult blood tests, with an AUC of 0.78 in a large-scale (more than 2,000 samples) bowel cancer screening program.⁶³ Our favorable results suggest the im-

portance of feature selection for bacterial markers in improving the performance of noninvasive diagnosis of colorectal tumors. Previous studies have shown the reduced discriminatory power of microbiome-based models to detect adenomas.²¹ Our analysis profiling gut microbiome-associated characteristics has the potential for the diagnosis of adenoma from polyps, including advanced adenoma. Adv_CRA was classified from CRP with 19 markers, with an AUC of approximately 0.80. Ten bacterial genera distinguished CRA from CRP, with an AUC of 0.70. To the best of our knowledge, this is the first study to explore microbial signatures between polyps and adenomas or advanced adenomas. Although the performance of the model in this study is lower than that of other models, it reflects the potential of identifying adenomas by bacterial markers.

The following limitations should be considered in this study. First, limited clinical metrics were collected due to the prospective collection of fecal samples in the hospital before colonoscopy. The addition of more clinical indexes and an independent cohort validation may further improve and verify the performance of the classifying models in the future. Second, nonbalanced samples were collected, with small sample sizes for the HS and CRC groups. Our original intention was to reveal the shifts, performance, and potentials of the bacterial communities in the noninvasive screening of colorectal tumors, with a particular focus (large sample size) on precancerous stages, including adenoma and polyps, since comparisons between CRC and HS have been well studied. Third, 16S rRNA gene sequencing was applied in this study. Metagenomic sequencing and metabolomics would provide more insights and further reveal the shifts of microbial features. This study should be extended in terms of sample size, multi-center verification with more baseline clinical characteristics as well as sample collection and storage, and shotgun metagenomic sequencing analysis for optimization to benefit patients in clinical practice.

Conclusions

In conclusion, we observed dynamic shifts in the fecal bacterial diversity, and the bacterial composition predicted the pathways across multistep colorectal tumorigenesis. Additionally, after feature selection based on genera, OTUs, pathways, and age, we built classifying models with a good performance for classifying overall colorectal tumors vs. healthy controls and precancerous tumors vs. CRC. More importantly, for the first time, we explored the differences in bacterial communities and the noninvasive models for benign polyps (hyperplastic or inflammatory) and precancerous adenomatous polyps, which is meaningful in the clinic for noninvasively identifying the risk of progression to cancer from polyps.

Supporting information

Supplementary material for this article is available at <https://doi.org/10.14218/CSP.2022.00017>.

Supplementary Fig. 1. Relative abundances of phyla (a), classes (b), and top 10 families (c), and top 20 pathways (d) among patients with colorectal tumors and healthy subjects.

Supplementary Fig. 2. Relative abundances of OTUs that were selected as important features for classifying colorectal lesions (non_HS) from HS group.

Supplementary Fig. 3. Relative abundances of families (a) and

genera (b) among patients with CRP, CRA, and Adv_CRA.

Supplementary Table 1. Relative abundances of the 81 genera that were significantly different among HS, CRP, CRA, Adv_CRA, and CRC.

Supplementary Table 2. Relative abundances of the 589 OTUs that were significantly different among HS, CRP, CRA, Adv_CRA, and CRC.

Supplementary Table 3. Relative abundances of the 157 pathways that were significantly different among HS, CRP, CRA, Adv_CRA, and CRC.

Supplementary Table 4. Taxonomy of the OTU features selected for models in classifying CRC, Adv_CRA, CRA and CRP from HS.

Supplementary Table 5. Taxonomy of the OTU features selected for models in classifying Adv_CRA, CRA and CRP from CRC.

Supplementary Table 6. Taxonomy of the OTU shared between differences from CRP vs Adv_CRA and from CRP Vs CRA.

Acknowledgments

The authors thank the lab members and clinicians of the Gastroenterology Department at Zhongshan Hospital of Xiamen University for thoughtful comments on the manuscript and for helping to manage the patients.

Funding

This work was supported by the Xiamen Key Programs of National Health (3502Z20204007), the Xiamen Priority Programs of Medical Health (3502Z20199172), the Fujian Provincial Natural Science Foundation (2021J011329), and the Fundamental Research Funds for the Central Universities.

Conflict of interest

Prof. Jianlin Ren has been an editorial board member of *Cancer Screening and Prevention* since March 2022. The authors have no other conflict of interests related to this publication.

Author contributions

Study concept and design (JLR, HZX, BZZ), acquisition of data (QYC, YYF, YQZ, CSY, XNY), analysis and interpretation of data (BZZ, MC, QYC, YYF), drafting of the manuscript (BZZ, QYC, MC, YYF), administrative and technical support (JLJ, YYF), and study supervision (JLR, HZX). All authors contributed significantly to this study and approved the final manuscript.

Ethical statement

The Ethics Committee of Zhongshan hospital, Xiamen University approved this study.

Data sharing statement

The raw sequences used to support the findings of this study were

deposited in the National Center for Biotechnology Information Sequence Read Archive under accession number PRJNA869338.

References

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71(3):209–249. doi:10.3322/caac.21660, PMID: 33538338.
- [2] Huang CS, Farraye FA, Yang S, O'Brien MJ. The clinical significance of serrated polyps. *Am J Gastroenterol* 2011;106(2):229–240. doi:10.1038/ajg.2010.429, PMID:21045813.
- [3] Snover DC. Update on the serrated pathway to colorectal carcinoma. *Hum Pathol* 2011;42(1):1–10. doi:10.1016/j.humpath.2010.06.002, PMID:20869746.
- [4] Fearon ER. Molecular genetics of colorectal cancer. *Annu Rev Pathol* 2011;6:479–507. doi:10.1146/annurev-pathol-011110-130235, PMID:21090969.
- [5] Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology* 2015;149(5):1177–1190.e3. doi: 10.1053/j.gastro.2015.06.047, PMID:26216840.
- [6] Strum WB. Colorectal Adenomas. *N Engl J Med* 2016;374(11):1065–1075. doi:10.1056/NEJMra1513581, PMID:26981936.
- [7] DeDecker L, Coppedge B, Avelar-Barragan J, Karnes W, Whiteson K. Microbiome distinctions between the CRC carcinogenic pathways. *Gut Microbes* 2021;13(1):1854641. doi:10.1080/19490976.2020.1854641, PMID:33446008.
- [8] Robert ME. The malignant colon polyp: diagnosis and therapeutic recommendations. *Clin Gastroenterol Hepatol* 2007;5(6):662–667. doi:10.1016/j.cgh.2007.04.001, PMID:17544991.
- [9] Rutter MD, East J, Rees CJ, Cripps N, Docherty J, Dolwani S, *et al*. British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines. *Gut* 2020;69(2):201–223. doi:10.1136/gutjnl-2019-319858, PMID:31776230.
- [10] Shah MS, DeSantis TZ, Weinmaier T, McMurdie PJ, Cope JL, Altrichter A, *et al*. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. *Gut* 2018;67(5):882–891. doi:10.1136/gutjnl-2016-313189, PMID:28341746.
- [11] Wu Y, Jiao N, Zhu R, Zhang Y, Wu D, Wang AJ, *et al*. Identification of microbial markers across populations in early detection of colorectal cancer. *Nat Commun* 2021;12(1):3063. doi:10.1038/s41467-021-23265-y, PMID:34031391.
- [12] Chen H, Li N, Ren J, Feng X, Lyu Z, Wei L, *et al*. Participation and yield of a population-based colorectal cancer screening programme in China. *Gut* 2019;68(8):1450–1457. doi:10.1136/gutjnl-2018-317124, PMID:30377193.
- [13] 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(3):171. doi:10.7326/M13-1484, PMID:24658694.
- [14] Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. *Ann Intern Med* 2009;150(3):162–169. doi:10.7326/0003-4819-150-3-200902030-00005, PMID:19189905.
- [15] Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, *et al*. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014;370(14):1287–1297. doi:10.1056/NEJMoa1311194, PMID:24645800.
- [16] Song M, Chan AT. Environmental Factors, Gut Microbiota, and Colorectal Cancer Prevention. *Clin Gastroenterol Hepatol* 2019;17(2):275–289. doi:10.1016/j.cgh.2018.07.012, PMID:30031175.
- [17] Vipperla K, O'Keefe SJ. Diet, microbiota, and dysbiosis: a 'recipe' for colorectal cancer. *Food Funct* 2016;7(4):1731–1740. doi:10.1039/c5fo01276g, PMID:26840037.
- [18] Janney A, Powrie F, Mann EH. Host-microbiota maladaptation in colorectal cancer. *Nature* 2020;585(7826):509–517. doi:10.1038/s41586-020-2729-3, PMID:32968260.

- [19] Dai Z, Zhang J, Wu Q, Chen J, Liu J, Wang L, *et al.* The role of microbiota in the development of colorectal cancer. *Int J Cancer* 2019;145(8):2032–2041. doi:10.1002/ijc.32017, PMID:30474116.
- [20] Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 2019;25(4):679–689. doi:10.1038/s41591-019-0406-6, PMID:30936547.
- [21] Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med* 2019;25(4):667–678. doi:10.1038/s41591-019-0405-7, PMID:30936548.
- [22] Sze MA, Schloss PD. Leveraging Existing 16S rRNA Gene Surveys To Identify Reproducible Biomarkers in Individuals with Colorectal Tumors. *mBio* 2018;9(3). doi:10.1128/mBio.00630-18, PMID:29871916.
- [23] Zhang B, Xu S, Xu W, Chen Q, Chen Z, Yan C, *et al.* Leveraging Fecal Bacterial Survey Data to Predict Colorectal Tumors. *Front Genet* 2019;10:447. doi:10.3389/fgene.2019.00447, PMID:31191599.
- [24] Brennan CA, Garrett WS. *Fusobacterium nucleatum* - symbiont, opportunist and oncobacterium. *Nat Rev Microbiol* 2019;17(3):156–166. doi:10.1038/s41579-018-0129-6, PMID:30546113.
- [25] Guo S, Li L, Xu B, Li M, Zeng Q, Xiao H, *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):1327–1337. doi:10.1373/clinchem.2018.289728, PMID:29914865.
- [26] He Z, Gharaibeh RZ, Newsome RC, Pope JL, Dougherty MW, Tomkovich S, *et al.* *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut* 2019;68(2):289–300. doi:10.1136/gutjnl-2018-317200, PMID:30377189.
- [27] Liang JQ, Li T, Nakatsu G, Chen YX, Yau TO, Chu E, *et al.* A novel faecal *Lachnospirillum* marker for the non-invasive diagnosis of colorectal adenoma and cancer. *Gut* 2020;69(7):1248–1257. doi:10.1136/gutjnl-2019-318532, PMID:31776231.
- [28] Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* 2019;25(6):968–976. doi:10.1038/s41591-019-0458-7, PMID:31171880.
- [29] Flemer B, Warren RD, Barrett MP, Cisek K, Das A, Jeffery IB, *et al.* The oral microbiota in colorectal cancer is distinctive and predictive. *Gut* 2018;67(8):1454–1463. doi:10.1136/gutjnl-2017-314814, PMID:28988196.
- [30] Mori G, Rampelli S, Orena BS, Rengucci C, De Maio G, Barbieri G, *et al.* Shifts of Faecal Microbiota During Sporadic Colorectal Carcinogenesis. *Sci Rep* 2018;8(1):10329. doi:10.1038/s41598-018-28671-9, PMID:29985435.
- [31] Baxter NT, Ruffin MT, Rogers MA, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med* 2016;8(1):37. doi:10.1186/s13073-016-0290-3, PMID:27056827.
- [32] Zhang Y, Yu X, Yu E, Wang N, Cai Q, Shuai Q, *et al.* Changes in gut microbiota and plasma inflammatory factors across the stages of colorectal tumorigenesis: a case-control study. *BMC Microbiol* 2018;18(1):92. doi:10.1186/s12866-018-1232-6, PMID:30157754.
- [33] Cai S, Fan Y, Zhang B, Lin J, Yang X, Liu Y, *et al.* Appendectomy Is Associated With Alteration of Human Gut Bacterial and Fungal Communities. *Front Microbiol* 2021;12:724980. doi:10.3389/fmicb.2021.724980, PMID:34603252.
- [34] Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011;27(21):2957–2963. doi:10.1093/bioinformatics/btr507, PMID:21903629.
- [35] Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 2013;10(10):996–998. doi:10.1038/nmeth.2604, PMID:23955772.
- [36] Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv [Preprint]* 2016. doi:10.1101/081257.
- [37] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7(5):335–336. doi:10.1038/nmeth.f.303, PMID:20383131.
- [38] Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, *et al.* The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res* 2014;42(Database issue):D643–D648. doi:10.1093/nar/gkt1209, PMID:24293649.
- [39] Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, *et al.* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 2020;38(6):685–688. doi:10.1038/s41587-020-0548-6, PMID:32483366.
- [40] Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’Hara RB, *et al.* *vegan*: community ecology package. R package 2015. Available from: <https://github.com/vegandevs/vegan>. Accessed August 15, 2022.
- [41] de Mendiburu F, Yaseen M. *agricolae*: statistical procedures for agricultural research. R package 2015. Available from: <https://myaseen208.github.io/agricolae/>. Accessed August 15, 2022.
- [42] Wickham H, Chang W, Henry L, Takahashi K, Wilke C, Woo K, *et al.* *ggplot2*: Create elegant data visualisations using the grammar of graphics. R package 2017. Available from: <https://github.com/tidyverse/ggplot2/blob/HEAD/R/ggplot2-package.R>. Accessed August 15, 2022.
- [43] VennDiagram: Generate high-resolution Venn and Euler plots. 2014. Available from: <https://cran.r-project.org/web/packages/VennDiagram/VennDiagram.pdf>. Accessed August 15, 2022.
- [44] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2018. Available from: <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing>. Accessed August 15, 2022.
- [45] Olesen SW, Alm EJ. Dysbiosis is not an answer. *Nature microbiology* 2016;1(12):16228. doi:10.1038/nmicrobiol.2016.228.
- [46] Wong SH, Zhao L, Zhang X, Nakatsu G, Han J, Xu W, *et al.* Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice. *Gastroenterology* 2017;153(6):1621–1633.e6. doi:10.1053/j.gastro.2017.08.022, PMID:28823860.
- [47] Peters BA, Dominianni C, Shapiro JA, Church TR, Wu J, Miller G, *et al.* The gut microbiota in conventional and serrated precursors of colorectal cancer. *Microbiome* 2016;4(1):69. doi:10.1186/s40168-016-0218-6, PMID:28038683.
- [48] Huipeng W, Lifeng G, Chuang G, Jiaying Z, Yuankun C. The differences in colonic mucosal microbiota between normal individual and colon cancer patients by polymerase chain reaction-denaturing gradient gel electrophoresis. *J Clin Gastroenterol* 2014;48(2):138–144. doi:10.1097/MCG.0b013e3182a26719, PMID:24162169.
- [49] Sanapareddy N, Legge RM, Jovov B, McCoy A, Burcal L, Araujo-Perez F, *et al.* Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. *ISME J* 2012;6(10):1858–1868. doi:10.1038/ismej.2012.43, PMID:22622349.
- [50] Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, *et al.* Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* 2015;6:6528. doi:10.1038/ncomms7528, PMID:25758642.
- [51] Thomas AM, Jesus EC, Lopes A, Aguiar S Jr, Begnami MD, Rocha RM, *et al.* Tissue-Associated Bacterial Alterations in Rectal Carcinoma Patients Revealed by 16S rRNA Community Profiling. *Front Cell Infect Microbiol* 2016;6:179. doi:10.3389/fcimb.2016.00179, PMID:28018861.
- [52] Ren Z, Li A, Jiang J, Zhou L, Yu Z, Lu H, *et al.* Gut microbiome analysis as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma. *Gut* 2019;68(6):1014–1023. doi:10.1136/gutjnl-2017-315084, PMID:30045880.
- [53] Vital M, Karch A, Pieper DH. Colonic Butyrate-Producing Communities in Humans: an Overview Using Omics Data. *mSystems* 2017;2(6):e00130-17. doi:10.1128/mSystems.00130-17, PMID:29238752.
- [54] Warren YA, Citron DM, Merriam CV, Goldstein EJ. Biochemical differentiation and comparison of *Desulfovibrio* species and other phenotypically similar genera. *J Clin Microbiol* 2005;43(8):4041–4045. doi:10.1128/JCM.43.8.4041-4045.2005, PMID:16081948.
- [55] Attene-Ramos MS, Nava GM, Muellner MG, Wagner ED, Plewa MJ, Gaskins HR. DNA damage and toxicogenomic analyses of hydrogen sulfide in human intestinal epithelial FHs 74 Int cells. *Environ Mol Mutagen* 2010;51(4):304–314. doi:10.1002/em.20546, PMID:20120018.
- [56] Wu F, Guo X, Zhang J, Zhang M, Ou Z, Peng Y. *Phascolarctobacterium faecium* abundant colonization in human gastrointestinal tract.

- Exp Ther Med 2017;14(4):3122–3126. doi:10.3892/etm.2017.4878, PMID:28912861.
- [57] Iglesias-Vázquez L, Van Ginkel Riba G, Arijia V, Canals J. Composition of Gut Microbiota in Children with Autism Spectrum Disorder: A Systematic Review and Meta-Analysis. *Nutrients* 2020;12(3):792. doi:10.3390/nu12030792, PMID:32192218.
- [58] Hung CC, Chang CC, Huang CW, Nouchi R, Cheng CH. Gut microbiota in patients with Alzheimer’s disease spectrum: a systematic review and meta-analysis. *Aging (Albany NY)* 2022;14(1):477–496. doi:10.18632/aging.203826, PMID:35027502.
- [59] Bi Y, Yang Q, Li J, Zhao X, Yan B, Li X, *et al*. The Gut Microbiota and Inflammatory Factors in Pediatric Appendicitis. *Dis Markers* 2022;2022:1059445. doi:10.1155/2022/1059445, PMID:35845131.
- [60] Carey MA, Medlock GL, Alam M, Kabir M, Uddin MJ, Nayak U, *et al*. *Megasphaera* in the Stool Microbiota Is Negatively Associated With Diarrheal Cryptosporidiosis. *Clin Infect Dis* 2021;73(6):e1242–e1251. doi:10.1093/cid/ciab207, PMID:33684930.
- [61] Evans RC, Fear S, Ashby D, Hackett A, Williams E, Van Der Vliet M, *et al*. Diet and colorectal cancer: an investigation of the lectin/galactose hypothesis. *Gastroenterology* 2002;122(7):1784–1792. doi:10.1053/gast.2002.33659, PMID:12055585.
- [62] Li Q, Hu W, Liu WX, Zhao LY, Huang D, Liu XD, *et al*. *Streptococcus thermophilus* Inhibits Colorectal Tumorigenesis Through Secreting β -Galactosidase. *Gastroenterology* 2021;160(4):1179–1193.e14. doi:10.1053/j.gastro.2020.09.003, PMID:32920015.
- [63] Young C, Wood HM, Fuentes Balaguer A, Bottomley D, Gallop N, Wilkinson L, *et al*. Microbiome Analysis of More Than 2,000 NHS Bowel Cancer Screening Programme Samples Shows the Potential to Improve Screening Accuracy. *Clin Cancer Res* 2021;27(8):2246–2254. doi:10.1158/1078-0432.ccr-20-3807, PMID:33658300.