Microbial Biomarkers for Colorectal Cancer Identified with Random Forest Model

Weili Sun¹, Lili Wang², Qiuyue Zhang² and Quanjiang Dong¹,²*

¹The Affiliated Hospital of Dalian Medical University, Shandong, China; ²Department of Central Laboratories and Gastroenterology, Qingdao Municipal Hospital Affiliated to Qingdao University, Shandong, China

Abstract

Colorectal cancer (CRC) is one of the most common cancers and a leading cause of cancer-related death. Gut microbiota are part of a complex microbe-based ecosystem of the human body, and changes in the microbiota can lead to a variety of diseases. All currently used CRC detection methods, including endoscopy, guaiac-based fecal occult blood test and fecal immunochemical test, have many limitations. Therefore, establishing novel screening methods which are accurate, inexpensive and non-invasive is indicated. Random forest models, as a superiority machine learning model, are increasingly used in research to select biomarkers. In this review, we summarized progresses of the diagnoses of CRC based on the random forest model of gut microbiota. We concluded that some cancer-associated bacteria in gut microbiota could be used as biomarkers for detecting early CRC. We also aimed to discuss how to select possible markers of colorectal diseases based on gut microbiota using the random forest model.

Introduction

Colorectal cancer (CRC) is one of the most common cancers and a leading cause of cancer-related death. The global annual incidence of CRC is approximately 1.2 million, and the annual death toll reaches 600,000.¹ CRC is characterized by a long preclinical phase, progressing over years from early adenoma to invasive cancer.²⁻⁴ Diagnosis of early-stage CRC can significantly improve patient prognosis, with a survival rate higher than 80%,⁴ which makes early diagnosis of CRC more critical for better patient outcomes.

Screening methods for CRC, including fecal occult blood test (referred to as FOBT), colonoscopy, determination of genetic mutation and gut microbiota tests, have received widespread attention; although, these methods significantly differ in specificity, accuracy, convenience, and universality with regard to clinical diagnosis. Colonoscopy, the gold standard for the diagnosis of CRC, can facilitate early prevention through adenoma resection; however, it is an invasive endoscopy, which causes patient discomfort and reduces their compliance to clinical examinations, relies on endoscopist skill, and requires patient bowel cleansing.⁵⁻⁶ The traditional guaiac-based FOBT, which is inexpensive and noninvasive, is the first-choice screening test for CRC. Both guaiac-based FOBT and fecal immunochemical test (FIT) can improve the positive rate of diagnosis by repeated detection over a long period of time. However, cancer bleeding is usually intermittent, which leads to lower sensitivity, specificity, and false-negative results with these two tests.⁷ As the abovementioned methods have many limitations, it is necessary to establish novel screening methods that are accurate, inexpensive, and noninvasive.

The human microbiota is a very large and complex microbial system. Gut microbiota includes bacteria, viruses, archaea, and fungi, but the main component is bacteria; thus, most studies of the microbiota focus on bacteria, as it is composed of approximately 10¹⁴ bacteria, which equates to 10 times the total number of human cells.⁸ The gut microbiota contains nearly 11,500 common bacteria.⁹ The total number of bacterial genes in the gut microbiota is 150 times the total number of human genes.¹⁰⁻¹² Coincidentally, the gut microbiota has been called a “forgotten organ”.¹⁰⁻¹² Dysbacteriosis plays an important role in the pathogenesis of several diseases. For example, the gut microbiota is associated with many diseases, such as obesity, type 2 diabetes, and atherosclerosis.¹³⁻¹⁷ Moreover, it is well-known that genomic alterations of the APC/Wnt pathway can potentially lead to carcinoma.¹⁸ Recent studies have increasingly indicated that the gut microbiota is closely related to CRC,¹⁹ with many differences in gut microbiota between healthy people and patients with CRC. Studies have also

Keywords: Gut microbiota; Colorectal cancer; Colorectal adenoma; Random forest model.

Abbreviations: AUC, area under the curve; CRA, colorectal adenoma; CRC, colorectal cancer; CEA, carcinoembryonic antigen; FIT, fecal immunochemical tests; FOBT, fecal occult blood test; OTU, operational taxonomic unit; RF, random forest; TNM, tumor-node-metastasis.

Received: October 24, 2019; Revised: November 21, 2019; Accepted: December 25, 2019

Correspondence to: Quanjiang Dong, Department of Central Laboratories and Gastroenterology, Qingdao Municipal Hospital Affiliated to Dalian Medical University, Shandong 266071, China; Present Address: Donghai Zhong Road 5, Qingdao, Shandong, China. Tel: +86-0532890589; E-mail: jianjacer@126.com

How to cite this article: Sun W, Wang L, Zhang Q, Dong Q. Microbial Biomarkers for Colorectal Cancer Identified with Random Forest Model. Exploratory Research and Hypothesis in Medicine 2020;5(1):19–26. doi: 10.14218/ERHM.2019.00026.
shown that biomarkers from gut microbiota may be used for the diagnosis of CRC.\textsuperscript{20–23} This review focuses on the relationship between gut microbiota and CRC as well as the possibility that gut microbiota-related markers may be novel biological screening markers for CRC.

**Random forest model**

Gut microbiota can be used as a marker to screen for CRC, but the screening capacity of a single flora is limited, and multibacterial models are needed to increase the screening capacity. If too many bacterial species are present, this model will be difficult to apply in the clinical setting; therefore, we need to screen for the most discriminating bacteria. As an integrated machine-learning algorithm, random forest (RF) is a classifier with multiple decision trees that are established through randomly repeated sampling, and a final voting result is obtained.\textsuperscript{24} The current classifier algorithms mainly include, Bayes net, simple logistic, TB-line (TB pedigree), machine learning, sequential minimum optimization algorithm for support vector, and RF.\textsuperscript{25,32}

In many applications, the RF algorithm, by far, has the best accuracy. Moreover, compared to other technologies, RF has many advantages, such as anti-interference, simple optimization, and efficient parallel processing, in processing highly nonlinear biological data.\textsuperscript{33,34} The RF model considers the interaction between nonlinear data and characteristics of the model, and can also carry out internal cross-validation within the model to prevent overfitting.\textsuperscript{35} A recent article reported that an RF model successfully screened microbial markers for early cirrhotic liver cancer.\textsuperscript{36} Another study on hepatocellular carcinoma used the same method to obtain predictive biomarkers for advanced hepatocellular carcinoma.\textsuperscript{37} An RF model based on intestinal gut microbiota was established to successfully predict the variation of Bifidobacteria after probiotic treatment, and it revealed the effects of probiotics on intestinal flora.\textsuperscript{38} Thus, RF models, as a superior machine-learning model, are increasingly used in research for the selection of biomarkers. Therefore, we reviewed the recent progress in the diagnosis of colorectal diseases by using an RF model based on gut microbiota as well as the selection of possible predictive markers of colorectal diseases.

**Dysbiosis of gut microbiota in CRC**

The gut microbiota and its hosts coparticipate in the establishment of a symbiotic relationship to maintain homeostasis in the digestive system. In healthy individuals, at the phyla level, the *Firmicutes* and *Bacteroidetes* phyla are predominant in the gut microbiota, despite remarkable interindividual differences.\textsuperscript{39} At the genus level, a meta-analysis indicated that a high abundance of the genera *Barnesiella, Ruminococcaceae UCG-005, Alistipes, Christensenellaceae R-7 group*, and an unclassified member of the *Lachnospiraceae* family correlated with the healthy state in their subjects.\textsuperscript{40} However, the most abundant bacterial genera in the gut microbiota include *Prevotella, Bacteroides*, and *Ruminococcus*. Based on genus compositional variations, the gut microbiota could be classified into different enterotypes, namely the *Prevotella* predominant enterotype, *Bacteroides* predominant, and *Ruminococcus*-related enterotype, respectively. Simultaneously, this research shows that the human intestinal microbiota has commonalities.\textsuperscript{41}

As usual, the features of bacterial populations are specified by tissue, colonic lumen, and feces, which themselves have different cellular and physiologic features. In the proximal colon, *Bacteroides, Actinomyces, Pseudomonas*, and *Enterobacteriaceae* show differential abundance between the lumen and mucosa.\textsuperscript{42} Similarly, *Enterobacteriaceae, Bacteroides*, and *Pseudomonas* enrich the proximal colonic mucosa, whereas there is increased relative abundance of *Finegoldia, Murdocchiella, Peptoniphilus, Porphyromonas*, and *Anaerococcus* in the distal colon.\textsuperscript{43} The abundance of *Turibacter*, *Finegoldia, Peptoniphilus*, and *Anaerococcus* was found to be different between the lumen and mucosa microbiota in the distal colon.\textsuperscript{44} Furthermore, there are differences in the microbiota compositions between different anatomical parts of the colon.\textsuperscript{42}

Colorectal adenoma (CRA) confers a high risk for the development of CRC but the gut microbiota is necessary for the formation of intestinal adenoma, and healthy gut microbiota are associated with a reduced risk of advanced CRA.\textsuperscript{45} In cases of CRA, the relative abundance of two bacterial genera (*Enterococcus* and *Streptococcus*) increased, whereas that of three genera (*Clostridium, Roseburia*, and *Eubacterium*) decreased.\textsuperscript{46} Another study found that some species belonging to *Ruminococcaceae*, *Clostridium*, *Pseudomonas*, and *Porphyromonadaceae* showed increased numbers in patients with CRA, whereas other species belonging to *Bacteroides, Lachnospiraceae*, *Clostridiales*, and *Clostridium* decreased.\textsuperscript{20} Analyses of fecal microbiota from 95 patients with CRA revealed substantial changes in the microbiota compositions. In CRA, the *Proteobacteria* phylum was found to have enriched the microbiota. These bacteria are associated with precancerous lesions.\textsuperscript{47} *Lachnospiraceae*, a potentially beneficial bacteria genus, was depleted, and the relative abundance of this genus had high accuracy in differentiating patients with CRA from normal individuals.\textsuperscript{23,46}

Dysbiosis of the gut microbiota and its substantial compositional alterations are closely related to the development of CRC.\textsuperscript{47} The microbial communities in tumor microhabitat are different from those in tumor-adjacent healthy tissue.\textsuperscript{48} Moreover, antibiotic intervention in the microbiota can significantly reduce the burden of colonic tumors.\textsuperscript{19} Virulence-associated genes in tumors may potentially depend upon the genomes *Fusobacterium* and *Providence*.\textsuperscript{48} A study on fecal microbiota from 120 patients with CRC showed the increased abundance of a number of species, including *Pophyromonas assacharolytica, Fusobacterium nucleatum, Parvimonas micra, Peptostreptococcus stomatis, Gemella spp.*, and *Prevotella* spp.\textsuperscript{23} Furthermore, in a mouse model of CRC, the abundance of *Lactobacillus* negatively correlated with the number of colonic tumors.\textsuperscript{49} Moreover, fecal microbiota from patients with CRC can promote tumorigenesis in both germ-free mice and conventional mice.\textsuperscript{50} Taking these findings together, it appears that some cancer-associated bacteria in gut microbiota can serve as biomarkers to detect CRC.

**RF model for identification of microbial biomarkers for CRC**

Alterations in the relative abundance of bacteria in CRC indicate they are potential predictive or diagnostic biomarkers for CRC or CRA. *Escherichia coli*, *Bacteroides fragilis*, and *Fusobacterium nucleatum* have been shown to directly influence tumor development in the colon.\textsuperscript{51,52} A recent small-sample study on the quantification of this bacterium in fecal samples found a great increase in the number of these bacteria in CRC; the conclusive findings of the study supported the use of *Fusobacterium nucleatum* as biomarker of CRC as well as a marker of early CRC.\textsuperscript{51–53} The supernatant of a *Fusobacterium nucleatum* culture exhibited strong
bactericidal activity against some probiotics, such as *Faecalibacterium prausnitzii* and *Bifidobacterium* strains, that may cause disease.54

A study found a stepwise increase in the abundance of *Clostridium* from normal tissues to adenoma and, finally, colonic cancer.55 Therefore, *Clostridium symbiosum* can be used singly as a biomarker for detecting CRC. The results showed that a stepped increase in *Clostridium* abundance outperformed all other conventional screening methods, such as carcinoembryonic antigen (referred to as CEA) and FIT, both of which were known to have greater sensitivity (area under the curve (AUC) = 0.73 vs. 0.38–0.54 for other methods). In combination with FIT, the predicting accuracy of *Clostridium symbiosum* increased significantly, with an AUC of 0.803. Moreover, with the combination of *Clostridium symbiosum* and *Fusobacterium nucleatum*, FIT (200 ng/mL) and CEA (3.3 ng/mL) achieved a performance an AUC of 0.876.56

A meta-analysis of a publicly available dataset showed that the depletion of *Faecalibacterium*, *Bacteroides*, and *Romboutsia* could be a potential biomarker for CRC.57 A Chinese study supported the identification of CRC and differentiation from the healthy group via 76 fecal potential biomarkers; the CRC group was enriched with 18 operational taxonomic units (OTUs); moreover, fecal metabolites in healthy patients and cancer groups are different.58 Another meta-analysis of eight studies from different countries and regions identified 29 species as biomarkers of CRC.59 Furthermore, the microbial species can predict taxonomic and functional microbiome CRC signatures as a basis for future diagnostics. (These data are summarized in Table 1.)60,61,62,63,64,65,66,67,68

There have been attempts to explore whether a combination of bacterial markers could increase the AUC value in predicting CRC. Zackular et al.69 analyzed fecal microbiota from healthy subjects and patients with CRA or CRC; they found substantial alterations in the gut microbiome of patients with CRA or CRC compared to healthy controls, with a classification accuracy for CRC of 0.798 AUC. Therefore, combining microbial markers with known clinical risk factors can significantly improve the differentiation ability of the tests.21

By using a LASSO logistic regression classifier, a model constructed with fecal microbiota could predict CRC with an accuracy of 0.82.22 Another study from China demonstrated that a model constructed with microbiota showed better value than FOBT.60 Szé et al.69 constructed an RF classification model using 8 taxa based on significant odds ratios obtained in a meta-analysis of 14 studies from various geographical regions. Their analysis included 1,737 fecal samples and 492 tissue samples. These encompassed *Fusobacterium*, *Parvimonas*, *Porphyromonas*, *Peptostreptococcus*, *Clostridium XI*, *Enterobacteriaceae*, *Ruminococcus*, and *Escherichia*. The combined model had an AUC of 0.75 based on fecal samples. Similarly, the AUC was 0.77 in tissue samples used in a combined model trained by *Dorea*, *Blautia*, and *Weisella*.69 Their model could successfully classify CRC with high accuracy when models trained using one data set were tested on other data sets.69

Nonetheless, a test that objectively reflects the early gut changes in CRA or CRC fully is needed. New noninvasive screening methods are needed to increase the sensitivity and specificity for CRC detection. Baxter et al.23 established an RF classification model by using the relative abundance of gut microbiota and FIT from stool samples of 490 patients. They observed that the sensitivity and specificity of a combination model of bacterial abundance and FIT, which they obtained by incorporating data on hemoglobin concentration (determined by FIT), and bacterial relative abundances (multitarget microbiota test) for CRC and CRA were better than those with FIT alone.23 Their model used 25 OTUs, including *Lachnospiraceae* (OTU87), *Lachnospiraceae* (OTU60), *Lachnospiraceae* (OTU32), *Lachnospiraceae* (OTU88), *Lachnospiraceae* (OTU44), *Lachnospiraceae* (OTU14), *Bacteroides* (OTU7), *Bacteroides* (OTU3), *Bacteroides* (OTU2), *Ruminococcus* (OTU11), *Ruminococcus* (OTU16), *Ruminococaceae* (OTU29), *Blautia* (OTU13), *Blautia* (OTU9), *Collinsella* (OTU19), *Firmicutes* (OTU282), *Enterobacteriaceae* (OTU28), *Parabacteroides* (OTU49), *Roseburia* (OTU5), *Clostridiales* (OTU10), *Faecalibacterium* (OTU6), *Anaerostipes* (OTU8), *Porphyromonas* (OTU105), and FIT with a 100 ng/mL cutoff. We infer that 16 of these were members of the *Firmicutes* phylum.21 The multtarget microbiota test detected 91.7% of cancers and 45.5% of adenomas, compared to 75.0% and 15.7% by FIT, respectively.23 Thus, screening methods for colorectal lesions need to be continually optimized to find the optimal screening program. A recent study analyzed a total of 969 fecal mate genomes, including 5 publicly available data sets, 2 new cohorts, and 2 validation cohorts.62 Twenty-four species with high RF accuracy features were selected; these were: *Actinomyces graevenitzii*, *Alstipes* spp., *Anaerotongus coliformis*, *Bifidobacterium longum*, *Clostridium hathewayi*, *Clostridium leptum*, *Clostridium symbiosum*, *Dialister invisus*, *Eubacterium eligens*, *Escherichia coli*, *Fusobacterium nucleatum*, *Gemella morbillorum*, *Lachnospiraceae* 3157F AA CT1, *Lachnospiraceae* 8157F AA, *Lachnospiraceae* 5163F AA, *Parvimonas* spp., *Peptostreptococcus stomaticis*, *Porphyromonas asaccharolytica*, *Pravimonas micra*, *Prevotella copri*, *Ruminococcus gnarus*, *Subdoligranulum* spp., *Streptococcus parasanguinis*, and *Streptococcus salivarius*. The predictive microbiome signatures trained on different data sets consistently showed high accuracy. Nonetheless, it appears their model has lower sensitivity and specificity values for predicting CRA.58

Despite significant advances in the study of the effects of gut microbiota on colorectal lesions, few studies have investigated the gut microbiota after the treatment of patients with colorectal lesions. The tumor-node-metastasis (commonly known as TNM) international staging system has always been considered the gold standard to determine CRC prognosis.63 In addition, findings of aneuploidy, tumor-infiltrating lymphocytes, allelic loss in DCC, TP53, APC and MCC genes, TP53 gene mutations, CD44 protein expression, high levels of thymidylate synthetase, microsatellite instability, and gene studies of both *RAS* and *BRAF* are independent, strong prognostic factors. In addition, C-reactive protein, overexpression of the CEA in tumors, and circulating free DNA are considered to be associated with the prognosis of patients with CRC.64 Ai et al.65 analyzed the composition of fecal microbiota in 124 samples from France and 99 samples from Austria. They excluded unrelated and redundant features during feature selection by mutual information, and trained an RF classifier on a large mate genomic data set of patients with CRC and healthy individuals. The RF classifier assembled from published reports as well as extracted and analyzed information from learned decision trees. *Porphyromonas asaccharolytica*, *Peptostreptococcus stomaticis*, *Fusobacterium*, *Parvimonas* spp., *Streptococcus vestibularis*, and *Flavonifractor plautii* were determined to be key microbial species associated with CRCs.65

By using an RF model based on fecal microbiota, Szé et al.61 found significant differences between the pre- and post-treatment samples of 67 individuals, including those with adenoma (*n* = 22), advanced adenoma (*n* = 19), and carcinoma (*n* = 26). *Fusobacterium*, *Porphyromonas*, and *Parvimonas* were significantly decreased in the post-treatment samples.61 Furthermore, in a mouse model, interventions of microbiota with antibiotics led to a dramatic decrease in the tumor burden in the color.65 In addition, as an important probiotic, *Bifidobacterium* has been shown to enrich the gut microbiome in healthy individuals.61 Moreover, studies
| Phyla                  | Genus                                | Abundance | Reference                                                      |
|-----------------------|--------------------------------------|-----------|----------------------------------------------------------------|
| **Firmicutes**        | **Clostridiales**                    | Increased | Xie YH (2017)                                                   |
| **Unknown Clostridiales** | **Clostridium bolteae/clostridioforme** | Increased | Wirbel J et al. (2019)                                          |
| **Clostridium symbiosum** | Increased | Wirbel J et al. (2019) | Thomas AM et al. (2019)                                        |
| **Clostridium leptum** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Clostridium hathewayi** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Unknown Clostridiales** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Subdoligranulum spp.** | Decreased | Thomas AM et al. (2019) |                                                                 |
| **Unknown Peptostreptococcaceae** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Peptostreptococcus stomatis** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Anaeroccoccusbesiensis/vaginalis** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Anaeructuncus colihominis** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Gemella morbillorum** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Unknown Dialister** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Hungatellahathewayi** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Parvimonas species** | Increased | Wirbel J et al. (2019) | Ai D et al. (2019)                                              |
| **Parvimonas spp.** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Praevimonas micra** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Ruminococcus torques** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Ruminococcus gnavus** | Decreased | Thomas AM et al. (2019) |                                                                 |
| **Uubdoligranulum species** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Lachnospiaceae 3157FAA CT1** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Lachnospiaceae 8157FAA** | Decreased | Thomas AM et al. (2019) |                                                                 |
| **Lachnospiaceae5163FAA** | Decreased | Thomas AM et al. (2019) |                                                                 |
| **Alistipes spp.** | Decreased | Sze MA et al. (2018) |                                                                 |
| **Dialister invisus** | Decreased | Thomas AM et al. (2019) |                                                                 |
| **Eubacterium eligens** | Decreased | Thomas AM et al. (2019) |                                                                 |
| **Streptococcus parasanguinis** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Streptococcus salivarius** | Decreased | Thomas AM et al. (2019) |                                                                 |
| **Streptococcus vestibularis** | Increased | Ai D et al. (2019) |                                                                 |
| **Bacteroidetes**     | **Bacteroides**                      | Decreased | Mangifesta M et al. (2018)                                      |
| **Unknown Porphyromonas** | **Porphyromonas uenonis** | Increased | Wirbel J et al. (2019)                                          |
| **Porphyromonas asaccharolytica** | Increased | Ai D et al. (2019) |                                                                 |
| **Prevotella intermedia** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Prevotella lingens** | Increased | Wirbel J et al. (2019) |                                                                 |
| **Prevotella copri** | Increased | Thomas AM et al. (2019) |                                                                 |
| **Flavonifractor plautii** | Increased | Ai D et al. (2019) |                                                                 |
| **Proteobacteria**    | **Faecalibacterium**                 | Decreased | Mangifesta M et al. (2018)                                      |
| **Escherichia coli**  | Increased | Thomas AM et al. (2019) |                                                                 |
have shown that *Bifidobacterium* can inhibit the growth of intestinal carcinogenic bacteria and protect intestinal mucosa, which makes it an important probiotic for clinical application (Table 2).\(^{20,23,42,44,49,59,61,62}\)

The gut microbiota includes bacteria, viruses, and fungi. Despite the close relation of bacteria to colorectal lesions, gut microbes can interact with each other. Through RF modeling, Hannigan *et al.*\(^{63}\) found that viruses indirectly affect cancer progression by altering bacterial host communities. Nakatsu *et al.*\(^{64}\) conducted a study on the survival prediction of CRC by viruses. Their study

| Phyla          | Genus                          | Abundance     | Reference                                                                 |
|----------------|--------------------------------|---------------|---------------------------------------------------------------------------|
| *Fusobacteria* | *Fusobacterium nucleatum*      | Increased     | Tunsjø HS *et al.* (2019)\(^{52}\); Mangifesta M *et al.* (2018)\(^{53}\); Xie YH (2017)\(^{53}\); Bullman S *et al.* (2017)\(^{54}\); Thomas AM *et al.* (2019)\(^{55}\); Ai D *et al.* (2019)\(^{59}\) |
|                | *F. nucleatum subspecies animalis* | Increased     | Wirbel J *et al.* (2019)\(^{57}\)                                             |
|                | *F. nucleatum subspecies nucleatum* | Increased     | Wirbel J *et al.* (2019)\(^{57}\)                                             |
|                | *F. nucleatum subspecies vincentii* | Increased     | Wirbel J *et al.* (2019)\(^{57}\)                                             |
|                | *Fusobacterium species oral taxon 370* | Increased     | Wirbel J *et al.* (2019)\(^{57}\)                                             |
| *Actinobacteria* | *Actinomyces graevenitzii*      | Increased     | Thomas AM *et al.* (2019)\(^{58}\)                                           |
| *Tenericutes*  | *Bifidobacterium longum*       | Decreased     | Thomas AM *et al.* (2019)\(^{58}\)                                           |
|                | *Solobacterium moorei*         | Increased     | Wirbel J *et al.* (2019)\(^{57}\)                                             |

Table 1. Characteristics of the bacteria species as potential biomarkers for CRC - (continued)

| Participants | Sample sources | Models | Bacteria, genus | Bacteria features |
|--------------|----------------|--------|-----------------|-------------------|
| Flynn *et al.* (2018)\(^{42}\) | Healthy (n = 20) | Mucosal, feces and luminal contents | Random forest classification models | *Enterobacteriaceae, Bacteroides and Pseudomonas* | Enriched in the proximal |
| Chen *et al.* (2013)\(^{44}\) | Healthy (n = 344), A-CRA groups (n = 344) | Feces | NA | *Enterococcus and Streptococcus* | Increased in A-CRA; Decreased in A-CRA |
| Zackular *et al.* (2014)\(^{20}\) | Healthy (n = 30), colonic adenoma (n = 30), and colonic adenocarcinoma (n = 30) | Feces | NA | *Ruminococcaceae, Clostridium, Pseudomonas, and Porphyromonadaceae* | Increase in CRA patients |
| Goedert *et al.* (2015)\(^{45}\) | Normal patients (n = 24), CRA (n = 20), CRC (n = 2), and other conditions (n = 15) | Feces | Random forest | *Proteobacteria* | Enriched in CRA |
| Ai D *et al.* (2019)\(^{59}\) | France (n = 124); Austria (n = 99) | Feces | Random forest | *Porphyromonas asaccharolytica, Eubacterium hallii, Parvimonas spp., Fusobacterium 7, Prevotella melaninogenica, Streptococcus vestibularis, Prevotellacapri, Peptostreptococcus stomatis, Fusobacterium nucleatum, Parvimonas micra, Gemella morbillorum, Flavonifractor placiti, Fusobacterium, Clostridium SS2* | Enriched in CRC |

Table 2. Important studies in dysbiosis of gut microbiota in CRC, bacterial features, detection methods and models
found a combination of four classification markers that reduce patient survival in CRC. Further research on CRC-related viral group characteristics could lead to the development of new tools to identify individuals with CRC or to predict outcomes.

There are some limitations of our meta-analysis. Despite the available data on CRC detection through the gut microbiome, there is a lack of consensus on which features are most informative. The contradictory reports from some studies could be attributed to differences inherent among study populations, procedures for fecal collection and storage, DNA extraction and amplification, sequencing, and bioinformatics processing methods. Moreover, recent studies used their models only to differentiate CRC from CRA or healthy individuals. Further studies are required to identify methods that can differentiate CRC from other colonic diseases, such as inflammatory bowel diseases.

**Future research directions**

Our analysis of recent studies on CRC biomarkers and the list of related genera showed the absence of an accepted biomarker for CRC. The RF model has features of anti-interference, simple optimization and efficient parallel processing, all of which imply it may be the best choice for screening biomarkers. Future research to develop a kit to accurately screen for CRA and CRC biomarkers through an RF model for fecal microbiota could accurately, quickly and conveniently improve early detection of these conditions.

**Conclusion**

In recent years, several studies have demonstrated that the gut microbiota in CRC patients differs substantially from that in healthy individuals. Fecal microbial markers have the potential to provide a noninvasive alternative method to diagnose CRC. RF models or other statistical models based on a collection of bacteria in the gut microbiota could help identify CRC with high accuracy. When combined with other conventional screening markers and clinical risk factors, the predictive accuracy for CRC increases dramatically. The findings in our review provide a new approach to identify powerful biomarkers in the gut microbiota. This will facilitate clinician decision-making for early intervention in CRC.

**Funding**

This work was supported by funding from the National Natural Science Foundation of China (No. 81602144, 31870777)

**Conflict of interest**

The authors have no financial interests or any conflict of interests to disclose.

**Author contributions**

Manuscript writing (WLS, QJD); critical revision of the manuscript for important intellectual content (WLS, LLW, QYZ); administrative, technical, or material support, study supervision (QJD).

**References**

[1] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61(2):69–90. doi:10.3322/caac.20107.

[2] Brenner H, Hoffmeister M, Stegmaier C, Brenner G, Altenhofen L, Haug U. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. Gut 2007;56(11):1585–1589. doi:10.1136/gut.2007.122739.

[3] Kunst KM, Lansdorp-Vogelaar I, Rutter CM, Knudsen AB, van Baal gooijen M, Savarino JE, et al. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. Med Decis Making 2011;31(4):530–539. doi:10.1177/0272989X11408730.

[4] O’Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst 2004;96(19):1420–1425. doi:10.1093/jnci/djh275.

[5] Kaminski MF, Regula J, Krasszewska E, Poikolli M, Wojciechowska U, Didkowska J, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010;362(19):1795–1803. doi:10.1056/NEJMoa090767.

[6] Lebwohl B, Kastrinos F, Glick M, Rosenbaum AJ, Wang T, Neugut AI. The impact of suboptimal bowel preparation on adenoma miss rates and the factors associated with early repeat colonoscopy. Gastrointest Endosc 2011;73(6):1207–1214. doi:10.1016/j.gie.2011.01.051.

[7] Kuipers EJ, Rösch T. Colorectal cancer screening—optimizing current...
Sun W. et al.: Microbial biomarkers of CRC

Explor Res Hypothesis Med

strategies and new directions. Nat Rev Clin Oncol 2013;10:130–142. doi:10.1038/nrclinonc.2013.12.

[8] Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diver-
sity, stability and resilience of the human gut microbiota. Nature 2012;489(7415):220–230. doi:10.1038/nature11505.

[9] Morgan XC, Segata N, Huttenhower C. Biodiversity and functional genomics in the human microbiome. Trends Genet 2013;29(1):51–
58. doi:10.1016/j.tig.2012.09.005.

[10] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbiome gene catalogue established by metagenomic
sequencing. Nature 2010;464:56–59. doi:10.1038/nature08821.

[11] O’Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO
Rep 2006;7:689–693. doi:10.1038/sj.embor.7400731.

[12] Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fager
Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, O’Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO
DOI: 10.14218/ERHM.2019.00026  |  V olume 5 Issue 1, March 2020 25

[13] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Zackular JP, Baxter NT, Chen GY, Schloss PD. Manipulation of the
human gut microbiota reveals role in colon tumorigenesis. Cancer Prev Res 2011;4(8):730–37. doi:10.1158/1940-6207.CAPR-11-0213.

[14] Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide
association study of gut microbiota in type 2 diabetes. Nature 2012;490(7418):55–60. doi:10.1038/nature11450.

[15] Koehler RA, Wang Z, Levison BS, Bufa JA, Org E, Shiehby BT, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, pro-
motes satherosclerosis. Nat Med 2013;19(5):576–585. doi:10.1038/nm.3145.

[16] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Zackular JP, Baxter NT, Chen GY, Schloss PD. Microbiota-based mod-

[17] Sun W. Manipulation of the gut microbiota metabolism of L-carnitine, a nutrient in red meat, pro-
motes satherosclerosis. Nat Med 2013;19(5):576–585. doi:10.1038/nm.3145.

[18] Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Comprehensive molecular charac-
terization of human colon and rectal cancer. Nature 2012;487(7407):
330–337. doi:10.1038/nature11252.

[19] Zackular JP, Baxter NT, Chen GY, Schloss PD. Mapping of the gut microbiota reveals role in colon tumorigenesis. mSphere 2015;1(1):e00001–15. doi:10.1128/mSphere.00001-15.

[20] Zackular JP, Rogers MA, Ruffin MT, Schloss PD. The human gut micro-
bio as a screening tool for colorectal cancer. Cancer Prev Res 2014;7:1112–1121. doi:10.1158/1940-6207.CAPR-14-0129.

[21] Zeller G, Tap J, Voigt AV, Sunagawa S, Kultima JR, Costea PI, et al. Poten-
tial of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 2014;10:766. doi:10.15252/msb.20145645.

[22] Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagen-
omic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma. Gut 2019;68(6):
1014–1023. doi:10.1136/gutjnl-2017-315084.

[23] Loomba R, Seguritan V, Li W, Long T, Kittgold N, Bhattacharya A, et al. Gut micro-
bio-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. Cell Metab 2017;25(5):1054–1062.e5. doi:10.1016/j.cmet.2017.04.001.

[24] Luo YM, Liu FT, Chen MX, Tang WL, Yang YL, et al. A machine learning model based on initial gut microbiota data for predicting changes of Bifidobacterium after probiotics consumption. Nan Fang Yi Ke Da Xue Bao 2018;38(3):251–260.

[25] Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbi-
ome. Genome Med 2016;8(1):51. doi:10.1186/s13073-016-0307-y.

[26] Mancabelli L, Milani C, Lugli GA, Turroni F, Cocconi D, van Sinderen D, et al. Identification of universal gut microbiomal signatures of common human intestinal diseases by meta-analysis. FEMS Microbiol Ecol 2017;93(12):fxz153. doi:10.1093/femsme/fxz153.

[27] Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mendez
DR, et al. Enterotypes of the human gut microbiome. Nature 2011;473(7346):174–180. doi:10.1038/nature09944.

[28] Flynn RJ, Ruffin MT, urgeson DK, Schloss PD. Spatial Variation of the Native Colon Microbiota in Healthy Adults. Cancer Prev Res (Phila) 2018;11(7):393–402. doi:10.1158/1940-6207.CAPR-17-0370.

[29] Dove WF, Clipson L, Gould KA, Loungo C, Marshall DJ, Moser AR, et al. Intestinal neoplasia in the ApCMin mouse: independence from
the microbial and natural killer (beige locus) status. Cancer Res 1997;57(5):812–814.

[30] Chen HM, Yu YN, Wang JL, Lin YW, Kong X, Yang CQ, et al. De-
creased dietary fiber intake and structural alteration of gut micro-
bio in patients with advanced colorectal adenoma. Am J Clin Nutr 2013;97(5):1045–1052. doi:10.3945/ajcn.112.046607.

[31] Goedert JJ, Gong Y, Hua X, Zhong H, He Y, Peng P, et al. Fecal microbiota characteristics of patients with colorectal adenoma detected by screening: a population-based study. EBioMedicine 2015;2(6):597–
603. doi:10.1016/j.ebiom.2015.04.010.

[32] Lepage P, Häsel R, Spehlmann ME, Rehman A, Zvirbliene A, Begun D, et al. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. Gastroenterology 2011;141(1):227–236. doi:10.1053/j.gastro.2010.04.011.

[33] Arthur JC, Gharabehiz RZ, Mühlbauer M, Perez-Chanona E, Uronis JM, McCafferty J, et al. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. Nat Commun 2014;5:4724. doi:10.1038/ncomms5724.

[34] Burns MB, Lynch J, Starr TK, Knights D, Blekman R. Virulence genes are a signature of the microbiome in the colorectal tumor microen-
vironment. Genome Med 2015;7(1):55. doi:10.1186/s13073-015-

DOI: 10.14218/ERHM.2019.00026  |  Volume 5 Issue 1, March 2020 25
0177-8.

[49] Sze MA, Schloss PD. Leveraging existing 16S rRNA gene surveys to identify reproducible biomarkers in individuals with colorectal tumors. mBio 2018;9(3):e00630–18. doi:10.1128/mBio.00630-18.

[50] 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.

[51] Guo SH, Li LF, Xu BL, Li MH, Zeng QY, Xiao H, et al. A simple and novel fecal biomarker for colorectal cancer: ratio of to probiotics populations, based on their antagonistic effect. Clin Chem 2018;64(9):1327–1337. doi:10.1373/clinchem.2018.289728.

[52] Tunsjø HS, Gundersen G, Rangnes F, Noone JC, Endres A, Bemanian V. Detection of Fusobacterium nucleatum in stool and colonic tissues from Norwegian colorectal cancer patients. Eur J Clin Microbiol Infect Dis 2019;38(7):1367–1376. doi:10.1007/s10096-019-03562-7.

[53] Mangifesta M, Mancabelli L, Milani C, Gaiani F, de'Angelis N, de'Angelis GL, et al. Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer. Sci Rep 2018;8(1):13974. doi:10.1038/s41598-018-32413-2.

[54] Bullman S, Pedamallu CS, Scinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science 2017;358(6369):1443–1448. doi:10.1126/science.aal5240.

[55] Xie YH, Gao QY, Cai GX, Sun XM, Zou TH, Chen HM, et al. Fecal Clostridium sym biosum for noninvasive detection of early and advanced colorectal cancer: test and validation studies. EBio Medicine 2017;25:32–40. doi:10.1016/j.ebiom.2017.10.005.

[56] Yang Y, Misra BB, Liang L, Bi D, Weng W, Wu W, et al. Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. Theranostics 2019;9(14):4101–4114. doi:10.7150/thno.35186.

[57] Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal meta genomes reveals global microbiota signatures that are specific for colorectal cancer. Nat Med 2019;25(4):679–689. doi:10.1038/s41591-019-0406-6.

[58] Thomas AM, Manghi P, Asnicar F, Pasolini 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.

[59] Ai D, Pan H, Han R, Li X, Liu G, Xia LC. Using decision tree aggregation with random forest model to identify gut microbes associated with colorectal cancer. Genes (Basel) 2019;10(2):E112. doi:10.3390/genes10020112.

[60] Ai L, Tian H, Chen Z, Chen H, Xu J, Fang JY. Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer. Oncotarget 2017;8(6):9546–9556. doi:10.18632/oncotarget.14488.

[61] Sze MA, Baxter NT, Ruffin MT4th, Rogers MAM, Schloss PD, et al. Normalization of the microbiota in patients after treatment for colonic lesions. Microbiome 2017;5(1):150. doi:10.1186/s40168-017-0366-3.

[62] Pinzone MR, Celesia BM, Di Rosa M, Cacopardo B, Nunnari G. Microbial translocation in chronic liver diseases. Int J Microbiol 2012;2012:694629. doi:10.1155/2012/694629.

[63] Hannigan GD, Duhaime MB, Ruffin MT4th, Koumpouras CC, Schloss PD. Diagnostic potential and interactive dynamics of the colorectal cancer virome. mBio 2018;9(6):e02248–18. doi:10.1128/mBio.02248-18.

[64] Nakatsu G, Zhou H, Wu WKK, Wong SH, Coker OO, Dai Z, et al. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. Gastroenterology 2018;155(2):529–541.e5. doi:10.1053/j.gastro.2018.04.018.