



# Microbial Biomarkers for Colorectal Cancer Identified with Random Forest Model

Weili Sun<sup>1</sup>, Lili Wang<sup>2</sup>, Qiuyue Zhang<sup>2</sup> and Quanjiang Dong<sup>1,2\*</sup>

<sup>1</sup>The Affiliated Hospital of Dalian Medical University, Shandong, China; <sup>2</sup>Department 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 progressions 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.<sup>1</sup> CRC is characterized by a long preclinical phase, progressing over years from early adenoma to invasive cancer.<sup>2,3</sup> Diagnosis of early-stage CRC can significantly improve patient prognosis, with a survival rate higher than 80%,<sup>4</sup> 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 di-

agnosis. 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.<sup>5,6</sup> 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.<sup>7</sup> 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 10<sup>14</sup> bacteria, which equates to 10 times the total number of human cells.<sup>8</sup> The gut microbiota contains nearly 11,500 common bacteria.<sup>9</sup> The total number of bacterial genes in the gut microbiota is 150 times the total number of human genes.<sup>10–12</sup> Coincidentally, the gut microbiota has been called a “forgotten organ”.<sup>10–12</sup>

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.<sup>13–17</sup> Moreover, it is well-known that genomic alterations of the APC/Wnt pathway can potentially lead to carcinoma.<sup>18</sup> Recent studies have increasingly indicated that the gut microbiota is closely related to CRC,<sup>19</sup> 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: +86-05328890589; E-mail: jiangacer@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.<sup>20–23</sup> 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.<sup>24</sup> 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.<sup>25–32</sup>

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.<sup>33,34</sup> 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.<sup>35</sup> A recent article reported that an RF model successfully screened microbial markers for early cirrhotic liver cancer.<sup>36</sup> Another study on hepatocellular carcinoma used the same method to obtain predictive biomarkers for advanced hepatocellular carcinoma.<sup>37</sup> 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.<sup>38</sup> 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.<sup>39</sup> 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.<sup>40</sup> 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.<sup>41</sup>

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.<sup>42</sup> Similarly, *Enterobacteriaceae*, *Bacteroides*, and *Pseudomonas* enrich the proximal colonic mucosa, whereas there is increased relative abundance of *Finegoldia*, *Murdochiella*, *Peptoniphilus*, *Porphyromonas*, and *Anaerococcus* in the distal colon.<sup>42</sup> The abundance of *Turicibacter*, *Finegoldia*, *Peptoniphilus*, and *Anaerococcus* was found to be different between the lumen and mucosa microbiota in the distal colon.<sup>42</sup> Furthermore, there are differences in the microbiota compositions between different anatomical parts of the colon.<sup>42</sup>

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.<sup>43</sup> 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.<sup>44</sup> 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.<sup>20</sup> 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.<sup>45</sup> *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.<sup>23,46</sup>

Dysbiosis of the gut microbiota and its substantial compositional alterations are closely related to the development of CRC.<sup>47</sup> The microbial communities in tumor microhabitat are different from those in tumor-adjacent healthy tissue.<sup>48</sup> Moreover, antibiotic intervention in the microbiota can significantly reduce the burden of colonic tumors.<sup>19</sup> Virulence-associated genes in tumors may potentially depend upon the genomes *Fusobacterium* and *Providencia*.<sup>48</sup> A study on fecal microbiota from 120 patients with CRC showed the increased abundance of a number of species, including *Porphyrromonas assaccharolytica*, *Fusobacterium nucleatum*, *Parvimonas micra*, *Peptostreptococcus stomatis*, *Gemella* spp., and *Prevotella* spp.<sup>23</sup> Furthermore, in a mouse model of CRC, the abundance of *Lactobacillus* negatively correlated with the number of colonic tumors.<sup>49</sup> Moreover, fecal microbiota from patients with CRC can promote tumorigenesis in both germ-free mice and conventional mice.<sup>50</sup> 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.<sup>19</sup> 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.<sup>51–53</sup> 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.<sup>54</sup>

A study found a stepwise increase in the abundance of *Clostridium* from normal tissues to adenoma and, finally, colonic cancer.<sup>55</sup> 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.<sup>55</sup>

A meta-analysis of a publicly available dataset showed that the depletion of *Faecalibacterium*, *Bacteroides*, and *Romboutsia* could be a potential biomarker for CRC.<sup>52</sup> 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.<sup>56</sup> Another meta-analysis of eight studies from different countries and regions identified 29 species as biomarkers of CRC.<sup>57</sup> 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.)<sup>49,52–55,57–59</sup>

There have been attempts to explore whether a combination of bacterial markers could increase the AUC value in predicting CRC. Zackular *et al.*<sup>20</sup> 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.<sup>21</sup>

By using a LASSO logistic regression classifier, a model constructed with fecal microbiota could predict CRC with an accuracy of 0.82.<sup>22</sup> Another study from China demonstrated that a model constructed with microbiota showed better value than FOBT.<sup>60</sup> Sze *et al.*<sup>49</sup> 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 *Weissella*.<sup>49</sup> Their model could successfully classify CRC with high accuracy when models trained using one data set were tested on other data sets.<sup>49</sup>

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.*<sup>23</sup> 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.<sup>23</sup> Their model used 23 OTUs, including *Lachnospiraceae*

(OTU87), *Lachnospiraceae* (OTU60), *Lachnospiraceae* (OTU32), *Lachnospiraceae* (OTU88), *Lachnospiraceae* (OTU44), *Lachnospiraceae* (OTU14), *Bacteroides* (OTU7), *Bacteroides* (OTU3), *Bacteroides* (OTU2), *Ruminococcus* (OTU11), *Ruminococcus* (OTU16), *Ruminococcaceae* (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.<sup>23</sup> The multitarget microbiota test detected 91.7% of cancers and 45.5% of adenomas, compared to 75.0% and 15.7% by FIT, respectively.<sup>23</sup> 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.<sup>58</sup> Twenty-four species with high RF accuracy features were selected; these were: *Actinomyces graevenitzi*, *Alistipes* spp., *Anaeruncus colihominis*, *Bifidobacterium longum*, *Clostridium hathewayi*, *Clostridium leptum*, *Clostridium symbiosum*, *Dialister invisus*, *Eubacterium eligens*, *Escherichia coli*, *Fusobacterium nucleatum*, *Gemella morbillorum*, *Lachnospiraceae* 3157FAA CTI, *Lachnospiraceae* 8157FAA, *Lachnospiraceae* 5163FAA, *Parvimonas* spp., *Peptostreptococcus stomatis*, *Porphyromonas asaccharolytica*, *Pravimonas micra*, *Prevotella copri*, *Ruminococcus gnavus*, *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.<sup>58</sup>

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. 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.<sup>19</sup> Ai *et al.*<sup>59</sup> 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 stomatis*, *Fusobacterium*, *Parvimonas* spp., *Streptococcus vestibularis*, and *Flavonifractor plautii* were determined to be key microbial species associated with CRCs.<sup>59</sup>

By using an RF model based on fecal microbiota, Sze *et al.*<sup>61</sup> 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.<sup>61</sup> Furthermore, in a mouse model, interventions of microbiota with antibiotics led to a dramatic decrease in the tumor burden in the colon.<sup>19</sup> In addition, as an important probiotic, *Bifidobacterium* has been shown to enrich the gut microbiome in healthy individuals.<sup>61</sup> Moreover, studies

**Table 1. Characteristics of the bacteria species as potential biomarkers for CRC**

| Phyla                                | Genus                                     | Abundance                    | Reference  |   |
|--------------------------------------|---|------------------------------|--|---|
| Firmicutes                           | <i>Clostridiales</i>                      | Increased                    | Xie YH (2017) <sup>55</sup>  |   |
|                                      | <i>Unknown Clostridiales</i>              | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Clostridium boltea/clostridioforme</i> | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Clostridium symbiosum</i>              | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup> ; Thomas AM <i>et al.</i> (2019) <sup>58</sup> |   |
|                                      | <i>Clostridium leptum</i>                 | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Clostridium hathewayi</i>              | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Unknown Clostridiales</i>              | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Subdoligranulum spp.</i>               | Decreased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Unknown Peptostreptococcaceae</i>      | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Peptostreptococcus stomatis</i>        | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup> ; Ai D <i>et al.</i> (2019) <sup>59</sup>     |   |
|                                      | <i>Anaerococcusobesiensis/vaginalis</i>   | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Anaertuncus colihominis</i>            | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Gemella morbillorum</i>                | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup> ; Thomas AM <i>et al.</i> (2019) <sup>58</sup> |   |
|                                      | <i>Unknown Dialister</i>                  | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Hungatellahathewayi</i>                | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Parvimonas species</i>                 | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup> ; Ai D <i>et al.</i> (2019) <sup>59</sup>      |   |
|                                      | <i>Parvimonas spp.</i>                    | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Pravimonas micra</i>                   | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Ruminococcus torques</i>               | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Ruminococcus gnavus</i>                | Decreased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Uubdoligranulum species</i>            | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
|                                      | <i>Lachnospiraceae 3157FAA CT1</i>        | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Lachnospiraceae 8157FAA</i>            | Decreased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Lachnospiraceae5163FAA</i>             | Decreased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Alistipes spp.</i>                     | Decreased                    | Sze MA <i>et al.</i> (2018) <sup>49</sup>  |   |
|                                      | <i>Dialister invisus</i>                  | Decreased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Eubacterium eligens</i>                | Decreased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Streptococcus parasanguinis</i>        | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Streptococcus salivarius</i>           | Decreased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
|                                      | <i>Streptococcus vestibularis</i>         | Increased                    | Ai D <i>et al.</i> (2019) <sup>59</sup>  |   |
|                                      | Bacteroidetes                             | <i>Bacteroides</i>           | Decreased  | Mangifesta M <i>et al.</i> (2018) <sup>53</sup> |
|                                      |   | <i>Unknown Porphyromonas</i> | Increased  | Wirbel J <i>et al.</i> (2019) <sup>57</sup>     |
|                                      |   | <i>Porphyromonas uenonis</i> | Increased  | Wirbel J <i>et al.</i> (2019) <sup>57</sup>     |
| <i>Porphyromonas somerae</i>         |   | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
| <i>Porphyromonas asaccharolytica</i> |   | Increased                    | Ai D <i>et al.</i> (2019) <sup>59</sup> ; Thomas AM <i>et al.</i> (2019) <sup>58</sup>     |   |
| <i>Prevotella intermedia</i>         |   | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
| <i>Prevotellan igrescens</i>         |   | Increased                    | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |   |
| <i>Prevotella copri</i>              |   | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |
| <i>Flavonifractor plautii</i>        |   | Increased                    | Ai D <i>et al.</i> (2019) <sup>59</sup>  |   |
| Proteobacteria                       | <i>Faecalibacterium</i>                   | Decreased                    | Mangifesta M <i>et al.</i> (2018) <sup>53</sup>  |   |
|                                      | <i>Escherichia coli</i>                   | Increased                    | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |   |

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

| Phyla          | Genus                                       | Abundance | Reference  |
|----------------|---|-----------|--|
| Fusobacteria   | <i>Fusobacterium nucleatum</i>              | Increased | Tunsgjød HS <i>et al.</i> (2019) <sup>52</sup> ; Mangifesta M <i>et al.</i> (2018) <sup>53</sup> ; Xie YH (2017) <sup>55</sup> ; Bullman S <i>et al.</i> (2017) <sup>54</sup> ; Thomas AM <i>et al.</i> (2019) <sup>58</sup> ; Ai D <i>et al.</i> (2019) <sup>59</sup> |
|                | <i>F. nucleatum subspecies animalis</i>     | Increased | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |
|                | <i>F. nucleatum subspecies nucleatum</i>    | Increased | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |
|                | <i>F. nucleatum subspecies vincentii</i>    | Increased | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |
|                | <i>Fusobacterium species oral taxon 370</i> | Increased | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |
| Actinobacteria | <i>Actinomyces graevenitzi</i>              | Increased | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |
|                | <i>Bifidobacterium longum</i>               | Decreased | Thomas AM <i>et al.</i> (2019) <sup>58</sup>   |
| Tenericutes    | <i>Solobacterium moorei</i>                 | Increased | Wirbel J <i>et al.</i> (2019) <sup>57</sup>  |

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).<sup>20,23,42,44,45,49,59,61,62</sup>

The gut microbiota includes bacteria, viruses, and fungi. De-

spite the close relation of bacteria to colorectal lesions, gut microbes can interact with each other. Through RF modeling, Hannigan *et al.*<sup>63</sup> found that viruses indirectly affect cancer progression by altering bacterial host communities. Nakatsu *et al.*<sup>64</sup> conducted a study on the survival prediction of CRC by viruses. Their study

**Table 2. Important studies in dysbiosis of gut microbiota in CRC, bacterial features, detection methods and models**

|   | Participants   | Sample sources                      | Models                              | Bacteria, genus   | Bacteria features                            |
|---|--|-------------------------------------|-------------------------------------|---|--|
| Flynn <i>et al.</i> (2018) <sup>42</sup>    | Healthy (n = 20)   | Mucosal, feces and luminal contents | Random forest classification models | <i>Enterobacteriaceae, Bacteroides</i> and <i>Pseudomonas</i>   | Enriched in the proximal                     |
|   |  |                                     |                                     | <i>Finegoldia, Murdochiella, Peptoniphilus, Porphyromonas</i> and <i>Anaerococcus</i>   | Colonic mucous increased in the distal colon |
| Chen <i>et al.</i> (2013) <sup>44</sup>     | Healthy (n = 344), A-CRA groups (n = 344)  | Feces                               | NA                                  | <i>Enterococcus</i> and <i>Streptococcus</i>  | Increased in A-CRA;                          |
|   |  |                                     |                                     | <i>Clostridium, Roseburia, and Eubacterium</i>  | Decreased in A-CRA                           |
| Zackular <i>et al.</i> (2014) <sup>20</sup> | Healthy (n = 30), colonic adenoma (n = 30), and colonic adenocarcinoma (n = 30)    | Feces                               | NA                                  | <i>Ruminococcaceae, Clostridium, Pseudomonas, and Porphyromonadaceae</i>  | Increase in CRA patients                     |
|   |  |                                     |                                     | <i>Bacteroides, Lachnospiraceae, Clostridiales, and Clostridium</i>   | Decreased in CRA patients                    |
| Goedert <i>et al.</i> (2015) <sup>45</sup>  | Normal patients (n = 24), CRA (n = 20), CRC (n = 2), and other conditions (n = 15) | Feces                               | Random forest                       | <i>Proteobacteria</i>   | Enriched in CRA                              |
| Ai D <i>et al.</i> (2019) <sup>59</sup>     | France (n = 124); Austria (n = 99)   | Feces                               | Random forest                       | <i>Porphyromonas asaccharolytica, Eubacterium hallii, Parvimonas spp., Fusobacterium 7, Prevotella melaninogenica, Streptococcus vestibularis, Prevotellacopri, Peptostreptococcus stomatis, Fusobacterium nucleatum, Parvimonas micra, Gemella morbillorum, Flavonifractor plautii, Fusobacterium, Clostridium SS2</i> | Enriched in CRC                              |

**Table 2. Important studies in dysbiosis of gut microbiota in CRC, bacterial features, detection methods and models - (continued)**

|   | Participants   | Sample sources   | Models        | Bacteria, genus   | Bacteria features          |
|---|--|------------------|---------------|---|----------------------------|
| Baxter <i>et al.</i> (2016) <sup>23</sup> | CRC ( <i>n</i> = 120), CRA ( <i>n</i> = 198), no colonic lesions ( <i>n</i> = 172) | Feces            | Random forest | <i>Pophyromonas assaccharolytica</i> , <i>Fusobacterium nucleatum</i> , <i>Parvimonas micra</i> , <i>Peptostreptococcus stomatis</i> , <i>Gemella</i> spp. and <i>Prevotella</i> spp.   | Increased in CRC           |
| Sze MA <i>et al.</i> (2018) <sup>49</sup> | Control ( <i>n</i> = 1145), CRA ( <i>n</i> = 521), CRC ( <i>n</i> = 536)           | Feces and tissue | Random forest | <i>Fusobacterium</i> , <i>Parvimonas</i> , <i>Porphyromonas</i> , <i>Peptostreptococcus</i> , <i>Clostridium XI</i> , <i>Enterobacteriaceae</i> , <i>Ruminococcus</i> and <i>Escherichia</i> , <i>Dorea</i> , <i>Blautia</i> and <i>Weissella</i> | Significant ORs            |
| Sze MA <i>et al.</i> (2017) <sup>61</sup> | Adenoma ( <i>n</i> = 22), advanced adenoma ( <i>n</i> = 19), CRC ( <i>n</i> = 26)  | Feces            | Random forest | <i>Fusobacterium</i> , <i>Porphyromonas</i> , <i>Parvimonas</i>   | Decrease in post-treatment |

CRC, colorectal cancer; CRA, colorectal adenoma; NA, not available; OR, odds ratio.

found a combination of four classification markers that reduce patient survival in CRC.<sup>43</sup> 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] Kuntz KM, Lansdorp-Vogelaar I, Rutter CM, Knudsen AB, van Ballegooijen 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, Kraszewska E, Polkowski 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/NEJMcoa0907667.
- [6] Lebowitz 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

- 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. Diversity, stability and resilience of the human gut microbiota. *Nature* 2012;489(7415):220–230. doi:10.1038/nature11550.
- [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 microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65. doi:10.1038/nature08821.
- [11] O'Hara AM, Shanahan F. The gut flora as a forgotten organ. *EMBO Rep* 2006;7:688–693. doi:10.1038/sj.embor.7400731.
- [12] Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* 2006;312:1355–1359. doi:10.1126/science.1124234.
- [13] Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, *et al.* Gut meta genome in European women with normal, impaired and diabetic glucose control. *Nature* 498:99–103. doi:10.1038/nature12198.
- [14] Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* 2013;500(7464):541–546. doi:10.1038/nature12506.
- [15] Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promote atherosclerosis. *Nat Med* 2013;19(5):576–585. doi:10.1038/nm.3145.
- [16] 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.
- [17] Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;444(7122):1027–1031. doi:10.1038/nature05414.
- [18] Cancer Genome Atlas Network. Comprehensive molecular characterization 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. Manipulation 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 microbiome 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 AY, Sunagawa S, Kultima JR, Costea PI, *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 2014;10:766. doi:10.15252/msb.20145645.
- [22] Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 2017;66(1):70–78. doi:10.1136/gutjnl-2015-309800.
- [23] 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.
- [24] Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- [25] Nassif H, Wu Y, Page D, Burnside E. Logical Differential Prediction Bayes Net, improving breast cancer diagnosis for older women. *AMIA Annu Symp Proc* 2012;2012:1330–1339.
- [26] Gevaert O, De Smet F, Timmerman D, Moreau Y, De Moor B. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics* 2006;22(14):e184–e190. doi:10.1093/bioinformatics/btl230.
- [27] Xue X, Zeng N, Gao Z, Du MQ. Diffuse large B-cell lymphoma: subclassification by massive parallel quantitative RT-PCR. *Lab Invest* 2015;95:113–120. doi:10.1038/labinvest.2014.136.
- [28] Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, *et al.* TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. *Infect Genet Evol* 2012;12(4):789–797. doi:10.1016/j.meegid.2012.02.010.
- [29] Habibi S, Ahmadi M, Alizadeh S. Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining. *Glob J Health Sci* 2015;7(5):304–310. doi:10.5539/gjhs.v7n5p304.
- [30] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2014;13:8–17. doi:10.1016/j.csbj.2014.11.005.
- [31] Lebedev AV, Westman E, Van Westen GJ, Kramberger MG, Lunder-vold A, Aarsland D, *et al.* Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between cohort robustness. *Neuroimage Clin* 2014;6:115–125. doi:10.1016/j.nicl.2014.08.023.
- [32] Takahashi N, Guo J, Nishi T. Global convergence of SMO algorithm for support vector regression. *IEEE Trans Neural Netw* 2008;19(6):971–982. doi:10.1109/TNN.2007.915116.
- [33] De Bruyn T, van Westen GJ, Ijzerman AP, Stieger B, de Witte P, Augustijns PF, *et al.* Structure-based identification of OATP1B1/3 inhibitors. *Mol Pharmacol* 2013;83(6):1257–1267. doi:10.1124/mol.112.084152.
- [34] Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, *et al.* A comparison of random forest and its Gini importance with standard che-mometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 2009;10:213. doi:10.1186/1471-2105-10-213.
- [35] Liaw A, Wiener M. Classification and regression by random Forest. *R News* 2002;2:18–22.
- [36] 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.
- [37] Loomba R, Seguritan V, Li W, Long T, Klitgord N, Bhatt A, *et al.* Gut microbiome-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.
- [38] Luo YM, Liu FT, Chen MX, Tang WL, Yang YL, *et al.* A machine learning model based on initial gut microbiome data for predicting changes of Bifidobacterium after prebiotics consumption. *Nan Fang Yi Ke Da Xue Xue Bao* 2018;38(3):251–260.
- [39] Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Med* 2016;8(1):51. doi:10.1186/s13073-016-0307-y.
- [40] Mancabelli L, Milani C, Lugli GA, Turrioni F, Cocconi D, van Sinderen D, *et al.* Identification of universal gut microbial biomarkers of common human intestinal diseases by meta-analysis. *FEMS Microbiol Ecol* 2017;93(12):fix153. doi:10.1093/femsec/fix153.
- [41] Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, *et al.* Enterotypes of the human gut microbiome. *Nature* 2011;473(7346):174–180. doi:10.1038/nature09944.
- [42] Flynn KJ, Ruffin MT4th, Turgeon 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.
- [43] 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.
- [44] Chen HM, Yu YN, Wang JL, Lin YW, Kong X, Yang CQ, *et al.* Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. *Am J Clin Nutr* 2013;97(5):1044–1052. doi:10.3945/ajcn.112.046607.
- [45] 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.
- [46] Lepage P, Häslér R, Sphelmann ME, Rehman A, Zvirbliene A, Begun A, *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.2011.04.011.
- [47] Arthur JC, Gharaibeh 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.
- [48] Burns MB, Lynch J, Starr TK, Knights D, Blekhan R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. *Genome Med* 2015;7(1):55. doi:10.1186/s13073-015-

- 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, Sicinska 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* 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 AI, Milanese A, *et al.* Meta-analysis of fecal meta genomes reveals global microbial 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, 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.
- [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.