



Human Gut Microbiome-Based Knowledgebase as a Biomarker Screening Tool to Improve the Predicted Probability for Colorectal Cancer

Zhongkun Zhou¹, Shiqiang Ge², Yang Li¹, Wantong Ma¹, Yuheng Liu¹, Shujian Hu¹, Rentao Zhang¹, Yunhao Ma¹, Kangjia Du¹, Ashikujaman Syed¹ and Peng Chen^{1*}

¹ School of Pharmacy, Lanzhou University, Lanzhou, China, ² Department of Electronic Information Engineering, Lanzhou Vocational Technical College, Lanzhou, China

OPEN ACCESS

Edited by:

Hyun-Seob Song,
University of Nebraska-Lincoln,
United States

Reviewed by:

Minsuk Kim,
Mayo Clinic, United States
Anita Voigt,
The Jackson Laboratory for Genomic
Medicine, United States

*Correspondence:

Peng Chen
chenpeng@lzu.edu.cn

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 18 August 2020

Accepted: 29 October 2020

Published: 19 November 2020

Citation:

Zhou Z, Ge S, Li Y, Ma W, Liu Y,
Hu S, Zhang R, Ma Y, Du K, Syed A
and Chen P (2020) Human Gut
Microbiome-Based Knowledgebase
as a Biomarker Screening Tool
to Improve the Predicted Probability
for Colorectal Cancer.
Front. Microbiol. 11:596027.
doi: 10.3389/fmicb.2020.596027

Colorectal cancer (CRC) is a common clinical malignancy globally ranked as the fourth leading cause of cancer mortality. Some microbes are known to contribute to adenoma-carcinoma transition and possess diagnostic potential. Advances in high-throughput sequencing technology and functional studies have provided significant insights into the landscape of the gut microbiome and the fundamental roles of its components in carcinogenesis. Integration of scattered knowledge is highly beneficial for future progress. In this study, literature review and information extraction were performed, with the aim of integrating the available data resources and facilitating comparative research. A knowledgebase of the human CRC microbiome was compiled to facilitate understanding of diagnosis, and the global signatures of CRC microbes, sample types, algorithms, differential microorganisms and various panels of markers plus their diagnostic performance were evaluated based on statistical and phylogenetic analyses. Additionally, prospects about current challenges and solution strategies were outlined for identifying future research directions. This type of data integration strategy presents an effective platform for inquiry and comparison of relevant information, providing a tool for further study about CRC-related microbes and exploration of factors promoting clinical transformation (available at: http://gsbios.com/index/experimental/dts_mben?id=1).

Keywords: biomarkers, colorectal cancer, database, diagnosis, microbiome

INTRODUCTION

Colorectal cancer (CRC) is a common malignancy worldwide accounting for about 1 in 10 cancer cases, with incidence and mortality rates of 6.1 and 9.2%, respectively (Bray et al., 2018). Various genetic and environmental factors contribute to CRC development from aberrant crypts to tumors. Overall, $\sim 3 \times 10^{13}$ bacteria colonize the human gut and abnormal microbiome composition has been shown to contribute to the initiation, progression and metastasis of CRC (Pitot, 1993; Qin et al., 2010; Wong et al., 2017c). In cases where patients are rapidly diagnosed and treated

with surgery at the early stages, survival exceeds 90%. However, the survival rate is significantly decreased to 13% in patients with advanced metastatic disease (Shah et al., 2018). The potential value of microorganisms in early diagnosis has attracted significant research attention over the last few decades.

The term “microbiome” refers to the entire habitat including microorganisms (bacteria, archaea, lower and higher eukaryotes, and viruses), their genomes, and the surrounding environmental conditions (Marchesi and Ravel, 2015). These factors are altered along the adenoma-carcinoma sequence, reflected by changes in abundance. Some microbes produce genotoxic compounds and induce inflammation while others proliferate in the tumor-associated niche, designated “driver” and “passenger” bacteria, respectively (Tjalsma et al., 2012). Systematic analysis of microbial communities and identification of those with differential abundance as biomarkers presents an effective diagnostic strategy. Further advances, such as next-generation sequencing, have generated massive amounts of data on the CRC microbiome. Bioinformatics as well as machine learning methods additionally provide powerful tools to advance our understanding (Tabib et al., 2020). Metagenomics and 16S rRNA sequencing studies have revealed different abundance of some microbes between patients and healthy populations and effective combinations of microbial biomarkers could be applied for CRC diagnosis (Sze and Schloss, 2018; Thomas et al., 2019b). Upon combination of these strategies with the fecal immunochemical test (FIT), superior sensitivity and area under the receiver operating characteristic curve (AUC) were obtained relative to standalone FIT, which facilitated advanced adenoma detection (Wong et al., 2017a). Several microbes have been linked with CRC development, including *Fusobacterium nucleatum* (Fn), *Peptostreptococcus anaerobius* (Pa), *Parvimonas micra* (Pm), *Enterotoxigenic Bacteroides fragilis* (ETBF), *Peptostreptococcus stomatis* (Ps) and *Escherichia coli* (Yu et al., 2017a; Pleguezuelos-Manzano et al., 2020). Recently, the ratio of pathogenic bacteria to probiotic populations with decreased abundance in CRC patients was used in a diagnostic model based on their antagonistic effect (Guo et al., 2018). Metabolomics and metagenomics studies have shown that shifts in pathogenicity island genes, short-chain fatty acids (SCFA), amino acids, butyrate and bile acids occur at the early stages of CRC development. Some of these factors possess health-promoting and antineoplastic properties, such as maintenance of mucosal integrity and suppression of inflammation and carcinogenesis. Thus, the shift, particularly the decrease of these health-promoting factors, could contribute to the malignant outgrowth of the tumors (O’Keefe, 2016; Yachida et al., 2019). Subsequent mechanistic research further confirmed their involvement in CRC. For instance, Fn harbors the FadA virulence factor, which binds E-cadherin and activates Wnt/ β -catenin and TLR4/MYD88 pathways to promote cancer initiation, proliferation and invasion (Rubinstein et al., 2013, 2019). Enterotoxigenic *Bacteroides fragilis* (ETBF) harbors the toxin BFT that causes inflammatory diarrhea, inflammation-related tumorigenesis and upregulation of spermine oxidase. Colibactin-producing *E. coli* alkylates DNA at adenine residues and induces double-stranded breaks, anaphase bridges and chromosome

aberrations (Cuevas-Ramos et al., 2010; Goodwin et al., 2011; Chung et al., 2018; Pleguezuelos-Manzano et al., 2020). Based on these omics and experimental data, a theoretical foundation for clinical translation was proposed, which requires validation with more economical methods, such as quantitative PCR (qPCR), or integration with other indices, such as FIT, to obtain optimal benefits (Wong et al., 2017a). More novel biomarkers should emerge with further research progress. However, effective diagnostic panels remain to be established.

While several meta-analyses and reviews based on large-scale, cross-cohort studies have revealed robust associations between microbiome and diseases, developing solutions from the perspective of integration remains a considerable problem due to a number of reasons. First, among the published studies, feces is the most common sample type owing to the non-invasive nature and convenience of sample collection. Other non-invasive types of samples, such as oral swabs, offer an alternative but still need more studies (Flemer et al., 2018). Second, a number of studies were based on 16S rRNA sequencing while others involved metagenomics analyses, which may generate different taxonomic resolutions and involve distinct bioinformatics methods (Wirbel et al., 2019). Third, robustness among different countries or regions is another key contributory factor in microbiome composition, including genetic background, dietary habits and the environment. Fourth, optimal numbers of microbial markers recorded are significantly variable among studies (Duvall et al., 2017). Fifth, specificity deserves further research attention, since only a few studies to date have included cases of other diseases. For example, *Helicobacter pylori* and human papillomavirus are specifically associated with gastric and cervical cancer types while other microbes, such as the order of *Clostridiales* (*Lachnospiraceae* and *Ruminococcaceae* families), are non-specifically associated with disease (Duvall et al., 2017). In general, integration of different types of markers may obtain higher sensitivity, yet specificity will decrease. Therefore, biomarkers that are specific to CRC are of great importance. Finally, classification basis, algorithms, costs and standardization are also worth noting, but systematic integration of the data is lacking.

In this study, a knowledgebase of CRC-related microbes was established by reviewing the relevant literature and extracting key information. Next, a web-based platform using structured query language (SQL) was constructed and statistical analysis were performed that included three classifications and more than seven hundred records of microbial markers. By integrating the scattered data, our novel database could be used to perform inquiry and comparison across different models or databases, such as SILVA, VFDB and the Human Microbiome Oral Database (HOMD), thus contributing to the study of microbiome-based diagnosis of CRC.

MATERIALS AND METHODS

Database Construction

Literature was retrieved from PubMed during September 2019 and April 2020 based on the relevant search criteria.

Two keyword groups were used, the first being “colorectal cancer” and second comprising “16S rDNA,” “metagenomics,” “sequencing,” “quantitative real-time PCR,” “biomarker,” “diagnosis,” “screening,” and “microbiome.” Studies that used blood samples or focused on prognosis, genes, methylation, proteins, small molecule metabolites and liquid biopsy biomarkers were excluded. Following a comprehensive search of the literature and supplementary materials, the relevant data, including names of microbes, sensitivity, specificity, changes in abundance, functions of microbes, technology, algorithm, number of cases, sources and links, were collected. Furthermore, information of the taxonomy of microbial markers was collected from NCBI (Taxonomy) and added into the database. Ultimately, biomarkers were classified into three categories. Microbes that displayed statistical significance in both high-throughput sequencing/pyrosequencing and qPCR experiments were defined as “Class One,” those confirmed with one of the above techniques as “Class Two,” and combinations of different microbes for diagnosis as “Class Three.” Notably, these candidates specifically refer to gut bacteria although the gut microbiome comprises bacteria, fungi, archaea, viruses and bacteriophages.

Data Query and Display

Integrated data were accessible through a web interface that indirectly generates MySQL queries. The interface supports query functions, such as “scientific name of the bacterium” and “taxonomy.” Additionally, basic statistics and visualization were performed according to personalized requirements. Article links for verification or further research are provided for interested authors. The organizational framework is presented in **Figure 1**.

Construction of the Phylogenetic Tree and Statistical Analysis of CRC-Associated Microbes

16S rRNA sequences of all the species (all CRC-associated overabundant and depleted species) in the database were aligned using MEGA-X v10.1.8 software (Kumar et al., 2018). Phylogenetic tree was constructed using the following settings: maximum likelihood as the statistical method, 500 bootstrap replications, Kimura two-parameter as the substitution model and Near-Neighbor-Interchange as the ML Heuristic method. Finally, the tree was adjusted and visualized in Interactive Tree Of Life (iTOL)¹ (Letunic and Bork, 2019). Other statistical analyses were performed with OriginPro software (OriginLab Corporation, United States).

RESULTS AND DISCUSSION

Global Signature of CRC-Related Microbes

In our database, 17 species belonged to Class One (microbes with statistical importance verified using both high-throughput sequencing/pyrosequencing and qPCR), 219 species/clusters

to Class Two (microbes confirmed via high-throughput sequencing/pyrosequencing or qPCR), including 11 phyla, 22 classes, 41 orders, 68 families and 117 genera (**Figure 2**), and 41 panels to Class Three (combinations of different microbes for diagnosis). Despite many microbes proposed for diagnosis and several confirmed conclusions, inconsistent results have been obtained by different research groups.

In healthy individuals, the most dominant phyla (over 90%) are *Firmicutes*, *Bacteroidetes*, *Proteobacteria* and *Verrucomicrobia* (Eckburg et al., 2005). Moreover, significant differences between healthy individuals and CRC patients are detected. Meanwhile, these differences of indices usually showed stepwise decreased or increased frequency from controls, to dysplasia to cancers, though some changes may not be statistically significant between healthy and adenoma groups. In addition to relative abundance, differences in other indices, such as alpha and beta diversity, have been identified. Feces of healthy controls generally contain microbial communities with higher diversity while tissue samples from CRC patients show greater alpha diversity. Earlier studies revealed greater microbial diversity in tumor samples compared with control and polyp samples, with a 75% higher estimated number of species than tissues from healthy sites (Mira-Pascual et al., 2015; Vogtman et al., 2016), characterized by increased levels of opportunistic pathogens. Chao1 and Shannon indices are commonly used to estimate microbial richness and diversity. Decreased Shannon and Chao1 indices were recently reported in fecal samples collected from CRC patients (Yang et al., 2019). Similarly, in an azoxymethane (AOM) mouse model, the CRC group showed significantly lower bacterial richness and Shannon-Weaver's diversity index (Wong et al., 2017b). Other analyses revealed no significant differences in either richness or biodiversity, which could be attributable to the relatively small study cohorts (Wu et al., 2013; Youssef et al., 2018). However, differences at the taxonomic levels (family, genus and species) were universally observed. For instance, patients with CRC usually have increased abundance of operational taxonomic units (OTU) assigned as *Ruminococcus*, *Porphyromonas*, *Peptostreptococcus*, *Parvimonas*, and *Fusobacterium*, while healthy individuals possess more beneficial butyrate-producing bacteria, such as *Bifidobacterium* and *Clostridium butyricum* (Flemer et al., 2017; Sacks et al., 2018). The collective results clearly demonstrate differences in microbial populations between CRC and healthy groups.

Biomarker Identification for Diagnosis Sample Types Used for Diagnosis

In studies on CRC-related microbes, fecal samples from CRC and adenoma patients and healthy volunteers were the most commonly used owing to the non-invasive nature and convenience of sample collection. Cancerous and adjacent non-cancerous normal tissues represent another type of sample that can effectively reveal the overall structure of microbiota in the tumor microenvironment but are unsuitable for early diagnosis (Gao et al., 2015). The microbial diversity in fecal samples is twice as high as that in tissue samples (Mira-Pascual et al., 2015). Oral

¹<https://itol.embl.de/>

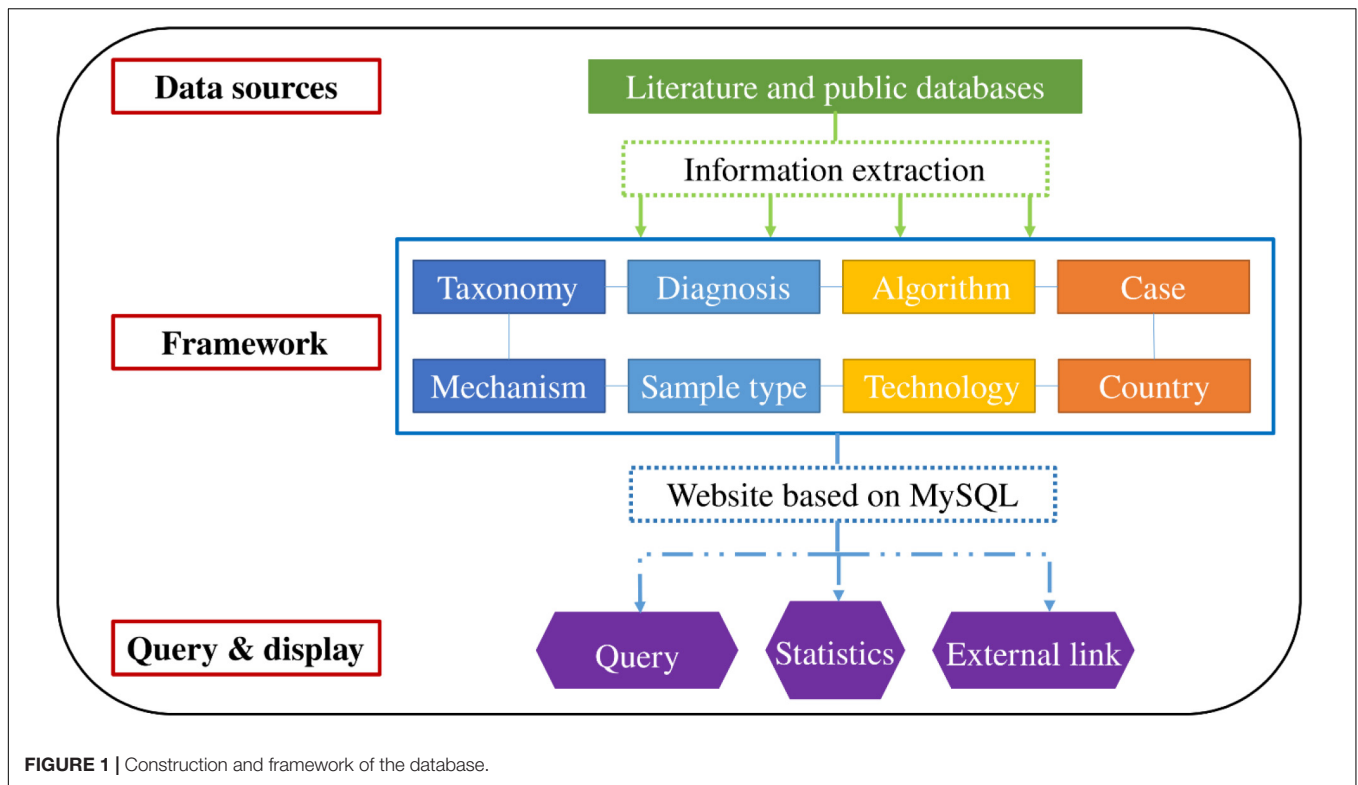


FIGURE 1 | Construction and framework of the database.

swabs represent another novel sample type. Previously identified biomarkers, such as *Fusobacterium nucleatum* and *Parvimonas micra*, are oral microbes. An earlier investigation profiled the oral microbiome as an alternative screening method for CRC (Flemer et al., 2018). Interestingly, a retrospective study on data obtained from adult patients diagnosed with bacteremia and subsequently CRC reported association with *Bacteroides fragilis*, *Streptococcus gallolyticus* and other intestinal microbes, thus providing a new perspective for clinicians (Kwong et al., 2018). Recently, (Poore et al., 2020) reported that predictions based on microbial DNA in blood could discriminate CRC from healthy, cancer-free individuals. However, blood samples were not included in this database due to the requirement for further exploration.

Diagnostic Techniques

This database involves five technical protocols, specifically, denaturing gradient gel electrophoresis (DGGE), qPCR, pyrosequencing, 16S rRNA sequencing and metagenomics sequencing, which have various advantages and disadvantages. Initially, the culture-dependent method was used to analyze CRC microbes as early as the 1960s, which led to significant underestimation of microbial diversity (Wong and Yu, 2019). Recently, a library containing 7,758 human gut bacterial isolates was constructed. Although culture-based methodologies provide access to data that both overlap and complement sequencing surveys, yet these protocols were both labor- and time-consuming compared with culture-independent methods (Poyet et al., 2019). Molecular analysis technology has developed

from DGGE and qPCR to high-throughput sequencing over the years. While the efficiency of analysis was improved by DGGE and qPCR, limitations of low throughput remained unresolved. In 2005, the introduction of next-generation sequencing (NGS) facilitated massive parallel, low-cost and rapid sequencing. 16S rRNA and metagenomics sequencing have further improved efficiency and are widely employed at present. The former procedure is based on the 16S rRNA gene amplicon and facilitates taxonomic and phylogenetic analyses. While the cost-effective feature enables its universal application, several limitations exist: (1) amplicon sequencing of 16S rRNA gene via PCR may miss OTU/taxa detection due to various biases associated with PCR, (2) possible overestimation of community diversity or species abundance, and (3) lack of ability to directly analyze the biological functions of associated taxa (Xia et al., 2018). Recently, potentially unbiased shotgun metagenomics analyses have been conducted, which provide higher taxonomic resolution, gene function and comparative analyses at a decreased cost (Wirbel et al., 2019). However, in terms of clinical transformation, the qPCR-based method is more economical and rapid.

Algorithms Used for Diagnosis

Algorithms include the processes of classification, biomarker identification and model prediction. The classification approaches comprise OTU-based, metagenomics linkage group (MLG)-based, integrated microbial genome (IMG)-based and co-abundant gene group (CAG)-based methods. The model prediction algorithms include random forest (RF), support

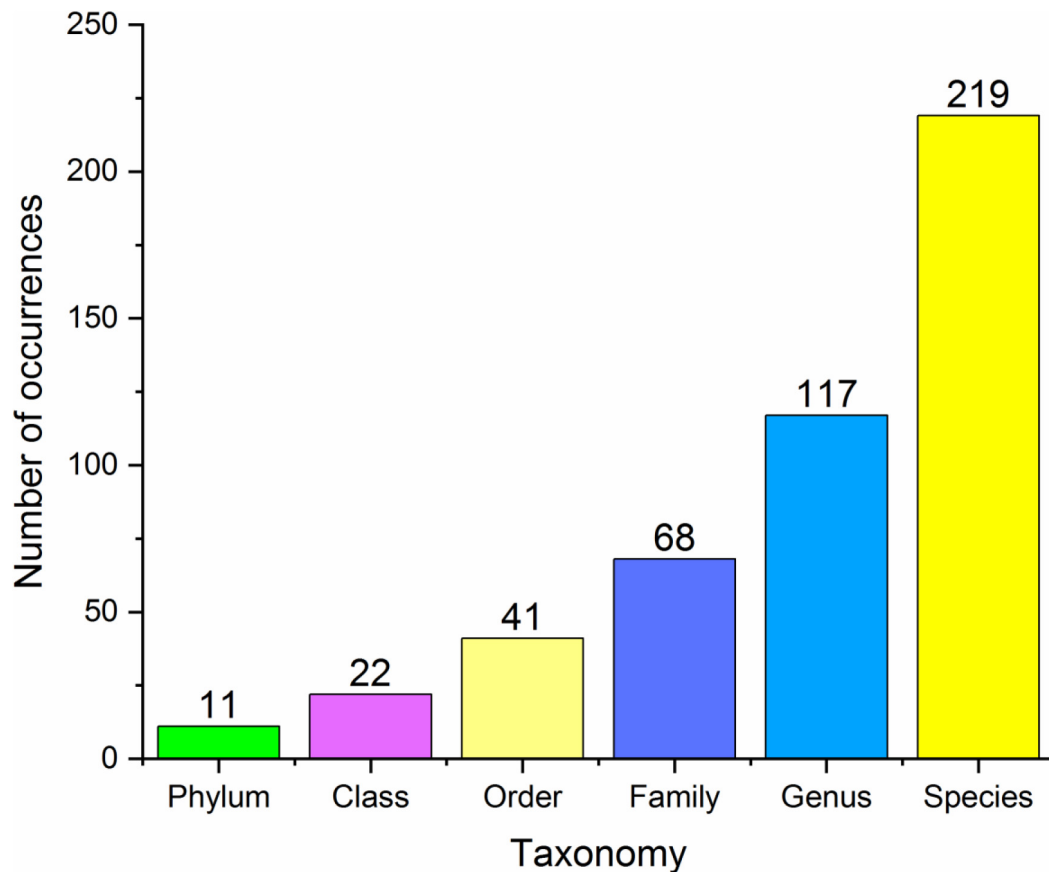


FIGURE 2 | Basic statistics at different taxonomy levels of all the microbial markers in the database.

vector machine (SVM), logistic regression (LR) and leave-one-dataset-out (LODO) analyses, among which random forest is the most widely used algorithm. For the biomarker identification process, relative abundance and Linear discriminant analysis Effect Size (LEfSe) methods are the most commonly used.

Random forest provides a measure of variable importance and out-of-bag (OOB) error when building a tree, making it suitable for prediction analysis. A recent meta-analysis employed the random forest classifier to determine accurate predictive models using a minimal microbial signature. The data showed that using 16 species, cross-validation of AUC > 0.80 was achieved for the majority of datasets (Thomas et al., 2019a). SVM is advantageous for classifying small data volumes and achieved an overall AUC of 0.80 for the combined population (Dai et al., 2018). Recent studies have examined different machine learning classifiers, including RF, Bayesian network, SVM, k-Nearest neighbor and general regression neural networks (Arabameri et al., 2020). LR, applied by most studies, is used to predict binary outcome from a set of numeric variables and aims to identify the most significant features (Wong et al., 2017a). Phylotype-based and OTU-based methods are the main approaches for sequence identification, with the latter being most widely used. However, the OTU-based method has a number of limitations, such as a computationally intensive protocol and

larger memory requirement (Schloss and Westcott, 2011). Other methods have been developed to overcome these drawbacks. For instance, CAGs have been proposed to mitigate the ultrahigh dimensionality challenge of gene-level metagenomics (Minot and Willis, 2019). In addition, CAG-based clusters could be used to determine CRC-associated microbe profiles (Flemer et al., 2017). Taking the collective factors (such as data quantity, number of cohorts and risk factors) into consideration, appropriate approaches and classifiers should be adopted.

Overview of Current Biomarkers for Diagnosis

More than 200 species belonged to the Class Two microbe group (confirmed using either high-throughput sequencing/pyrosequencing or qPCR), among which only 17 were verified as statistically significant with both high-throughput sequencing/pyrosequencing and qPCR (Class One). *Fn* is a known opportunistic pathogen showing increased abundance in feces of CRC patients with a sensitivity range of 69.2–82.9%, specificity of 52.8–90.8% and AUC of 0.675–0.875. Combined with FIT or fecal occult blood test (FOBT), sensitivity, specificity and AUC values reached 92.3, 94.4% and 0.95, respectively. Recently, a number of novel markers have been shown to perform well in CRC diagnosis. *Pa* was increased in four different cohorts and induced carcinogenesis in mice via a

PCWBR2-integrin $\alpha 2/\beta 1$ -PI3K-Akt-NF- κ B signaling axis with a sensitivity of 79.8% and specificity of 98% in combination with FIT (Yu et al., 2017a; Long et al., 2019). *Lachnoclostridium* sp. (designated m3) sharing 97% (1883/1935) DNA sequence similarity with *Lachnoclostridium* sp. YL32 was significantly enriched in adenoma. m3 showed specificity of 78.5% and sensitivity of 48.3% for adenoma and 62.1% for CRC. However, its role in tumorigenesis warrants further research (Liang et al., 2019). The other 15 biomarkers are presented in **Table 1** (4 were decreased and 11 were enriched in patients).

With regard to Class Two microbes, basic statistics are shown in **Figure 3** and phylogenetic tree in **Figure 4**. The majority of enriched microbes were classified into *Fusobacteriaceae*, *Peptoniphilaceae*, *Lachnospiraceae*, *Porphyromonadaceae*, *Peptostreptococcaceae*, *Bacteroidaceae*, *Prevotellaceae*, *Ruminococcaceae*, *Streptococcaceae*, and *Bacillales incertae sedis* at the family level (**Figure 3A**). Among the group of decreased microbes, most were classified into *Lachnospiraceae*, *Ruminococcaceae*, *Bacteroidaceae*, *Streptococcaceae*, *Bifidobacteriaceae*, and *Eubacteriaceae* (**Figure 3B**). In the Venn diagram, only a small overlap of increased and decreased microbes was observed, supporting the reliability of most microbial markers despite some inconsistencies (**Figure 3C**). At the species level, phylogenetic tree showed details of current CRC-related biomarkers as well as their evolutionary relationships. Additionally, species belonging to oral microbes were marked with stars.

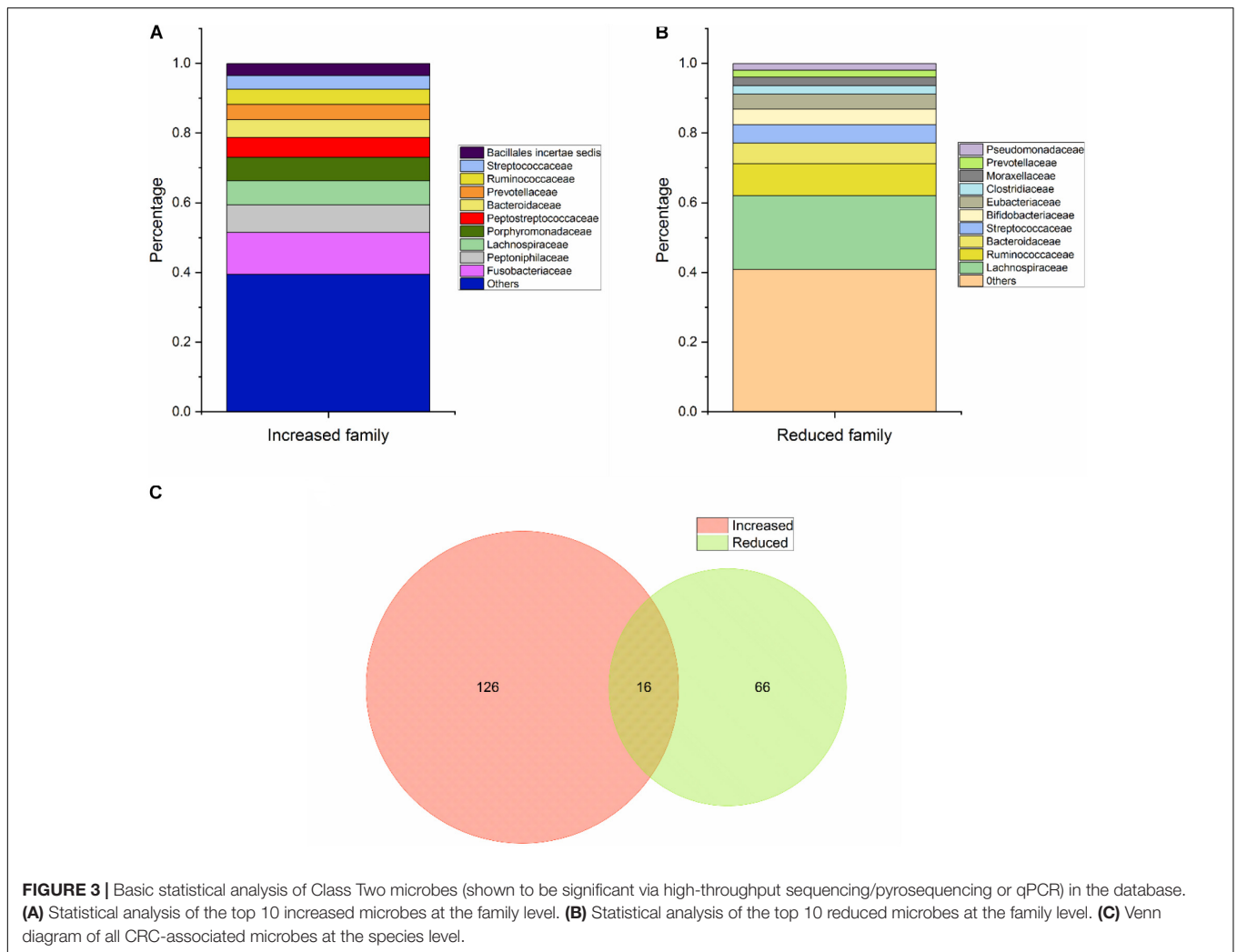
The functions of gut microbes include fermenting complex carbohydrates to produce large amounts of metabolites, maintaining epithelial homeostasis, serving as an endocrine organ and participating in the development, maturation and differentiation of the immune system of the host (Villéger et al., 2018; Rastelli et al., 2019). In a sense, intestinal metabolites

directly affect the occurrence of CRC and not intestinal flora. The majority of nutrients from food are absorbed in the small intestine with protein residues and complex nutrients, such as fiber moving to the colon, and consequently metabolized by the microbial populations (O'Keefe, 2016). Therefore, from the perspective of microbial function, the majority are associated with protein fermentation, bile acid biotransformation, decomposition of polysaccharides and polyphenols and energy metabolism. For example, *Faecalibacterium prausnitzii* (Fp), *Bifidobacterium* (Bb), *Roseburia* spp. (Rb), *Eubacterium rectale* (EUB), *Clostridium butyricum* (Cb), *Lactobacillus* spp. (Lc), *Akkermansia muciniphila* (Akk), *Ruminococcus*, and *Lachnospiraceae* were found to be more abundant in healthy controls compared with CRC patients. Fp is a butyrate producer decreased in Crohn's disease (CD) patients, whose metabolites exert anti-inflammatory effects via blocking NF- κ B activation and IL-8 production (Sokol et al., 2008). Bb and Lc are used as probiotics for human consumption and benefit the gut through inducing cancer cell apoptosis, inhibiting cell proliferation, modulating host immunity and inactivating carcinogenic toxins (Wong and Yu, 2019). An earlier study reported that determination of Fn/Bb and Fn/Fp ratios could improve diagnostic performance for CRC based on their antagonistic effect (Rezasoltani et al., 2018). Both Rb and EUB are butyrate-producing *Firmicutes* and metabolize dietary fibers to provide energy sources and achieve anti-inflammatory effects (Paramsothy et al., 2019). Their capabilities as a non-invasive tool were additionally evaluated but not included in the final model (Malagón et al., 2019). More recently, the utility of other widely recognized markers, including Fn, colibactin-producing *E. coli* and ETBF, in diagnosis of CRC has been systematically analyzed (Chung et al., 2018; Malagón et al., 2019; Wu et al., 2019; Pleguezuelos-Manzano et al., 2020). However, several

TABLE 1 | Diagnostic performance of Class One microbials.

Name	Sensitivity%	Specificity%	AUC	Algorithm	Sample	Case	Region	References
Fn	73.1	90.8	0.860	Relative Abundance	Feces	490	China	Wong et al., 2017a
Pa	56.7	86.3	0.720	Logistic regression	Feces	390	China	Wong et al., 2017a
Pm	45.2	97.1	0.730	Logistic regression	Feces	390	China	Wong et al., 2017a
Gm	39.0	76.0	0.622	Relative Abundance	Feces	333	Spain	Malagón et al., 2019
Ps	53.0	76.0	0.710	Relative Abundance	Feces	333	Spain	Malagón et al., 2019
Bf	33.0	0.76	0.571	Relative Abundance	Feces	333	Spain	Malagón et al., 2019
pks	56.4	82.0	NA	Relative Abundance	Feces	238	Sweden	Eklöf et al., 2017
Fp	81.8	62.6	0.741	Abundance Rate	Feces	549	China	Rezasoltani et al., 2018
Bb	90.4	76.4	0.870	Abundance Rate	Feces	549	China	Rezasoltani et al., 2018
Cs	73.3	66.1	0.736	logistic regression	Feces	781	China	Xie et al., 2017
Ap	NA	NA	NA	Relative abundance	Feces	146	Meta	Yachida et al., 2019
Gl	NA	NA	NA	Relative abundance	Mucosa	207	China	Nakatsu et al., 2015
m3	62.1	79.0	0.741	Relative Abundance	Feces	1012	China	Liang et al., 2019
Bd	NA	NA	NA	Relative abundance	Feces	179	French	Sobhani et al., 2011
afaC	NA	NA	NA	Relative abundance	Tissue	55	South Africa	Viljoen et al., 2015
Akk	NA	NA	NA	Relative abundance	Feces	112	China	Wang et al., 2020
Cb	NA	NA	0.930	Random forest	Feces	60	China	Yang et al., 2020

NA, non-available; Meta, meta-analysis; *Gemella morbillorum* (Gm); *Bacteroides fragilis* (Bf); pks + clbA + *Escherichia coli* (pks); *Clostridium symbiosum* (Cs); *Atopobium parvulum* (Ap); *Granulicatella* (Gl); *Bacteroides* (Bd); afaC-positive *E. coli* (afaC).



issues require further clarification. Although the pathogenesis and benefits of ETBF and Bb have been validated, inconsistencies exist among different samples. ETBF was shown to be increased in tumor tissues and form a biofilm in the gut. However, this pathogenic bacterium displayed no significant differences in abundance in patient fecal samples and was not detectable using qPCR targeting the toxin-producing gene, making it difficult to discriminate between patients and healthy controls (Zackular et al., 2014; Kosumi et al., 2018; Sze and Schloss, 2018; Malagón et al., 2019; Saffarian et al., 2019). Finally, *Lachnospiraceae* and *Ruminococcaceae* families were associated with multiple diseases (known as non-specific responders), which inspired us to obtain non-gastrointestinal cancer samples for future experimental design (Duvall et al., 2017; Rezasoltani et al., 2018).

Diagnostic Strategy and Performance

Combinations of Different Microbial Markers

Class Three (combinations of different microbes for diagnosis) included 41 panels verified using various methods (Table 2). The combinations ranged from two species to 63 OTUs, with AUC ranging from 0.531 to 0.998. Twelve panels were based on

qPCR, whose algorithms usually link with logistic regression or relative abundance. Meanwhile, 16 panels and 12 combinations were based on 16S rRNA and metagenomics sequencing data, predominantly using the random forest-based model. Based on AUC, qPCR-based models could achieve comparable outcomes to the two other technologies with limited biomarkers (usually no more five species). Nevertheless, 16S rRNA and metagenomics-based models show performance advantages at the cost of the number of markers (more than 10 OTUs on average). In the random forest and Minimum Redundancy Maximum Relevance (mRMR) models, both OOB and error rate parameters demonstrated that panels comprising ~16–20 biomarkers achieved the best prediction accuracy (Flemer et al., 2018; Wirbel et al., 2019).

Combination of microbes may be operative, rather than representing a strain that is increased or decreased in the intestine (Tilg et al., 2018). In addition, prediction models from single dataset may lead to reduced accuracy and be sensitive to both technique and heterogeneity (Thomas et al., 2019a). An earlier study identified 63 OTUs (29 from oral swabs and 34 from fecal samples) to predict CRC. While the final AUC value

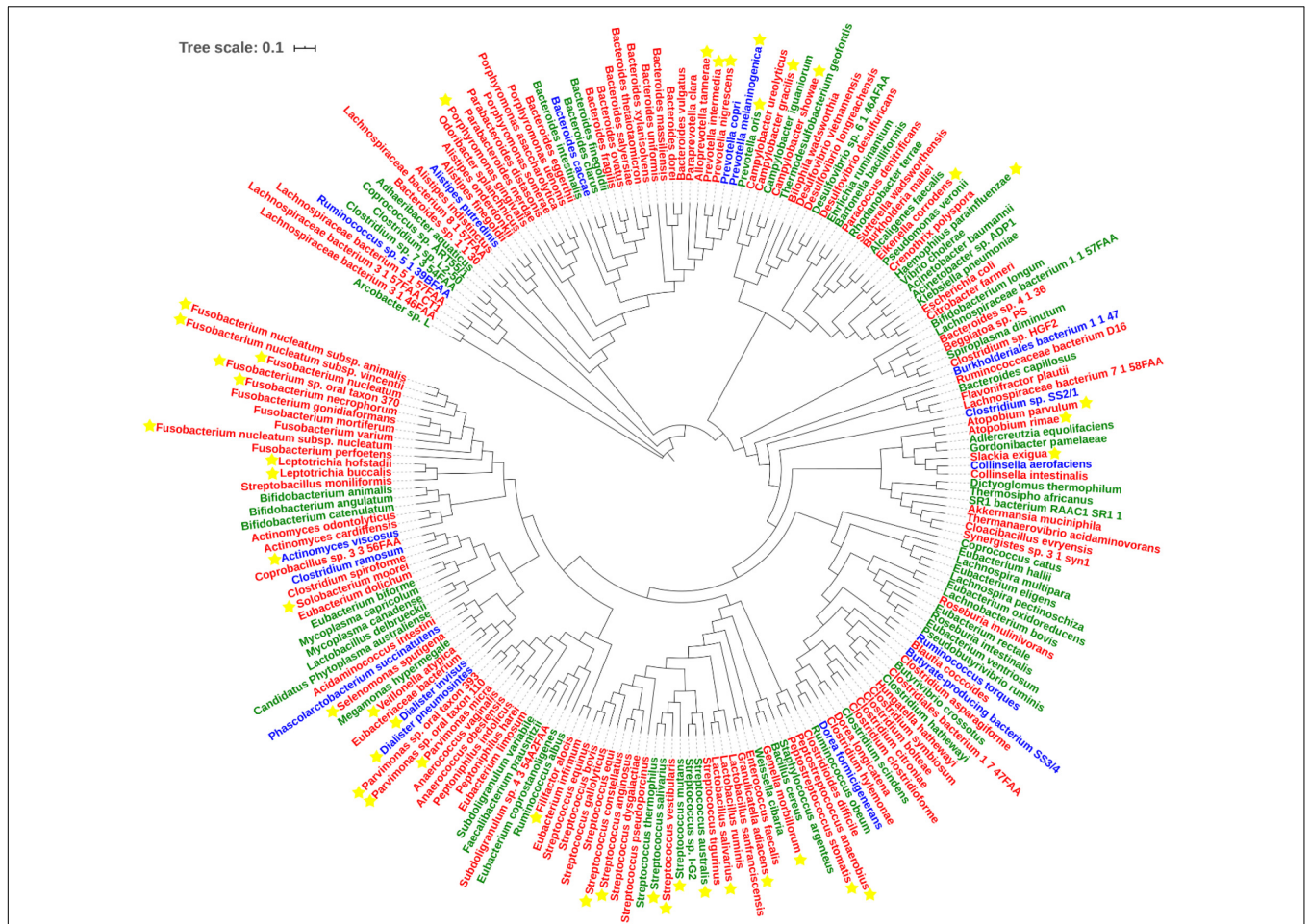


FIGURE 4 | Phylogenetic tree of all CRC-related microbes in the database. Species marked in red and green refer to the increased and decreased microbes, and species marked in blue refer to the microbes that show up in both increased and decreased groups. Species marked with yellow stars refer to oral microbes according to HMOD (16S rRNA sequences of m7 and *Sulfurovum* sp. SCGC AAA036-O23 are not available, which also belong to the increased group).

was up to 0.98, its application in clinical examination remains a challenge (Flemer et al., 2018). Several other researchers used more than 30 OTUs/phylotypes/MLGs to construct a random forest classifier and obtained AUC values > 0.80 (Nakatsu et al., 2015; Baxter et al., 2016a; Yu et al., 2017a). Previous studies suggest that the *Firmicutes/Bacteroidetes* ratio responds to health and disease states, such as obesity and CRC (Ley et al., 2006; Saffarian et al., 2019). Interactions between bacteria provide an ecological perspective for screening, and increase in pathogenic bacteria is always accompanied by decrease in beneficial microbes (Dai et al., 2018). Some researchers observed an association of the group of *Bacteroides* and *Prevotella* with elevated IL17-producing cells in colon cancer and demonstrated that supernatant from Fn inhibited the bactericidal activities of Fp and Bb (Sobhani et al., 2011; Guo et al., 2018). Furthermore, beneficial microbes can contribute to several intestinal functions and protect the organ from pathogenic microorganisms, and the “pathogenic bacteria:probiotics” ratio generates a better effect than single organism model (Eslami et al., 2019; Malagón et al., 2019; Yang et al., 2020). Thus, the complementary

effects between enriched and reduced microbes should be highlighted for further investigation. Clearly, combinations of different microbial markers exhibit better predictive performance than single markers.

Integration With FIT

In the database, FIT was also presented when available. FIT has been extensively tested and recommended by National Comprehensive Cancer Network guidelines. The method involves direct detection of globin rather than heme, and shows greater sensitivity than the highly sensitive guaiac fecal occult blood test. Retrospective analysis showed that replacing 3-year colonoscopy surveillance with annual FIT could reduce the requirement for colonoscopy and provide economic benefits. However, sensitivity was relatively low for advanced neoplasms, ranging from 21.8 to 46.3% at the preset thresholds (Gies et al., 2018; Cross et al., 2019a). Combining microbe analysis with FIT could enhance the detection of advanced precancerous lesions, as validated in numerous experiments. Taking results from Class One and Three as representative cases, combined quantitation

TABLE 2 | Different panels for CRC screening.

Name	Sensitivity%	Specificity%	AUC	Technique	Algorithm	Sample	Case	Region	References
Fn, Pa, Pm	89.4	93.0	0.950	qPCR	LR	Feces	390	China	Wong et al., 2017a
Ps/EUB, Bt/EUB, Bt/EUB	80.0	90.0	0.837	qPCR	LR	Feces	333	Spain	Malagón et al., 2019
pks, Fn	89.7	61.0	NA	qPCR	DA	Feces	238	Sweden	Eklöf et al., 2017
Fn/Fp	95.0	71.3	0.914	qPCR	AR	Feces	549	China	Rezasoltani et al., 2018
Fn/Bb	84.6	92.3	0.911	qPCR	AR	Feces	549	China	Rezasoltani et al., 2018
Fn/Fp, Fn/Bb	80.8	85.6	0.910	qPCR	AR	Feces	549	China	Rezasoltani et al., 2018
Fn, Fp, Bb	92.5	83.5	0.943	qPCR	AR	Feces	549	China	Rezasoltani et al., 2018
5 OTUs	90.0	80.0	0.896	16SrDNA	LR	Feces	90	America	Zackular et al., 2014
6 OTUs	90.0	83.0	0.922	16SrDNA	LR	Feces	90	America	Zackular et al., 2014
22OTUs	81.2	97.1	0.673	16SrDNA	RF	Feces	490	Canada, United States	Baxter et al., 2016b
34 OTUs	51.7	97.1	0.847	16SrDNA	RF	Feces	490	Canada, United States	Baxter et al., 2016b
23 OTUs	70.0	92.8	0.829	16SrDNA	RF	Feces	490	Canada, United States	Baxter et al., 2016b
16 OTUs	53.0	96.0	0.900	16SrDNA	RF	Oral swabs	60	Ireland	Flemer et al., 2018
28 OTUs (16 oral swabs, 12 feces)	74.0	94.0	0.940	16SrDNA	RF	Feces and oral swabs	60	Ireland	Flemer et al., 2018
63 OTUs (29 oral swabs, 34 feces)	88.0	94.0	0.980	16SrDNA	RF	Feces and oral swabs	60	Ireland	Flemer et al., 2018
22 OTUs	58.0	92.0	0.840	Metagenomics	LR	Feces	156	France, Germany	Zeller et al., 2014
7 OTUs	87.0	83.7	0.886	Metagenomics	RF	Feces	128	China	Yu et al., 2017a
15 MLGs	NA	NA	0.983	Metagenomics	RF	Feces	96	Austria	Feng et al., 2015
16 OTUs	NA	NA	0.860	Metagenomics	RF	Feces	969	Meta	Thomas et al., 2019a
17 OTUs	60.1	84.8	0.804	Metagenomics	RF	Feces	424	Meta	Shah et al., 2018
30 OTUs	NA	NA	0.830	Metagenomics	RF	Feces	208	Meta	Yachida et al., 2019
8 taxa	NA	NA	0.750	16SrDNA	RF	Feces	492	Meta	Yachida et al., 2019
12 genus	NA	NA	0.846	16SrDNA	RF	Feces	1674	Meta	Sze and Schloss, 2018
18 OTUs	NA	NA	0.831	16SrDNA	RF	Feces	404	Canada, United States	Baxter et al., 2016a
32 OTUs	NA	NA	0.853	16SrDNA	RF	Feces	404	Canada, United States	Baxter et al., 2016a
41 OTUs	NA	NA	0.686	16SrDNA	RF	Feces	404	Canada, United States	Baxter et al., 2016a
12 phylotypes	NA	NA	0.831	16SrDNA	LEfSe	Mucosa	160	China	Nakatsu et al., 2015
18 OTUs	NA	NA	0.871	16SrDNA	RF	Mucosa	160	China	Nakatsu et al., 2015
38 phylotypes	NA	NA	0.846	16SrDNA	Dirichlet MM	Mucosa	160	China	Nakatsu et al., 2015
m3, Fn, Ch, Bc	85.2	80.2	0.907	qPCR	LR	Feces	1012	China	Liang et al., 2019
m3, Fn	NA	NA	0.891	qPCR	LR	Feces	1012	China	Liang et al., 2019
Fn, Ch, m7, Bc	92.8	79.8	0.886	qPCR	SLC	Feces	370	China	Liang et al., 2017
Fn, Ch, m7, Bc, Ri	74.3	88.9	0.843	qPCR	LR	Feces	128	China	Liang et al., 2017
17 IMG species	NA	NA	0.860	Metagenomics	IMG	Feces	128	China	Liang et al., 2017
7 species-level mOTUs	NA	NA	0.890	Metagenomics	mOTUs	Feces	128	China	Liang et al., 2017
27 MLG	NA	NA	0.960	Metagenomics	MLG	Feces	128	China	Liang et al., 2017
Fn, Pa, Pm (4 genes)	NA	NA	0.770	Metagenomics	CRC index	Feces	96	China, Denmark, Austrian, French	Yu et al., 2017a
22 genes	NA	NA	0.998	Metagenomics	RF	Feces	107	China	Yang et al., 2020
Cb, Cs	NA	NA	0.935	qPCR	RF	Feces	60	China	Yang et al., 2020
7 CRC-enriched bacteria	NA	NA	0.800	Metagenomics	SVM	Feces	526	Meta	Dai et al., 2018
55 species	NA	NA	0.830	Metagenomics	RF	Feces	181	Meta	Sze and Schloss, 2018

DA, Decision Abundance; AR, Abundance Rate; SLC, Simple linear combination; Dirichlet MM, Dirichlet multinomial mixtures; Clostridium hathewayi (Ch); Unclassified species (m7); Roseburia intestinalis (Ri); Bacteroides thetaiotaomicron (Bt).

of Fn and FIT showed superior sensitivity to FIT alone, leading to detection of lesions missed by FIT alone (Wong et al., 2017a). Similarly, Pa, Pm, Cs, and m3 displayed an obvious improvement in both sensitivity and AUC, with a slight decrease in specificity (Xie et al., 2017; Liang et al., 2019). This complementary role was also illustrated using biomarker panels. Upon combining 22 OTUs identified using the penalized linear model with FIT, sensitivity increased from 58 to 72% at the same specificity (Zeller et al., 2014). In another study, combination of *Bacteroides clarus* (Bc), Fn, Ch, and m7 showed an increase of 9 percentage points when integrated with FIT in a logistic regression model (Liang et al., 2017). In conclusion, clinical screening programs based on both microbial markers and FIT/FOBT are cost-effective and present a promising diagnostic tool.

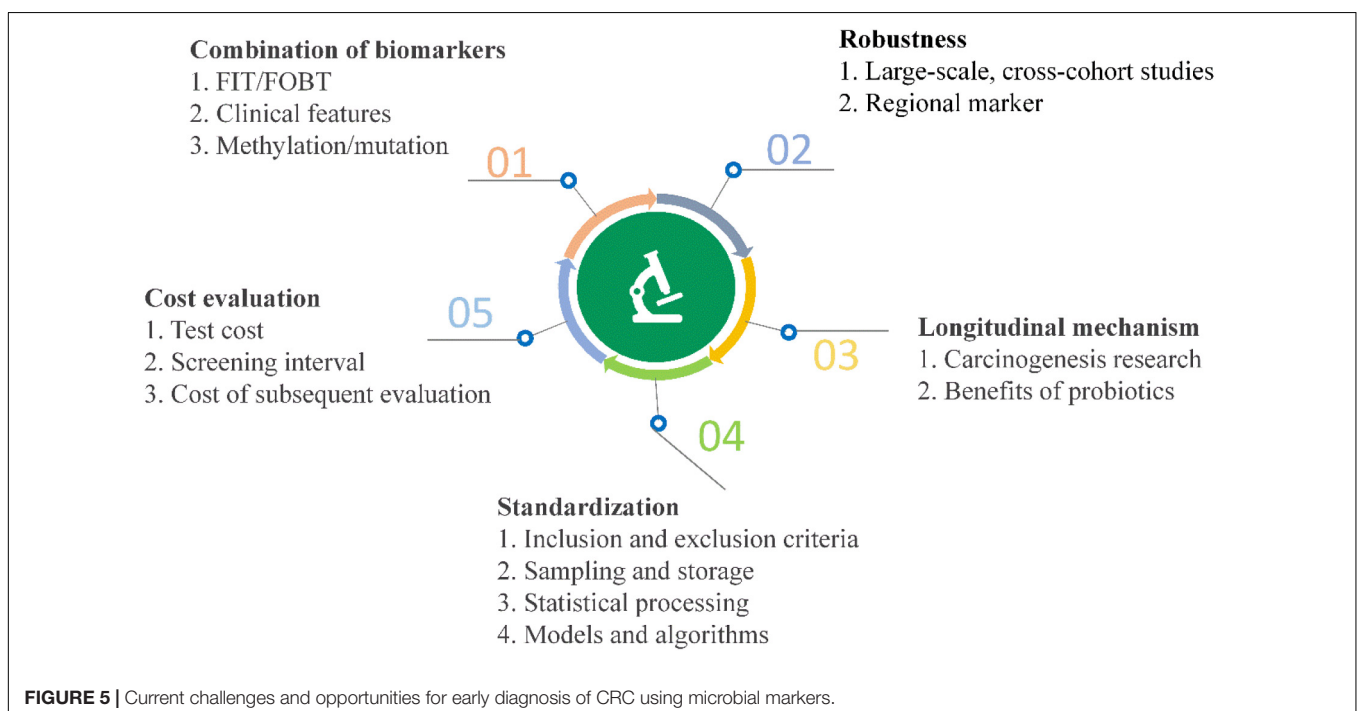
Prospects and Challenges

High-throughput sequencing and other analyses over the past decade have facilitated significant advances and gradual elucidation of the role of microbes in CRC. Current research on the value of clinical transformation of microbial markers in CRC diagnosis highlights the continued challenges of using available data effectively for making a contribution to precision medicine. Inspiration from other fields may additionally facilitate novel breakthroughs (Figure 5).

Formation of CRC is a multifactorial process and potential complementary effects between molecular markers require further attention. More than 80% CRC results from chromosomal instabilities, including mutation of the adenomatous polyposis coli (APC) gene and K-ras oncogene. APC gene-deficient mice can spontaneously grow tumors in the intestine and patients carrying the KRAS mutation show chemotherapeutic resistance (Colnot et al., 2004; Kuipers et al., 2015). Fecal DNA samples

have been used to detect colorectal neoplasia (Imperiale et al., 2004). Septin 9 gene methylation has been shown to be effective as a biomarker and approved by the FDA (Lofton-Day et al., 2008). Meanwhile, methylation of bone morphogenic protein 3 and N-Myc downstream-regulated gene 4 displayed high specificity as an early and frequent event in colorectal tumors (Melotte et al., 2009; Loh et al., 2010). In 2014, multitarget stool DNA testing of combined KRAS, BMP3, NDRG4, and FIT achieved significantly higher detection of cancers, which led to FDA approval of Cologuard (Imperiale et al., 2014). Therefore, integration of genomics with microbiome analysis presents a promising direction. A recent study discussed this issue, suggesting that associations between tumor genomics and the microbiome could be beneficial in diagnostics (Burns and Blekhan, 2018). Since about 11% CRC cases result from overweight and obesity, other researchers used clinical data, such as body mass index (BMI) representing overall body fat, which displayed excellent discriminatory ability. However, no statistical significance was observed in a number of other analyses (Bardou et al., 2013; Zackular et al., 2014). To extract data from plain text files, Natural Language Processing methods or software have been employed for effective use of clinical features (Yim et al., 2016). Overall, these findings offer possible solutions and important directions for future research.

Universality is another key challenge, since differing opinions exist with regard to universal microbial markers. On the one hand, cross-cohort studies and meta-analyses have provided practicable and effective strategies that could overcome heterogeneity and ethnic differences with unbiased bioinformatics and statistical analysis. For instance, an earlier metagenomics analysis involving five ethnically different cohorts identified not only known biomarkers such as Fn,



Ps, Pm, and *Solobacterium moorei*, but also a novel strain, *Peptostreptococcus anaerobius*, with subsequently confirmed roles in carcinogenesis using a $Apc^{Min/+}$ mouse model (Yu et al., 2017a; Long et al., 2019). Numerous meta-analyses also leveraged 16S rRNA or metagenomics data sets to reveal altered microbiome. Wirbel et al. (2019) identified a core set of 29 species while Dai et al. (2018) found 69 CRC-associated bacteria with metagenomic analysis. Similarly, two other teams identified 25 microbial OTUs and 12 common genera based on a random forest model using 16S rRNA sequencing datasets (Shah et al., 2018; Sze and Schloss, 2018). On the other hand, (Yang et al., 2020) proposed a strategy from a new angle, which inferred that regional biomarkers display high accuracy in specific populations. This theory was also supported by another study, which identified multiple *Fusobacterium* taxa (including *F. varium* and *F. ulcerans*) in Southern Chinese populations as disease biomarkers or targets that could be tailored according to discrepancies (Yeoh et al., 2020). Both alternative strategies provide well-powered assessments.

One of the significant challenges of clinical transformation is insufficient mechanistic analysis. While efficient computational frameworks and tools based on feature selection have been developed, machine learning requires further research (Tabib et al., 2020). Distinct from FIT/FOBT and fecal DNA tests, these semi-supervised or supervised learning methods are more like a “black box” with unclear mechanism. To date, hundreds of microorganisms have been shown to be linked with CRC, among which limited numbers have been further investigated. As a case in point, Fn was shown to be overabundant in tumor versus matched normal tissue and its potential role in CRC attracted widespread research attention (Castellarin et al., 2012; Kostic et al., 2012). Over the last few years, numerous studies have supported a role of Fn in promoting colorectal carcinogenesis through various functions such as inducing inflammatory cell infiltration, modulating E-cadherin/ β -catenin signaling, activating immune cells, mediating interactions between bacteria, and binding to tumor-expressing Gal-GalNAc (Rubinstein et al., 2013, 2019; Abed et al., 2016; Yang et al., 2017). These advances have enhanced our knowledge of the potential relationships between Fn and chemoresistance, metastasis and poor prognosis (Mima et al., 2016; Yu et al., 2017b; Chen et al., 2020). Therefore, detection of Fn for early screening or exploitation of inhibitors targeting related pathways may be efficacious in clinical practice. In terms of methodological aspects, Bertrand Routy proposed a viable solution involving five steps: (1) microbial metagenomics should be standardized, (2) different “omics” analyses should be integrated, (3) the amount of cultivable microbial species should be increased, (4) non-invasive sampling methods should be combined with capsule endoscopy, and (5) Avatar mouse models should be standardized and investigated (Routy et al., 2018). Overall, longitudinal profiling of etiological and protection mechanisms of microorganisms achieves higher information richness and pave the way to take advantage of gut microbiome for diagnosis.

Development of standardized methods should also attenuate inconsistency of data. Inclusion and exclusion criteria have

been gradually established, including diet, treatment, genetic background, disease history, antibiotic usage history and colonoscopy, aiming to avoid intestinal microbiota changes (O'Brien et al., 2013). During transportation and storage, a low temperature of -80°C and preservative buffer, such as RNAlater or EDTA, are effective to maintain DNA stability and integrity (Carozzi and Sani, 2013). In particular, compared to freezing for preservation, smaller technical variability was introduced without disrupting subject- and time-point specificity of the gut microbiome (Voigt et al., 2015). DNA extraction exerted the most significant effect on outcome of metagenomics analysis, highlighting the standardized DNA extraction method for human fecal samples (Costea et al., 2017). To address the complex challenges posed by large-scale studies, a protocol involving collection of microbiome samples at home and shipping to laboratories for molecular analysis was developed by Franzosa et al. (2014). Furthermore, for library preparation, PCR-free based methods were recommended to reduce PCR bias and improve assembly for accurate taxonomic assignment (Jones et al., 2015). Nevertheless, lack of standardization with regard to data access, metadata and analysis tools remain a barrier to acquisition of accurate and comparative results (Laudadio et al., 2018). Data integration and system-level modeling from multiple omics platforms is one of the most promising directions of microbiome research (Nayfach and Pollard, 2016). To improve the *status quo*, comprehensive platforms, such as MicrobiomeAnalyst and gcMeta, were recently constructed for downstream statistical analysis and functional interpretation (Dhariwal et al., 2017; Shi et al., 2019). Notably, the International Human Microbiome Standards (IHMS) project is committed to coordinate the development of standard operating procedures designed to optimize data quality and comparability in the human microbiome field. SYBR Green and probe-based qPCR are two common choices toward application, the former being more economical and the latter achieving greater accuracy for absolute quantification.

Cost-effectiveness is the ultimate challenge, including the costs of testing, screening intervals and subsequent evaluations resulting from the initial test (Dickinson et al., 2015). Due to high-cost resources, colonoscopy is not generally employed as a screening tool, except in a few countries like the United States, Germany and Austria. In low-income or middle-income countries with a low incidence of CRC, colonoscopy screening strategies may not be sufficiently cost-effective for implementation (Keum and Giovannucci, 2019). Taking FIT and Cologuard as examples, although incremental costs per additional advanced adenoma (AA) and CRC detected using colonoscopy versus FIT were £7,354 and £180,778, respectively, annual FIT reduced the colonoscopy incidence by 71% in intermediate-risk patients compared to three-yearly colonoscopy surveillance (Cross et al., 2019b). Cologuard shows superior performance for screening of AA, but carries a higher cost. In terms of the rate of screening compliance, stool DNA test is associated with higher patient acceptance owing to its simplicity. A preliminary calculation showed that combination of FIT and bacterial markers would avert up to 30% of total colonoscopies

as well as save an estimated 77 million € per 100,000 participants (Malagón et al., 2019). Meanwhile, usage of residual buffer from FIT cartridges is feasible for microbiota-based analysis and could greatly ameliorate the cost (Baxter et al., 2016a; Gudra et al., 2019).

Considering the collective findings, bacteriophages, viruses, archaea and fungi will be integrated into this database as biomarkers in the future. In addition, with advances in elucidation of mechanisms and omics analyses (such as transcriptomics, proteomics, and metabolomics), corresponding function descriptions should be more systematic. Systems biology and computational biology play crucial roles in mass data integration, and machine learning-based algorithms are under development for analysis of metadata to facilitate CRC diagnosis.

CONCLUSION

Development of colorectal cancer is a multifactorial process in which gut microbes play an important role. Determination of dysbiosis of microbial communities and differential patterns of abundance of microorganisms as biomarkers based on sequencing, algorithms and experimental data may aid in diagnosis and reduce morbidity and mortality. Except for a few pathogenic bacteria, the relationships between several microorganisms and colorectal cancer remain to be established, which are reflected by inconsistencies among different studies. Here, a database of CRC-related microbes was constructed using SQL and basic statistical analyses were conducted to outline biomarkers at different taxon levels. Diagnostic performance and mechanisms are discussed in detail. This

REFERENCES

- Abed, J., Emgård, J. E., Zamir, G., Faroja, M., Almogly, G., Grenov, A., et al. (2016). Fap2 mediates *Fusobacterium nucleatum* colorectal adenocarcinoma enrichment by binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe* 20, 215–225. doi: 10.1016/j.chom.2016.07.006
- Arabameri, A., Asemani, D., and Teymourpour, P. (2020). Detection of colorectal carcinoma based on microbiota analysis using generalized regression neural networks and nonlinear feature selection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17, 547–557. doi: 10.1109/tcbb.2018.2870124
- Bardou, M., Barkun, A. N., and Martel, M. (2013). Obesity and colorectal cancer. *Gut* 62, 933–947. doi: 10.1136/gutjnl-2013-304701
- Baxter, N. T., Koumpouras, C. C., Rogers, M. A. M., Ruffin, M. T., and Schloss, P. D. (2016a). DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model. *Microbiome* 4:59. doi: 10.1186/s40168-016-0205-y
- Baxter, N. T., Ruffin, M. T., Rogers, M. A. M., and Schloss, P. D. (2016b). Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med.* 8:37. doi: 10.1186/s13073-016-0290-3
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492
- Burns, M. B., and Blekman, R. (2018). Integrating tumor genomics into studies of the microbiome in colorectal cancer. *Gut Microbes* 10, 547–552. doi: 10.1080/19490976.2018.1549421

type of knowledge integration is important for understanding and monitoring CRC. Moreover, this database can be used to perform inquiries and comparisons across different models and databases, contributing to further study of CRC-related microbes and promotion of cost-effective and non-invasive CRC screening strategies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

PC and ZZ contributed to the study design and drafted the manuscript. ZZ, SG, YaL, WM, YuL, SH, RZ, YM, KD and AS performed the statistical analysis and interpretation. All authors contributed to critical revision of the final manuscript and approved the final version of the manuscript.

FUNDING

This work was supported by Special Funding for Open and Shared Large-Scale Instruments and Equipments of Lanzhou University (Grant No. LZU-GXJJ-2019C012), the Project of Lanzhou City for Innovative and Entrepreneurial Talents (Grant No. 2017-RC-73), and Science and Technology Project of Lanzhou City (Grant No. 2018-4-59).

- Carozzi, F. M., and Sani, C. (2013). Fecal collection and stabilization methods for improved fecal DNA test for colorectal cancer in a screening setting. *J. Cancer Res.* 2013:818675. doi: 10.1155/2013/818675
- Castellari, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., et al. (2012). *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res.* 22, 299–306. doi: 10.1101/gr.126516.111
- Chen, Y., Chen, Y., Zhang, J., Cao, P., Su, W., Deng, Y., et al. (2020). *Fusobacterium nucleatum* promotes metastasis in colorectal cancer by activating autophagy signaling via the Upregulation of CARD3 expression. *Theranostics* 10, 323–339. doi: 10.7150/thno.38870
- Chung, L., Thiele Orberg, E., Geis, A. L., Chan, J. L., Fu, K., DeStefano Shields, C. E., et al. (2018). *Bacteroides fragilis* toxin coordinates a pro-carcinogenic inflammatory cascade via targeting of colonic epithelial cells. *Cell Host Microbe* 23, 203–214.e5. doi: 10.1016/j.chom.2018.01.007
- Colnot, S., Niwa-Kawakita, M., Hamard, G., Godard, C., Le Plenier, S., Houbbron, C., et al. (2004). Colorectal cancers in a new mouse model of familial adenomatous polyposis: influence of genetic and environmental modifiers. *Lab. Invest.* 84, 1619–1630. doi: 10.1038/labinvest.3700180
- Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., et al. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* 35, 1069–1076. doi: 10.1038/nbt.3960
- Cross, A. J., Wooldrage, K., Robbins, E. C., Kralj-Hans, I., MacRae, E., Piggott, C., et al. (2019a). Faecal immunochemical tests (FIT) versus colonoscopy for surveillance after screening and polypectomy: a diagnostic accuracy and cost-effectiveness study. *Gut* 68, 1642–1652. doi: 10.1136/gutjnl-2018-317297
- Cross, A. J., Wooldrage, K., Robbins, E. C., Kralj-Hans, I., MacRae, E., Piggott, C., et al. (2019b). Faecal immunochemical tests (FIT) versus colonoscopy

- for surveillance after screening and polypectomy: a diagnostic accuracy and cost-effectiveness study. *Gut* 68, 1642–1652. doi: 10.1136/gutjnl-2018-317297
- Cuevas-Ramos, G., Petit, C. R., Marcq, I., Boury, M., Oswald, E., and Nougayrède, J. P. (2010). *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 107, 11537–11542. doi: 10.1073/pnas.1001261107
- Dai, Z., Coker, O. O., Nakatsu, G., Wu, W. K., Zhao, L., Chen, Z., et al. (2018). Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* 6:70. doi: 10.1186/s40168-018-0451-2
- Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., and Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* 45, W180–W188. doi: 10.1093/nar/gkx295
- Dickinson, B. T., Kisiel, J., Ahlquist, D. A., and Grady, W. M. (2015). Molecular markers for colorectal cancer screening. *Gut* 64, 1485–1494. doi: 10.1136/gutjnl-2014-308075
- Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., and Alm, E. J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* 8:1784. doi: 10.1038/s41467-017-01973-8
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638. doi: 10.1126/science.1110591
- Eklöf, V., Löfgren-Burström, A., Zingmark, C., Edin, S., Larsson, P., Karling, P., et al. (2017). Cancer-associated fecal microbial markers in colorectal cancer detection. *Int. J. Cancer* 141, 2528–2536. doi: 10.1002/ijc.31011
- Eslami, M., Yousefi, B., Kokhaei, P., Hemati, M., Nejad, Z. R., Arabkari, V., et al. (2019). Importance of probiotics in the prevention and treatment of colorectal cancer. *J. Cell Physiol.* 234, 17127–17143. doi: 10.1002/jcp.28473
- Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., et al. (2015). Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat. Commun.* 6:6528. doi: 10.1038/ncomms7528
- Flemer, B., Lynch, D. B., Brown, J. M., Jeffery, I. B., Ryan, F. J., Claesson, M. J., et al. (2017). Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* 66, 633–643. doi: 10.1136/gutjnl-2015-309595
- Flemer, B., Warren, R. D., Barrett, M. P., Cisek, K., Das, A., Jeffery, I. B., et al. (2018). The oral microbiota in colorectal cancer is distinctive and predictive. *Gut* 67, 1454–1463. doi: 10.1136/gutjnl-2017-314814
- Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. U. S. A.* 111, E2329–E2338. doi: 10.1073/pnas.1319284111
- Gao, Z., Guo, B., Gao, R., Zhu, Q., and Qin, H. (2015). Microbiota dysbiosis is associated with colorectal cancer. *Front. Microbiol.* 6:20. doi: 10.3389/fmicb.2015.00020
- Gies, A., Cuk, K., Schrotz-King, P., and Brenner, H. (2018). Direct comparison of diagnostic performance of 9 quantitative fecal immunochemical tests for colorectal cancer screening. *Gastroenterology* 154, 93–104. doi: 10.1053/j.gastro.2017.09.018
- Goodwin, A. C., Destefano Shields, C. E., Wu, S., Huso, D. L., Wu, X., Murray-Stewart, T. R., et al. (2011). Polyamine catabolism contributes to enterotoxigenic *Bacteroides fragilis*-induced colon tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* 108, 15354–15359. doi: 10.1073/pnas.1010203108
- Gudra, D., Shoaie, S., Fridmanis, D., Klovinis, J., Wefer, H., Silamikelis, I., et al. (2019). A widely used sampling device in colorectal cancer screening programmes allows for large-scale microbiome studies. *Gut* 68, 1723–1725. doi: 10.1136/gutjnl-2018-316225
- Guo, S., Li, L., Xu, B., Li, M., Zeng, Q., Xiao, H., et al. (2018). A Simple and novel fecal biomarker for colorectal cancer: ratio of *Fusobacterium Nucleatum* to probiotics populations. *Based on Their Antagonistic Effect.* *Clin. Chem.* 64, 1327–1337. doi: 10.1373/clinchem.2018.289728
- Imperiale, T. F., Ransohoff, D. F., and Itzkowitz, S. H. (2014). Multitarget stool DNA testing for colorectal-cancer screening. *N. Engl. J. Med.* 371, 187–188. doi: 10.1056/NEJMc1405215
- Imperiale, T. F., Ransohoff, D. F., Itzkowitz, S. H., Turnbull, B. A., and Ross, M. E. (2004). Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N. Engl. J. Med.* 351, 2704–2714. doi: 10.1056/nejmoa033403
- Jones, M. B., Highlander, S. K., Anderson, E. L., Li, W., Dayrit, M., Klitgord, N., et al. (2015). Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc. Natl. Acad. Sci. U. S. A.* 112, 14024–14029. doi: 10.1073/pnas.1519288112
- Keum, N., and Giovannucci, E. (2019). Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* 16, 713–732. doi: 10.1038/s41575-019-0189-8
- Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., et al. (2012). Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* 22, 292–298. doi: 10.1101/gr.126573.111
- Kosumi, K., Hamada, T., Koh, H., Borowsky, J., Bullman, S., Twombly, T. S., et al. (2018). The amount of bifidobacterium genus in colorectal carcinoma tissue in relation to tumor characteristics and clinical outcome. *Am. J. Pathol.* 188, 2839–2852. doi: 10.1016/j.ajpath.2018.08.015
- Kuipers, E. J., Grady, W. M., Lieberman, D., Seufferlein, T., Sung, J. J., Boelens, P. G., et al. (2015). Colorectal cancer. *Nat. Rev. Dis. Primers* 1:5065. doi: 10.1038/nrdp.2015.65
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kwong, T. N. Y., Wang, X., Nakatsu, G., Chow, T. C., Tipoe, T., Dai, R. Z. W., et al. (2018). Association between bacteremia from specific microbes and subsequent diagnosis of colorectal cancer. *Gastroenterology* 155, 383–390.e8. doi: 10.1053/j.gastro.2018.04.028
- Laudadio, I., Fulci, V., Palone, F., Stronati, L., Cucchiara, S., and Carissimi, C. (2018). Quantitative assessment of shotgun metagenomics and 16S rDNA amplicon sequencing in the study of human gut microbiome. *Omic* 22, 248–254. doi: 10.1089/omi.2018.0013
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Ley, R. E., Turnbaugh, P. J., Klein, S., and Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* 444, 1022–1023. doi: 10.1038/4441022a
- Liang, J. Q., Chiu, J., and Chen, Y. (2017). Fecal bacteria act as novel biomarkers for non-invasive diagnosis of colorectal cancer. *Clin. Cancer Res.* 23, 2061–2070. doi: 10.1158/1078-0432.ccr-16-1599
- Liang, J. Q., Li, T., Nakatsu, G., Chen, Y. X., Yau, T. O., Chu, E., et al. (2019). A novel faecal *Lachnospirillum* marker for the non-invasive diagnosis of colorectal adenoma and cancer. *Gut* 69, 1248–1257. doi: 10.1136/gutjnl-2019-318532
- Lofton-Day, C., Model, F., Devos, T., Tetzner, R., Distler, J., Schuster, M., et al. (2008). DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin. Chem.* 54, 414–423. doi: 10.1373/clinchem.2007.095992
- Loh, K., Chia, J., and Greco, S. (2010). Bone morphogenic protein 3 inactivation is an early and frequent event in colorectal cancer development. *Genes Chromosomes Cancer* 47, 449–460. doi: 10.1002/gcc.20552
- Long, X., Wong, C. C., Tong, L., Chu, E. S. H., Ho Szeto, C., Go, M. Y. Y., et al. (2019). *Peptostreptococcus anaerobius* promotes colorectal carcinogenesis and modulates tumour immunity. *Nat. Microbiol.* 4, 2319–2330. doi: 10.1038/s41564-019-0541-3
- Malagón, M., Ramió-Pujol, S., Serrano, M., Serra-Pagès, M., Amoedo, J., Oliver, L., et al. (2019). Reduction of faecal immunochemical test false-positive results using a signature based on faecal bacterial markers. *Aliment. Pharmacol. Ther.* 49, 1410–1420. doi: 10.1111/apt.15251
- Marchesi, J. R., and Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome* 3:31. doi: 10.1186/s40168-015-0094-5
- Melotte, V., Lentjes, M. H. F. M., Van, d. B., and Sandra, M. (2009). N-Myc Downstream-Regulated Gene 4 (NDRG4): a candidate tumor suppressor gene and potential biomarker for colorectal cancer. *J. Nat. Cancer Inst.* 101, 916–927. doi: 10.1093/jnci/djp131
- Mima, K., Nishihara, R., Qian, Z. R., Cao, Y., Sukawa, Y., Nowak, J. A., et al. (2016). *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut* 65, 1973–1980. doi: 10.1136/gutjnl-2015-310101
- Minot, S. S., and Willis, A. D. (2019). Clustering co-abundant genes identifies components of the gut microbiome that are reproducibly associated with colorectal cancer and inflammatory bowel disease. *Microbiome* 7:110. doi: 10.1186/s40168-019-0722-6

- Mira-Pascual, L., Cabrera-Rubio, R., Ocon, S., Costales, P., Parra, A., Suarez, A., et al. (2015). Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. *J. Gastroenterol.* 50, 167–179. doi: 10.1007/s00535-014-0963-x
- Nakatsu, G., Li, X., Zhou, H., Sheng, J., Wong, S. H., Wu, W. K., et al. (2015). Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat. Commun.* 6:8727. doi: 10.1038/ncomms9727
- Nayfach, S., and Pollard, K. S. (2016). Toward accurate and quantitative comparative metagenomics. *Cell* 166, 1103–1116. doi: 10.1016/j.cell.2016.08.007
- O'Brien, C. L., Allison, G. E., Grimpen, F., and Pavli, P. (2013). Impact of colonoscopy bowel preparation on intestinal microbiota. *PLoS One* 8:e62815. doi: 10.1371/journal.pone.0062815
- O'Keefe, S. J. (2016). Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* 13, 691–706. doi: 10.1038/nrgastro.2016.165
- Paramsothy, S., Nielsen, S., Kamm, M. A., Deshpande, N. P., Faith, J. J., Clemente, J. C., et al. (2019). Specific bacteria and metabolites associated with response to fecal microbiota transplantation in patients with ulcerative colitis. *Gastroenterology* 156, 1440–1454.e2. doi: 10.1053/j.gastro.2018.12.001
- Pitot, H. C. (1993). The molecular biology of carcinogenesis. *Cancer* 72(3 Suppl.), 962–970.
- Pleguezuelos-Manzano, C., Puschhof, J., Huber, A. R., van Hoesck, A., Wood, H. M., Nomburg, J., et al. (2020). Mutational signature in colorectal cancer caused by genotoxic pks(+) *E. coli*. *Nature* 580, 7269–7273. doi: 10.1038/s41586-020-2080-8
- Poore, G. D., Kopylova, E., Zhu, Q., Carpenter, C., Fraraccio, S., Wandro, S., et al. (2020). Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* 579, 567–574. doi: 10.1038/s41586-020-2095-1
- Poyet, M., Groussin, M., Gibbons, S. M., Avila-Pacheco, J., Jiang, X., Kearney, S. M., et al. (2019). A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat. Med.* 25, 1442–1452. doi: 10.1038/s41591-019-0559-3
- Qin, J., Li, R., Raes, J., Arumugam, M., and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821
- Rastelli, M., Cani, P. D., and Knauf, C. (2019). The gut microbiome influences host endocrine functions. *Endocr. Rev.* 40, 1271–1284. doi: 10.1210/er.2018-00280
- Rezasoltani, S., Sharafkhan, M., Asadzadeh Aghdaei, H., Nazemalhosseini Mojarad, E., Dabiri, H., Akhavan Sepahi, A., et al. (2018). Applying simple linear combination, multiple logistic and factor analysis methods for candidate fecal bacteria as novel biomarkers for early detection of adenomatous polyps and colon cancer. *J. Microbiol. Methods* 155, 82–88. doi: 10.1016/j.mimet.2018.11.007
- Routy, B., Gopalakrishnan, V., Daillère, R., Zitvogel, L., Wargo, J. A., and Kroemer, G. (2018). The gut microbiota influences anticancer immunosurveillance and general health. *Nat. Rev. Clin. Oncol.* 15, 382–396. doi: 10.1038/s41571-018-0006-2
- Rubinstein, M. R., Baik, J. E., Lagana, S. M., Han, R. P., Raab, W. J., Sahoo, D., et al. (2019). *Fusobacterium nucleatum* promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO Rep.* 20:e47638. doi: 10.15252/embr.201847638
- Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y. W. (2013). *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* 14, 195–206. doi: 10.1016/j.chom.2013.07.012
- Sacks, D., Baxter, B., Campbell, B. C. V., Carpenter, J. S., Cognard, C., Dippel, D., et al. (2018). Multisociety consensus quality improvement revised consensus statement for endovascular therapy of acute ischemic stroke. *Int. J. Stroke* 13, 612–632. doi: 10.1177/1747493018778713
- Saffarian, A., Mulet, C., Regnault, B., Amiot, A., Tran-Van-Nhieu, J., Ravel, J., et al. (2019). Crypt- and mucosa-associated core microbiotas in humans and their alteration in colon cancer patients. *mBio* 10, e1315–e1319. doi: 10.1128/mBio.01315-19
- Schloss, P. D., and Westcott, S. L. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl. Environ. Microbiol.* 77, 3219–3226. doi: 10.1128/aem.02810-10
- Shah, M. S., Desantis, T. Z., Weinmaier, T., Mcmurdie, P. J., Cope, J. L., Altrichter, A., et al. (2018). Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. *Gut* 67, 882–891. doi: 10.1136/gutjnl-2016-313189
- Shi, W., Qi, H., Sun, Q., Fan, G., Liu, S., Wang, J., et al. (2019). gcMeta: a Global Catalogue of Metagenomics platform to support the archiving, standardization and analysis of microbiome data. *Nucleic Acids Res.* 47, D637–D648. doi: 10.1093/nar/gky1008
- Sobhani, I., Tap, J., Roudot-Thoraval, F., Roperch, J. P., Letulle, S., Langella, P., et al. (2011). Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS One* 6:e16393. doi: 10.1371/journal.pone.0016393
- Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., Gratadoux, J. J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16731–16736. doi: 10.1073/pnas.0804812105
- Sze, M. A., and Schloss, P. D. (2018). Leveraging Existing 16S rRNA gene surveys to identify reproducible biomarkers in individuals with colorectal tumors. *mBio* 9, e630–e618. doi: 10.1128/mBio.00630-18
- Tabib, N. S. S., Madgwick, M., Sudhakar, P., Verstockt, B., Korcsmaros, T., and Vermeire, S. (2020). Big data in IBD: big progress for clinical practice. *Gut* 69, 1520–1532. doi: 10.1136/gutjnl-2019-320065
- Thomas, A. M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., et al. (2019a). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* 25, 667–678. doi: 10.1038/s41591-019-0405-7
- Thomas, A. M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., et al. (2019b). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* 25, 667–678. doi: 10.1038/s41591-019-0405-7
- Tilg, H., Adolph, T. E., Gerner, R. R., and Moschen, A. R. (2018). The intestinal microbiota in colorectal cancer. *Cancer Cell* 33, 954–964. doi: 10.1016/j.ccell.2018.03.004
- Tjalsma, H., Boleij, A., Marchesi, J. R., and Dutilh, B. E. (2012). A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat. Rev. Microbiol.* 10, 575–582. doi: 10.1038/nrmicro2819
- Viljoen, K. S., Dakshinamurthy, A., Goldberg, P., and Blackburn, J. M. (2015). Quantitative profiling of colorectal cancer-associated bacteria reveals associations between fusobacterium spp., enterotoxigenic *Bacteroides fragilis* (ETBF) and clinicopathological features of colorectal cancer. *PLoS One* 10:e0119462. doi: 10.1371/journal.pone.0119462
- Villéger, R., Lopès, A., Veziat, J., Gagnière, J., Barnich, N., Billard, E., et al. (2018). Microbial markers in colorectal cancer detection and/or prognosis. *World J. Gastroenterol.* 24, 2327–2347. doi: 10.3748/wjg.v24.i22.2327
- Vogtmann, E., Hua, X., Zeller, G., Sunagawa, S., Voigt, A. Y., Hercog, R., et al. (2016). Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS One* 11:e0155362. doi: 10.1371/journal.pone.0155362
- Voigt, A. Y., Costea, P. I., Kultima, J. R., Li, S. S., Zeller, G., Sunagawa, S., et al. (2015). Temporal and technical variability of human gut metagenomes. *Genome Biol.* 16:73. doi: 10.1186/s13059-015-0639-8
- Wang, L., Tang, L., Feng, Y., Zhao, S., Han, M., Zhang, C., et al. (2020). A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium blunts colitis associated tumorigenesis by modulation of CD8(+) T cells in mice. *Gut* 69, 1988–1997. doi: 10.1136/gutjnl-2019-320105
- Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* 25, 679–689. doi: 10.1038/s41591-019-0406-6
- Wong, S. H., Kwong, T. N. Y., Chow, T. C., Luk, A. K. C., Dai, R. Z. W., Nakatsu, G., et al. (2017a). Quantitation of faecal *Fusobacterium* improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut* 66, 1441–1448. doi: 10.1136/gutjnl-2016-312766
- Wong, S. H., and Yu, J. (2019). Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* 16, 690–704. doi: 10.1038/s41575-019-0209-8
- Wong, S. H., Zhao, L., Zhang, X., Nakatsu, G., Han, J., Xu, W., et al. (2017b). Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. *Gastroenterology* 153, 1621–1633.e6. doi: 10.1053/j.gastro.2017.08.022

- Wong, S. H., Zhao, L., Zhang, X., Nakatsu, G., and Yu, J. (2017c). Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. *Gastroenterology* 153, 1621–1633. doi: 10.1053/j.gastro.2017.08.022
- Wu, J., Li, Q., and Fu, X. (2019). *Fusobacterium nucleatum* contributes to the carcinogenesis of colorectal cancer by inducing inflammation and suppressing host immunity. *Transl. Oncol.* 12, 846–851. doi: 10.1016/j.tranon.2019.03.003
- Wu, N., Yang, X., Zhang, R., Li, J., Xiao, X., Hu, Y., et al. (2013). Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb. Ecol.* 66, 462–470. doi: 10.1007/s00248-013-0245-9
- Xia, Y., Sun, J., and Chen, D.-G. (2018). *Statistical Analysis of Microbiome Data with R*. Singapore: Springer Singapore.
- Xie, Y. H., Gao, Q. Y., and Cai, G. X. (2017). Fecal clostridium symbiosum for noninvasive detection of early and advanced colorectal cancer: test and validation studies. *Ebiomedicine* 25, 32–40. doi: 10.1016/j.ebiom.2017.10.005
- Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., et al. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* 25, 968–976. doi: 10.1038/s41591-019-0458-7
- Yang, J., Li, D., Yang, Z., Dai, W., Feng, X., Liu, Y., et al. (2020). Establishing high-accuracy biomarkers for colorectal cancer by comparing fecal microbiomes in patients with healthy families. *Gut Microbes* 13, 918–929. doi: 10.1080/19490976.2020.1712986
- Yang, Y., Misra, B. B., Liang, L., Bi, D., Weng, W., Wu, W., et al. (2019). Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. *Theranostics* 9, 4101–4114. doi: 10.7150/thno.35186
- Yang, Y., Weng, W., Peng, J., Hong, L., Yang, L., Toiyama, Y., et al. (2017). *Fusobacterium nucleatum* increases proliferation of colorectal cancer cells and tumor development in mice by activating toll-like receptor 4 signaling to nuclear factor- κ B, and up-regulating expression of MicroRNA-21. *Gastroenterology* 152, 851–866.e24. doi: 10.1053/j.gastro.2016.11.018
- Yeoh, Y. K., Chen, Z., Wong, M. C. S., Hui, M., Yu, J., Ng, S. C., et al. (2020). Southern Chinese populations harbour non-nucleatum *Fusobacteria* possessing homologues of the colorectal cancer-associated FadA virulence factor. *Gut* 69, 1998–2007. doi: 10.1136/gutjnl-2019-319635
- Yim, W. W., Yetisgen, M., Harris, W. P., and Kwan, S. W. (2016). Natural language processing in oncology: a review. *JAMA Oncol.* 2, 797–804. doi: 10.1001/jamaoncol.2016.0213
- Youssef, O., Lahti, L., Kokkola, A., Karla, T., Tikkanen, M., Ehsan, H., et al. (2018). Stool microbiota composition differs in patients with stomach. *Colon, and Rectal Neoplasms. Dig. Dis. Sci.* 63, 2950–2958. doi: 10.1007/s10620-018-5190-5
- Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. Y., Qin, Y., et al. (2017a). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 66, 70–78. doi: 10.1136/gutjnl-2015-309800
- Yu, T., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., et al. (2017b). *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* 170, 548–563.e16. doi: 10.1016/j.cell.2017.07.008
- Zackular, J. P., Rogers, M. A., Ruffin, M. T. T., and Schloss, P. D. (2014). The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev. Res. (Phila)* 7, 1112–1121. doi: 10.1158/1940-6207.Capr-14-0129
- Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* 10:766. doi: 10.15252/msb.20145645

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhou, Ge, Li, Ma, Liu, Hu, Zhang, Ma, Du, Syed and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.