



OPEN ACCESS

EDITED BY

Xuehua Wan,
Nankai University, China

REVIEWED BY

Valentin Pavlovich Volobaev,
Sirius University, Russia
Marcela Pereira,
Karolinska Institutet (KI), Sweden

*CORRESPONDENCE

Ping Jiang

✉ hsyblk1987@hotmail.com

SPECIALTY SECTION

This article was submitted to
Gastrointestinal Cancers:
Colorectal Cancer,
a section of the journal
Frontiers in Oncology

RECEIVED 05 December 2022

ACCEPTED 30 December 2022

PUBLISHED 23 January 2023

CITATION

Zhu A, Liu Y, Li Z, He Y, Bai L, Wu Y,
Zhang Y, Huang Y and Jiang P (2023)
Diagnosis and functional prediction of
microbial markers in tumor tissues of
sporadic colorectal cancer patients
associated with the MLH1 protein
phenotype.

Front. Oncol. 12:1116780.

doi: 10.3389/fonc.2022.1116780

COPYRIGHT

© 2023 Zhu, Liu, Li, He, Bai, Wu, Zhang,
Huang and Jiang. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Diagnosis and functional prediction of microbial markers in tumor tissues of sporadic colorectal cancer patients associated with the MLH1 protein phenotype

Anchao Zhu^{1,2}, Yingying Liu³, Zongmin Li¹, Ying He⁴, Lijing Bai⁵,
Youtian Wu¹, Yuying Zhang¹, Ying Huang¹ and Ping Jiang^{1*}

¹Department of Pathology, Harbin First Hospital, Harbin, China, ²Department of Pathology, Harbin Medical University, Harbin, China, ³Department of Pathology, Heilongjiang Provincial Hospital, Harbin, China, ⁴Department of Gastroenterology, Harbin First Hospital, Harbin, China, ⁵Department of Laboratory Diagnosis, First Affiliated Hospital of Harbin Medical University, Harbin, China

Objective: Most patients with sporadic colorectal cancer (SCRC) develop microsatellite instability because of defects in mismatch repair (MMR). Moreover, the gut microbiome plays a vital role in the pathogenesis of SCRC. In this study, we assessed the microbial composition and diversity of SCRC tumors with varying MutL protein homolog 1 (MLH1) status, and the effects of functional genes related to bacterial markers and clinical diagnostic prediction.

Methods: The tumor microbial diversity and composition were profiled using high-throughput sequencing of the 16S ribosomal RNA (rRNA) gene V4 region. Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) software and BugBase tool were used to predict the functional roles of the microbiome. We aimed to construct a high-accuracy model to detect and evaluate the area under the receiver operating characteristic curve with candidate biomarkers.

Results: The study included 23 patients with negative/defective MLH1 (DM group) and 22 patients with positive/intact MLH1 (IM group). Estimation of alpha diversity indices showed that the Shannon index ($p = 0.049$) was significantly higher in the DM group than in the controls, while the Simpson index ($p = 0.025$) was significantly lower. At the genus level, we observed a significant difference in beta diversity in the DM group versus the IM group. Moreover, the abundance of *Lachnospirillum* spp. and *Coproccoccus* spp. was significantly more enriched in the DM group than in the IM group ($q < 0.01$ vs. $q < 0.001$). When predicting metagenomes, there were 18 Kyoto Encyclopedia of Genes and Genomes

pathways and one BugBase function difference in both groups (all $q < 0.05$). On the basis of the model of diagnostic prediction, we built a simplified optimal model through stepwise selection, consisting of the top two bacterial candidate markers (area under the curve = 0.93).

Conclusion: In conclusion, the genera *Lachnoclostridium* and *Coprococcus* as key species may be crucial biomarkers for non-invasive diagnostic prediction of DM in patients with SCRC in the future.

KEYWORDS

colorectal cancer, 16S rRNA gene, biomarker, MLH1, microbiome

Introduction

According to Cancer Statistics, 2021, colorectal carcinoma (CRC) is the third most prevalent cancer in the United States, with sporadic and microsatellite-stable (MSS) CRC accounting for approximately 80–85% of the cases and microsatellite-instability (MSI) accounting for the rest (1, 2). In addition, sporadic CRC (SCRC, a subtype of CRC) development is related to several potential etiological factors, such as lifestyle, methylation, genetics (BRAF or MSI mutation), and the gut microbiome, which affect the prognosis and survival of patients with SCRC (3–5). However, the relationship between the mismatch repair (MMR) system and microbial markers, which is the main factor affecting the prognosis and survival of patients with SCRC, remains unclear. Accumulating evidence indicates that the gut microbiome and its metabolites are closely related to the development and progression of SCRC (6).

According to the National Comprehensive Cancer Network® (NCCN) Clinical Practice Guidelines in Oncology (NCCN Guidelines), bevacizumab is suitable as the first-line treatment before maintenance therapy for patients with unresectable tumors such as metastatic CRC (7, 8). Retrospective studies have found that although patients with high-MSI (MSI-H) CRC generally have a better prognosis than those with low MSI (MSI-L) or MSS CRC, they do not benefit from postoperative adjuvant 5-fluorouracil based chemotherapy (9). Recently, cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed death 1 (PD-1) blockers have attracted widespread attention as immunotherapy agents and have been proven to be effective treating several types of solid tumors (10). The gut microbiome has a prominent influence on the clinical effects of solid tumor treatment using immunotherapy (11, 12). However, patients with SCRC have not benefitted from immunotherapy, which is likely to be effective only if the tumor exhibits MSI (13). Therefore, an in-depth study of the composition and function of the gut microbiome may help determine the therapeutic direction for patients with SCRC.

Currently, defective MMR genes such as MLH1, MSH2, MSH6, and PMS2 can be detected in tumor and normal tissues (usually blood or adjacent normal tissues) using immunohistochemistry (IHC) and polymerase chain reaction (PCR) (14). However, researchers have discovered that PCR-based MSI detection methods are subpar (15); hence, the IHC method is preferred in clinical pathology practice.

According to mutation patterns, deficient mismatch repair (DMMR) proteins are generally considered equivalent to MSI-H, and proficient mismatch repair (PMMR) proteins are considered equivalent to MSI-L or MSS. Therefore, MSI is a molecular manifestation of defects in the MMR system and acquired abnormal methylation of the MLH1 gene promoter region is considered the main cause of SCRC (16). MSI-H/DMMR is a tumor marker in patients with Lynch syndrome; however, data from examining the MMR gene somatic defect revealed a sporadic event in 70–85% (14) of cases. With increasing evidence of a subtle link between the phenotypes of specific proteins in patients with SCRC and the gut microbiome, differences in gut microbiome diversity between MMR-deficient and MMR-nondeficient SCRC cohorts need further exploration.

The number of gut microorganisms in humans is over 100 trillion, and some of them participate in the host's metabolic activities and immune regulatory functions (17). Research has shown that intestinal bacteria play a critical role in determining intestinal health and that specific microbes, such as *Fusobacterium (F.) nucleatum* (18, 19) and *Peptostreptococcus anaerobius* (20), promote colon tumor development. Using 16S rRNA gene sequencing, several microbial species have been found to pose a carcinogenic risk to host intestinal epithelial cells. Previously, *Escherichia coli* NC101 was shown to regulate metabolic pathways and activate oncogenes, which can cause inflammatory changes in intestinal tissue (21). Furthermore, *F. nucleatum* was shown to influence host immune responses according to the MSI status (22, 23). Thus, microbial infection, and chronic inflammation are potential risk factors for intestinal malignancy, but the interactions of related molecular phenotypes (such as DNA repair proteins) with the gut microbiome remain unclear.

This study explored the diversity and composition of the gut microbiome from different perspectives and determined the possible functional pathways to gather relevant data for subsequent microbiome research related to MLH1 deficiency. PCR and IHC are used to methods evaluate tissues obtained after surgical resection, requiring patients to undergo invasive procedures. Therefore, we explored the feasibility of developing a non-invasive prediction method with high predictive value and designed an experimental study to identify bacterial gene markers in the tumor tissue that are strongly associated with MLH1 protein deficiency.

Materials and methods

Human specimens and procedure

In this study, BRAF mutations were an important aspect of screening for patients with MSI-positive SCRC after CRC MLH1 deletion but not for patients with MSI-negative (24–27). Subsequently, the data on family history were combined with the screening data to determine whether these patients had SCRC. Accordingly, patients with SCRC were screened on the basis of the Chinese expert consensus on clinical diagnosis, treatment, and pedigree management of hereditary CRC as recommended by the Colorectal Cancer Professional Committee of the Chinese Anti-Cancer Association (28) (Figure 1A). Following the recommended procedure, we designed the scheme for this study (Figure 1B). From 2020 to 2021, we retrospectively evaluated the pathology data of 491 patients from Northeast China, including 112 and 379 patients from the Harbin First Hospital and from the Heilongjiang Provincial Hospital, respectively. Basic clinical demographics and clinicopathological data for all the patients were obtained from the hospital's electronic medical records. Formalin-fixed paraffin-embedded (FFPE) tissue specimens were obtained by surgical resection in untreated patients with SCRC. The inclusion criteria were as follows: no history of drug consumption such as antibiotics or probiotics within the past 3 months, no family history of Lynch syndrome, no surgical history, or no treatment history. This study was approved by the Ethics Committee of Harbin First Hospital (Approval No. Q2021-064).

IHC and gram staining

For IHC staining, an SP (Rabbit) IHC Kit (Gene Tech, Shanghai, China) was used, according to the manufacturer's instructions. Staining was performed as previously described (29), and all surgically resected tumor tissues were immediately fixed in 10% neutral formaldehyde solution. For the IHC procedure, 4- μ m sections were used. MLH1, PMS2, MSH2, MSH6, and BRAF monoclonal antibodies were added to each sample, and the sections were fixed with the neutral resin for observation. Xylene was used in the deparaffinizing process of the slides, followed by graded ethanol for hydrated. Gram staining (Baso, Zhuhai, China) was performed on all tissue sections of equal thickness by referring to the reagent preparation instructions. The stained tissue slides were observed using the LEICA DM3000 light microscope at 400 \times and 1000 \times magnifications. The diagnostic results of all clinicopathological sections were confirmed through double-blind observation by three pathologists with deputy senior or higher titles. External positive control tissue was included on each slide to ensure the reliability of the experimental results.

DNA extraction and 16s rRNA gene sequencing

Bacterial genomic DNA was extracted from each sample using the FastDNA[®] SPIN Kit for Soil (MP Biomedicals, United States). Ten 8- μ m FFPE tissue sections were dewaxed, bacterial genomic DNA was

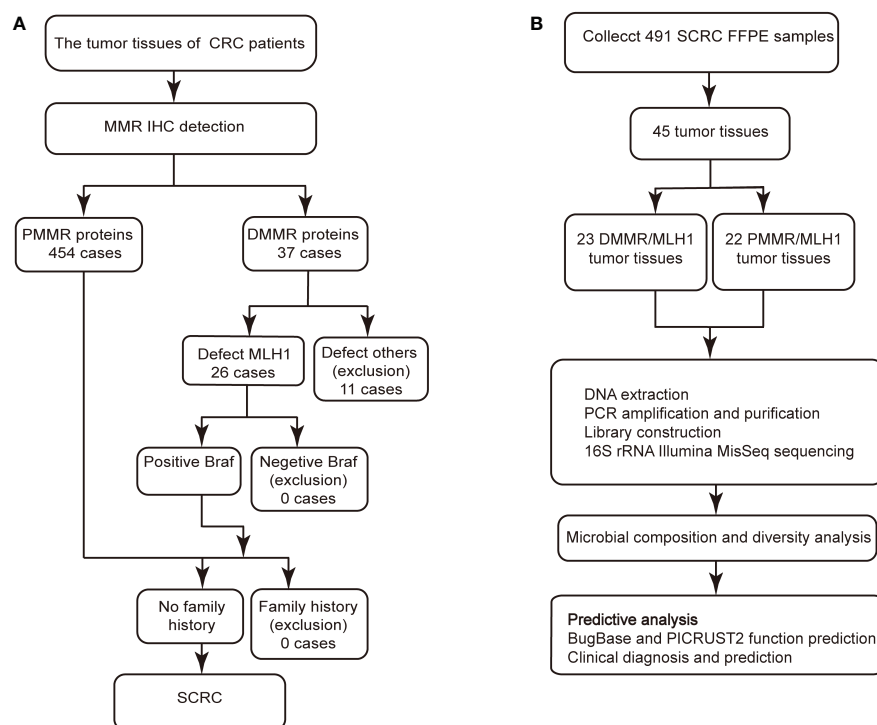


FIGURE 1

IHC staining and Gram staining were performed between the two groups. (A) Screening procedure for SCRC patients by MMR immunohistochemistry. (B) Study design and flow diagram. SCRC, sporadic colorectal carcinoma; MMR, mismatch repair; DMMR, deficient mismatch repair; PMMR, proficient mismatch repair.

extracted, and the extracted DNA was purified in a 1.5 mL Eppendorf tube using 100% ethanol precipitation. DNA quantification and purity was performed using the NanoDrop 2000 UV-vis spectrophotometer (Thermo Scientific, Wilmington, USA). The integrity and size of the DNA fragments were determined using 1% agarose gel electrophoresis. All 16S rRNA genes were extracted using previously developed methods (30–32) and were stored at -20°C until further processing.

PCR was used to amplify the 16S (variable region 4 [v4]) rRNA genes using the 515F–806R primer, which is a general primer (forward: 5'-GTGCCAGCMGCCGCGG-3'; reverse: 5'-GGACTACHVGGGT WTCTAAT-3'). The PCR products were extracted from 2% agarose gel electrophoresis and purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA). Subsequently, the amplicons were quantified using QuantusTM Fluorometer (Promega, USA). We used the TruSeq Nano DNA LT Library Prep Kit to prepare DNA library. In accordance with the manufacturer's instructions, the prepared library was checked using the 2x300 bp paired-end protocol (Majorbio Bio-Pharm Technology, Shanghai, China) for subsequent raw sequencing data on the Illumina MiSeqPE300 platform (Illumina, San Diego, USA).

The raw data were demultiplexed and quality-filtered using the FASTQ format (33) and were merged using FLASH (version 1.2.11¹) (34) according to the minimum identity of the sample sequence within a cluster. Refer to Yu et al. (35), the data processing needed to proceed the corresponding criteria. The methods of USEARCH (version 11.0²) were used for the operational taxonomic unit (OTU) clustering analysis with a 97% threshold. The ribosomal database project classifier (36) (version 2.13³) was used to against the SILVA (version 138⁴)/16S rRNA database. The OTU representative sequences were annotated with a confidence threshold of 70% (37), indicating that the similarity of more than 70% is the same species.

Prediction of functional modules

The annotated OTU matrix of microbial characteristics was predicted using the BugBase⁵ microbiome analysis tool based on nine phenotypic categories. Moreover, by integrating the standard functional gene content of species using the PICRUST2 software (version 2.2.0⁶), the proportional abundance spectrum of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database in SCRC samples was determined. The Wilcoxon signed-rank test was used as the inspection method.

1 <https://ccb.jhu.edu/software/FLASH/index.shtml>

2 <http://www.drive5.com/usearch/>

3 <http://rdp.cme.msu.edu/>

4 <https://www.arb-silva.de/>

5 <https://bugbase.cs.umn.edu/index.html>

6 <https://github.com/picrust/picrust2/>

Clinical diagnosis prediction analysis

Random forest (RF) analysis was used to classify and predict the accuracy and adaptability of the data, and the receiver operating characteristic (ROC) validation algorithm was applied. The software used in this model prediction analysis included the random forest package (R) and plotROC package (R). Based on previous descriptions (38), the number of decision trees was set to 500, and the relative abundance of OTUs was used for data standardization processing. The method of 10-fold cross-validation was simultaneously used to evaluate the important ranking variables to determine the minimum predicted number of taxa. The ROC curve was used to predict clinical diagnostic biomarkers in the microbiome of patients with CRC, and confidence intervals (CIs) were set at 95% for the CRC risk. The area under the ROC curve was interpreted using the following criteria: the area under the curve (AUC) had a low accuracy when it was between 0.5 and 0.7, a certain accuracy between 0.7 and 0.9, and a high accuracy above 0.9. An AUC of 0.5 indicated that the diagnostic method was completely ineffective and had no diagnostic value.

Statistical analysis

Categorical variables in quantitative data were compared using the chi-square and one-way analysis of variance tests, with the SPSS software (version 21.0). Statistical significance was set at $p < 0.05$. The dilution curve was plotted using the R software, with the number of OTUs corresponding to each sample. Alpha diversity analysis was used to determine the diversity and richness of the microbial colonies using the Mothur (version 1.30.2) software package. Principal coordinate analysis (PCOA) was used to evaluate the structural distribution of microbial colony samples through dimensionality reduction so as to simplify the structure of sequencing data on the matrix of Bray Curtis and weighted UniFrac distances. The heatmap was generated according to the abundance of genera using R. In this study, CIs were calculated using a two-tailed test. To assess the differences in bacterial abundance of various groups, we performed the Mann–Whitney U test for different microbial system classification units, and the adjusted p value (q value) < 0.01 , which was obtained through the false discovery rate (FDR) step-up procedure, was considered significant for multiple comparisons.

Results

MMR mutation and BRAF mutation analysis with CRC tumor tissues

We conducted IHC tests on all surgically resected specimens obtained from adult patients with CRC in which the positive or negative expression of MMR proteins (MLH1, PMS2, MSH2, or MSH6) was observed in the cell nuclei. The positive expression of MMR proteins showed staining of tumor nuclei and a brown color, whereas the negative expression showed no staining (Figure 2A). The loss of heterodimeric binding of MLH1 with PMS2 (or of MSH2 with MSH6) indicates an underlying mutation in PMMR, dominated by MLH1 (or MSH2). Of the 491 SCRC cases, 454 (92.5%) and 37 (7.5%) involved PMMR and DMMP proteins, respectively. In DMMP proteins,

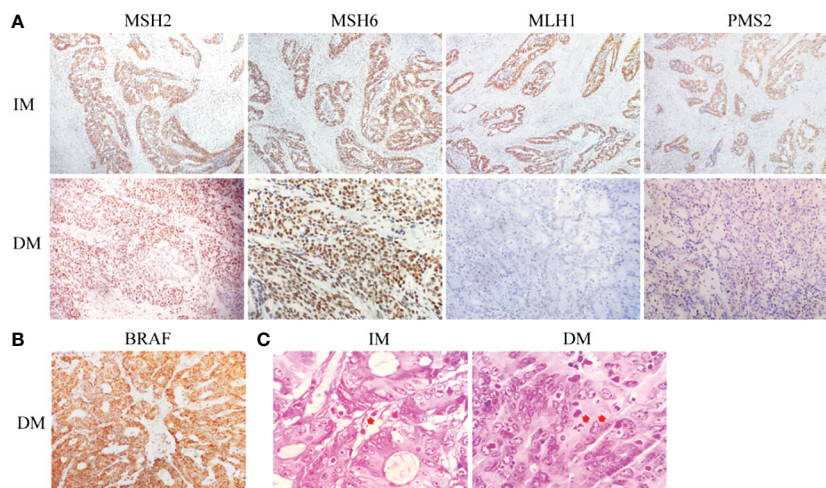


FIGURE 2
Histological staining of tumor specimens in both groups. **(A)** The corresponding IHC staining, followed by MSH2, MSH6, MLH1, and PMS2. **(B)** BRAF staining. **(C)** Gram staining (the arrow indicates piles of bacterial infections).

MLH1 deficiency (simultaneous deletion of PMS2) was the most prevalent mutation and was identified in 70.3% (26/37) of the SCRC cases, whereas only 7 cases (18.9%) showed PMS2 deficiency (Table 1). Of these, the expression patterns of 3 cases with MLH1 deficiency were not representative, leaving 23 cases for further analysis. Meanwhile, the proportion of protein deletions in MSH2 and MSH6 was lower, accounting for only 10.8% (4/37) of the cases. Subsequently, we performed a supplementary study of BRAF markers (unpublished data), showing that all defective MLH1 cases had BRAF-positive expression (Figure 2B); that is, the presence of BRAF mutations can be considered as an inclusion criterion for patients with SCRC.

Demographic characteristics and screening of SCRC patients with different MLH1 levels

The role of tumor microbiome composition in the clinical prognosis of SCRC tumor tissues that manifest with varying MLH1 status was explored. This study included 23 DMMR/MLH1 patients and 22 stage-matched PMMR/MLH1 controls. Additionally, patients in both groups were matched with the provided general information (age and sex) and clinicopathologic characteristics (including tumor diameter, differentiation, TNM stage, and American Joint Committee on Cancer stage), as shown in Table 2. There were no significant

TABLE 1 Amount of the patterns in 37 SCRC patients with DMMR.

DMMR proteins	SCRC (No. of patients)	Percentage (%)
MLH1(-)	26	70.3
PMS2(-)	7	18.9
MSH2(-)	3	8.1
MSH6(-)	1	2.7
Total	37	100

DMMR, deficient mismatch repair; SCRC, sporadic colorectal cancer.

differences in these variables between the two groups (all $p > 0.05$), except for tumor size ($p < 0.01$). Gram staining showed the morphology and structure of bacteria in the surrounding and adjacent stroma of cancer cells (Figure 2C). Subsequently, bacterial DNA was extracted from these specimens, and taxonomic profiling via 16S rRNA gene sequencing was performed.

Comparison of microbial diversity between defective and intact MLH1

A total of 2,421,631 high-quality sequences with an average length of 256 base pairs were generated. To determine whether the overall mean microbial community richness and diversity were different in negative/defective MLH1 (DM) versus positive/intact MLH1 (IM), we compared six measures of alpha diversity by sampling the following OTUs: Sobs and Chao1, which measure community richness; Shannon, Simpson, and phylogenetic diversity (PD) indices which estimate community diversity; and the Heip index, which describes the metric of community evenness (Figure 3A). The higher the value of Sobs and Chao1 index, the more the total number of species. There was a negative correlation between the Simpson index and community diversity, while a positive correlation between the Shannon index and community diversity. There were no significant differences in the Sobs, Chao1, Heip, and PD indices between DM and IM ($p = 0.078, 0.071, 0.061,$ and 0.082 , respectively); however, the Shannon and Simpson indices were significantly different ($p = 0.049$ and 0.025 , respectively). The Shannon index indicated that the DM group showed a larger variation in α -diversity values than the IM controls, and the Simpson index showed the opposite trend; however, the two groups had more constant values in the other indices. To detect whether the overall enriched microbial communities were different between the two groups, we performed PCOA for each group. At the phylum level, analysis of SCRC cases stratified by immune phenotype revealed that there was no significant difference in bacterial relative OTUs between the two groups ($p > 0.05$, Figure 3B), as distinct clusters could not be noticed for both types of tumors in the comparison. At the genus level, the PCOA

TABLE 2 Clinicopathological features of SCRC patients.

Characteristics	Total (n=45)	DM (n=23)	IM (n=22)	p-value
Age(years)	65.00 ± 11.04	64.30 ± 13.10	65.73 ± 8.97	0.674
Gender				0.884
Male	23	12	11	
Female	22	11	11	
Tumor Diameter(cm)	4.9 ± 1.6	5.5 ± 1.8	4.2 ± 1.0	0.005**
Differentiation				0.936
Well to Moderate	35	18	17	
Poor	10	5	5	
Primary Tumor				0.699
Low T-stage(T1/T2)	7	3	4	
High T-stage (T3)	38	20	18	
Lymph Node Metastasis				0.314
Node Negative(N0)	34	19	15	
Node Positive(N1/N2)	11	4	7	
Distant Metastasis				0.233
M0	43	23	20	
M1	2	0	2	
AJCC Stage				0.314
I-II	34	19	15	
III-IV	11	4	7	

DM, defective MLH1; IM, intact MLH1. **p < 0.01.

results illustrated a statistically significant difference for both Bray Curtis ($p = 0.038$) and weighted UniFrac ($p = 0.038$) measures (Figure 3C). Several sample nodes between the two groups overlapped in the PCOA distribution. On the basis of these observations, we plotted the proportion of microorganisms in each sample in the DM and IM groups to show the corresponding relationship and abundance of information more intuitively between both groups. Community investigation at the phylum (Figure 4A) and genus levels (Figure 4B) showed that some cases in both groups had similar sample nodes, which corresponded to their PCOA distribution. Collectively, clustering of community composition at different taxonomic levels was observed between the two groups of OTUs, indicating that the tumor microbiome of each group was similar in phylogeny.

Tumor microbial communities are significantly different between DM and IM

After the taxa subsampling step, 54 phyla, 165 classes, 404 orders, 687 families, 1467 genera, 2842 species, and 6846 OTUs were identified in the 45 patients with SCRC, most of which were identified in the two sample types. With respect to the relationship

between the intratissue bacterial diversity and protein phenotype of SCRC, we aimed to determine whether there were differences in tumor microbiome composition between the two groups. We first assessed the overall taxonomic profile of the tumor microbiota in all the patients, uncovering similar communities. Figures 5A and 5B depict the results at the taxonomic level; the Venn diagram shows 42 phyla at the phylum level in both groups. Eight of the fifty phyla were unique to the DM group, compared with the IM group, whereas 384 of the 1327 genera were unique to the DM group.

Microbiomes with 1% relative abundance at different taxonomic levels were classified as “others”. At the phylum level, *Proteobacteria*, *Actinobacteria*, *Firmicutes*, and *Bacteroidetes* were the most common in the two groups during gut microbiome analysis, accounting for 98.51% and 99.42% of the total communities in DM and IM, respectively (Figure 5C). Seven members of the genera *Rhodococcus*, *Pseudomonas*, *Bordetella*, *Aquamicrobium*, *Delftia*, *Stenotrophomonas*, and *Acinetobacter*, were the most abundant in both groups, with proportions of 94.91% and 96.47% in DM and IM groups, respectively (Figure 5D). Subsequent analyses were adjusted for the top 20 genera with total abundance according to the hierarchical clustering of the microbiome belonging to seven phyla, including four dominant phyla (Figure 5E). Among them, *Rhodococcus* was the dominant genus from the phylum *Actinobacteria*, and the remaining dominant genera were from the phylum *Proteobacteria*. In addition, *Firmicutes* and *Bacteroidetes* were the predominant phyla with no corresponding dominant taxonomic composition at the genus level. Hence, all the above-mentioned species belonged to the predominant phyla (or genera) in the comparison of microbial composition in both groups and could comprise the specific microbiomes of both groups with different protein phenotypes.

To further compare whether there were microbial differences between the two groups, we performed the Wilcoxon rank-sum test. Because phyla-level clusters were relatively similar, no difference was reported between the two groups in this regard. In the analysis of the gut microbiome genome, we compared the relative abundance of individual taxa of the top seven dominant genera between the two groups and found three dominant genera (i.e., *Rhodococcus*, *Delftia*, and *Acinetobacter*) enriched by the Wilcoxon signed-rank test ($p < 0.05$, $q > 0.05$; Table 3). Notably, Figure 5F shows that these dominant genera were not significantly different after adjusting for FDR, and only the genera *Lachnospirillum* and *Coprococcus* were significantly different in the comparison of all genera ($q < 0.01$ and $q < 0.001$, respectively). Correspondingly, the DM group showed markedly higher levels of *Lachnospirillum* spp. and *Coprococcus* spp. as key species than the IM group, suggesting that these genera are associated with an increased MLH1 deficiency at the tumor site. These results suggest that the presence of these key species in the tumor may contribute to MLH1 protein deletion. Therefore, we propose a hypothesis that the lower abundance of bacteria in the DM group plays a more important role than that in the IM group, which may affect the development of SCRC.

Prediction of tumor microbiome phenotypes and various potential functions

The bacterial phenotype for each group was assessed using BugBase (Figures 6A-I). The statistical calculation of phenotypic

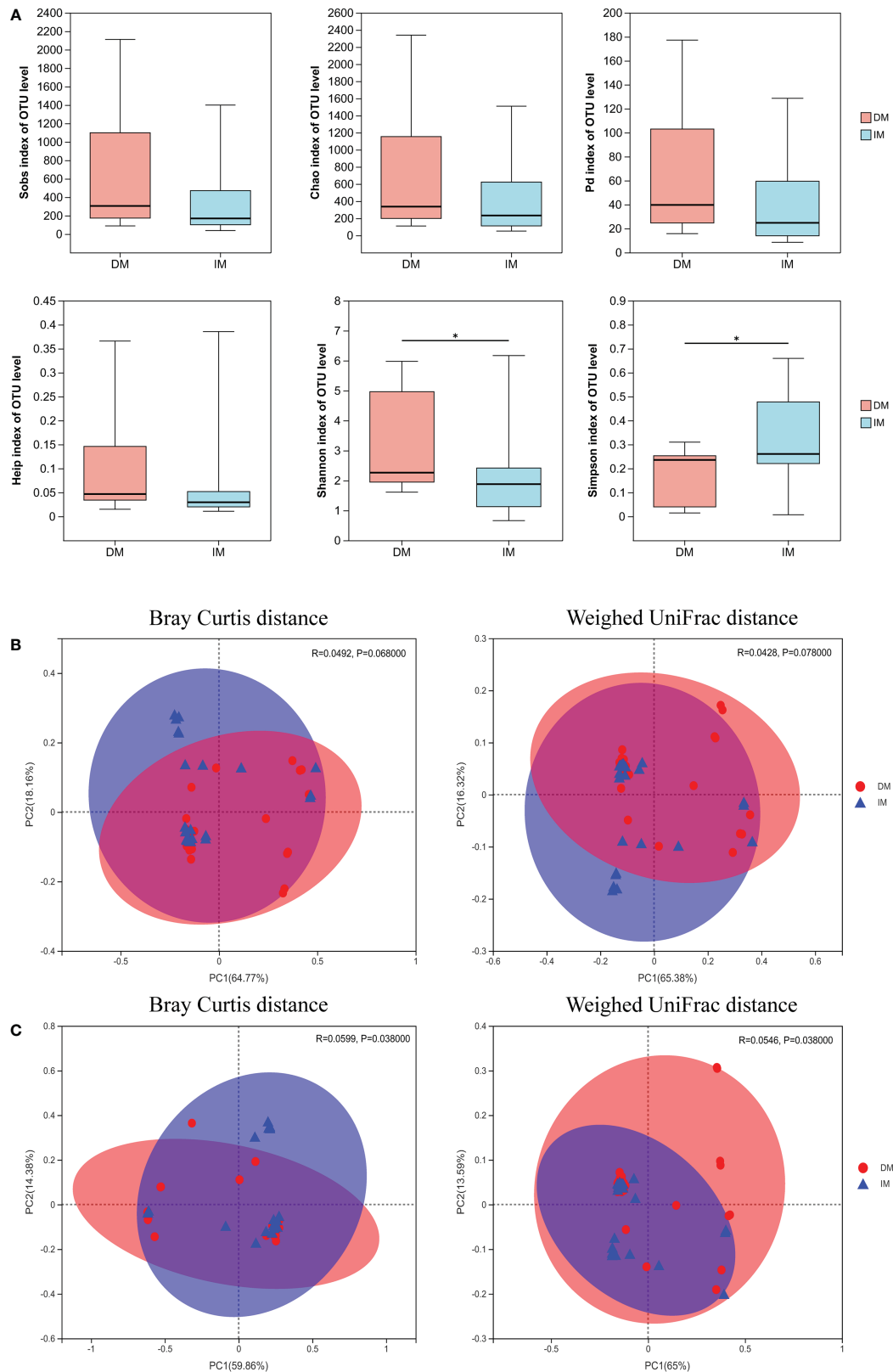


FIGURE 3

Tumor microbial alpha- and beta-diversity analysis based on influencing factors in both groups. **(A)** Microbial alpha diversity boxplot (observed OTUs, Sobs, Chao1, Pd, Heip, Shannon and Simpson indices) in SCRC patients. The DM (red) and IM (blue) groups (on the x-axis) and the different index types (on the y-axis). The PCOA plots of Bray Curtis and weighted UniFrac metric distances of beta diversity, **(B)** at the phylum level and **(C)** at the genus level. In the panel, each point represents a single DM (red circle) or IM (blue triangle), with ellipses (red ellipses= DM, blue ellipses = IM) representing the fitted mean and 95% confidence interval of each group, respectively. * $p < 0.05$.

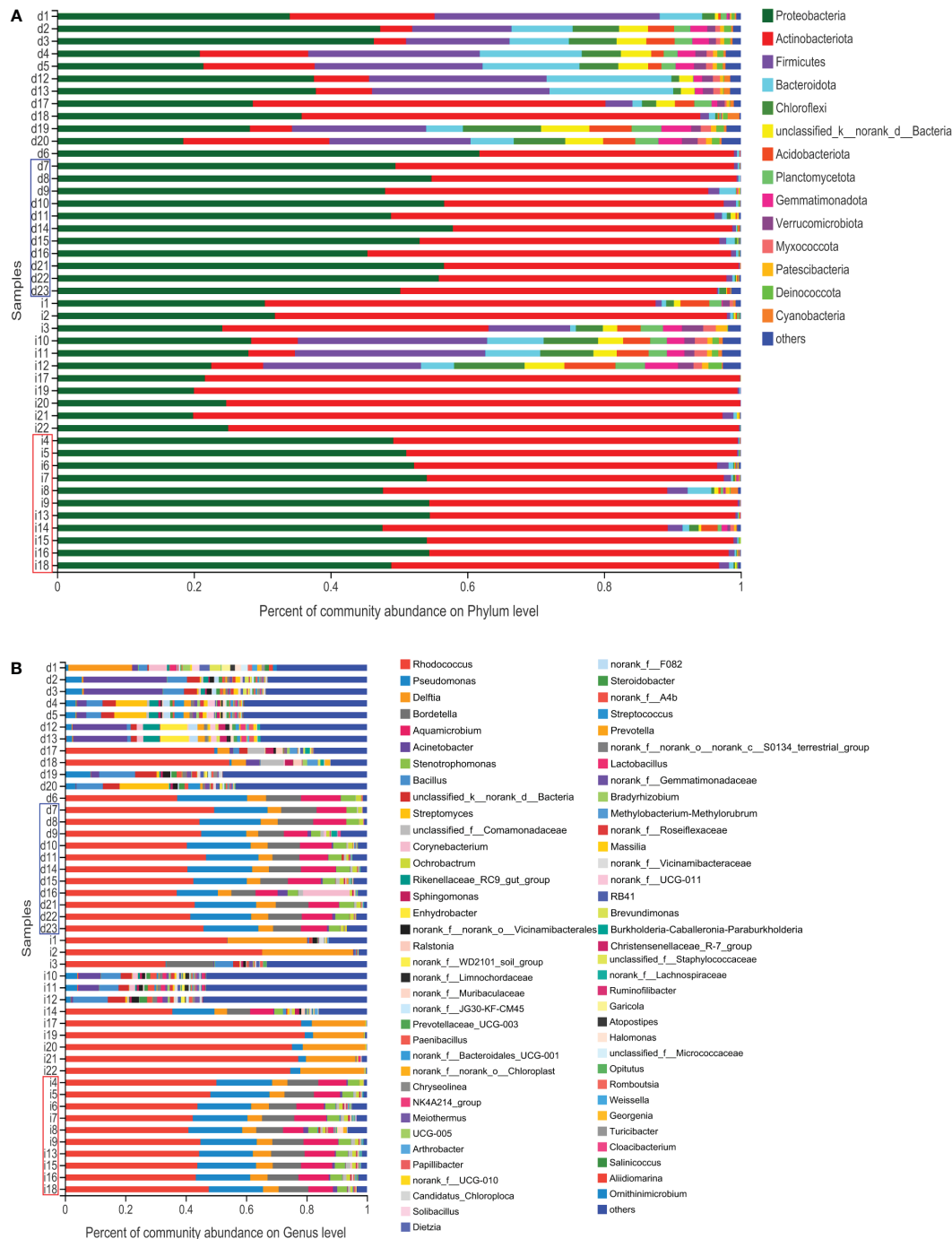


FIGURE 4 Histogram of the microbial community. The abundance of the community in each sample in the two groups. (A) At the phylum level. (B) At the genus level. The DM (red) and IM samples (blue) indicate similar abundances of microbial distribution.

abundance (Figure 7A) showed that after FDR correction, oxidative stress tolerance was significantly reduced in the DM group ($q < 0.05$). In addition, we used PICRUSt2 functional analysis to predict the metagenome and identified 18 level 2 KEGG pathways between the DM and IM groups ($q < 0.05$, Figure 7B). These functional genes were significantly different in both groups, belonging to five, level 1 KEGG pathways: metabolism, environmental information processing, genetic information processing, human disease, and biological system. When applied to the level 1 KEGG pathways, all SCRC samples had a coverage

of more than 90% and $q < 0.05$. The DM group exhibited enrichment in pathways related to signal transduction, translation, replication and repair, biosynthesis of other secondary metabolites, antimicrobial drug resistance, bacterial infectious disease, neurodegenerative disease, antineoplastic drug resistance, endocrine and metabolic disease, cardiovascular disease, immune system, development, and regeneration. In contrast, the IM group showed enrichment in amino acid metabolism, xenobiotic biodegradation and metabolism, lipid metabolism, metabolism of other amino acids, terpenoids, and

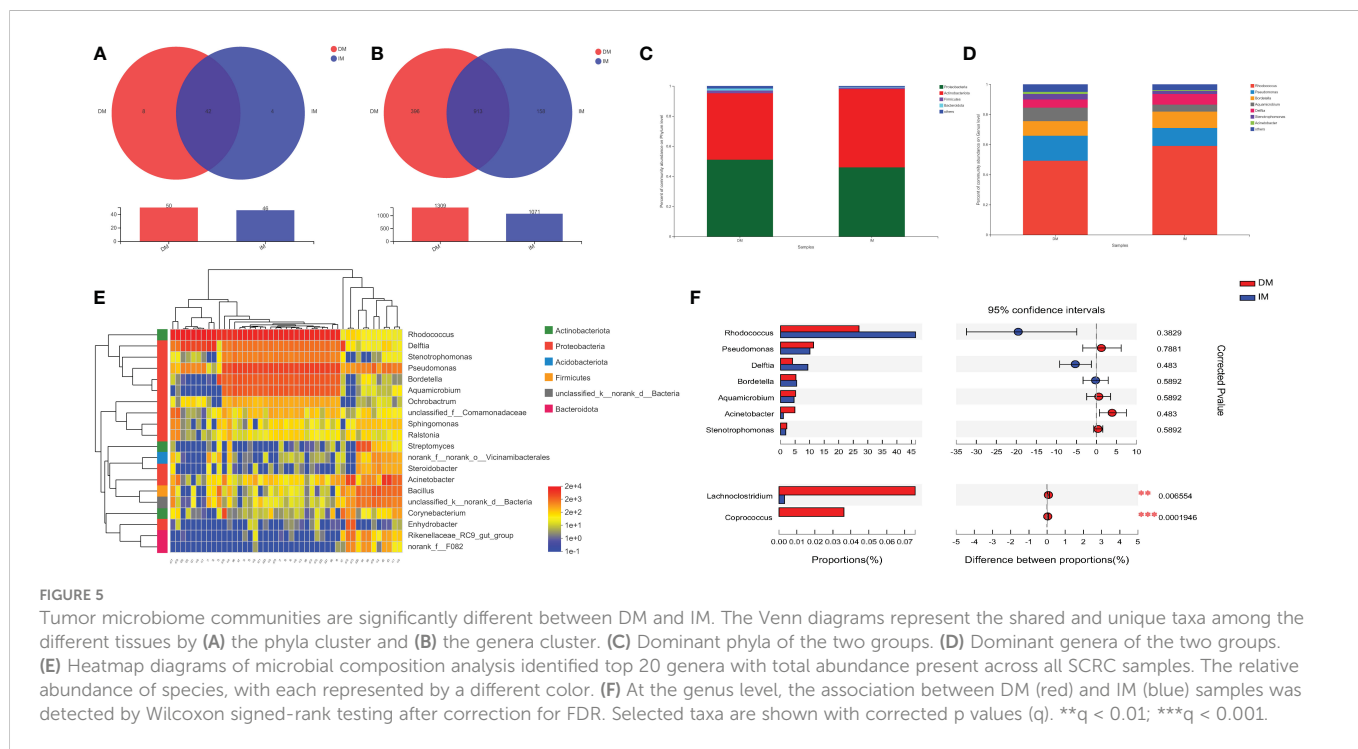


TABLE 3 The relative abundances at the genus level and statistical significance between the DM and IM groups by FDR.

Genera	Proportion of Sequences (Means ± SD)			
	DM	IM	p-value	corrected p-value (q-value)
<i>Rhodococcus</i>	26.96 ± 22.22	46.49 ± 23.80	0.025*	0.383
<i>Pseudomonas</i>	11.36 ± 8.36	10.14 ± 8.12	0.563	0.788
<i>Bordetella</i>	5.34 ± 5.24	5.57 ± 5.38	0.334	0.589
<i>Aquamicrobium</i>	5.24 ± 5.22	4.66 ± 4.82	0.320	0.589
<i>Delftia</i>	4.10 ± 4.41	9.40 ± 8.98	0.044*	0.483
<i>Stenotrophomonas</i>	2.20 ± 2.12	1.83 ± 1.84	0.291	0.589
<i>Acinetobacter</i>	4.95 ± 8.49	1.05 ± 2.01	0.042*	0.483

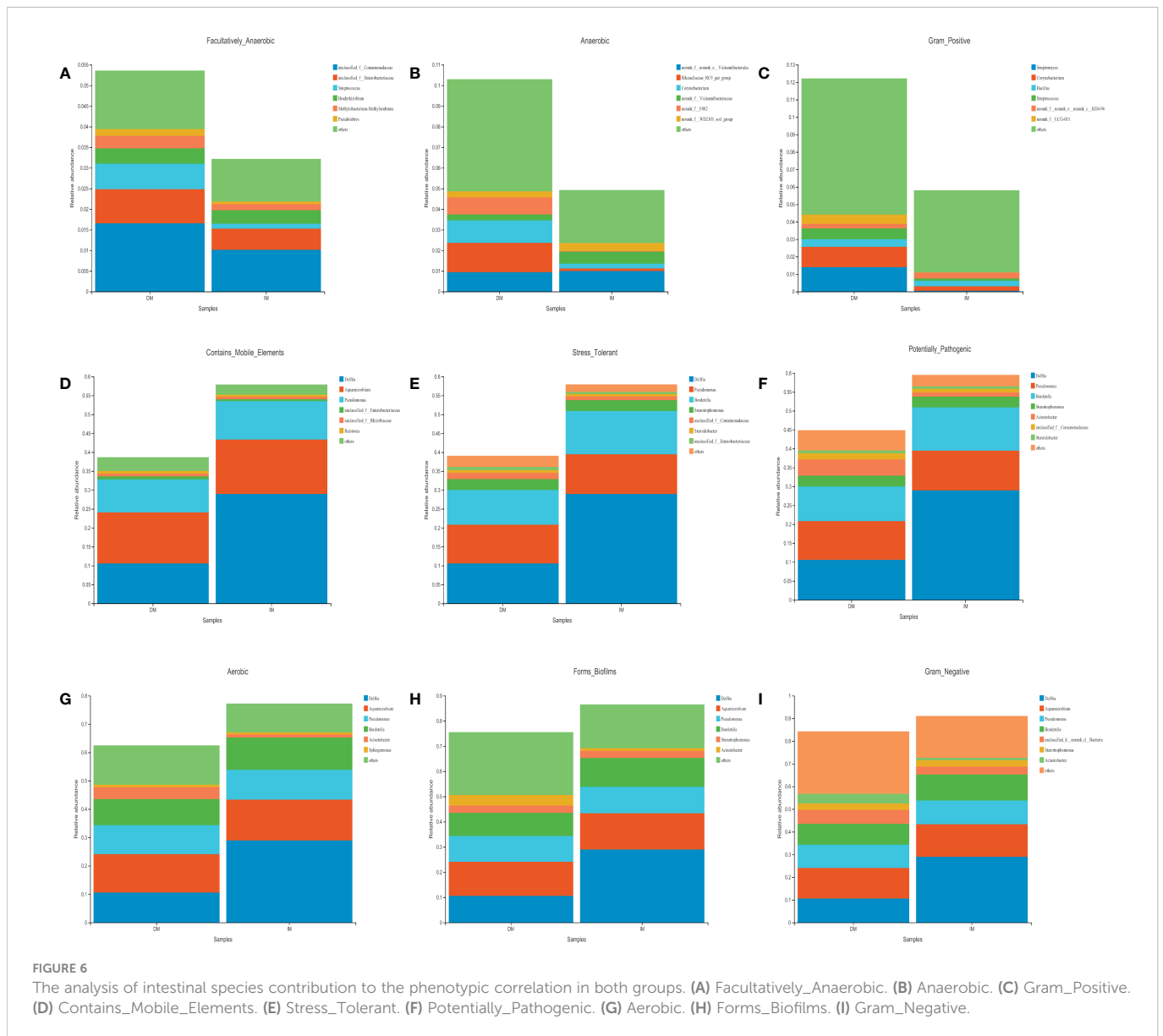
FDR, false discovery rate. **p* < 0.05.

polyketides, and substance dependence. In the functional prediction analysis, it was found that the homologous genes of some prokaryotic microbiomes may be annotated to the metabolic pathway of eukaryotic organisms (39–41). In this case, we assume that the genes of these bacteria may be related to the two-component regulatory system, which will eventually reveal certain pathways that are not related to bacterial metabolism. The PICRUSt2 taxonomic and functional relationships indicate that the microbial community determines the differential enrichment of functional pathways between DM and IM, which may influence the MLH1 gene expression.

Microbiome combination is significantly better than a single microbial factor in the diagnostic model

This study explored whether key species can be used as biomarkers for disease diagnosis. The RF first classifies and predicts data information using the ROC method, which can then be used to evaluate the applicability and accuracy of the model classification. This study used the RF-ROC algorithm to determine the optimal model with maximum efficacy. As shown in Figure 8A, when the lowest error rate of the ROC curve was used, 15 species were selected for importance ranking. On sorting the top 15 model variations (Figure 8B), the genera *Coprococcus* (phylum *Firmicutes*) was ranked first, followed by remaining microbes in the microbiome, including *Firmicutes* phyla (including *Lachnoclostridium* spp., *Blautia* spp., *unclassified_f_Ruminococcaceae* spp., and *Bacillus* spp.), *Proteobacteria* phyla (including *Parasutterella* spp., *Pseudomonas* spp., *Escherichia-Shigella* spp., *Stenotrophomonas* spp., *Sphingomonas* spp., *Paracoccus* spp. and *Ochrobactrum* spp.), *Desulfobacterota* phyla (*Bilophila* spp.), and *Bacteroidetes* phyla (*Prevotellaceae* UCG-001 spp. and *Prevotella* spp.).

Next, we attempted to assess the optimal biomarker candidates between single diagnostic prediction and multinomial combined model prediction (Figure 8C). We created Model 1, which illustrated the top 15 enriched genera, and had an AUC of 0.92 (95% CI: 0.83–1). We found that *Coprococcus* and *Lachnoclostridium*, as the top two genera, were significantly enriched in the DM group, with AUCs of 0.91 (95% CI: 0.83–0.99) and 0.85 (95% CI: 0.74–0.96), respectively. The results of the logistic regression model showed that Model 2 constituted the genera *Coprococcus* and *Lachnoclostridium*, with an AUC of 0.93 (95% CI: 0.84–1). These results demonstrated



that Model 2 was the simplified optimal diagnostic model and was constructed by combining the top two biomarkers.

Discussion

In this retrospective study, we sought to describe the gut microbiome of individuals with SCRC and compare the local microbiome of tumor tissues with different MLH1 expression levels. The comparison of SCRC patients with different protein expression levels revealed significant differences in the Shannon or Simpson indices, but not in the other indices. Importantly, this study showed two tumor microbial taxa as key genera, with *Lachnospirillum* and *Coprococcus* being significantly enriched in DM patients. We compared the differences in the gut microbiome and predicted the gene function between the two groups. These bacteria were useful as biomarkers for DM SCRC, and a new panel of potential bacterial markers was designed, which may improve the clinical diagnosis of SCRC.

Recently, studies have shown that DMMR protein is present in approximately 15% of patients with CRC, and the predominant defective protein is MLH1 (42), which is similar to our study findings. Consequently, this study mainly focused on the microbiomes of tumors with MLH1 molecular characteristics assessed using 16S rRNA gene sequencing.

In line with the results of a previous study (43), the dominant gut microbiome composition in this study included *Actinobacteria*, *Proteobacteria*, *Firmicutes*, and *Bacteroidetes*. Some scholars have found that the degradation of dietary fiber by some species of the phylum Firmicutes could produce short-chain fatty acids (SCFAs), and butyric acid was one of the main components of SCFAs in animal assays (44, 45). Zhao et al. reported that the genus *Acinetobacter* (*Proteobacteria*) was positively correlated with butyrate for food microbes (46) and may be involved in reducing the risk of SCRC. Yang et al. performed a metagenomic analysis of the gut microbiome to assess the relationship between dietary habits and CRC, showing that grain-based diets reduced the risk of CRC and finding a significant reduction in the abundance of *Acinetobacter* as a

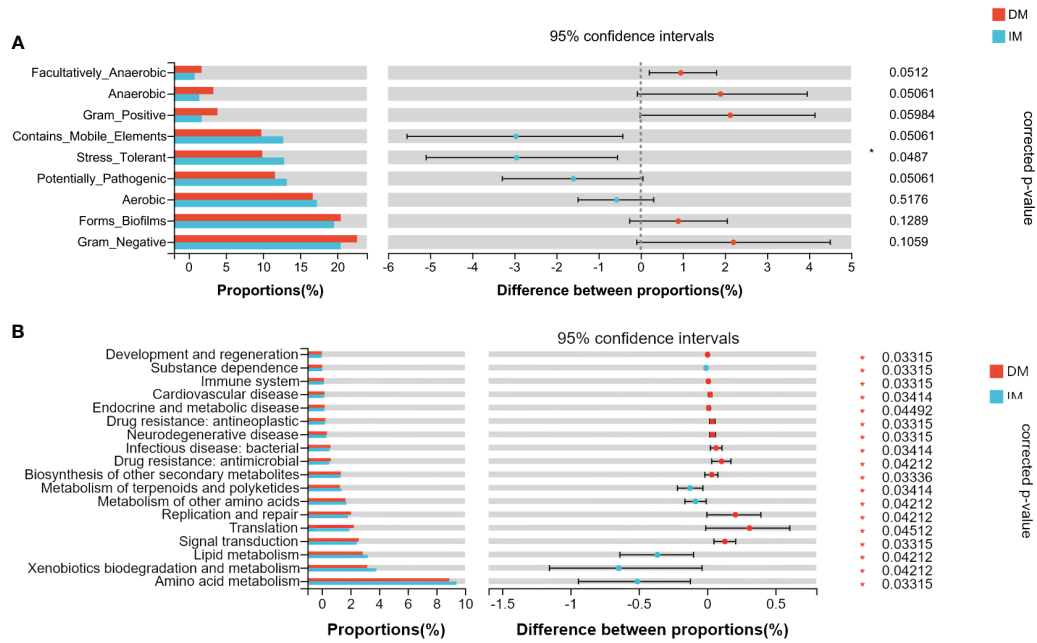


FIGURE 7 The difference of gut microbial phenotype and KEGG functional pathway between the two groups of SCRC patients. **(A)** The results of microbiome phenotype prediction by BugBase. **(B)** The significantly different KEGG pathways between the two groups by PICRUSt2. * $q < 0.05$.

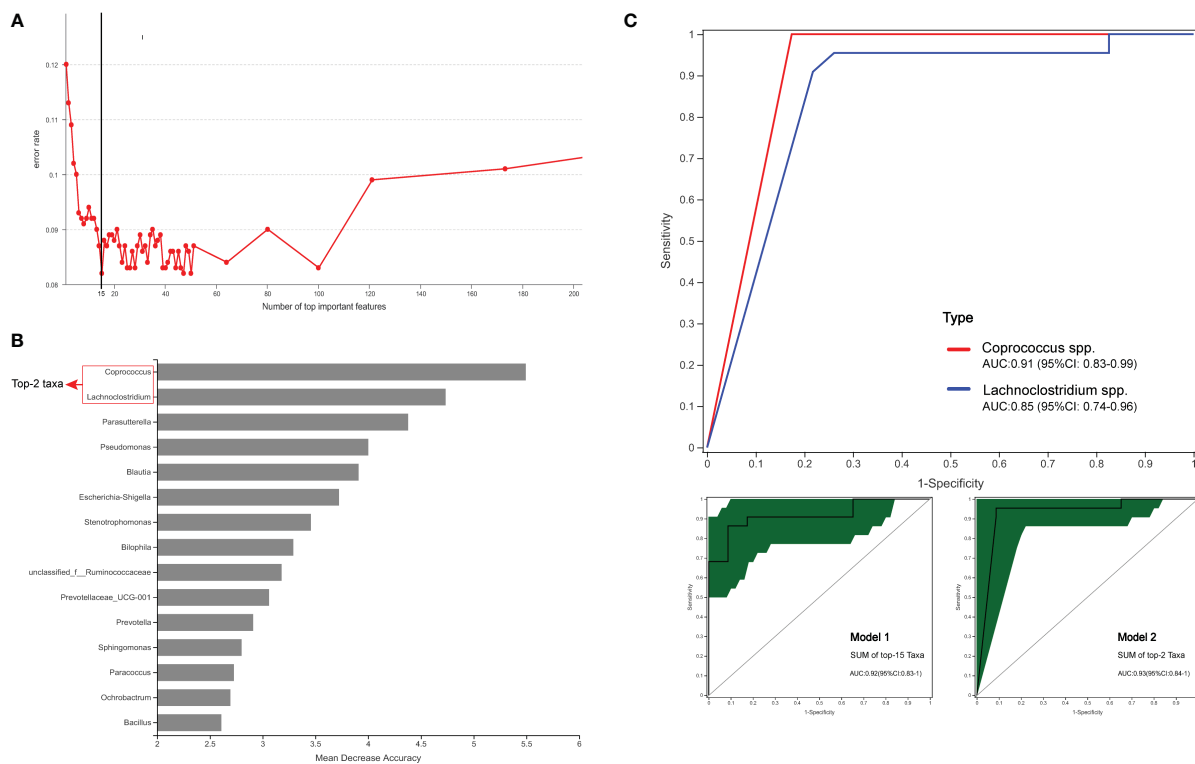


FIGURE 8 Comparison of bacterial (genus) candidate markers for the diagnostic prediction of DM SCRC. **(A)** The method of 10-fold cross-validation was used to determine the model with the lowest error rate of the ROC curve. **(B)** At the genus level, the relative abundance of the top15 hits, including *Lachnospiridium* and *Coprococcus*. **(C)** ROC analysis of taxa relative abundance as predictive of DM status. The above figure shows the diagnostic performances of the bacteria (genus) *Lachnospiridium* and *Coprococcus*. The figure below shows that Model 1 consists of the top 15 hits (left panel) and Model 2 consists of the top 2 hits (right panel). Different colors represent different groups of species.

candidate biomarker (47). Taken together, these findings demonstrate a complex system of associations between diet, the gut microbiome, and SCRC incidence, highlighting the role of the gut microbiome as a mediator.

Although gut microbial diversity is associated with the risk of SCRC, the key microbial roles that affected defective MLH1 in these tumors are not entirely clear. In this comparative study of the dominant genera, we found that the genera *Lachnospirillum* and *Coproccoccus* were positively correlated with defective MLH1 status. These microbes may play a role in inhibiting the MMR gene inactivation via the MLH1 expression pathway. Several studies have implicated that the genus-level microbiomes of *Lachnospirillum* and *Coproccoccus* are significantly different between patients with CRC and healthy cohorts (48, 49). Furthermore, Yuan et al. found a close correlation between *Lachnospirillum* spp. and KRAS mutations using high-throughput sequencing of stool samples from patients with CRC and healthy controls (49). Previous studies have reported that some butyrate-producing bacteria of *Coproccoccus* spp. (*Firmicutes*) could stimulate aberrant hyperproliferation of colonic epithelial cells with MSH2 or MLH1 deficiency and enhance β -catenin activity by increasing butyrate levels (50–52). Dronamraju et al. confirmed that butyrate has antitumor properties and can inhibit the proliferation of CRC cells with MLH1 gene deletion (53). As stated above, specific bacteria in the gut microbial community may induce MLH1 mutation to facilitate the occurrence of SCRC.

We speculate that the characteristics of tumor-associated microorganisms may trigger MLH1 protein deficiency in patients with SCRC and further promote the development of MSI. An understanding of relevant functional genes is required to explain this phenomenon and predict the role of these bacteria in SCRC for different MLH1 phenotypes. With the BugBase phenotypic prediction, we observed that, in patients with SCRC, oxidative stress tolerance was negatively correlated with DM and positively associated with IM. The oxidative stress response in the human body results in the enrichment of reactive oxygen species and destroys the biological molecular structure, which may eventually lead to cancer (54, 55). Qiao et al. indicated that the imbalance of flora in the diet might be caused by oxidative stress, and the changes in lactic acid bacteria, *Lactobacillus*, *E. coli*, and *Enterococci* in the intestine played a significant role (56). In contrast, our results indicate that, at the genus level, the microorganism most likely to be associated with oxidative stress is *Delftia*. In terms of functional prediction, the study showed that functional genes with the highest abundance in units with significant differences were those involved in amino acid metabolism. Previously, few researchers have reported that among the most enriched pathways, the pathway related to amino acid metabolism is closely associated with CRC (57). Based on the prediction analysis of the KEGG molecular function database, we retrieved partially functional genes in *Pseudomonas* spp. that were strongly linked to multiple level 2 KEGG pathways (such as replication and repair, translation, and signal transduction); this finding is supported by the results of Shen et al. (58). Regarding the mechanism of MLH1 protein deletion in patients with SCRC, the current relatively established theory suggests that functional pathways such as protein biosynthesis, KEGG metabolism pathways, the ammonia cycle, and galactose metabolism are the factors leading to

MLH1 promoter hypermethylation (59, 60). Hence, multiple functional predictions elucidated the relationship between microbiome sequencing abundance and tumor-related MLH1 protein phenotype and may be beneficial for further determination of metabolic biomarkers derived from patients with DM.

Testing analysis of fecal samples from 1012 Asian participants showed that the abundance of *Lachnospirillum* spp. was increased in intestinal adenoma and was a novel non-invasive genetic biomarker for the diagnosis and prediction of intestinal tumors (61). As a validation, Liang et al. identified that the combined predictive analysis of *Lachnospirillum* and two bacterial markers (*Fusobacterium nucleatum* and *Clostridium hathewayi*) could improve the sensitivity and specificity of recurrent adenoma diagnosis (62). However, to date, it is not clear whether microbial communities affect clinical practice in CRC tissues with an MLH1 protein phenotype. We found that patients who had SCRC with candidate gene biomarkers had a greater probability and higher biological susceptibility of developing MSI-H/DMMR. To this end, we further identified a new dataset of bacterial markers for patients with DM; the AUC value of the most highly predictive model was 0.92, obtained using RF-ROC analysis. By stepwise testing of the species model at the genus level, compared with the microbiome alone, the combination of *Lachnospirillum* and *Coproccoccus* demonstrated an increased AUC-ROC, indicating the impact of the bacterial combination on improving other candidate biomarkers for diagnostic prediction of DM SCRC. Surprisingly, we developed a simplified optimal model in the present study that required only two biomarkers to maintain an AUC of 0.93, demonstrating that these genera have a high discriminative power to evaluate the risk of patients developing DM. Additionally, we believe that if *Lachnospirillum* spp. and *Coproccoccus* spp. are the major variants, they can be used for clinical diagnosis prediction in patients with DM and have a high reference value. The results of such random forest analysis need to be interpreted with caution when small sample sizes are constructed for model predictive analysis.

Unfortunately, this retrospective study had some limitations. First, the low incidence of DM status in intestinal tumor tissue limited the number of patients with DM-SCRC who were enrolled; thus, the findings of this study need to be validated. Second, more clinical samples need to be included in future studies because differences in regions, eating habits, races, and ethnicities could potentially limit the generalizability of the study. Third, we analysed the microbiome within tumor tissues only in patients with SCRC without examining paired adjacent normal tissues or synchronously examining other microbiome samples (such as stool). Fourth, the causality between disease-related protein phenotypes and microbial diversity could not be well established. We continuously collected tissue samples from patients with DM, obtaining sequence data with as large a sample as possible to validate candidate bacterial markers in future studies of microbiomes based on disease diagnostic models.

Conclusion

We found that MLH1 gene silencing is the predominant role of DMMR in the tumor tissues of patients with SCRC. The relative

abundance of the gut microbiome differs significantly in the tumor microenvironment between DM and IM patients with SCRC. We also evaluated the bacterial phenotype and flora function relative to tumor microbiota in both groups. More importantly, we also designed a simplified optimal model (i.e., the combination of *Lachnospirillum* spp. and *Coproccoccus* spp.), which is a new non-invasive index to predict the clinical adjuvant efficacy and prognosis of SCRC by predicting the independent risk factors for MLH1 mutation phenotype and may provide a novel direction for future prognostic diagnosis and immunotherapy in patients with SCRC.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Harbin First Hospital (Approval No. Q2021-064). The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors made substantive intellectual contributions to the present study and approved the final manuscript. AZ designed the study and wrote the draft manuscript. YL, AZ, ZL and YHe coordinated patient recruitment and sample collection. AZ and LB conducted bioinformatic analysis. YW and YZ conducted 16S rRNA

gene sequencing. PJ and YHu contributed to the data interpretation and the revision of the manuscript. All authors participated in the coordination of the study and approved the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported, in part, by the Science and Technology Program Health Commission of Heilongjiang Province (2021, Nos. 20210101040241).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.1116780/full#supplementary-material>

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654
2. Sayed IM, Chakraborty A, Abd El-Hafeez AA, Sharma A, Sahan AZ, Huang WJM, et al. The DNA glycosylase NEIL2 suppresses fusobacterium-infection-induced inflammation and DNA damage in colonic epithelial cells. *Cells* (2020) 9. doi: 10.3390/cells9091980
3. Fidelle M, Yonekura S, Picard M, Cogdill A, Hollebecque A, Roberti MP, et al. Resolving the paradox of colon cancer through the integration of genetics, immunology, and the microbiota. *Front Immunol* (2020) 11:600886. doi: 10.3389/fimmu.2020.600886
4. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet (London England)* (2018) 391:1023–75. doi: 10.1016/s0140-6736(17)33326-3
5. Molina-Cerrillo J, San Román M, Pozas J, Alonso-Gordoa T, Pozas M, Conde E, et al. BRAF mutated colorectal cancer: New treatment approaches. *Cancers (Basel)* (2020) 12(6). doi: 10.3390/cancers12061571
6. Johnson CH, Dejea CM, Edler D, Hoang LT, Santidrian AF, Felding BH, et al. Metabolism links bacterial biofilms and colon carcinogenesis. *Cell Metab* (2015) 21:891–7. doi: 10.1016/j.cmet.2015.04.011
7. Botrel TEA, Clark LGO, Paladini L, Clark OAC. Efficacy and safety of bevacizumab plus chemotherapy compared to chemotherapy alone in previously untreated advanced or metastatic colorectal cancer: a systematic review and meta-analysis. *BMC Cancer* (2016) 16:677. doi: 10.1186/s12885-016-2734-y
8. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw JNCCN* (2021) 19:329–59. doi: 10.6004/jnccn.2021.0012
9. Yamamoto H, Imai K. Microsatellite instability: an update. *Arch Toxicol* (2015) 89:899–921. doi: 10.1007/s00204-015-1474-0
10. Bender E. Cancer immunotherapy. *Nature* (2017) 552:S61. doi: 10.1038/d41586-017-08699-z
11. Elkrief A, Derosa L, Zitvogel L, Kroemer G, Routy B. The intimate relationship between gut microbiota and cancer immunotherapy. *Gut Microbes* (2019) 10:424–8. doi: 10.1080/19490976.2018.1527167
12. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Sci (New York N.Y.)* (2018) 359:91–7. doi: 10.1126/science.aan3706
13. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* (2019) 25:1054–6. doi: 10.1038/s41591-019-0462-y
14. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* (2019) 16:361–75. doi: 10.1038/s41575-019-0126-x
15. Zhang X, Li J. Era of universal testing of microsatellite instability in colorectal cancer. *World J Gastrointest Oncol* (2013) 5:12–9. doi: 10.4251/wjgo.v5.i2.12
16. Nakayama Y, Iijima T, Inokuchi T, Kojika E, Takao M, Takao A, et al. Clinicopathological features of sporadic MSI colorectal cancer and lynch syndrome: a single-center retrospective cohort study. *Int J Clin Oncol* (2021) 26:1881–9. doi: 10.1007/s10147-021-01968-y

17. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* (2014) 12:661–72. doi: 10.1038/nrmicro3344
18. Gao Y, Bi D, Xie R, Li M, Guo J, Liu H, et al. Fusobacterium nucleatum enhances the efficacy of PD-L1 blockade in colorectal cancer. *Signal Transduct Target Ther* (2021) 6:398. doi: 10.1038/s41392-021-00795-x
19. Chen Y, Chen Y, Zhang J, Cao P, Su W, Deng Y, et al. Fusobacterium nucleatum promotes metastasis in colorectal cancer by activating autophagy signaling via the upregulation of CARD3 expression. *Theranostics* (2020) 10:323–39. doi: 10.7150/thno.38870
20. Tsoi H, Chu ESH, Zhang X, Sheng J, Nakatsu G, Ng SC, et al. Peptostreptococcus anaerobius induces intracellular cholesterol biosynthesis in colon cells to induce proliferation and causes dysplasia in mice. *Gastroenterology* (2017) 152:1419–1433.e1415. doi: 10.1053/j.gastro.2017.01.009
21. 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
22. Lee JA, Yoo SY, Oh HJ, Jeong S, Cho NY, Kang GH, et al. Differential immune microenvironmental features of microsatellite-unstable colorectal cancers according to fusobacterium nucleatum status. *Cancer Immunol Immunother* (2021) 70:47–59. doi: 10.1007/s00262-020-02657-x
23. Hamada T, Zhang X, Mima K, Bullman S, Sukawa Y, Nowak JA, et al. Fusobacterium nucleatum in colorectal cancer relates to immune response differentially by tumor microsatellite instability status. *Cancer Immunol Res* (2018) 6:1327–36. doi: 10.1158/2326-6066.Cir-18-0174
24. Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* (2002) 418:934. doi: 10.1038/418934a
25. Oliveira C, Pinto M, Duval A, Brennetot C, Domingo E, Espin E, et al. BRAF mutations characterize colon but not gastric cancer with mismatch repair deficiency. *Oncogene* (2003) 22:9192–6. doi: 10.1038/sj.onc.1207061
26. Domingo E, Espin E, Armengol M, Oliveira C, Pinto M, Duval A, et al. Activated BRAF targets proximal colon tumors with mismatch repair deficiency and MLH1 inactivation. *Genes Chromosomes Cancer* (2004) 39:138–42. doi: 10.1002/gcc.10310
27. Domingo E, Laiho P, Ollikainen M, Pinto M, Wang L, French AJ, et al. BRAF screening as a low-cost effective strategy for simplifying HNPCC genetic testing. *J Med Genet* (2004) 41:664–8. doi: 10.1136/jmg.2004.020651
28. Genetics Group of the Committee of Colorectal Cancer CA-cA. Consensus on clinical diagnosis, treatment and pedigree management of hereditary colorectal cancer in China. *Zhonghua zhong liu za zhi* (2018) 40(1):64–77. doi: 10.3760/cma.j.issn.0253-3766.2018.01.013
29. Huang L, Zhai E, Cai S, Lin Y, Liao J, Jin H, et al. Stress-inducible protein-1 promotes metastasis of gastric cancer via wnt/ β -catenin signaling pathway. *J Exp Clin Cancer Res CR* (2018) 37:6. doi: 10.1186/s13046-018-0676-8
30. C. Human Microbiome Project. A framework for human microbiome research. *Nature* (2012) 486:215–21. doi: 10.1038/nature11209
31. C. Human Microbiome Project. Structure, function and diversity of the healthy human microbiome. *Nature* (2012) 486:207–14. doi: 10.1038/nature11234
32. Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, et al. Tumor microbiome diversity and composition influence pancreatic cancer outcomes. *Cell* (2019) 178:795–806 e712. doi: 10.1016/j.cell.2019.07.008
33. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinf* 34 i884–i890 (2018). doi: 10.1093/bioinformatics/bty560
34. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* (2011) 27:2957–63. doi: 10.1093/bioinformatics/btr507
35. Yu X, Jiang W, Kosik RO, Song Y, Luo Q, Qiao T, et al. Gut microbiota changes and its potential relations with thyroid carcinoma. *J Adv Res* (2022) 35:61–70. doi: 10.1016/j.jare.2021.04.001
36. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* (2007) 73:5261–7. doi: 10.1128/aem.00062-07
37. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* (2013) 41:D590–596. doi: 10.1093/nar/gks1219
38. Subramanian S, Huq S, Yatsunenkov T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* (2014) 510:417–21. doi: 10.1038/nature13421
39. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, et al. Sequencing of aspergillus nidulans and comparative analysis with a. fumigatus A. oryzae *Nat* (2005) 438:1105–15. doi: 10.1038/nature04341
40. Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, et al. Complete genome sequence of the myxobacterium sorangium cellulosum. *Nat Biotechnol* (2007) 25:1281–9. doi: 10.1038/nbt1354
41. Mesa V, Monot M, Ferraris L, Popoff M, Mazuet C, Barbut F, et al. Core-, pan- and accessory genome analyses of clostridium neonatale: insights into genetic diversity. *Microbial Genomics* (2022) 8. doi: 10.1099/mgen.0.000813
42. Ghanipour L, Jirstrom K, Sundstrom M, Glimelius B, Birgisson H. Associations of defect mismatch repair genes with prognosis and heredity in sporadic colorectal cancer. *Eur J Surg Oncol (EJSO)* (2017) 43:311–21. doi: 10.1016/j.ejso.2016.10.013
43. Jin M, Shang F, Wu J, Fan Q, Chen C, Fan J, et al. Tumor-associated microbiota in proximal and distal colorectal cancer and their relationships with clinical outcomes. *Front Microbiol* (2021) 12:727937. doi: 10.3389/fmicb.2021.727937
44. Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. *Nat Med* (2016) 22:1079–89. doi: 10.1038/nm.4185
45. Shuwen H, Miao D, Quan Q, Wei W, Zhongshan Z, Chun Z, et al. Protective effect of the "food-microorganism-SCFAs" axis on colorectal cancer: from basic research to practical application. *J Cancer Res Clin Oncol* (2019) 145:2169–97. doi: 10.1007/s00432-019-02997-x
46. Zhao Z, Ning C, Chen L, Zhao Y, Yang G, Wang C, et al. Impacts of manufacture processes and geographical regions on the microbial profile of traditional Chinese cheeses. *Food Res Int (Ottawa Ont.)* (2021) 148:110600. doi: 10.1016/j.foodres.2021.110600
47. Yang J, McDowell A, Kim EK, Seo H, Lee WH, Moon CM, et al. Development of a colorectal cancer diagnostic model and dietary risk assessment through gut microbiome analysis. *Exp Mol Med* (2019) 51:1–15. doi: 10.1038/s12276-019-0313-4
48. Ai D, Pan H, Li X, Gao Y, Liu G, Xia LC, et al. Identifying gut microbiota associated with colorectal cancer using a zero-inflated lognormal model. *Front Microbiol* (2019) 10:826. doi: 10.3389/fmicb.2019.00826
49. Yuan D, Tao Y, Wang H, Wang J, Cao Y, Cao W, et al. A comprehensive analysis of the microbiota composition and host driver gene mutations in colorectal cancer. *Invest New Drugs* (2022) 40:884–94. doi: 10.1007/s10637-022-01263-1
50. Van den Abbeele P, Belzer C, Goossens M, Kleerebezem M, De Vos WM, Thas O, et al. Butyrate-producing clostridium cluster XIVa species specifically colonize mucins in an *in vitro* gut model. *ISME J* (2013) 7:949–61. doi: 10.1038/ismej.2012.158
51. Belcheva A, Irrazabal T, Robertson SJ, Streutker C, Maughan H, Rubino S, et al. Gut microbial metabolism drives transformation of MSH2-deficient colon epithelial cells. *Cell* (2014) 158:288–99. doi: 10.1016/j.cell.2014.04.051
52. Low L, Suleiman K, Shamdas M, Bassilious K, Poonit N, Rossiter AE, et al. Gut dysbiosis in ocular mucous membrane pemphigoid. *Front Cell Infect Microbiol* (2022) 12:780354. doi: 10.3389/fcimb.2022.780354
53. Dronamraju SS, Coxhead JM, Kelly SB, Mathers JC. Differential antineoplastic effects of butyrate in cells with and without a functioning DNA mismatch repair. *Nutr Cancer* (2010) 62:105–15. doi: 10.1080/01635580903191486
54. Vasquez EC, Pereira TMC, Peotta VA, Baldo MP, Campos-Toimil M. Probiotics as beneficial dietary supplements to prevent and treat cardiovascular diseases: Uncovering their impact on oxidative stress. *Oxid Med Cell Longevity* (2019) 2019:3086270. doi: 10.1155/2019/3086270
55. Feng T, Wang J. Oxidative stress tolerance and antioxidant capacity of lactic acid bacteria as probiotic: a systematic review. *Gut Microbes* (2020) 12:1801944. doi: 10.1080/19490976.2020.1801944
56. Qiao Y, Sun J, Ding Y, Le G, Shi Y. Alterations of the gut microbiota in high-fat diet mice is strongly linked to oxidative stress. *Appl Microbiol Biotechnol* (2013) 97:1689–97. doi: 10.1007/s00253-012-4323-6
57. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* (2019) 25:968–76. doi: 10.1038/s41591-019-0458-7
58. Shen H, Huang L, Dou H, Yang Y, Wu H. Effect of trilobatin from lithocarpus polystachyus rehd on gut microbiota of obese rats induced by a high-fat diet. *Nutrients* (2021) 13. doi: 10.3390/nu13030891
59. Yang Y, Misra BB, Liang L, Bi D, Weng W, Wu W, et al., et al. Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. *Theranostics* (2019) 9:4101–14. doi: 10.7150/thno.35186
60. Xia X, Wu WKK, Wong SH, Liu D, Kwong TNY, Nakatsu G, et al. Bacteria pathogens drive host colonic epithelial cell promoter hypermethylation of tumor suppressor genes in colorectal cancer. *Microbiome* (2020) 8:108. doi: 10.1186/s40168-020-00847-4
61. Liang JQ, Li T, Nakatsu G, Chen YX, Yau TO, Chu E, et al. A novel faecal lachnospirillum marker for the non-invasive diagnosis of colorectal adenoma and cancer. *Gut* (2020) 69:1248–57. doi: 10.1136/gutjnl-2019-318532
62. Liang JQ, Zeng Y, Kwok G, Cheung CP, Suen BY, Ching JYL, et al. Novel microbiome signatures for non-invasive diagnosis of adenoma recurrence after colonoscopic polypectomy. *Aliment Pharmacol Ther* (2022) 55:847–55. doi: 10.1111/apt.16799