TISSUE MICROBIOME IN COLORECTAL CANCER COMPARED DIFFERENT MICROSATELLITE CONDITIONS

Author	Student number	Email
Tong Yu	1210661	tong.yu@wur.nl

Supervisors Erasmus University Medical Centre

Nikolaos Strepis n.strepis@erasmusmc.nl

Andrew Stubbs a.stubbs@erasmusmc.nl

Willem de Koning w.dekoning.1@erasmusmc.nl

Supervisor Wageningen University & Research

Peter van Baarlen peter.vanbaarlen@wur.nl

Erasmus MC
University Medical Center Rotterdam

WAGENINGEN
UNIVERSITY & RESEARCH

ABSTRACT

Colorectal cancer (CRC) is a prevalent and deadly disease with a significant need for improved therapies. CRC can be classified into two main subtypes: **deficient DNA mismatch repair (MSI)** and **proficient DNA mismatch repair (MSS)**. In metastatic cases, MSI tumors are significantly less common than MSS tumors. In this study, MSI status was identified in 33 out of 162 patients (20%). These MSI patients often exhibit heterogeneous tumor characteristics, varying responses to immunotherapy, and distinct survival outcomes compared to patients with MSS tumors. Previous research has suggested that the tumor microbiome may differ between these two CRC subtypes. To our knowledge, this is the first study to investigate tumor microbiome profiling in MSI and MSS statuses using bulk RNA-sequencing. In this report, we explore the differences in microbiome diversity, composition, and structure between the two subtypes. The characterization of the microbial communities revealed significant differences in both alpha and beta diversity. Additionally, nine genera, including ***Fusobacterium*** and ***Leptotrichia***, were identified as potential biomarkers capable of modulating host gene expression. KEGG pathway analysis was conducted, revealing active functional pathways in MSI tumors, particularly in carbohydrate metabolism, amino acid metabolism, and genetic replication and repair. The study also investigated antimicrobial resistance (AMR) genes, with findings indicating the prevalence of genes conferring resistance to broad-spectrum antibiotics. These results demonstrate that CRC patients with different cancer subtypes harbor distinct microbiomes, which could have a profound impact on subtype detection and treatment outcomes in the future.

Table of Contents

INTRODUCTION

Colorectal cancer (CRC) is the third most common type of cancer with the second highest mortality. In 2020, two million new cases and one million deaths worldwide were reported (Morgan et al. 2023). CRC generally starts from the development of localized precancerous adenomatous polyps (adenomas) in the colon and gradually may progress into invasive and metastatic cancerous tumors over time (Zackular et al. 2014). Early detection of CRC cancer at the precancerous stage could lead the 5-year relative survival rate to around 90%. However, 60% of patients are identified at the metastatic phase with a 5-year survival rate of approximately 12-14% (Hashemi et al. 2023). This low survival rate can be caused by ineffective treatment regimens which include chemotherapy (e.g. oxaliplatin), targeted therapy (e.g. bevacizumab that targets vascular endothelial growth factor) and combined radiation therapy for late stages (Schmitt and Greten 2021). CRC is generally sporadic (60%) where the accumulation of point mutation on loci is heavily impacted by multiple risk factors including lifestyles, low physical activity and alcohol consumption (Malki et al. 2021). In comparison, familial CRC (20%) refers to patients with familial disposition however without a clear pattern of inherence; and inherited CRC (5%) only accounts for a small part of patients.

CRC and the two MS status

There are at least three molecular pathways known to lead to CRC, including chromosomal instability (CIN), CpG island methylator phenotype (CIMP) and microsatellite instability (MSI). Among these, the MSI pathway accounts for about 15% of patients. It is characterized by a high quantity of mutations caused by the inactivation of one of the four mismatch repair (MMR) genes: *MSH2*, *MLH1*, *MSH6* and *PMS2* in tumor cells or defects in the process of replication repair (Gupta, Sinha, and Paul 2018). Based on the MSI status, CRC can be classified into two subtypes: mismatch-repair-deficient (dMMR)/microsatellite instability-high (MSI) and mismatch-repair-proficient (pMMR)/microsatellite stability (MSS) (Figure 1).

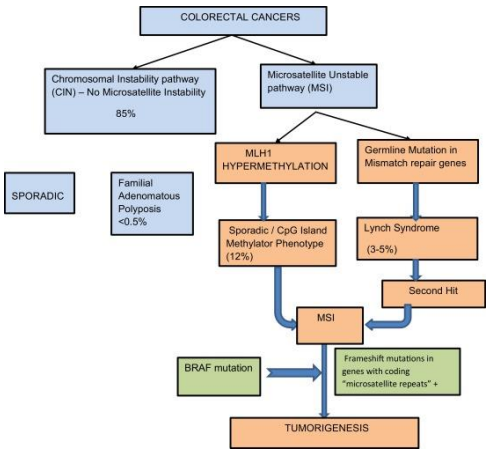


Figure 1: Schematic classification of CRCs and mechanism of tumor development via MSI pathway (Gupta et al. 2018)

In normal cells, the large number of microsatellite regions - usually over 100000 areas that

consist of short repeated sequences. These regions are particularly susceptible to DNA slippage during replication which can result in multiple insertions and deletions (indels). Proficient MMR genes can detect DNA mismatch errors and help form new, correct DNA strands. Damage to the MMR process can lead to the diffusion of mutation and form a 'hypermutator' phenotype, also known as an MSI tumor or dMMR tumor. In contrast, CRC tumors formed with proficient MMR genes can be referred to as microsatellite stable (MSS) tumors, indicating a proficient DNA repair mechanism that prevents the widespread accumulation of mutations (Li et al. 2020). The majority of MSI cases are caused by epigenetic changes, in which the promoter of *hMLH 1* becomes methylated, leading to the inactivation of gene expression. As tumors could potentially characterized by multiple pathways, the methylation can occur sporadically or be acquired through the CIMP pathway. In contrast, a small amount of MSI cases (3-5%) gained through germline MMR mutation, most commonly in *MLH1* or *MLH2*) in heredity form. These hereditary cases are associated with Lynch syndrome (LS).

Roles and Impact introduced by tumor microbiome in CRC

The presence of microbiome has been identified in a variety of cancer tissues in low-biomass for over a century. Recently, the concept of tumoral microbiome in tumor tissue is proposed (Wang, He, and Wang 2023). In CRC, tumorigenesis can be initiated by the invasion of commensal and pathogenic microorganisms after the disruption of the intestinal barrier (Schmitt and Greten 2021). It is hypothesized that this process begins with the local mucosal colonization of specific possible pathogens such as *F. nucleatum, Bacteroides fragilis* (ETBF), and *Enterobacteriaceae*. These potential pathogenic bacteria could induce changes in the tumor microenvironment, allowing opportunistic bacteria (e.g. *Bacillus*, *Bradyrhizobium*, and *Methylobacterium*) to colonize and facilitate cancer progression (Avril and DePaolo n.d.; Wang et al. 2020). Overtime, the changed microbiome can contribute to the reduction of epithelial cell E-cadherin, and increase epithelial permeability. The existence of the pathogens could further activate signal pathways, initiate inflammation, and ultimately lead to tumors (Dejea et al. 2014).

To date, most of the CRC studies regarding human microbiota have focused on analyzing fecal-luminal microbiota due to the non-invasive nature and convenience of sample collection. While these features make fecal samples suitable for earlier diagnosis, there is growing concern that they may not reflect the tumor microenvironment (Shen et al. 2020). In comparison, local tumor microbiota are significantly different and more stable, as they are less influenced by environmental factors. Thus, they are better suited for assessing the roles of the microbiome in the chronic inflammation process and more suitable for assessing physiopathology of CRC (Wirth et al. 2020).

Table 1. Association of the gut Microbiome With CRC, ↑ and ↓ indicates increasing or decreasing of relative abundance in corresponding studies.

Author, Year	Country	Sample size	Biospecimen type	Sequencing method	Main findings
(L. Zhao, Cho,	Spain, United	353	tissue	16S rRNA	*Clostridia and Bacteroidia*↓

and Nicolls 2021)	States, China, German, New Zealand				*Fusobacterium nucleatum and Parvimonas micra)*↑
(Xu et al. 2023)	TCGA samples from (Poore et al. 2020)	533	Tumor tissue sample of (Poore et al. 2020)	Whole genome sequencing Whole transcriptome sequencing	Proteobacteria, Firmicutes, and Bacteroidetes, was significantly associated with survival in CRC patients. Besides
(Feng et al. 2023)	China	10	Tissue samples	16S rRNA gene sequencing LC–MS analysis	*Caulobacterales*↑ *Brevundimonas*↑
(Sheng et al. 2020)	China	66	tumor and adjacent normal tissue	16S rRNA gene sequencing	*Fusobacterium*↑ *Faecalibacterium*↑ *Akkermansia*↑ *Ruminococcus2*↑ *Parabacteroides*↑ *Streptococcus*↑ *f_Ruminococcaceae*↑
(Mo et al. 2020)	China, Ireland, United States	2099	Normal tissue, CRC tissue, Adenoma tissue, Adjacent CRC tissue, Adjacent normal tissue	RNA-seq	1. Similar microbial ecology between adjacent tissue of CRC and adenoma tissue. 2. Cohort heterogeneity observed 3. Dysbiosis of the off-site tissue was distinctive and predictive
(Dejea et al. 2014)	Malaysia, United States	71	CRC tissue, Normal tissue, Polyp tissue	RNA-seq	Observed tumor biofilm aggregates with *Clostridium Ruminococcus, Butyrivibrio*
(Drewes et al. 2017)	China, Malay, India, African American, Caucasian	200	Tumor tissue, Paired normal samples, Healthy Biopsy	16S rRNA gene sequencing	CRC tissues are enriched for invasive biofilm with oral microbes including: *Bacteroides fragilis Fusobacterium nucleatum Parvimonas micra Peptostreptococcus stomatis*
(Roelands et al. 2023)	Netherland TCGA-COAD and AC-ICAM datasets	348	Snap-frozen tumor	RNA, whole-exome, deep T cell receptor and 16S bacterial rRNA gene sequencing	*Ruminococcus bromii*

The composition and abundance of microbiome in tumor tissues differ significantly from those in healthy tissues. Compared to healthy individuals, patients with CRC, as well as those with

polyps, showed an increased abundance of *Bacteroides*, *Roseburia*, *Ruminococcus*, *Oscillibacter*, *Porphyromonas*, *Peptostreptococcus*, *Parvimonas* and *Fusobacterium.* Conversely, microbes such as *Akkermansia*, *Parabacteroide*, *Lactobacillus, Prevotella* were observed at lower levels in CRC patients (Sheng et al. 2020). Due to these distinctive microbial patterns, these CRC-associated species could potentially serve as prognostic tools and be applied in precision oncology.

Previous studies have shown that a high relative abundance of genus *Fusobacterium* or a high proportion of reads mapped to *Bacteriodete fragilis* are correlated with an increased risk of mortality. In contrast, a high relative abundance of *Faecalibacterium* appears to be protective (Debelius et al. 2023). Additionally, *Parvimonas micra* which is enriched in CRC patients, has been shown to enhance the oncogenic Wnt signaling pathway, leading to poorer survival outcomes (Zhao et al. 2022). Conversely, a high abundance of *Lachnospiraceae* has been demonstrated to increase immunoscore and is associated with a better five-year survival rate (Hexun et al. 2023).

Manipulating the tumor microbiome could also have significant clinical implications for CRC patients. Various microbes, including well-known pathogens and probiotics such as *Akkermansia*, *Clostridium* and *Flavonifractor* are associated with immune checkpoint blockade therapy (ICB) through the mediation of tumor-infiltrating immune cells. For instance, the transplantation of *Akkermansia muniphila* has been shown to restore sensitivity to immunotherapy (Liu et al. 2023). Recently, fecal microbiota transplantation (FMT) was applied in a clinical trial. FMT, defined as the transplantation of gut microbiota from healthy donors to patients, aims to restore local microbial diversity. When combined with treatments like tislelizumab and fruquintinib, FMT has shown promising antitumor activity with a manageable toxicity profile. Another benefit of FMT is its potential to inhibit colonization of antibiotic resistant bacteria (ARB), thereby enhancing treatment efficiency (Zhao et al. 2023). Microbial changes could also impact the effectiveness of radiotherapy. For example, certain strains like *C.glabrata* and *F.canifelinum* have been linked to worse outcomes in rectal cancer. In contrast, the presence of Lactobacillus or *Streptococcus* genera might predict a better response to radiotherapy (Benej et al. 2024).

Microbiome in CRC patients between MSI and MSS status

Although the detailed mechanisms of pathogenesis in MSI are not fully elucidated, MSI status is considered a strong predictor of microbial community variance, significantly influencing cancer development (Hale et al. 2018). MSI status has been closely linked to CRC-associated microbes, including *F. nucleatum*, *F. periodonticum* and *B. fragilis* (Yin et al. 2022). Chronic inflammation caused by *F. nucleatum* could also increase the production of reactivate oxygen species (ROS) therefore potentially leading to MSI CRC development (Hale et al. 2018). In a mouse model where *Msh2* and *Tgfbr2* genes were inactivated (genes disrupted over 70% MSI cases), the development of CRC could be strongly modulated by pathogen species like *Desulfovibrio, Bacteroides, Parabacteroides* (Tosti et al. 2022). These pathogens, especially those producing genotoxins, have been experimentally confirmed to interfere with MMR, leading to MSI (Gagnière et al. 2017).

Different microbiomes could induce different metabolic patterns, which in turn impact immune

response. In MSI CRC, enrichment of retinoic acid (RA) can potentially promote CD8+ T cell response, which may improve the effectiveness of ICI therapy. In comparison, MSS CRC is characterized by a high level of lactic acid produced by the microbiome, which has been shown to exert an immunosuppressive effect by inducing anti-inflammatory gene transcription, thus reducing the impact of immunotherapy (Li et al. 2023). However, it has been suggested that PD-L1 antibody immunotherapy could be improved by the enrichment of several species including *Akkermansia* and *Provotella* through metabolism of glycolipid (Xu et al. 2020).

Roles of Antibiotics in different MS status

The usage of antibiotics in CRC has become a controversial problem, heavily impacted by the microbiome. Antibiotics have been identified as a dose-dependent risk factor in CRC development (Dik et al. 2016). Individuals with the highest level of antibiotics exposure, especially anti-anaerobic antibiotics were 10% more likely to develop colorectal neoplasia and CRC (Aneke-Nash et al. 2021). In CRC cancer, antibiotics can alter the composition of the tumor microbiome, disrupt critical host immune responses and affect the microbiome's functions. These impacts could be strong and persistent, even influencing distal tumor locations (Simin et al. 2020).

For MSS CRC, the use of antibiotics such as vancomycin could impact the antitumor effect of PD-L1 antibody therapy (Xu et al. 2020). In contrast, the impact of antibiotics on MSI CRC appears to be less significant (Serpas Higbie et al. 2022). Additionally, the treatment of broad-spectrum antibiotics has been shown to reduce the growth of Fusobacteriales-positive tumors in vivo, although the efficacy of antibiotics can be limited by the penetration of microbes into deep tumor layers (Salvucci et al. 2022). It is important to note that the effect of antibiotic usage could be impacted by multiple factors, including different doses, treatment duration, tumor locations and the number of cancerous cells, leading to significant variability in results (Bullman et al. 2017).

Another consequence of antibiotic usage is the increase in the richness of antibiotic-resistant species and the abundance of antimicrobial-resistant genes (AMR genes) in the microbiome. In the study of (Liu et al. 2021), AMR genes such as *adeF, TolC*, *E. coli soxS* were found in 13 CRC-associated species including *E. coli*, *A. onderdonkii*, *B. fragilis*, *A. muciniphila*, and *M. torques*. These genes confer multidrug resistance against multiple types of antibiotics, including those in the Reserve Class, which represent the last resort options in treatment. *Enterbacteriaceae* family and several members from ESKAPE group were found to possess a significant proportion of AMR genes. Environmental changes, such as exposure to organic acid, osmotic and oxidative pressures and inflammation of CRC may promote horizontal gene transfer among these bacteria, selectively upregulate certain AMR genes like *catA* in *A. baumannii* and *phoP* in *K. pneumoniae* (Lamaudière et al. 2023).

Overall, the impact of antibiotics on CRC remains a complex issue. There is substantial heterogeneity in the magnitude and pattern of antibiotic effects, highlighting the need for more clinical studies to better understand these interactions.

Characterization of tumor microbiome with next-generation sequencing and its challenges

With the recent advancements in human microbiome analysis technologies, next-generation sequencing (NGS) methods like RNA-seq have enabled the detailed characterization of microbial information in specific tissues. The most frequently used methods are 16s ribosomal RNA (rRNA) sequencing and shotgun sequencing. 16S rRNA sequencing allows for the relative quantitation of bacterial composition, primarily at the genus level, while shotgun sequencing analyzes all genomic content in the sample, providing insights into microbial composition and abundance with high genomic resolution (Saus et al. 2019).

However, genomic information may provide limited clinical potential since DNA sequences seldom transcript and translate directly to phenotype due to transcript modulation or post-transcriptional alteration. In this context, transcriptomics could better identify changes in microbial function, which can deepen our understanding of the interaction between humans and microbiomes. Despite these advancements, obtaining high-quality characterizations of cancer microbiomes remains challenging. One major issue is contamination, samples are exposed to contaminants at various steps, from biopsy acquisition, and processing to sequencing. This contamination complicates the task of distinguishing genuine microbial signals from background noise. Additionally, host tissues are generally microbe-poor habitats, containing only 2% of microbial reads (Gihawi et al. 2023). This low yield is partly because transcripts adapt and degrade rapidly upon collection, and much of the sequencing capacity is lost on host RNAs, which are irrelevant for microbial analysis (Ojala, Kankuri, and Kankainen 2023).

RESEARCH AIM

In previous research on cancer microbiomes, the majority of studies have relied solely on 16S rRNA sequencing, with transcriptomics being used in only a small proportion of cases. Moreover, not all of these studies focus on analyzing the microbiome community or deriving active functions from data. Due to the high sensitivity of the RNA-sequencing, active species can be recognized and their relationship with cancer cold be elucidated more clearly.

In this project, we utilized bulk RNA sequencing data from tissue biopsies of 162 patients with CRC to investigate the differences in microbial communities between the two microsatellite statuses. Microbial diversity was studied and the co-abundance networks were inferred. Microbial signatures were recognized during the process. Additionally, we looked into the types and quantities of AMR genes present in the tumor microbiome samples. To our knowledge, this research represents the first attempt to use transcriptomics data to explore the differences in microbial communities between MSI and MSS statuses. Through this project, we aim to enhance the understanding of tumor subtypes, microbial community interactions, functional consequences and ecology of tumor microbiome in CRC. Furthermore, our findings on antibiotic resistance may provide valuable insights into the appropriate use of antibiotics in future treatment.

MATERIAL AND METHODS

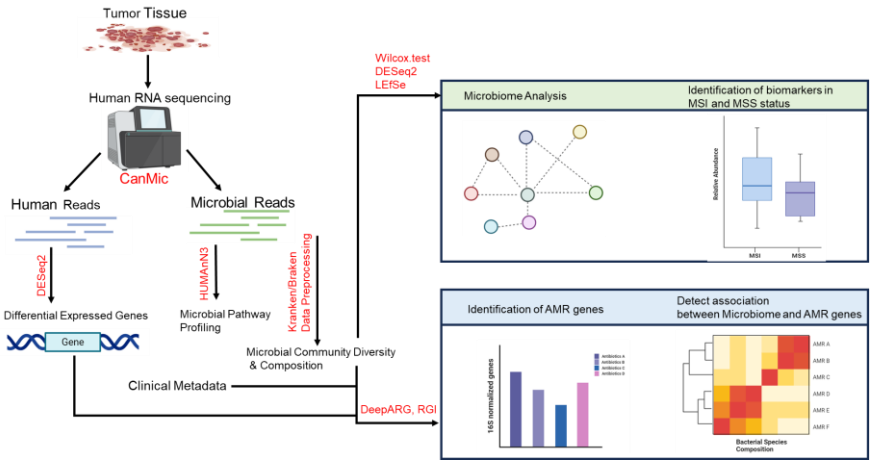


Figure 2. Scheme of the general workflow in this report

Sample Acquisition and Decontamination

In this study, RNA sequencing data from the tissues of a total of 162 patients with CRC were collected from research (Joanito et al. 2022). In the research, tissues were cut into 5μm tissue sections with standard microtomy techniques, and the RNA-sequencing libraries were prepared using the KAPA Standard RNA-seq kit with RiboErase and sequenced to a target depth of 200 million reads on the Illumina HiSeq platform. As the paper introduced, MS status was determined by immunohistochemistry for *MLH1*, *MSH2*, *MSH6* and *PMS2*.

Sequence decontamination

The obtained data was then treated with the workflow mentioned in https://github.com/BirgitRijvers/EMC-CanMic. In the workflow, the reads were mapped to the human genome GRCh38 with BWA, so that host genes could be removed. The remaining unmapped reads (**microbial reads**) were then used as input for Kraken2 for taxonomy classification. The Kraken system uses k-mer methods to achieve an extremely high speed and accuracy. However, complication arises when many reads align equally well to more than one species. Therefore, Bracken which uses the result of Kraken was further used to improve the classification accuracy. Also, for metatranscriptome analysis of human tissues with very low microbial contents, an optimized setting of combination Kraken2/Braken could minimize false positives and improve precision. The resulting microbial abundance data was further filtered for decontamination. In the filtration, bacteria existed in more than half of the samples are remained. External contaminants (n=94) were eliminated by using a blacklist of microorganisms commonly identified from reagents in the sequencing kit (Salter et al. 2014). Also, unclassified species at the genus level were discarded to ensure data quality. At last, only 449 species from 225 genera were maintained for further analysis.

MSI and MSS microbiota analysis

R package mia was used for analysis and visualization of relative abundance, alpha diversity and beta diversity. PERMANOVA was performed on the bray-curtis matrix to assess the effects of MMR status on variance between microbial communities. We performed a co-abundance network analysis on the cohort, which can more accurately infer the relationship between components. SparCC method (Friedman and Alm 2012) was used to calculate the co-occurrence between the 225 genera in tumor samples. The method can robustly estimate correlation values of compositional data with bootstrapping. Other than the aforementioned filtration, samples with fewer than 1000 reads were removed. A total of 50 inference and 10 exclusion iterations were used to estimate the median correlation of each pairwise. The statistical significance of the correlations was calculated using a bootstrapping procedure to generate 100 simulated data. For each component pair, pseudo-*P* values were assigned and adjusted with the Benjamini-Hochberg method. Also, only an absolute correlation value above 0.3 and an adjusted *p*-value less than 0.05 in the correlation matrix will be used for network construction. All co-occurrence networks were made and visualized using the R package 'NetCoMi (v1.1.0)' (Peschel et al. 2021).

Identification of MSI and MSS-associated microbiota

As various tools could lead to drastically different results depending on data pre-processing, it is suggested to use a consensus approach based on multiple different methods to ensure a robust biological interpretation (Nearing et al. 2022). Microbes associated with MSI or MSS samples were further calculated in R. In the analysis, differential abundant taxa were recognized by DESeq2 (Love, Huber, and Anders 2014), Wilcoxon's rank-sums ($p<0.05$) test and LEfSe ($p<0.05$, LDA>2) (Segata et al. 2011) with package microeco (v1.7.1). Key taxa that agree with at least two methods are used to estimate their impacts on host gene expression. The human expression data (in transcripts per million) were log10 normalized to ensure a normal distribution. Wilcoxon's rank-sums ($p<0.05$, *p*-value adjusted by FDR correction) was used for differentially expressed genes (DEGs) selection. As a result, a total of 265 genes were recognized as DEGs. Also, Gene Oncology (GO) enrichment test was applied to the genes to investigate related biological processes. The analysis was processed with the clusterProfiler v4.2.2 (Yu et al. 2012) package in R. Spearman correlations between key taxa and DEGs were calculated. Before the analysis, the abundance of microbiome was normalized using pseudocounts and a centered log-ratio (CLR) transform.

Functional profiling

Reads unmapped to the human genome (microbial reads) were further processed in FastQC (version 0.11.9) (Lo and Chain 2014), Trimmomatic (version 0.33) (Bolger, Lohse, and Usadel 2014), SortMeRNA (Kopylova, Noé, and Touzet 2012), to do quality check, quality filtering and rRNA sequences removal, respectively. The reads were trimmed based on a sliding window trimming approach. In the method, reads will be cut when the average base Phred quality within

a 4-base sliding window is less than 15. The reads were discarded when the length was less than 100 and more than half of the read length was trimmed. Further, SortMeRNA was used to remove rRNA reads. After trimming, a quality check was applied with FastQC. HUMAnN3 (HMP Unified Metabolic Analysis Network version 3.6) (Beghini et al. 2021) was then applied for functional analysis to determine the metabolic potential of the members of a microbial community. The software could estimate the relative abundances of functional features in the metagenome (transcriptome) by aligning prescreened reads against native UniRef90 protein databases. Gene family abundance is reported in reads per kilobase (RPK) units and then transformed to copies per million (CPM) units to account for sequencing depth. The databases used for the analysis were the full ChocoPhlAn pangenome database (v31) and UniRef90 EC filtered database (v201901b). Obtained gene family abundance was then annotated with KEGG Orthology (KO) terms in samples. Using 'humann_join_table', annotated files were joined in a single table. Differential abundant KO terms were then selected based on the Wilcoxon rank sum test with *p*-value <0.05 and log2FC>1.5 between MSI and MSS status.

Analysis of Antimicrobial Resistant Genes

In the research, the AMR genes were explored with the software RGI (6.0.3) (Alcock et al. 2023) and DeepARG (1.0.2) (Arango-Argoty et al. 2018). RGI is a software able to recognize antibiotic resistance genes from sequence data. In our research, high-quality reads after trimming were used as input of RGI, which was able to align the short DNA sequences using KWA against a comprehensive antibiotics resistance database (CARD) database. As a result, only genes predicted with over 50% percent of reference alleles covered by reads remained (percentage coverage). To further confirm the existence of the genes, DeepARG which is a deep-learning-based approach in AMR genes prediction was used. In DeepARG, a short-reads pipeline (DeepARG-SS model) was applied with the reads. So only reads agreed with both softwares were considered as existing AMR genes to reduce false positive rate. Finally, only 32 AMRs from 51 samples remained. These AMRs were then classified into two categories: high coverage (percentage coverage over 85%) and middle coverage (percentage coverage from 50% to 85%), and used for further analysis. Statistics analysis was performed with R.

RESULTS

Changes in Tissue microbiome between MSI and MSS status

After filtration of potential contamination, the rarefaction curve (Figure S1) showed that the sequencing depth approached saturation in most of the samples at the sequence depth of 65000. According to the classification result, we analyzed the species abundance at the phylum and genus levels. Most of the species in tissue samples in MSI and MSS patients belonged to Bacteroidota (61.2%), Bacillota (18.5%), Fusobacteriota (11.2%), accounting for about 91.02% of the total microbiome (Figure 3). Other main bacteria phylum includes Pseudomonadota (3.06%), Actinomycetota (2.16%), Campylobacterota (1.21%), Thermodesulfobacteriota (1.06%), Verrucomicrobiota (0.88%), Spirochaetota (0.43%), Synergistota (0.16%) in the

samples. There were significant differences found in the proportion of phyla between MSI and MSS, with Fusobacteriota accounting for 21.3% and 7.5%. (p=0.0247), Pseudomonadota accounting for 3.56% and 1.71% (p=0.0023), and Verrucomicrobiota accounting for 0.21% and 1.19% (p=5*10^{-5}), observed in the two groups respectively (FigureS2).

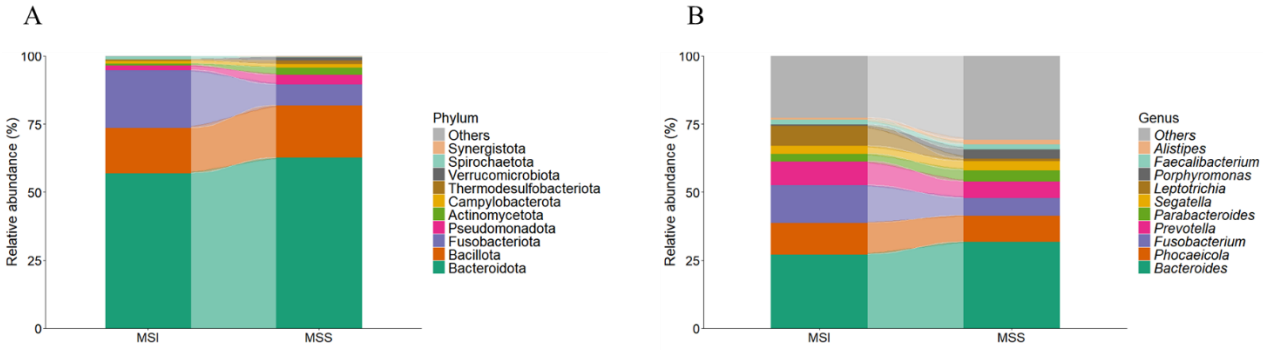


Figure3. Taxonomic characteristics of the microbiome among the groups.Structure and abundance of microbiome at the (A) phylum level and (B) Genus level between MSI and MSS

We next investigated variation in genus level between MSI and MSS status. At the genus level, the mean abundance of top ten species including *Bacteroides* (30.58%), *Phocaeicola* (10.14%), *Fusobacterium* (8.5%), *Prevotella* (6.77%), *Parabaceroides* (3.81%), *Segatella* (3.12%), *Leptotrichia* (2.77%), *Porphyromonas* (2.57%), *Faecalibacterium* (1.79%) and *Alistipes* (1.45%). In these genera, two species were found enriched which are *Leptotrichia* and *Alistipe*. *Leptotrichia* accounts for 7.45% in MSI and 1.04% in MSS (p=0.0049), *Alistipe* accounts for 0.72% in MSI and 1.72% in MSS (p=0.04) (Figure S3).

Tumor microbiome diversity analysis

Compared to the MSI, alpha diversity estimated by observed count and Shannon index were significantly higher in MSS (p=0.038 and p=0.023 respectively). There was no significant difference found in other indexes including Chao2, ACE and Simpson index. Interestingly, these indexes were indicated to be higher in MSI in the paper of (M. Jin et al. 2022). In another paper, the Simpson index is higher for patients with complete *MLH1*, which is an important gene for MSI cancer (Zhu et al. 2023). The different results from papers indicated a highly complex microbial condition that could be impacted by multiple factors including patient status, treatment before surgery and deficiency of genes.

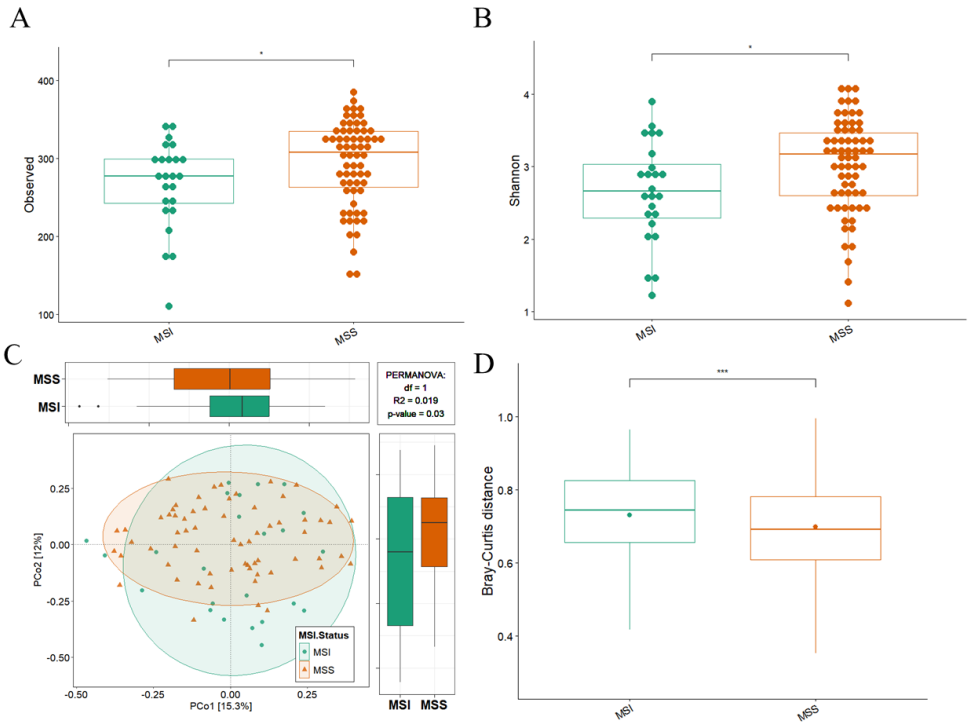


Figure4: Bacterial diversity of the tissue microbiome associated with MSI and MSS, p values are calculated by Wilcoxon rank sum p<0.05 (A). Tissue microbial diversity estimated by observed count (p=0.038) (B) Tissue microbial diversity estimated by Shannon index (p=0.023). (C) Beta diversity calculated by PCoA of bray-curtis distances and PERMANOVA (D) Differences in bray-Curtis distances between MSI and MSS group

Principal coordinates analysis (PcoA) demonstrating beta diversity calculated by bray-Curtis distance. As shown in Figure 4C, the status of MSI and MSS cannot be separated at the first two axes (p=0.54 for the first axis, p=0.11 for the second axis between groups). However, the results presented a significantly different distribution of centroid among groups using permutational multivariate analysis of variance (PERMANOVA) analysis (p=0.036). The result further agreed with the difference in bray-Curtis distance between the two groups based on the Wilcoxon rank sum test (p=4.629535e-13). The alpha diversity data suggested higher alpha diversity in MSS status. The difference in terms of PERMANOVA and distance indicated a strong correlation between microbial diversity correlated with MSI sand MSS CRC.

Analysis of microbial co-occurrence networks and topological properties between MSI status

To examine if the co-occurrence networks were different in MSI status and MSS status, we performed network analysis by SparCC. The topological properties of networks, nodes and co-occurrence patterns were calculated to describe the complex pattern of inter-relationships between genera (Figure 5). In both networks of the two conditions, 25 and 36 nodes with a high variance that satisfied the requirements (see Methods) were selected and each contains 60 and 145 edges respectively. The difference in node and edge number suggests tighter interaction among genera in MSS patients than in MSI patients. Modularity could quantify the connectivity

of networks and whether networks can be broken into smaller components. This may indicate the ecological processes governing community structure. MSI and MSS networks had modularity of 0.37 and 0.25 (Table S1). The higher modularity in MSI may suggest a more complex network however less stability. The fewer modules in MSS could indicate fewer constrained from perturbations and could easily get impacted by the environment.

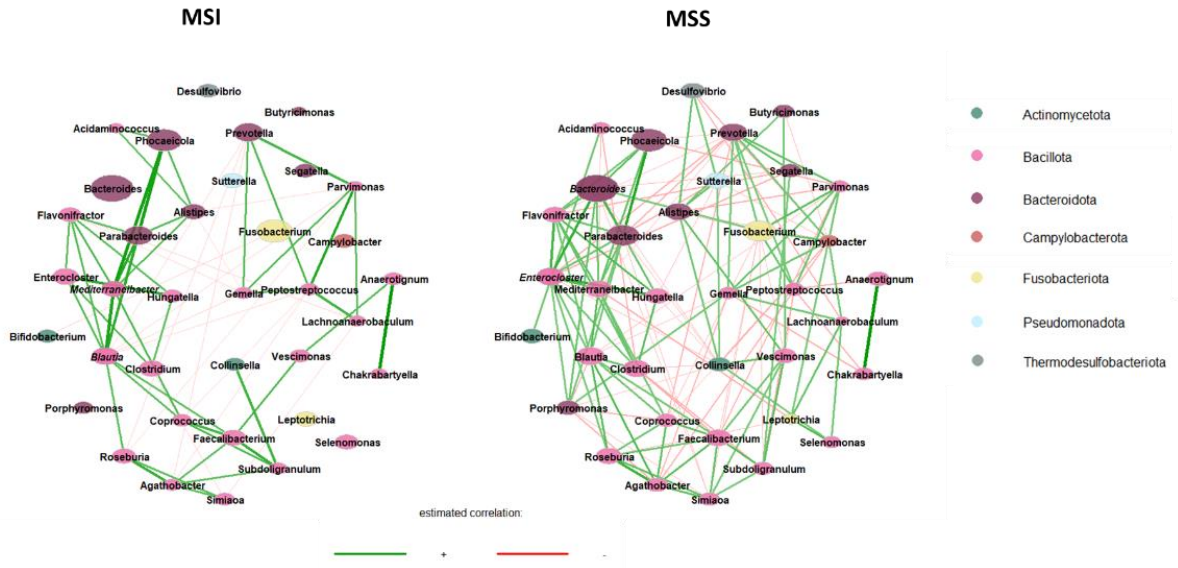


Figure 5. Co-occurrence network of bacterial genera detected in MSI (A) and MSS status (B). Each node corresponds to a genus, and node size reflects their transformed abundance values. Nodes are colored according to different phylum. Edges between nodes indicate either positive (green) or negative (red) correlations inferred from microbial abundance profiled by the SparCC analysis

Distinct taxonomic profiles were observed between the two statuses. Nodes in MSI networks mainly belonged to phylum Actimomycetota, Bacillota and Bacteroidota, while MSS status was composed of diverse phyla including Actinomycetota, Bacillota, Bacteroidota, Campylobacterota, Fusobacteriota, Psudomonadota and Thermpdesulfobacteriota. Nodes were analyzed with their eigenvector centrality for their potential roles in the co-occurrence network topologies. High eigenvector centrality nodes (nodes with centrality levels over 95% quantile, also called hubs) indicate their center position in the networks. Genera *Blautia* and *Mediterraneibacter* were recognized as hubs in MSI status. Genera *Enterocloster* and *Bacteroides* contributed as hubs in MSS status. As indicated in Figure 5, a positive correlation between genera was the main trend for both statuses (positive edge percentages were 73.3 and 58.6 respectively), which potentially results in the prevalence of mutualism interactions. The abundance of genera pairs between *Anaerotignum* and *Chakrabartyella, Phocaeicola* and *Parabacteroides*, *Enterocloster* and *Hungatella*, *Agathobacter* and *Roseburia* were strongly correlated with each other under both conditions in a positive manner. In particular, *Anaerotignum* and *Chakrabartyella* pair was suggested to have the highest positive association (correlation=0.91 and 0.84 respectively). For MSI status, a strong positive correlation pattern of abundance was observed between *Mediterraneibacter* with *Phocaeicola, Parabacteroides* and *Blautia*. *Faecalibacterium* and *Subdoligranulum*, *Peptostreptococcus* and *Parvimonas*

were also co-occurrent in the network. In comparison, strong co-existence patterns were found based on the two hubs genera *Bacteroides* and *Enterocloster* in MSS status. Also, the genera *Enterocloster* and *Flabonifractor*, *Bacteroides* and *Parabacteroides* were suggested connected. The correlations between the pairs could describe the tendency of different species to co-occur under two cancer conditions. Interestingly, most of these co-abundant pairs have the ability to produce short-chain fatty acids (SCFAs) which indicates a potential functional role of the microbiome (Abdugheni et al. 2022). Overall, the topology of network indicates a true ecological interaction and provides a better understanding of the causes of behavior.

Phylogenetic profiles of tissue microbial communities in MSI status

Identification of differentially abundant microbes is one of the main goals of microbiome studies. In the LEfSe method, 41 bacterial taxa showed distinct relative abundances between MSI and MSS (linear discriminant analysis (LDA) score>2.0, p<0.05). For the Wilcoxon rank sum test (p<0.05) and DESeq2 (log2FC>1.5, p<0.05), a total of 6 and 38 different abundant species were recognized respectively. 9 of these genera remained significantly different in over two methods. The increase of *Fusobacterium, Granulicella, Leptotrichia, Lanchnoanaerobaculum, Treponema, and Selenomonas* were observed in the MSI group, and genera *Duodenibacillus, Akkermansia, Morganella* contained the key taxa in MSS group (Figure 4). The statistical distribution and relative abundance of these genus-level biomarkers support the prevalence of bacteria in most of the samples.

Considering that tumor microbes could also be impacted by tumor stages (Cai et al. 2023) and tumor location (Kim et al. 2018), the microbial signatures across these two MSI statuses were also discussed. In terms of tumor stage in MSI status, genus *Actinomyces* were enriched in stage I-III, while *Desulfovibrio* and *Mogibacterium* were the significantly distinct bacteria dominant in stage IV (Figure S4A). For MSS status, *Lacrimispora* was observed to increase in stage I-III. Genera including *Porphyromonas, Rhodoluna* and *Fluviicola* were dominant in the late stage (stage IV) (Figure S4B). We further investigated changes in microbial signature in two MSI statuses according to location. In MSI, no genus was found enriched in left-side colon cancer (LCC), but pathogens *Salmonella* and *Shewanella* were found in right-side colon cancer (RCC). *Megamona, Roseburia* and *Agathobacter* contained main phylotype in LCC in MSS status. *Prevotella, Cellulosilyticum* and *Leptotrichia* were the key species in RCC of MSS status (Figure S4D).

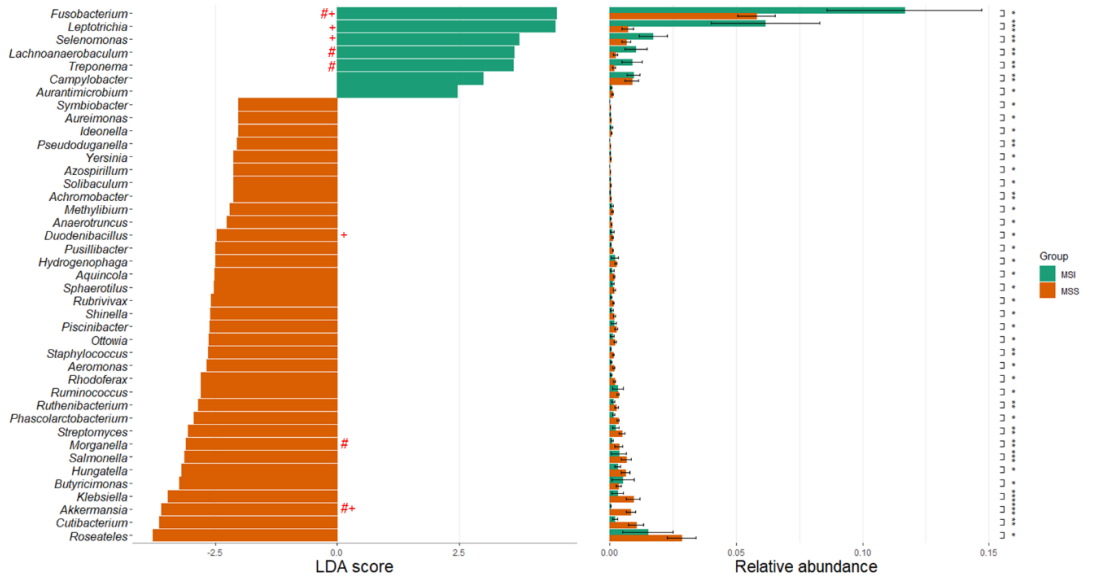


Figure 6. Histogram of LDA coupled with effective size measurement between MSI and MSS. '#' indicated differential abundance species between groups detected using DESeq2, '+' suggested significantly differential species recognized with Wilcoxon rank sum test

Key species are predictive of MSI and MSS status of host tissue molecular environment

Another piece of evidence that indicates a deep relationship between microbial communities and the two cancer subtypes was the result of the host Gene Ontology (GO) enrichment analysis. As shown in Figure 7A, the most enriched biological process was feeding behavior which could control body weight and increase the overall risk of CRC (Baroudi and Benammar-elgaaied 2016). Interestingly, this biological process is especially enriched in MSS CRC samples (Figure S5), which potentially illustrates a closer relationship between lifestyles and MSS CRC. All the other significant enriched processes were indicated related to antimicrobial behaviors including GO:0019730 antimicrobial humoral response, GO:0061844 antimicrobial humoral immune response mediated by antimicrobial peptide, GO0031640 killing of cells of other organism.

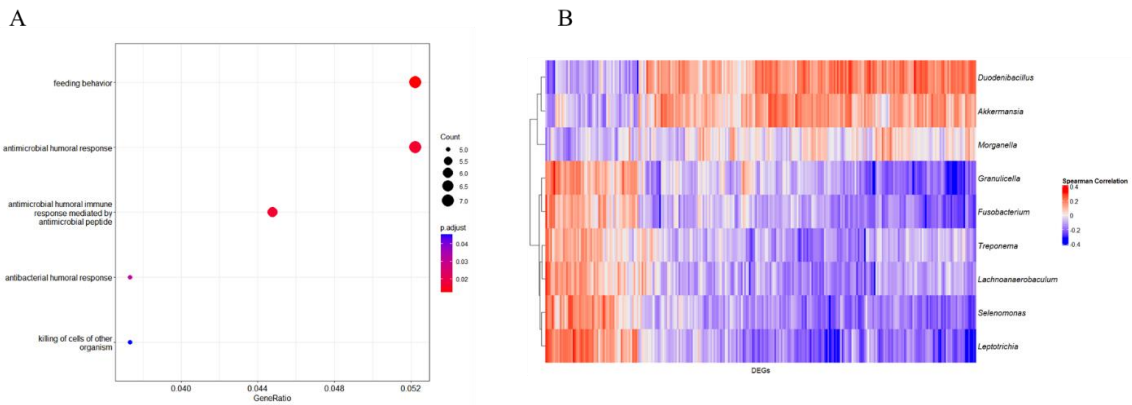


Figure 7. (A) Gene Ontology (GO) enrichment plot (biological process) differential expressed gene in two MS conditions (p.adj<0.05, q<0.2) (B) Correlation between host differential gene

(columns) expression (p<0.05, log2FC>1.5) in two MSI status and CLR-transformed species abundance (rows).

Next, we explored whether the 9 key species that we identified as significantly associated with two cancer subtypes had identifiable effects on host gene expression. Comparing normalized abundances of the species with matched mRNA expression data from 162 samples, we computed Spearman correlations and found clear transcriptional patterns that were associated with two MSI statuses (Figure 7B). Therefore, our results suggest that MSI status could be heavily impacted by microbes.

The alteration of microbial KEGG orthology genes from MSI and MSS CRC

Compared to other omics data types, transcriptomics enables to demonstration of the functions of microbial communities from a potential repertoire of bacteria that are actually in use in the environment (Bashiardes, Zilberman-Schapira, and Elinav 2016). Therefore, we further used HUMAnN3 to analyze the metabolic function of tissue microbiome results from MSI and MSS patients and compare the microbial genes annotated from the KEGG Orthology database. A total of 5216 KEGG annotations were observed and fell into 104 pathways. Features that differed in abundance were determined between MSI and MSS status were recognized with the Wilcoxon rank sum test with *p*-value<0.05 and log2 Fold Change>1.5. There were 67 differential expressed KO terms found in the samples, all of the terms were overrepresented in MSI CRC, with no significantly overrepresented functions found in MSS condition. Most terms upregulated in MSI status were related to carbohydrate metabolism (16% of annotated functions), amino acid metabolism (10%), signal and cellular processes and genetic information processing (13%).

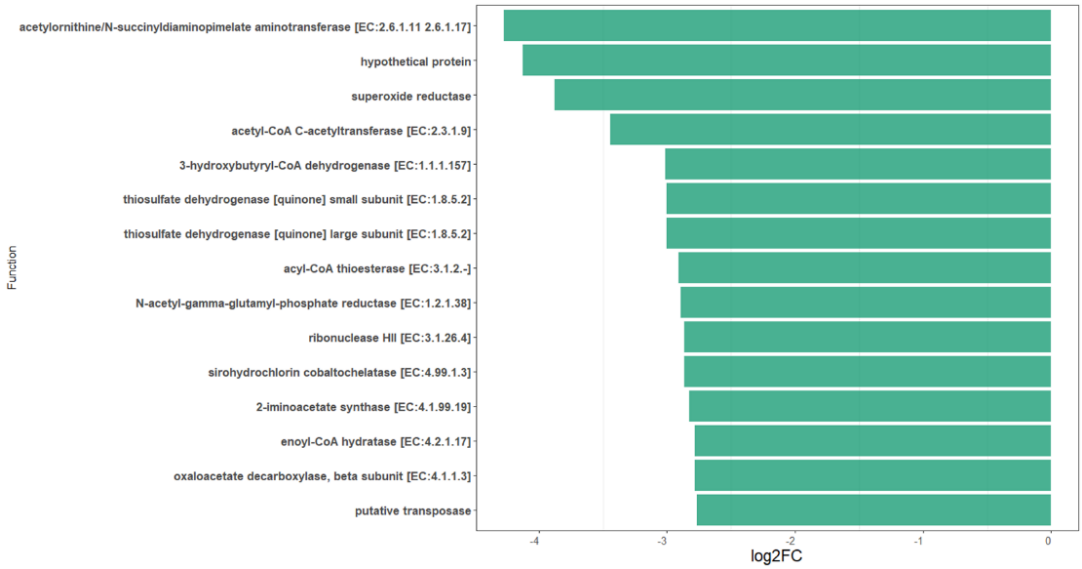


Figure 8. Bar plot of top15 differential abundant KO term identified with Wilcoxon rank sum test (p<0.05). Features are sorted by the absolute number of log2 fold change.

The top 15 significant expressed KO terms in MSI conditions were further explored for a more detailed description of the functions (Figure 8). These significant highly abundant terms in MSI

condition including Acetylornithine/N-succinyldiaminopimelate aminotransferase, N-acetyl-gamma-glutamyl-phosphate reductase and 2-iminoacetate synthase that related to amino acid metabolism pathway, Acyl-CoA thioesterase, Acetyl-CoA C-acetyltransferase, Enoyl-CoA hydratase, Acyl-CoA thioesterase that related to SCFAs metabolism, Superoxide reductase that related to oxidative stress response and Thiosulfate dehydrogenase which relates to sulfur metabolism. Also, terms like Ribonuclease HII with the function related DNA and chromosome were found significantly abundant under MSI condition. Globally, most of these pathways were found to be main functional pathways in CRC, which suggest an extra active functional activity under MSI condition (Debesa-Tur et al. 2021).

Antimicrobial resistance gene in the tissue samples sequences

In the previous functional analysis, KO terms related to antimicrobial resistance has been also potentially found in multiple pathogens including *Bacteroides fragilis*, *Fretibacterium fastidiusum*, *Streptococcus agalatiae* and *Porphyromonas gingivalis*. These KO terms were closely correlated to genes *erm, tolC*, *mar*, and membrane fusion proteins like *arcA*. Although the gene families were not significantly expressed, the phenomenon of antimicrobial resistance called our caution. Therefore, in the report, we also explored the existence of AMR genes within samples to provide very useful information for further CRC treatment and prescription. Using tools DeepARG and RGI, a total of 118 genes from 51 samples were detected after filtration. The abundance of genes was further normalized using mapped reads of a gene/number of bacteria in samples. Several AMR genes were found relatively higher expressed in samples, including *CTX-M* (β-lactamases), histone-like nucleoid-structuring protein (*H-NS*), *acrB* (tripartite efflux), *sul2* (Sulfonamide Resistance Gene), *APH(6)-Id* (aminoglycoside phosphotransferase) and MCR (colistin resistance gene) . Genes *ErmB, ErmF, TolC, acrA* which correspond with forementioned KO terms were also recognized in sample.

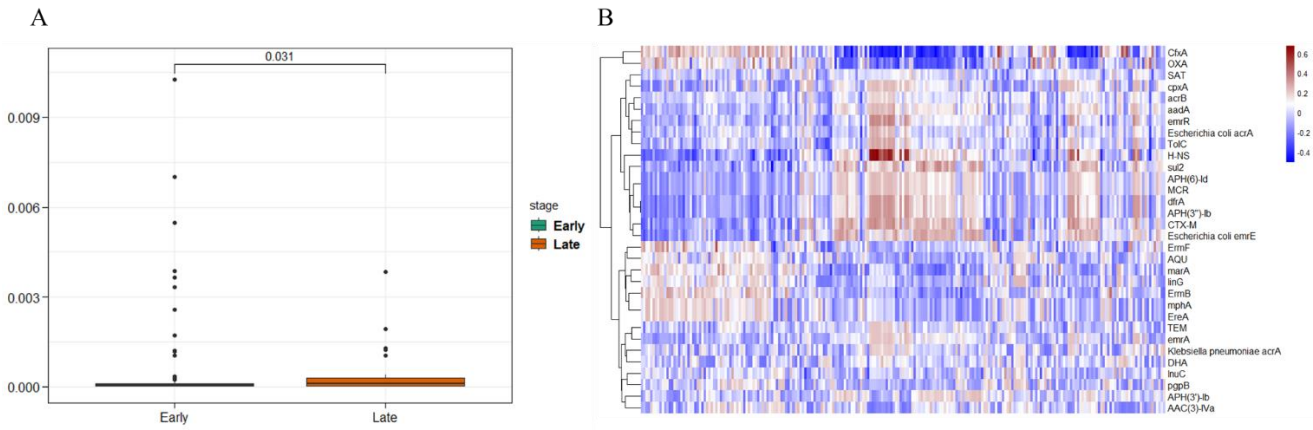


Figure 9. Significant differences in Normalized Reads Number for CRC stages (Early: stage I-III; Late: Stage IV). (B) Spearman correlation between abundance of genera (column, CLR transformed) and abundance of AMR genes

AMR genes that related to antibiotics including cephalantiosporin (mean normalized abundance:

0.015), aminoglycosides (mean normalized abundance: 0.0077) and sulfonamides (mean normalized abundance: 0.004) were suggested with high mean abundance among samples. Based on the Wilcoxon rank sum test, no significant difference was found between MSI and MSS samples (Figure S9). In comparison, a slight difference was indicated between the early stage (stage I-III) and late stage (stage IV) (Figure 9A). Using Spearman correlation, the relationship between antimicrobial genes and species could be revealed. In the research, a significant correlation was found between *CfxA* and genera *Hoylesella* (correlation=0.516), *OXA* and genus *Dialister* (correlation=0.404). Both *CfxA* and *OXA* are β-lactamases, they are responsible for resistance of cephamycin and penam respectively (Figure 9B).

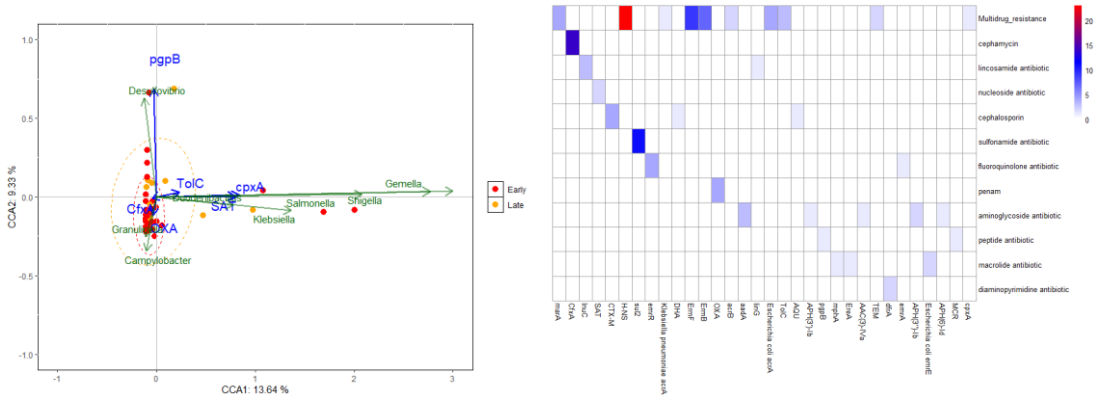


Figure 10. (A) Canonical correspondence analysis of the relationship between bacterial communities and AMR genes distribution. Samples were separated based on cancer stage based on different colors (B) Distribution of resistance drug types in samples

To further determine the detailed influence of AMR genes on sample distribution, canonical correspondence analysis was performed. After filtering with variance inflation factor analysis, the CCA results shown in Figure 10A demonstrated that bacterial community (Genus) was significantly shaped by several key AMR genes, including *pgpB, cpxA, SAT, OXA, TolC* and *CfxA*. Our results confirmed the conclusion driven by Spearman correlation analysis and found further relationship between genes and genera. Among these genes, *cpxA, SAT and TolC* were significantly related to CCA1 (explaining 13.64% of the variations). Potential correlations were indicated between the AMR genes and bacteria including *Gemella, Shigella, Klesiella* and *Duodenibacillus*. AMR genes *pgpB* and *OXA* were separated on CCA2 which explains 9.33% of the variations. *OXA* was suggested potentially related to *Granulicella* and *Campylobacter*, *pgpB* was suggests better correlated with *Desulfovibrio*.

DISCUSSION

Colorectal cancer tumors with microsatellite instability (MSI) are known to respond more favorably to immune therapy compared to those with microsatellite stability (MSS) phenotype. Recent research suggests tumor microbiome differs between these two subtypes and plays an important role in modulating host response (Byrd et al. 2023). Our study supports the notion, as our GO analysis of host differential expressed genes (DEGs) indicated an active antimicrobial host response. Thus, understanding the microbiome differences embedded in the two subtypes could offer valuable insight for improving CRC treatment and prognosis.

In the report, we analyzed RNA-sequencing data of 162 patients from a Singapore cohort to reveal the microbiome differences between MSI and MSS tumors. We identified that Bacteroidota and Bacillota, two dominant phyla in the human gut, (MclCallum and Tropini 2024) were the most abundant in both subtypes. However, Fusobacteriota and Acidobacteriota were significantly more abundant in MSI patients, whereas MSS tumor tissues were predominantly associated with phyla Pseudomonadota, Synergistota and Therdesulfobacteriota. At the genus level, *Leptotrichia* was found to be enriched in MSI patients, while *Alistipes* was more prevalent in MSS patients. These findings are consistent with previous studies by (Vanderbilt et al. 2020), which also highlighted the strong association of these genera with CRC. Contrary to multiple previous studies (M. Jin et al. 2022; Li et al. 2023), which reported higher alpha diversity in MSI tumors, our analysis found that MSS exhibited significantly higher values in several alpha diversity indices, such as the observed count and Shannon index. Several factors might account for these discrepancies. Variations in data preprocessing strategies and the impact of antibiotic usage, which can alter microbial burden are potential explanations (M. Jin et al. 2022). Additionally, the real species number in MSS tumors might have been underestimated in other studies, particularly those relying on 16S rRNA-sequencing. The use of 'universal' primers in PCR amplification can lead to the loss of a substantial fraction of reads, many of which constitute a significant portion of the 'shadow biosphere' (Yan et al. 2018).

Our study aligns with previous research in revealing significant differences in beta diversity between MSI and MSS tumors, indicating distinct microbial compositions. A total of 9 genera were recognized as differential abundant across at least two methods (Wilcoxon rank sum test, DESeq2 and LEfSe). These genera significantly influence host gene modulation, highlighting their potential role in CRC pathogenesis. Among them, *Fusobacterium, Leptotrichia, Selenomonas* and *Lanchnoanaerobaculum* – bacteria commonly found in the oral cavity and closely associated with oral diseases such as oral leukoplakia were found (Amer et al. 2017). This suggests an oral-gut translocation route associated with CRC. Typically, oral commensals are present in very low abundance in a healthy gut, compromising only about 2% of the gut microbiome. However, certain drug treatments and chronic inflammation could lead to the spread of these oral bacteria, reshaping the original microbiome, and potentially driving diseases like CRC (S. Jin, Wetzel, and Schirmer 2022). In our study, the mean abundance of these genera represents about 17.6% of the microbiome in the MSI group compared to 7.7% in MSS group, underscoring a strong correlation between MSI statuses, especially MSI CRC and oral commensal bacteria.

Fusobacterium is a well-known CRC-associated bacterium that is notably abundant in MSI

CRC tissues. It can penetrate deeply into tumors, internalized within endosomes and lysosomes and persist existence in the tissue. This bacterium is linked to a higher expression of proinflammatory markers and plays an active role in mediating inflammation (Salvucci et al. 2022). Moreover, *Fusobacterium*-enriched tissues was symbolized with increased tumor growth and invasion, and are correlated with pro-tumoral immune responses (Lee et al. 2021). Compared to *Fusobacterium*, molecular roles of *Leptotrichia*, *Selenomonas* and *Lanchnoanaerobaculum* were less studied, however, still with the indication of their significant abundance in high-grade CRC tissue samples (Fukuoka et al. 2023; J. Zhang et al. 2021). The former two genera, in particular, have been associated with short overall survival (OS) (An et al. 2024). Additionally, *Treponoma*, an intra-tumor pathogen was also found abundant in MSI. The genus is closely associated with high tumor mutation burden (TMB) and host genes SPEN and IGF2R, which are predictive biomarkers for MSI status (Dong et al. 2024; Okuda et al. 2021). The most enigmatic genus, *Granulicella,* is known for its acidophilic nature and iron utilization, potentially correlated to microbial metabolism for better adaptation in acidic tumor microenvironment (Costa et al. 2020). It has also been correlated with other cancers such as lung cancer (Z. Zhao et al. 2021).

Oral pathogens formed a cluster in MSS status

Our analysis of correlation networks revealed a more complex cross-talk between genera in MSS CRC compared to MSI CRC. In MSI, several oral bacteria including *Fusobacterium, Selenomonas, Leptotrichia* were known enriched. However, these bacteria did not co-occurrent within the MSI network, which appeared segregated into multiple components. In contrast, in MSS status, these bacteria are more tightly connected and form an oral pathogen cluster. This cluster pattern was also observed in CRC samples in other papers (Flemer et al. 2017; Warren et al. 2013). Previous research suggests that biofilms are often associated with right-side tumors, which are more frequently MSI-positive (Sugai et al. 2006). Therefore, in our samples, we hypothesized that biofilms were potentially formed at early stages or there could be a spatial distribution that affects cell-to-cell signals in the development (Kim et al. 2018). Additionally, the prevalence and ability of some oral bacteria such as *Fusobacterium* to invade mucosa does not necessarily depend on the presence of biofilm (DeDecker et al. 2021). As dominant species in tumors, these microbes may have indirect effects on each other instead of direct contact. Compared to MSI, the existence of more negative correlations could be a symbol of competition instead of synergism in MSS. In the nutrient-deprived environment in the tumors, microbial connections might intensify as a strategy to adapt to the harsh conditions and overcome host defenses (Baishya and Wakeman 2019). This competitive behavior may contribute to the distinct microbial dynamics observed between MSI and MSS tumors.

Different functions behind two CRC subtypes

The functional profiles between MSI and MSS patients showed limited differentiation yet the data suggested more active functional activities in MSI patients. In MSI CRC, a notable preference for oral microbes indicates a tissue hypoxia and low pH tumor microenvironment (Younginger et al. 2023). This altered transcriptome might enable better adaptation to acidic

conditions, reactive oxygen species and metabolite availability. For instance, elevated levels of superoxide reductase and activities related pentose phosphate pathway suggest enhanced ROS reduction in our case. Metabolic activities related to amino acids were suggested as markers of CRC at various stages. In our study, the phenomenon was highly related to the biosynthesis of lysine, arginine, and phenylalanine, aligning with the research (Mizutani, Yamada, and Yachida 2020). Amino acids play important roles in several steps of molecular biosynthesis, maintaining redox balance and serving as energy sources. Abundant amino acids were able to drive cancer cell proliferation, hence associated with carcinogenesis (Coker et al. 2022). Additionally, glycolysis-derived precursors for SCFAs like butyrate and acetate production were associated with enhanced MSI CRC (e.g. Enoyl-CoA Hydratase). From bacteria to cancer cells, glycolysis plays a fundamental role in rapid cell proliferation and cell growth when sugar is lacking in the environment. A lot of microbes rely primarily on glucose fermentation even in regardless of oxygen (Lunt and Heiden 2011). These SCFAs could be rapidly adsorbed from the lumen and used as energy sources by epithelial cells. Also, butyrate is suggested to significantly control intestinal inflammation, inducing apoptosis of CRC cells and keeping the homeostasis of microbiota. Which, could be a possible reason for a better prognostic outcome of MSI patients. However, a low concentration of butyrate could reversely promote the pro-inflammatory impact, therefore the overall biological impact could be more complicated and need further research (Louis, Hold, and Flint 2014). Another intriguing finding is the enrichment of sulfur metabolism in *Bacteroides fragilis* in MSI samples. This aligns with the research of (Hale et al. 2018), where hydrogen sulfide (H_2S) was the final product that could encourage cancer cell proliferation. Furthermore, DNA-related proteins were overrepresented in MSI patients. This may indicate the chromosomal instability of microbes under selective pressure within the tumor, where a high mutation rate is needed for rapid adaptation and drug resistance (de Bruin, Taylor, and Swanton 2013). It is very interesting that in the paper of (Khan 2015), they mentioned the possibility that the epigenetic regulators of microbes could reprogram the gene expression of host cells. Besides, intracellular microbes can potentially impact host DNA repair function through horizontal gene transfer of segment DNA. Considering MSI CRC could be caused by DNA methylation and mutation, this phenomenon may suggest a hidden genetic correlation between the two.

Antimicrobial resistance genes in the samples

Recent studies have demonstrated that intratumor bacteria may persist even after treatment with impermeable antibiotics like ampicillin and gentamicin (Wang et al. 2023). The fact may indicate the prevalence of antimicrobial resistance in tumor microbiome. In the report, we applied mapping-based AMR gene identification methods. Compared to assembled-based transcriptomics, mapping transcripts back to reference database is the most common approach to profile gene features and it has higher sensitivity. However, this high sensitivity is usually at the expense of lower specificity (Y. Zhang et al. 2021). To address this, we applied both RGI and DeepARG in AMR gene identification, and only genes recognized in both tools with high coverage were considered. Interestingly, our previous functional analysis with HUMAnN3 also corroborated the presence of AMR genes in intratumor tissues, which suggests that bacteria on tissue do have resistance against antibiotic usage. In the research, multidrug resistance and

aminoglycoside were the two most prevalent resistance drug types observed in our samples. Several identified AMR genes including *acrB*, *H-NS*, *Escherichia coli acrA, marA* were observed in high abundance and confer resistance to broad-spectrum antibiotics such as cephalosporins, aminoglycosides, sulfonamides. These findings are consistent with those reported by (Liu et al. 2021). A slight difference was observed in the AMR gene abundance between the early and late stages of cancer, although no gene was found significantly abundant in particular stages. This variation might be attributed to long-term antibiotic use by patients, leading to the accumulation and horizontal transfer of AMR genes. β-lactamases were indicated to play crucial roles in shaping the microbial community in CRC patients. They are one of the most used antibiotics in CRC treatment. In the study, they were found highly related to significant abundant gut species (for example *Granulicella*), together with other research, suggesting the prevalence of β-lactamases in the gut. Some AMR genes such as *cpxA* and *SAT* which are associated with aminoglycosides, were linked to clinical oral pathogens such as *Gemella* and *Schaalia*. These pathogens are known to acquire resistance determinants through horizontal transfer of mobile genetic elements (MGE), and poses high risk of MGEs dissemination (Lamaudière et al. 2023). Notably, the presence of *MCR* gene which conders resistance to the "last resort" polymyxin antibiotics was found, underscores the potential consequences of antibiotic usage.

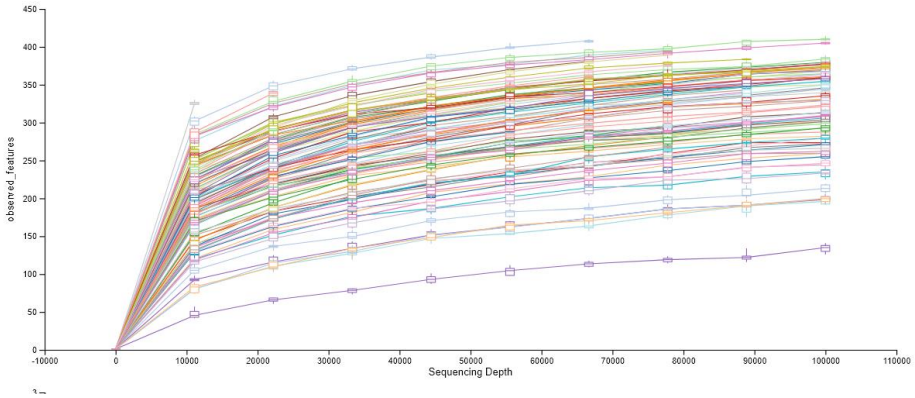
CONCLUSION

Our study provides valuable insights into the relationship between tumor microbiomes and colorectal cancer (CRC) subtypes, specifically microsatellite instability (MSI) and microsatellite stability (MSS). However, there are several limitations and areas for improvement that need to be addressed.

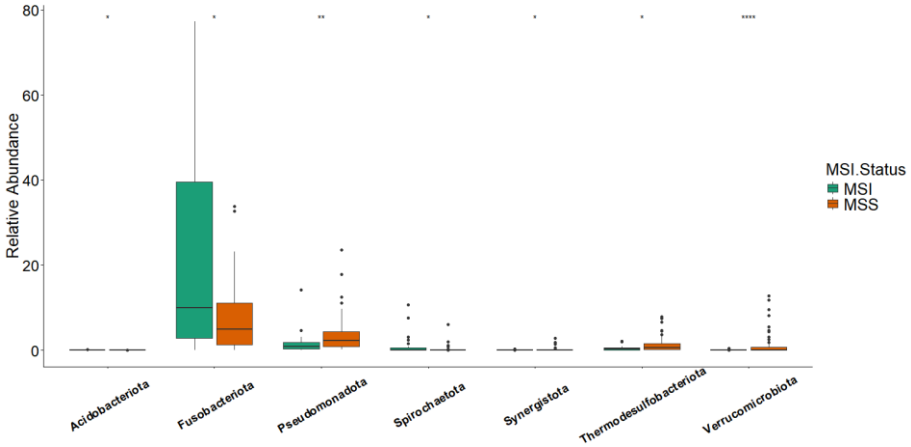
First, the significant disparity in sample sizes between MSI (33 samples) and MSS (126 samples) subtypes may introduce potential bias. This imbalance may affect the robustness and generalizability of the findings. Second, although multiple methods were applied during the recognition of MSI status biomarkers, in our study design, we cannot distinguish causes from consequences. Therefore, further experimental validation is needed to confirm the result and reduce false positive predictions. Finally, the lack of detailed antibiotic administration information limits our ability to explore its impact on AMR within tumor microbiomes comprehensively.

Overall, our study illustrated a close relationship between microbiomes and two subtypes of CRC: MSI and MSS status. The composition of *Fusobacterium*, *Leptotrichia*, *Selenomonas* and *Lanchnoanaerobaculu* differ significantly between MSI and MSS patients. These differences in microbiome profiles may contribute to different immunotherapeutic responses and clinical outcomes of CRC. In our research, pathways of carbohydrate metabolism, amino acid metabolic, signal and cellular processes and genetic information processing were vital and more active for MSI patients. Also, AMR gens against multiple wide-spectrum antibiotics were found in patients. Future studies on the detailed mechanisms of how the differences between tumor microbiomes in cancer subtypes could be crucial for immunotherapy. Also, plan-wised antibiotics usage should be considered during cancer treatment.

APPENDIX



FigureS1. Rarefaction curve of a total of 162 samples used in the study. Most of curves reached their asymptotes and started to plateau, suggesting that saturation in sequencing was achieved and most of species were sampled.



FigureS2. The significant difference in top 10 phyla between MSI and MSS status

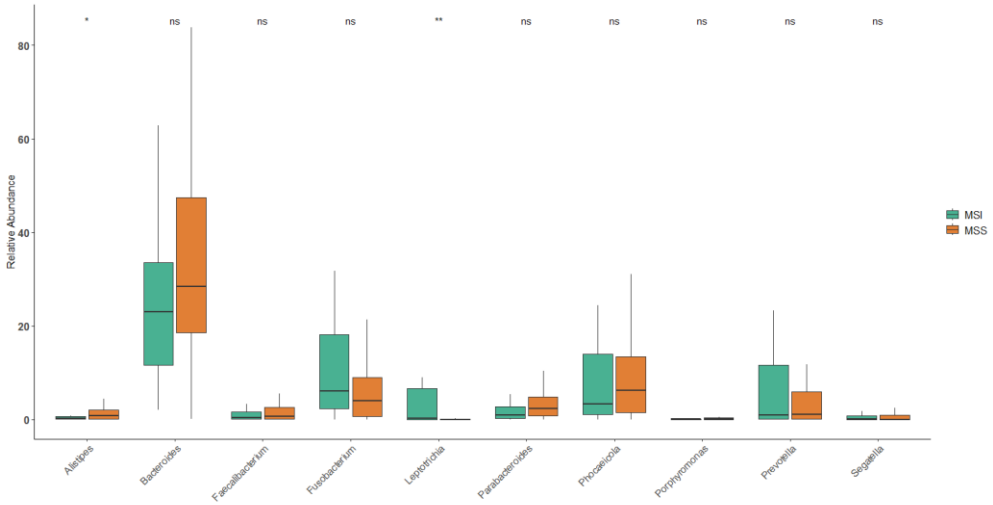


Figure S3. The significant difference in top 10 Genera between MSI and MSS status

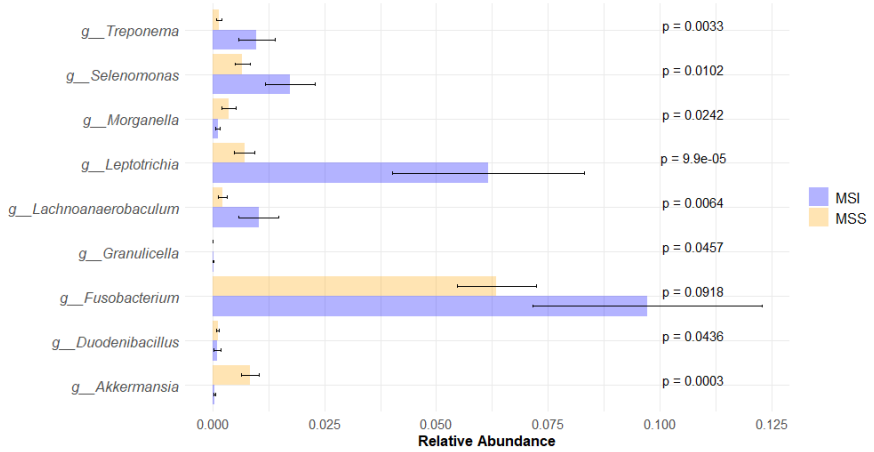


Figure S4. Relative abundance of key species (differential abundant species agreed with over two tools in LEfse, DESeq2 and Wilcoxon rank sum test). P values were calculated based on the Wilcoxon rank sum test.

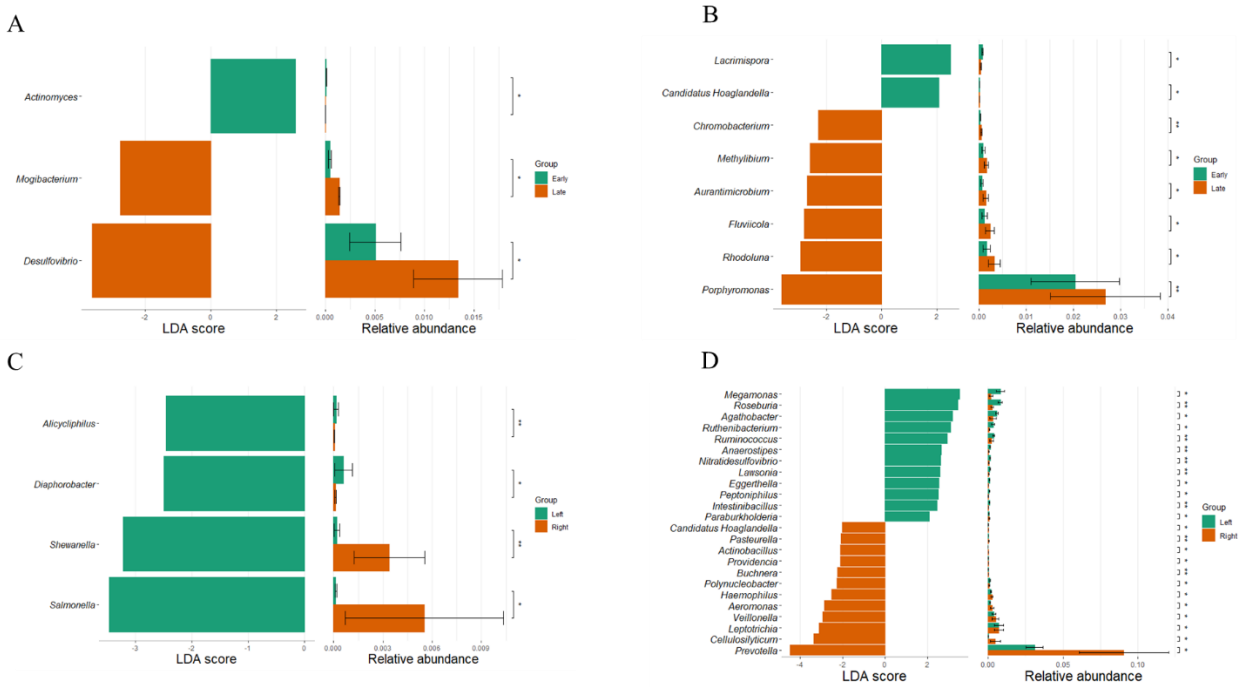
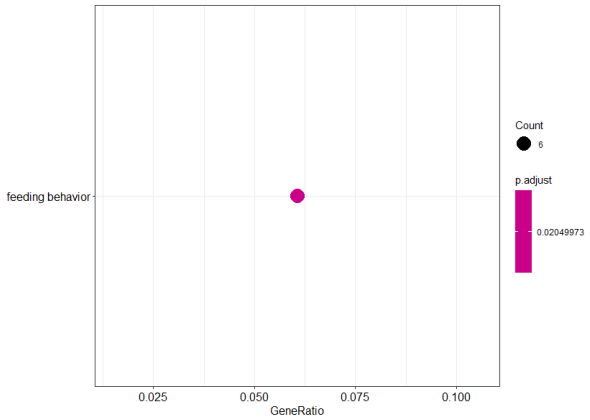


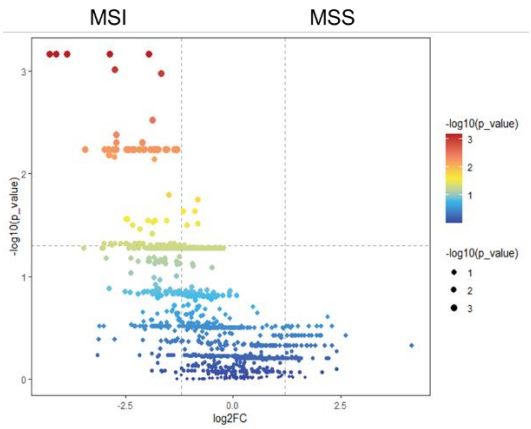
Figure S4. (A)Histogram of LEfSe measurement of 33 MSI status in different tumor stages (early: stage I-III, later: stage IV). p values are calculated by Kruskal-Wallis test, LDA score >2.0, $p < 0.05$. (B) Histogram of LEfSe measurement in 123 MSS status in different tumor stages (C) Histogram of LEfSe measurement of 33 MSI status in different tumor locations (Left: left-side colon cancer, Right: right-side colon cancer) (D) Histogram of LEfSe measurement of 123 MSS status in different tumor locations.

TableS1: Topological properties of whole networks of MSI and MSS

	MSI status	MSS status
Number of components	11	1
Clustering coefficient	0.44	0.43
Modularity	0.36	0.24
Positive edge percentage	73.3	58.6
Edge density	0.09	0.23
Natural connectivity	0.04	0.05



FigureS5. Gene Ontology (GO) enrichment analysis for host differential expressed genes in MSS samples



Figur S6. Volcano plot for the abundance of KEGG Orthology (KO) terms in the samples. Differential abundant terms were selected using the Wilcoxon rank sum test between MSI and MSS status ($p<0.05$, $log2FC>1.5$). Significantly abundant KO terms were only observed in MSI status.

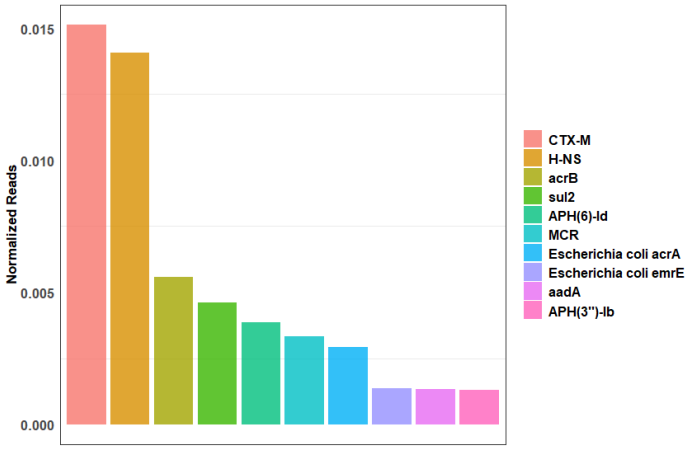


Figure S7. Top10 most abundant AMR genes among filtered samples.

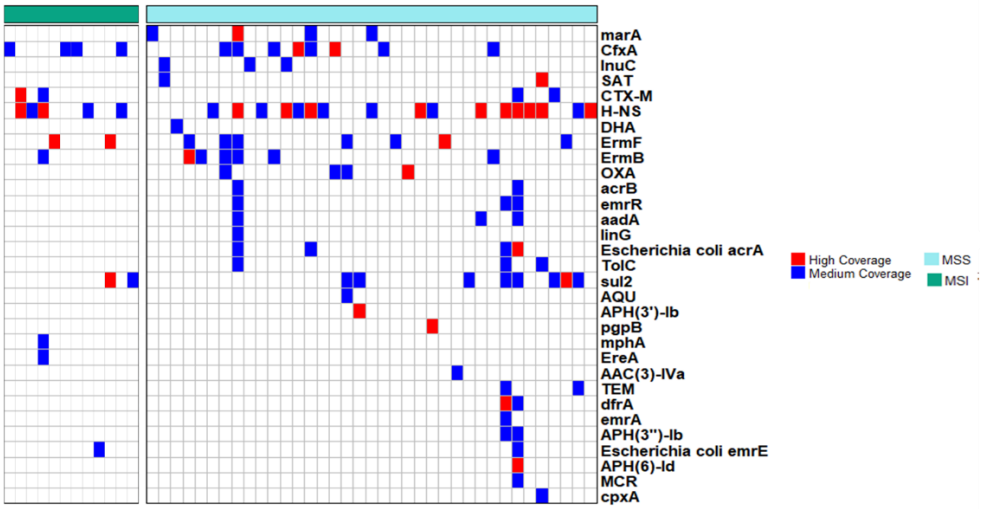
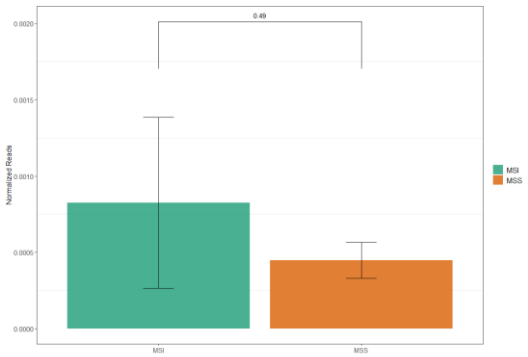


Figure S8: Distribution of AMR genes across MSI and MSS status. High-quality reads were mapped back to AMR genes database, where high coverage (red) indicate over 85% AMR genes were covered with reads. Medium Coverage indicates 50%-85% of genes were covered with reads (blue).



FigureS9. Comparison of the normalized reads (number reads covered on genes/number of observed microbes in the sample) based on student t test. No significance were observed between MSI and MSS status

REFERENCES

Abdugheni, Rashidin, Wen-Zhao Wang, Yu-Jing Wang, Meng-Xuan Du, Feng-Lan Liu, Nan Zhou, Cheng-Ying Jiang, Chang-Yu Wang, Linhuan Wu, Juncai Ma, Chang Liu, and Shuang-Jiang Liu. 2022. "Metabolite Profiling of Human-Originated Lachnospiraceae at the Strain Level." *iMeta* 1(4):e58. doi: 10.1002/imt2.58.

Alcock, Brian P., William Huynh, Romeo Chalil, Keaton W. Smith, Amogelang R. Raphenya, Mateusz A. Wlodarski, Arman Edalatmand, Aaron Petkau, Sohaib A. Syed, Kara K. Tsang, Sheridan J. C. Baker, Mugdha Dave, Madeline C. McCarthy, Karyn M. Mukiri, Jalees A. Nasir, Bahar Golbon, Hamna Imtiaz, Xingjian Jiang, Komal Kaur, Megan Kwong, Zi Cheng Liang, Keyu C. Niu, Prabakar Shan, Jasmine Y. J. Yang, Kristen L. Gray, Gemma R. Hoad, Baofeng Jia, Timsy Bhando, Lindsey A. Carfrae, Maya A. Farha, Shawn French, Rodion Gordzevich, Kenneth Rachwalski, Megan M. Tu, Emily Bordeleau, Damion Dooley, Emma Griffiths, Haley L. Zubyk, Eric D. Brown, Finlay Maguire, Robert G. Beiko, William W. L. Hsiao, Fiona S. L. Brinkman, Gary Van Domselaar, and Andrew G. McArthur. 2023. "CARD 2023: Expanded Curation, Support for Machine Learning, and Resistome Prediction at the Comprehensive Antibiotic Resistance Database." *Nucleic Acids Research* 51(D1):D690–99. doi: 10.1093/nar/gkac920.

Amer, Abdrazak, Sheila Galvin, Claire M. Healy, and Gary P. Moran. 2017. "The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for Fusobacterium, Leptotrichia, Campylobacter, and Rothia Species." *Frontiers in Microbiology* 8. doi: 10.3389/fmicb.2017.02391.

An, Ho Jung, Mira A. Partha, HoJoon Lee, Billy T. Lau, Dmitri S. Pavlichin, Alison Almeda, Anna C. Hooker, Giwon Shin, and Hanlee P. Ji. 2024. "Tumor-Associated Microbiome Features of Metastatic Colorectal Cancer and Clinical Implications." *Frontiers in Oncology* 13. doi: 10.3389/fonc.2023.1310054.

Aneke-Nash, Chino, Garrett Yoon, Mengmeng Du, and Peter Liang. 2021. "Antibiotic Use and Colorectal Neoplasia: A Systematic Review and Meta-Analysis." *BMJ Open Gastroenterology* 8(1):e000601. doi: 10.1136/bmjgast-2021-000601.

Arango-Argoty, Gustavo, Emily Garner, Amy Pruden, Lenwood S. Heath, Peter Vikesland, and Liqing Zhang. 2018. "DeepARG: A Deep Learning Approach for Predicting Antibiotic Resistance Genes from Metagenomic Data." *Microbiome* 6(1):23. doi: 10.1186/s40168-018-0401-z.

Avril, Marion, and R. William DePaolo. n.d. "'Driver-Passenger' Bacteria and Their Metabolites in the Pathogenesis of Colorectal Cancer." *Gut Microbes* 13(1):1941710. doi: 10.1080/19490976.2021.1941710.

Baishya, Jiwasmika, and Catherine A. Wakeman. 2019. "Selective Pressures during Chronic Infection Drive Microbial Competition and Cooperation." *Npj Biofilms and Microbiomes* 5(1):1–9. doi: 10.1038/s41522-019-0089-2.

Baroudi, Olfa, and Amel Benammar-elgaaied. 2016. "Involvement of Genetic Factors and Lifestyle on the Occurrence of Colorectal and Gastric Cancer." *Critical Reviews in Oncology/Hematology* 107:72–81. doi: 10.1016/j.critrevonc.2016.08.014.

Bashiardes, Stavros, Gili Zilberman-Schapira, and Eran Elinav. 2016. "Use of Metatranscriptomics in Microbiome Research." *Bioinformatics and Biology Insights* 10:BBI.S34610. doi: 10.4137/BBI.S34610.

Beghini, Francesco, Lauren J. McIver, Aitor Blanco-Míguez, Leonard Dubois, Francesco Asnicar, Sagun Maharjan, Ana Mailyan, Paolo Manghi, Matthias Scholz, Andrew Maltez Thomas, Mireia

Valles-Colomer, George Weingart, Yancong Zhang, Moreno Zolfo, Curtis Huttenhower, Eric A. Franzosa, and Nicola Segata. 2021. “Integrating Taxonomic, Functional, and Strain-Level Profiling of Diverse Microbial Communities with bioBakery 3” edited by P. Turnbaugh, E. Franco, and C. T. Brown. *eLife* 10:e65088. doi: 10.7554/eLife.65088.

Benej, Martin, Rebecca Hoyd, McKenzie Kreamer, Caroline E. Wheeler, Dennis J. Grencewicz, Fouad Choueiry, Carlos H. F. Chan, Yousef Zakharia, Qin Ma, Rebecca D. Dodd, Cornelia M. Ulrich, Sheetal Hardikar, Michelle L. Churchman, Ahmad A. Tarhini, Lary A. Robinson, Eric A. Singer, Alexandra P. Ikeguchi, Martin D. McCarter, Gabriel Tinoco, Marium Husain, Ning Jin, Aik C. Tan, Afaf E. G. Osman, Islam Eljilany, Gregory Riedlinger, Bryan P. Schneider, Katarina Benejova, Martin Kery, Ioanna Papandreou, Jiangjiang Zhu, Nicholas Denko, Daniel Spakowicz, and for the exORIEN Consortium. 2024. “The Tumor Microbiome Reacts to Hypoxia and Can Influence Response to Radiation Treatment in Colorectal Cancer.” *Cancer Research Communications* 4(7):1690–1701. doi: 10.1158/2767-9764.CRC-23-0367.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30(15):2114–20. doi: 10.1093/bioinformatics/btu170.

de Bruin, Elza C., Tiffany B. Taylor, and Charles Swanton. 2013. “Intra-Tumor Heterogeneity: Lessons from Microbial Evolution and Clinical Implications.” *Genome Medicine* 5(11):101. doi: 10.1186/gm505.

Bullman, Susan, Chandra S. Pedamallu, Ewa Sicinska, Thomas E. Clancy, Xiaoyang Zhang, Diana Cai, Donna Neuberg, Katherine Huang, Fatima Guevara, Timothy Nelson, Otari Chipashvili, Timothy Hagan, Mark Walker, Aruna Ramachandran, Begoña Diosdado, Garazi Serna, Nuria Mulet, Stefania Landolfi, Santiago Ramon y Cajal, Roberta Fasani, Andrew J. Aguirre, Kimmie Ng, Elena Élez, Shuji Ogino, Josep Tabernero, Charles S. Fuchs, William C. Hahn, Paolo Nuciforo, and Matthew Meyerson. 2017. “Analysis of Fusobacterium Persistence and Antibiotic Response in Colorectal Cancer.” *Science* 358(6369):1443–48. doi: 10.1126/science.aal5240.

Byrd, Doratha A., Wenyi Fan, K. Leigh Greathouse, Michael C. Wu, Hao Xie, and Xuefeng Wang. 2023. “The Intratumor Microbiome Is Associated with Microsatellite Instability.” *JNCI: Journal of the National Cancer Institute* 115(8):989–93. doi: 10.1093/jnci/djad083.

Cai, Ping, Jinbo Xiong, Haonan Sha, Xiaoyu Dai, and Jiaqi Lu. 2023. “Tumor Bacterial Markers Diagnose the Initiation and Four Stages of Colorectal Cancer.” *Frontiers in Cellular and Infection Microbiology* 13. doi: 10.3389/fcimb.2023.1123544.

Coker, Olabisi Oluwabukola, Changan Liu, William Ka Kei Wu, Sunny Hei Wong, Wei Jia, Joseph J. Y. Sung, and Jun Yu. 2022. “Altered Gut Metabolites and Microbiota Interactions Are Implicated in Colorectal Carcinogenesis and Can Be Non-Invasive Diagnostic Biomarkers.” *Microbiome* 10(1):35. doi: 10.1186/s40168-021-01208-5.

Costa, Ohana Y. A., Chidinma Oguejiofor, Daniela Zühlke, Cristine C. Barreto, Christine Wünsche, Katharina Riedel, and Eiko E. Kuramae. 2020. “Impact of Different Trace Elements on the Growth and Proteome of Two Strains of Granulicella, Class ‘Acidobacteriia.’” *Frontiers in Microbiology* 11. doi: 10.3389/fmicb.2020.01227.

Debelius, Justine W., Lars Engstrand, Andreas Matussek, Nele Brusselaers, James T. Morton, Margaretha Stenmarker, and Renate S. Olsen. 2023. “The Local Tumor Microbiome Is Associated with Survival in Late-Stage Colorectal Cancer Patients.” *Microbiology Spectrum* 11(3):e05066-22. doi: 10.1128/spectrum.05066-22.

Debesa-Tur, Gabriela, Vicente Pérez-Brocal, Susana Ruiz-Ruiz, Adela Castillejo, Amparo Latorre,

José Luis Soto, and Andrés Moya. 2021. “Metagenomic Analysis of Formalin-Fixed Paraffin-Embedded Tumor and Normal Mucosa Reveals Differences in the Microbiome of Colorectal Cancer Patients.” *Scientific Reports* 11(1):391. doi: 10.1038/s41598-020-79874-y.

DeDecker, Lauren, Bretton Coppedge, Julio Avelar-Barragan, William Karnes, and Katrine Whiteson. 2021. “Microbiome Distinctions between the CRC Carcinogenic Pathways.” *Gut Microbes* 13(1):1–12. doi: 10.1080/19490976.2020.1854641.

Dejea, Christine M., Elizabeth C. Wick, Elizabeth M. Hechenbleikner, James R. White, Jessica L. Mark Welch, Blair J. Rossetti, Scott N. Peterson, Erik C. Snesrud, Gary G. Borisy, Mark Lazarev, Ellen Stein, Jamuna Vadivelu, April C. Roslani, Ausuma A. Malik, Jane W. Wanyiri, Khean L. Goh, Iyadorai Thevambiga, Kai Fu, Fengyi Wan, Nicolas Llosa, Franck Housseau, Katharine Romans, XinQun Wu, Florencia M. McAllister, Shaoguang Wu, Bert Vogelstein, Kenneth W. Kinzler, Drew M. Pardoll, and Cynthia L. Sears. 2014. “Microbiota Organization Is a Distinct Feature of Proximal Colorectal Cancers.” *Proceedings of the National Academy of Sciences* 111(51):18321–26. doi: 10.1073/pnas.1406199111.

Dik, Vincent K., Martijn G. H. van Oijen, Hugo M. Smeets, and Peter D. Siersema. 2016. “Frequent Use of Antibiotics Is Associated with Colorectal Cancer Risk: Results of a Nested Case–Control Study.” *Digestive Diseases and Sciences* 61(1):255–64. doi: 10.1007/s10620-015-3828-0.

Dong, Yuanmei, Sisi Ye, Huizi Li, Juan Li, Rongrui Liu, and Yanyun Zhu. 2024. “The Role of SPEN Mutations as Predictive Biomarkers for Immunotherapy Response in Colorectal Cancer: Insights from a Retrospective Cohort Analysis.” *Journal of Personalized Medicine* 14(2):131. doi: 10.3390/jpm14020131.

Drewes, Julia L., James R. White, Christine M. Dejea, Payam Fathi, Thevambiga Iyadorai, Jamuna Vadivelu, April C. Roslani, Elizabeth C. Wick, Emmanuel F. Mongodin, Mun Fai Loke, Kumar Thulasi, Han Ming Gan, Khean Lee Goh, Hoong Yin Chong, Sandip Kumar, Jane W. Wanyiri, and Cynthia L. Sears. 2017. “High-Resolution Bacterial 16S rRNA Gene Profile Meta-Analysis and Biofilm Status Reveal Common Colorectal Cancer Consortia.” *Npj Biofilms and Microbiomes* 3(1):1–12. doi: 10.1038/s41522-017-0040-3.

Feng, Jiahui, Zhizhong Gong, Zhangran Sun, Juan Li, Na Xu, Rick F. Thorne, Xu Dong Zhang, Xiaoying Liu, and Gang Liu. 2023. “Microbiome and Metabolic Features of Tissues and Feces Reveal Diagnostic Biomarkers for Colorectal Cancer.” *Frontiers in Microbiology* 14. doi: 10.3389/fmicb.2023.1034325.

Flemer, Burkhardt, Denise B. Lynch, Jillian M. R. Brown, Ian B. Jeffery, Feargal J. Ryan, Marcus J. Claesson, Micheal O’Riordain, Fergus Shanahan, and Paul W. O’Toole. 2017. “Tumour-Associated and Non-Tumour-Associated Microbiota in Colorectal Cancer.” *Gut* 66(4):633–43. doi: 10.1136/gutjnl-2015-309595.

Friedman, Jonathan, and Eric J. Alm. 2012. “Inferring Correlation Networks from Genomic Survey Data.” *PLOS Computational Biology* 8(9):e1002687. doi: 10.1371/journal.pcbi.1002687.

Fukuoka, Hironori, Dieter M. Tourlousse, Akiko Ohashi, Shinsuke Suzuki, Kazuya Nakagawa, Mayumi Ozawa, Atsushi Ishibe, Itaru Endo, and Yuji Sekiguchi. 2023. “Elucidating Colorectal Cancer-Associated Bacteria through Profiling of Minimally Perturbed Tissue-Associated Microbiota.” *Frontiers in Cellular and Infection Microbiology* 13. doi: 10.3389/fcimb.2023.1216024.

Gagnière, Johan, Virginie Bonnin, Anne-Sophie Jarrousse, Emilie Cardamone, Allison Agus, Nancy Uhrhammer, Pierre Sauvanet, Pierre Déchelotte, Nicolas Barnich, Richard Bonnet, Denis Pezet, and

Mathilde Bonnet. 2017. “Interactions between Microsatellite Instability and Human Gut Colonization by Escherichia Coli in Colorectal Cancer.” *Clinical Science* 131(6):471–85. doi: 10.1042/CS20160876.

Gihawi, Abraham, Yuchen Ge, Jennifer Lu, Daniela Puiu, Amanda Xu, Colin S. Cooper, Daniel S. Brewer, Mihaela Pertea, and Steven L. Salzberg. 2023. “Major Data Analysis Errors Invalidate Cancer Microbiome Findings.” *mBio* 14(5):e01607-23. doi: 10.1128/mbio.01607-23.

Gupta, Ruby, Surajit Sinha, and Rabindra N. Paul. 2018. “The Impact of Microsatellite Stability Status in Colorectal Cancer.” *Current Problems in Cancer* 42(6):548–59. doi: 10.1016/j.currproblcancer.2018.06.010.

Hale, Vanessa L., Patricio Jeraldo, Jun Chen, Michael Mundy, Janet Yao, Sambhawa Priya, Gary Keeney, Kelly Lyke, Jason Ridlon, Bryan A. White, Amy J. French, Stephen N. Thibodeau, Christian Diener, Osbaldo Resendis-Antonio, Jaime Gransee, Tumpa Dutta, Xuan-Mai Petterson, Jaeyun Sung, Ran Blekhman, Lisa Boardman, David Larson, Heidi Nelson, and Nicholas Chia. 2018. “Distinct Microbes, Metabolites, and Ecologies Define the Microbiome in Deficient and Proficient Mismatch Repair Colorectal Cancers.” *Genome Medicine* 10(1):78. doi: 10.1186/s13073-018-0586-6.

Hashemi, Mehrdad, Sahar Abbaszadeh, Mohsen Rashidi, Nafisesadat Amini, Kasra Talebi Anaraki, Motahhar Motahhary, Ensi Khalilipouya, Abdorrahman Harif Nashtifani, Sasan Shafiei, Marzieh Ramezani Farani, Noushin Nabavi, Shokooh Salimimoghadam, Amir Reza Aref, Rasoul Raesi, Afshin Taheriazam, Maliheh Entezari, and Wenliang Zha. 2023. “STAT3 as a Newly Emerging Target in Colorectal Cancer Therapy: Tumorigenesis, Therapy Response, and Pharmacological/Nanoplatform Strategies.” *Environmental Research* 233:116458. doi: 10.1016/j.envres.2023.116458.

Hexun, Zhang, Toru Miyake, Takeru Maekawa, Haruki Mori, Daiki Yasukawa, Masashi Ohno, Atsushi Nishida, Akira Andoh, and Masaji Tani. 2023. “High Abundance of Lachnospiraceae in the Human Gut Microbiome Is Related to High Immunoscores in Advanced Colorectal Cancer.” *Cancer Immunology, Immunotherapy* 72(2):315–26. doi: 10.1007/s00262-022-03256-8.

Jin, Min, Jingjing Wu, Linli Shi, Bin Zhou, Fumei Shang, Xiaona Chang, Xiaochuan Dong, Shenghe Deng, Li Liu, Kailin Cai, Xiu Nie, Tao Zhang, Jun Fan, and Hongli Liu. 2022. “Gut Microbiota Distinct between Colorectal Cancers with Deficient and Proficient Mismatch Repair: A Study of 230 CRC Patients.” *Frontiers in Microbiology* 13. doi: 10.3389/fmicb.2022.993285.

Jin, Shen, Daniela Wetzel, and Melanie Schirmer. 2022. “Deciphering Mechanisms and Implications of Bacterial Translocation in Human Health and Disease.” *Current Opinion in Microbiology* 67:102147. doi: 10.1016/j.mib.2022.102147.

Joanito, Ignasius, Pratyaksha Wirapati, Nancy Zhao, Zahid Nawaz, Grace Yeo, Fiona Lee, Christine L. P. Eng, Dominique Camat Macalinao, Merve Kahraman, Harini Srinivasan, Vairavan Lakshmanan, Sara Verbandt, Petros Tsantoulis, Nicole Gunn, Prasanna Nori Venkatesh, Zhong Wee Poh, Rahul Nahar, Hsueh Ling Janice Oh, Jia Min Loo, Shumei Chia, Lih Feng Cheow, Elsie Cheruba, Michael Thomas Wong, Lindsay Kua, Clarinda Chua, Andy Nguyen, Justin Golovan, Anna Gan, Wan-Jun Lim, Yu Amanda Guo, Choon Kong Yap, Brenda Tay, Yourae Hong, Dawn Qingqing Chong, Aik-Yong Chok, Woong-Yang Park, Shuting Han, Mei Huan Chang, Isaac Seow-En, Cherylin Fu, Ronnie Mathew, Ee-Lin Toh, Lewis Z. Hong, Anders Jacobsen Skanderup, Ramanuj DasGupta, Chin-Ann Johnny Ong, Kiat Hon Lim, Emile K. W. Tan, Si-Lin Koo, Wei Qiang Leow, Sabine Tejpar, Shyam Prabhakar, and Iain Beehuat Tan. 2022. “Single-Cell and Bulk

Transcriptome Sequencing Identifies Two Epithelial Tumor Cell States and Refines the Consensus Molecular Classification of Colorectal Cancer." *Nature Genetics* 54(7):963–75. doi: 10.1038/s41588-022-01100-4.

Khan, Shahanavaj. 2015. "Potential Role of *Escherichia Coli* DNA Mismatch Repair Proteins in Colon Cancer." *Critical Reviews in Oncology/Hematology* 96(3):475–82. doi: 10.1016/j.critrevonc.2015.05.002.

Kim, Kwangmin, Ernes John T. Castro, Hongjin Shim, John Vincent G. Advincula, and Young-Wan Kim. 2018. "Differences Regarding the Molecular Features and Gut Microbiota Between Right and Left Colon Cancer." *Annals of Coloproctology* 34(6):280–85. doi: 10.3393/ac.2018.12.17.

Kopylova, Evguenia, Laurent Noé, and Hélène Touzet. 2012. "SortMeRNA: Fast and Accurate Filtering of Ribosomal RNAs in Metatranscriptomic Data." *Bioinformatics* 28(24):3211–17. doi: 10.1093/bioinformatics/bts611.

Lamaudière, Matthew T. F., Ramesh Arasaradnam, Gareth D. Weedall, and Igor Y. Morozov. 2023. "The Colorectal Cancer Gut Environment Regulates Activity of the Microbiome and Promotes the Multidrug Resistant Phenotype of ESKAPE and Other Pathogens." *mSphere* 8(2):e00626-22. doi: 10.1128/msphere.00626-22.

Lee, Ji Ae, Seung-Yeon Yoo, Hyeon Jeong Oh, Seorin Jeong, Nam-Yun Cho, Gyeong Hoon Kang, and Jung Ho Kim. 2021. "Differential Immune Microenvironmental Features of Microsatellite-Unstable Colorectal Cancers According to Fusobacterium Nucleatum Status." *Cancer Immunology, Immunotherapy* 70(1):47–59. doi: 10.1007/s00262-020-02657-x.

Li, Jinming, Yangyang Guo, Jianqiang Liu, Fanying Guo, Lutao Du, Yongzhi Yang, Xinxiang Li, and Yanlei Ma. 2023. "Depicting the Landscape of Gut Microbial-Metabolic Interaction and Microbial-Host Immune Heterogeneity in Deficient and Proficient DNA Mismatch Repair Colorectal Cancers." *Journal for Immunotherapy of Cancer* 11(8):e007420. doi: 10.1136/jitc-2023-007420.

Li, Kai, Haiqing Luo, Lianfang Huang, Hui Luo, and Xiao Zhu. 2020. "Microsatellite Instability: A Review of What the Oncologist Should Know." *Cancer Cell International* 20(1):16. doi: 10.1186/s12935-019-1091-8.

Liu, Chuanfa, Zhiming Li, Jiahong Ding, Hefu Zhen, Mingyan Fang, and Chao Nie. 2021. "Species-Level Analysis of the Human Gut Microbiome Shows Antibiotic Resistance Genes Associated With Colorectal Cancer." *Frontiers in Microbiology* 12. doi: 10.3389/fmicb.2021.765291.

Liu, Zhi, Xuemei Zhang, Haoding Zhang, Hong Zhang, Zhongyuan Yi, Qingqing Zhang, Qisha Liu, and Xingyin Liu. 2023. "Multi-Omics Analysis Reveals Intratumor Microbes as Immunomodulators in Colorectal Cancer." *Microbiology Spectrum* 11(2):e05038-22. doi: 10.1128/spectrum.05038-22.

Lo, Chien-Chi, and Patrick S. G. Chain. 2014. "Rapid Evaluation and Quality Control of next Generation Sequencing Data with FaQCs." *BMC Bioinformatics* 15(1):366. doi: 10.1186/s12859-014-0366-2.

Louis, Petra, Georgina L. Hold, and Harry J. Flint. 2014. "The Gut Microbiota, Bacterial Metabolites and Colorectal Cancer." *Nature Reviews Microbiology* 12(10):661–72. doi: 10.1038/nrmicro3344.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15(12):550. doi: 10.1186/s13059-014-0550-8.

Lunt, Sophia Y., and Matthew G. Vander Heiden. 2011. "Aerobic Glycolysis: Meeting the Metabolic

Requirements of Cell Proliferation." *Annual Review of Cell and Developmental Biology* 27(Volume 27, 2011):441–64. doi: 10.1146/annurev-cellbio-092910-154237.

Malki, Ahmed, Rasha Abu ElRuz, Ishita Gupta, Asma Allouch, Semir Vranic, and Ala-Eddin Al Moustafa. 2021. "Molecular Mechanisms of Colon Cancer Progression and Metastasis: Recent Insights and Advancements." *International Journal of Molecular Sciences* 22(1):130. doi: 10.3390/ijms22010130.

McCallum, Giselle, and Carolina Tropini. 2024. "The Gut Microbiota and Its Biogeography." *Nature Reviews Microbiology* 22(2):105–18. doi: 10.1038/s41579-023-00969-0.

Mizutani, Sayaka, Takuji Yamada, and Shinichi Yachida. 2020. "Significance of the Gut Microbiome in Multistep Colorectal Carcinogenesis." *Cancer Science* 111(3):766–73. doi: 10.1111/cas.14298.

Mo, Zongchao, Peide Huang, Chao Yang, Sihao Xiao, Guojia Zhang, Fei Ling, and Lin Li. 2020. "Meta-Analysis of 16S rRNA Microbial Data Identified Distinctive and Predictive Microbiota Dysbiosis in Colorectal Carcinoma Adjacent Tissue." *mSystems* 5(2):10.1128/msystems.00138-20. doi: 10.1128/msystems.00138-20.

Morgan, Eileen, Melina Arnold, A. Gini, V. Lorenzoni, C. J. Cabasag, Mathieu Laversanne, Jerome Vignat, Jacques Ferlay, Neil Murphy, and Freddie Bray. 2023. "Global Burden of Colorectal Cancer in 2020 and 2040: Incidence and Mortality Estimates from GLOBOCAN." *Gut* 72(2):338–44. doi: 10.1136/gutjnl-2022-327736.

Nearing, Jacob T., Gavin M. Douglas, Molly G. Hayes, Jocelyn MacDonald, Dhwani K. Desai, Nicole Allward, Casey M. A. Jones, Robyn J. Wright, Akhilesh S. Dhanani, André M. Comeau, and Morgan G. I. Langille. 2022. "Microbiome Differential Abundance Methods Produce Different Results across 38 Datasets." *Nature Communications* 13(1):342. doi: 10.1038/s41467-022-28034-z.

Ojala, Teija, Esko Kankuri, and Matti Kankainen. 2023. "Understanding Human Health through Metatranscriptomics." *Trends in Molecular Medicine* 29(5):376–89. doi: 10.1016/j.molmed.2023.02.002.

Okuda, Shujiro, Yoshifumi Shimada, Yosuke Tajima, Kizuki Yuza, Yuki Hirose, Hiroshi Ichikawa, Masayuki Nagahashi, Jun Sakata, Yiwei Ling, Nobuaki Miura, Mika Sugai, Yu Watanabe, Shiho Takeuchi, and Toshifumi Wakai. 2021. "Profiling of Host Genetic Alterations and Intra-Tumor Microbiomes in Colorectal Cancer." *Computational and Structural Biotechnology Journal* 19:3330–38. doi: 10.1016/j.csbj.2021.05.049.

Peschel, Stefanie, Christian L. Müller, Erika von Mutius, Anne-Laure Boulesteix, and Martin Depner. 2021. "NetCoMi: Network Construction and Comparison for Microbiome Data in R." *Briefings in Bioinformatics* 22(4):bbaa290. doi: 10.1093/bib/bbaa290.

Poore, Gregory D., Evguenia Kopylova, Qiyun Zhu, Carolina Carpenter, Serena Fraraccio, Stephen Wandro, Tomasz Kosciolek, Stefan Janssen, Jessica Metcalf, Se Jin Song, Jad Kanbar, Sandrine Miller-Montgomery, Robert Heaton, Rana Mckay, Sandip Pravin Patel, Austin D. Swafford, and Rob Knight. 2020. "Microbiome Analyses of Blood and Tissues Suggest Cancer Diagnostic Approach." *Nature* 579(7800):567–74. doi: 10.1038/s41586-020-2095-1.

Roelands, Jessica, Peter J. K. Kuppen, Eiman I. Ahmed, Raghvendra Mall, Tariq Masoodi, Parul Singh, Gianni Monaco, Christophe Raynaud, Noel F. C. C. de Miranda, Luigi Ferraro, Tatiana C. Carneiro-Lobo, Najeeb Syed, Arun Rawat, Amany Awad, Julie Decock, William Mifsud, Lance D. Miller, Shimaa Sherif, Mahmoud G. Mohamed, Darawan Rinchai, Marc Van den Eynde, Rosalyn

W. Sayaman, Elad Ziv, Francois Bertucci, Mahir Abdulla Petkar, Stephan Lorenz, Lisa Sara Mathew, Kun Wang, Selvasankar Murugesan, Damien Chaussabel, Alexander L. Vahrmeijer, Ena Wang, Anna Ceccarelli, Khalid A. Fakhro, Gabriele Zoppoli, Alberto Ballestrero, Rob A. E. M. Tollenaar, Francesco M. Marincola, Jérôme Galon, Souhaila Al Khodor, Michele Ceccarelli, Wouter Hendrickx, and Davide Bedognetti. 2023. "An Integrated Tumor, Immune and Microbiome Atlas of Colon Cancer." *Nature Medicine* 29(5):1273–86. doi: 10.1038/s41591-023-02324-5.

Salter, Susannah J., Michael J. Cox, Elena M. Turek, Szymon T. Calus, William O. Cookson, Miriam F. Moffatt, Paul Turner, Julian Parkhill, Nicholas J. Loman, and Alan W. Walker. 2014. "Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses." *BMC Biology* 12(1):87. doi: 10.1186/s12915-014-0087-z.

Salvucci, Manuela, Nyree Crawford, Katie Stott, Susan Bullman, Daniel B. Longley, and Jochen H. M. Prehn. 2022. "Patients with Mesenchymal Tumours and High Fusobacteriales Prevalence Have Worse Prognosis in Colorectal Cancer (CRC)." *Gut* 71(8):1600–1612. doi: 10.1136/gutjnl-2021-325193.

Saus, Ester, Susana Iraola-Guzmán, Jesse R. Willis, Anna Brunet-Vega, and Toni Gabaldón. 2019. "Microbiome and Colorectal Cancer: Roles in Carcinogenesis and Clinical Potential." *Molecular Aspects of Medicine* 69:93–106. doi: 10.1016/j.mam.2019.05.001.

Schmitt, Mark, and Florian R. Greten. 2021. "The Inflammatory Pathogenesis of Colorectal Cancer." *Nature Reviews Immunology* 21(10):653–67. doi: 10.1038/s41577-021-00534-x.

Segata, Nicola, Jacques Izard, Levi Waldron, Dirk Gevers, Larisa Miropolsky, Wendy S. Garrett, and Curtis Huttenhower. 2011. "Metagenomic Biomarker Discovery and Explanation." *Genome Biology* 12(6):R60. doi: 10.1186/gb-2011-12-6-r60.

Serpas Higbie, Victoria, Jane Rogers, Hyunsoo Hwang, Wei Qiao, Lianchun Xiao, Arvind Dasari, Kerri Mola-Rudd, Van K. Morris II, Robert A. Wolff, Kanwal Raghav, Ryan Huey, Christine Parseghian, Jason Willis, Scott Kopetz, Michael J. Overman, and Benny Johnson. 2022. "Antibiotic Exposure Does Not Impact Immune Checkpoint Blockade Response in MSI-H/dMMR Metastatic Colorectal Cancer: A Single-Center Experience." *The Oncologist* 27(11):952–57. doi: 10.1093/oncolo/oyac162.

Shen, Weitao, Jiayu Sun, Fen Yao, Kaihuang Lin, Yumeng Yuan, Yexi Chen, Hui Han, Zhiyang Li, Juan Zou, and Xiaoyang Jiao. 2020. "Microbiome in Intestinal Lavage Fluid May Be A Better Indicator in Evaluating The Risk of Developing Colorectal Cancer Compared with Fecal Samples." *Translational Oncology* 13(5):100772. doi: 10.1016/j.tranon.2020.100772.

Sheng, Qin-Song, Kang-Xin He, Jian-Jiong Li, Zi-Feng Zhong, Fei-Xia Wang, Le-Lin Pan, and Jian-Jiang Lin. 2020. "Comparison of Gut Microbiome in Human Colorectal Cancer in Paired Tumor and Adjacent Normal Tissues." *OncoTargets and Therapy* 13:635–46. doi: 10.2147/OTT.S218004.

Simin, Johanna, Romina Fornes, Qing Liu, Renate Slind Olsen, Steven Callens, Lars Engstrand, and Nele Brusselaers. 2020. "Antibiotic Use and Risk of Colorectal Cancer: A Systematic Review and Dose–Response Meta-Analysis." *British Journal of Cancer* 123(12):1825–32. doi: 10.1038/s41416-020-01082-2.

Sugai, Tamotsu, Wataru Habano, Yu-Fei Jiao, Mitsunori Tsukahara, Yuichiro Takeda, Koki Otsuka, and Shin-ichi Nakamura. 2006. "Analysis of Molecular Alterations in Left- and Right-Sided Colorectal Carcinomas Reveals Distinct Pathways of Carcinogenesis." *The Journal of Molecular Diagnostics : JMD* 8(2):193–201. doi: 10.2353/jmoldx.2006.050052.

Tosti, Elena, Ana S. Almeida, Tam T. T. Tran, Mariel Barbachan e Silva, Pilib Ó. Broin, Robert Dubin, Ken Chen, Amanda P. Beck, Andrew S. Mclellan, Eduardo Vilar, Aaron Golden, Paul W. O'Toole, and Winfried Edelmann. 2022. "Loss of MMR and TGFBR2 Increases the Susceptibility to Microbiota-Dependent Inflammation-Associated Colon Cancer." *Cellular and Molecular Gastroenterology and Hepatology* 14(3):693–717. doi: 10.1016/j.jcmgh.2022.05.010.

Vanderbilt, Chad Michael, Ajaratu Keshinro, Chin-Tung Chen, Esther Babady, Michael F. Berger, Ahmet Zehir, Marc Ladanyi, Jaclyn F. Hechtman, Zsofia K. Stadler, Jinru Shia, and Martin R. Weiser. 2020. "Abstract 6095: Unique Tumor Microbiome in Microsatellite Instability High (MSI-H) Colon Carcinoma." *Cancer Research* 80(16_Supplement):6095. doi: 10.1158/1538-7445.AM2020-6095.

Wang, Gang, Xianli He, and Qian Wang. 2023. "Intratumoral Bacteria Are an Important 'Accomplice' in Tumor Development and Metastasis." *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1878(1):188846. doi: 10.1016/j.bbcan.2022.188846.

Wang, Yijia, Chunze Zhang, Shaobin Hou, Xiaojing Wu, Jun Liu, and Xuehua Wan. 2020. "Analyses of Potential Driver and Passenger Bacteria in Human Colorectal Cancer." *Cancer Management and Research* 12:11553–61. doi: 10.2147/CMAR.S275316.

Warren, René L., Douglas J. Freeman, Stephen Pleasance, Peter Watson, Richard A. Moore, Kyla Cochrane, Emma Allen-Vercoe, and Robert A. Holt. 2013. "Co-Occurrence of Anaerobic Bacteria in Colorectal Carcinomas." *Microbiome* 1(1):16. doi: 10.1186/2049-2618-1-16.

Wirth, Ulrich, Debora Garzetti, Lara M. Jochum, Stefanie Spriewald, Florian Kühn, Matthias Ilmer, Serene M. L. Lee, Hanno Niess, Alexandr V. Bazhin, Joachim Andrassy, Jens Werner, Barbara Stecher, and Tobias S. Schiergens. 2020. "Microbiome Analysis from Paired Mucosal and Fecal Samples of a Colorectal Cancer Biobank." *Cancers* 12(12):3702. doi: 10.3390/cancers12123702.

Xu, Xinjian, Ji Lv, Fang Guo, Jing Li, Yitao Jia, Da Jiang, Na Wang, Chao Zhang, Lingyu Kong, Yabin Liu, Yanni Zhang, Jian Lv, and Zhongxin Li. 2020. "Gut Microbiome Influences the Efficacy of PD-1 Antibody Immunotherapy on MSS-Type Colorectal Cancer via Metabolic Pathway." *Frontiers in Microbiology* 11. doi: 10.3389/fmicb.2020.00814.

Xu, Yixin, Jing Zhao, Yu Ma, Jia Liu, Yingying Cui, Yuqing Yuan, Chenxi Xiang, Dongshen Ma, and Hui Liu. 2023. "The Microbiome Types of Colorectal Tissue Are Potentially Associated with the Prognosis of Patients with Colorectal Cancer." *Frontiers in Microbiology* 14. doi: 10.3389/fmicb.2023.1100873.

Yan, Yong-Wei, Qiu-Yue Jiang, Jian-Gong Wang, Ting Zhu, Bin Zou, Qiong-Fen Qiu, and Zhe-Xue Quan. 2018. "Microbial Communities and Diversities in Mudflat Sediments Analyzed Using a Modified Metatranscriptomic Method." *Frontiers in Microbiology* 9. doi: 10.3389/fmicb.2018.00093.

Yin, Han, Zhuangzhuang Miao, Lu Wang, Beibei Su, Chaofan Liu, Yu Jin, Bili Wu, Hu Han, and Xianglin Yuan. 2022. "Fusobacterium Nucleatum Promotes Liver Metastasis in Colorectal Cancer by Regulating the Hepatic Immune Niche and Altering Gut Microbiota." *Aging (Albany NY)* 14(4):1941–58. doi: 10.18632/aging.203914.

Younginger, Brett S., Oleg Mayba, Jens Reeder, Deepti R. Nagarkar, Zora Modrusan, Matthew L. Albert, and Allyson L. Byrd. 2023. "Enrichment of Oral-Derived Bacteria in Inflamed Colorectal Tumors and Distinct Associations of Fusobacterium in the Mesenchymal Subtype." *Cell Reports Medicine* 4(2). doi: 10.1016/j.xcrm.2023.100920.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters." *OMICS: A Journal of*

Integrative Biology 16(5):284–87. doi: 10.1089/omi.2011.0118.

Zackular, Joseph P., Mary A. M. Rogers, Mack T. Ruffin IV, and Patrick D. Schloss. 2014. "The Human Gut Microbiome as a Screening Tool for Colorectal Cancer." *Cancer Prevention Research* 7(11):1112–21. doi: 10.1158/1940-6207.CAPR-14-0129.

Zhang, Jiali, Ji Tao, Ruo-Nan Gao, Zhi-Yuan Wei, Yu-Shan He, Chun-Yan Ren, Qi-Chun Li, Yan-Shan Liu, Ke-Wei Wang, Gong Yang, Chengjia Qian, and Jian-Huan Chen. 2021. "Cytotoxic T-Cell Trafficking Chemokine Profiles Correlate With Defined Mucosal Microbial Communities in Colorectal Cancer." *Frontiers in Immunology* 12. doi: 10.3389/fimmu.2021.715559.

Zhang, Yancong, Kelsey N. Thompson, Tobyn Branck, Yan Yan, Long H. Nguyen, Eric A. Franzosa, and Curtis Huttenhower. 2021. "Metatranscriptomics for the Human Microbiome and Microbial Community Functional Profiling." *Annual Review of Biomedical Data Science* 4(Volume 4, 2021):279–311. doi: 10.1146/annurev-biodatasci-031121-103035.

Zhao, Lan, William C. Cho, and Mark R. Nicolls. 2021. "Colorectal Cancer-Associated Microbiome Patterns and Signatures." *Frontiers in Genetics* 12. doi: 10.3389/fgene.2021.787176.

Zhao, Liuyang, Xiang Zhang, Yunfei Zhou, Kaili Fu, Harry Cheuk-Hay Lau, Tommy Wai-Yiu Chun, Alvin Ho-Kwan Cheung, Olabisi Oluwabukola Coker, Hong Wei, William Ka-Kei Wu, Sunny Hei Wong, Joseph Jao-Yiu Sung, Ka Fai To, and Jun Yu. 2022. "Parvimonas Micra Promotes Colorectal Tumorigenesis and Is Associated with Prognosis of Colorectal Cancer Patients." *Oncogene* 41(36):4200–4210. doi: 10.1038/s41388-022-02395-7.

Zhao, Wensi, Jun Lei, Shaobo Ke, Yuan Chen, Jiping Xiao, Ze Tang, Li Wang, Yiping Ren, Mohammed Alnaggar, Hu Qiu, Wei Shi, Lei Yin, and Yongshun Chen. 2023. "Fecal Microbiota Transplantation plus Tislelizumab and Fruquintinib in Refractory Microsatellite Stable Metastatic Colorectal Cancer: An Open-Label, Single-Arm, Phase II Trial (RENMIN-215)." *eClinicalMedicine* 66. doi: 10.1016/j.eclinm.2023.102315.

Zhao, Zhe, Kailun Fei, Hua Bai, Zhijie Wang, Jianchun Duan, and Jie Wang. 2021. "Metagenome Association Study of the Gut Microbiome Revealed Biomarkers Linked to Chemotherapy Outcomes in Locally Advanced and Advanced Lung Cancer." *Thoracic Cancer* 12(1):66–78. doi: 10.1111/1759-7714.13711.

Zhu, Anchao, Yingying Liu, Zongmin Li, Ying He, Lijing Bai, Youtian Wu, Yuying Zhang, Ying Huang, and Ping Jiang. 2023. "Diagnosis and Functional Prediction of Microbial Markers in Tumor Tissues of Sporadic Colorectal Cancer Patients Associated with the MLH1 Protein Phenotype." *Frontiers in Oncology* 12. doi: 10.3389/fonc.2022.1116780.