

Tumor Tissue-Specific Bacterial Biomarker Panel for Colorectal Cancer: *Bacteroides Massiliensis*, *Alistipes Species*, *Alistipes Onderdonkii*, *Bifidobacterium Pseudocatenulatum*, *Corynebacterium Appendicis*

Rizwana Hasan

Sir Ganga Ram Hospital

Rahul Roy

Sir Ganga Ram Hospital

Debarati Paul

Amity University

Saumitra Rawat

Sir Ganga Ram Hospital

Pravin Nilwe

Thermo Fisher Scientific Inc

Neeraj K Chauhan

Thermo Fisher Scientific Inc

Sudeep Bose

Amity University Amity Institute of Biotechnology

Sangeeta Choudhury (✉ dr.sangeeta.sgrh@gmail.com)

Sir Ganga Ram Hospital <https://orcid.org/0000-0001-5075-778X>

Research Article

Keywords: CRC, microbiome, next-generation sequencing (NGS), 16s rRNA, microbial species, dysbiosis

Posted Date: January 5th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1211053/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Human microbiome studies have shown diversity to exist among different ethnic populations. However, studies pertaining to the microbial composition of CRC among the Indian population have not been well explored. We aimed to decipher the microbial signature in tumor tissues from North Indian CRC patients. Next-generation sequencing of tumor and adjacent tissue derived bacterial 16s rRNA V3-V4 hypervariable regions was performed to investigate the abundance of specific microbes. The expression profile analysis deciphered a decreased diversity among the tumor-associated microbial communities, and at the phyla level, Proteobacteria was differentially expressed in CRC tissues than adjacent normal. Further, DESeq2 normalization identified 4 out of 79 distinct species ($p < 0.005$) only in CRC, *Bacteroides massiliensis*, *Alistipes onderdonkii*, *Bifidobacterium pseudocatenulatum*, and *Corynebacterium appendicis*. Thus, our findings suggest the use of these microbial signatures as putative biomarkers that can distinguish CRC tissues from their adjacent normal, which may shed light on the pathogenesis of CRC.

1. Introduction

Colorectal cancer (CRC) is a challenging health problem and ranks third in terms of incidence and second in terms of mortality; third most common cancer among men and second most common in women. Overall 1.9 million new cases of CRC and 935,000 deaths were estimated to occur worldwide in 2020 (Sung et al. 2021). The incidence rate is steadily increasing in many countries in Eastern Europe, South Eastern and South Central Asia, and South America (Arnold et al. 2020; Arnold et al. 2017). India contributes to the top 5 frequent cases reported worldwide. In countries undergoing a significant transition, with an increasing Human Development Index (HDI), there is a constant rise in the incidence rate of CRC (Fidler et al. 2016). Since CRC symptoms are difficult to identify because early indicators include constipation, diarrhea, changes in stool color, blood in the stool, bleeding from the rectum, excessive gas, abdominal cramps, and abdominal pain (Siegel et al. 2020), therefore, the key strategies are its timely screening, early diagnosis, and prevention.

In this direction over the last decade, microbiome patterns have been explored as biomarkers to diagnose and predict prognosis (Chen et al. 2012; Raskov et al. 2017). Most literature cited are from the Western population (high incidence countries), and only a few handful studies have been reported from the Indian perspective. Gupta et al. identified a panel of potential microbial taxonomy and gene markers that discriminated Indian CRC fecal samples from that of the normal samples (Gupta et al. 2019). There are only a handful of studies that have been conducted in fecal samples of a normal healthy Indian population. Das et al. examined the gut microbiota in 84 healthy individuals from the northern part of India (high altitude of Ladakh). This study demonstrated region-specific differences in α - β -diversity, with Firmicutes dominance followed by Bacteroidetes, Actinobacteria, and Proteobacteria (Das et al. 2018). A study by Dhakan et al. evaluated the gut microbiota in two geographically and socio-demographically distinct cohorts from central (Bhopal) and southern (Kerala) India. The study observed enrichment of genus *Prevotella* among the vegetarian population and *Bacteroides*, *Ruminococcus*, and *Faecalibacterium* among the omnivorous population (Dhakan et al. 2019). In 2018, Dubey AK et al. carried out a study including 14 geographical locations across India [Landscape of Gut Microbiome-Pan India Exploration (LogMPIE)]. The study revealed 36 unique organisms in northern India, 149 in southern, 95 in western, and 62 in eastern Indian population. The most predominant organisms were *Prevotella copri* and *Faecalibacterium prausnitzii* (Dubey et al. 2018).

Although diverse taxonomical microbiome structure has been shown to be associated with different regions of India among healthy individuals, there is a lack of studies pertaining to Indian CRC tissues. Thus, in the present study, the V3-V4 region of 16S rRNA was sequenced using an Ion Torrent™ platform to investigate the microbiome pattern in CRC tumor tissue. This microbial composition derived may provide a clue to CRC diagnosis.

2. Materials And Methods

2.1 Patient selection

CRC patients (histopathologically confirmed) undergoing resection, above 18 years, either sex and were not on antibiotics < 30 days were included in the study. Tissue samples from both tumor area (n=5) and adjacent normal (AN) area (n=5) were aseptically collected and stored at -80°C. Patients who received radiotherapy or chemotherapy before the surgery and/or on immunosuppressive drug treatment were excluded from the study. Ethical consent was obtained, and each patient provided written informed consent.

2.2 DNA extraction

Microbial DNA was extracted from tumor/adjacent normal tissue samples using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA) according to the modified protocol from ~25mg of tissue. In brief, each sample was incubated in 180µl lysozyme (10 mg/ml, Sigma-Aldrich, USA) for 40 min at 37°C, and further incubated at 56°C for 16 hours with 20 µl proteinase K, until the tissue was completely lysed. Buffer AL was added to this mixture and incubated for 10 minutes at 70°C; then, DNA has been eluted according to the manufacturer's instructions. Microbial DNA quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). DNA samples were stored at -20°C until further processing.

2.3 16S rRNA Amplicon Library Generation

16S rRNA hypervariable regions were amplified from extracted DNA by using primer pair Probio_Uni and Probio_Rev (Milani et al. 2013), which targets the V3 - V4 region of 16S rRNA gene sequence [520F/802R] (Ventura et al. 2009). The 5'ends of the forward primers were fused with the A-Adaptor (ThermoFisher Scientific), 10 base pair barcode key sequence plus barcode adaptor, whereas the reverse primers were fused with the truncated P1-adaptor sequence (trP1), respectively. The complete list of the primers which is used in this study is mentioned in Table 1.

Table 1
Primer Sequence

Primer Name	Adapter sequence	Key	Barcode	Barcode Adapter	Primer Sequence (5'-3')
Probio_Uni	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TTACAACCTC	GAT	CCTACGGGRSGCAGCAG
Probio_Rev	CCTCTCTATGGGCAGTCGGTGAT				ATTACCGCGGCTGCT
520F	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TTACAACCTC	GAT	AYTGGGYDTAAAGNG
802R	CCTCTCTATGGGCAGTCGGTGAT				TACNVGGGTATCTAATCC

The primers were diluted in a low Tris-EDTA solution (1:10). The sample prepared fusion amplicons with 5 ng of bacterial DNA from each sample. In brief, Amplitaq Gold 360 MM (ThermoFisher Scientific) containing premixed and premeasured components and 5 pmol of fusion primers were used for 25 ul PCR amplification reaction. The PCR conditions were maintained at 5 min at 94°C, 35 cycles of 30s at 94°C, 30s at 55°C and 90s at 72°C, followed by 10 min at 72°C. Amplification was carried out by using a 9700 Thermocycler (Applied Biosystems, USA). The PCR amplicon was purified using AMPure XP reagent (Beckman Coulter, USA), and the concentration was determined with Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). The respective size distribution of the amplicon was verified with Agilent 2100 Bioanalyzer using a high-sensitivity DNA kit (Agilent, Santa Clara, CA). Amplicon libraries were diluted to 100 pM, and an equimolar pool was prepared for clonal amplification.

2.4 Template preparation and sequencing

Template preparation for libraries on Ion Sphere™ was performed as per OneTouch 2 protocols and reagents (Thermo Fisher Scientific, USA). Briefly, library amplicons were clonally amplified onto ion sphere particles (ISPs) through emulsion PCR and then enriched for template-positive ISPs. Ion S5 system emulsion PCR reactions utilized the Ion 520TM (Thermo Fisher Scientific, USA). Following recovery, enrichment was completed by selectively binding the ISPs containing amplified library amplicons to streptavidin-coated magnetic beads, removing empty ISPs through washing steps, and denaturing the template strands to allow the collection of template-positive ISPs. For all reactions, these steps were accomplished using the Thermo Fisher Scientific's ES module of Ion OneTouch 2.

The templated ISPs were loaded on Ion 520 chip kit using the manufacturer's protocol (Thermo Fisher Scientific, USA). Sequencing was performed with the Ion 520 OT2 (Thermo Fisher Scientific, USA) using the 200 bp chemistry with 500 flow (125cycle) for the V3 region and 400 bp chemistry with 800 flow (200 cycles) for the V4 region run format.

2.5 Ion Reporter platform—metagenomics workflow

Combining the two primer pools (V3-V4) resulted in broad-range sequence-based identification of bacteria from mixed samples. Ion Reporter Software enabled rapid identification (at genus or species level) of microbes present in samples, using both curated Greengenes and premium curated MicroSEQ ID 16S rRNA reference databases. The Ion Reporter metagenomics workflow provided primer information, classification information, and mapping information.

2.6 16S rRNA sequence data analysis

Ion Reporter Software 5.10, 16S Metagenomics workflow version 1.0. (Thermo Fisher Scientific, USA) and QIIME2 (<https://qiime2.org>) were used for DNA sequences processing. Analysis of the sample data was performed utilizing the Ion Reporter Software, which is based on the QIIME's open-source bioinformatics platform to perform diversity analyses (α and β diversity) and visualizations in the form of a Krona diagram. Unaligned binary data files (Binary Alignment Map) created by the Ion Torrent PGM were transferred to an in-house Ion Reporter server (Thermo Fisher Scientific, USA) and analyzed using default settings. Each respective sequence was assigned based on Ribosomal Database Project (RDP) taxonomy. Consensus reads were normalized by converting OTUs for each sample to a percentage of the reads for a given sample. Taxonomy was assigned with a minimum alignment threshold of 97%. The identity score of sequence alignment for genus and species was set at 99% or higher.

2.7 Alpha diversity analysis

Chao1 index determined the species richness in the samples based on the number of OTUs. The Shannon and Simpson index was utilized to reflect community diversity, including species richness and species evenness. Relative abundance of phyla and α -diversity metrics were analyzed using the Paleontological Statistics Software Package (PAST). The univariate frequency distribution was conducted to verify data normalization. Boxplots showing the Shannon, Simpson, and Chao1 diversity indices, the upper whisker starts from 75th percentile to the highest value within the 1.5 x interquartile range (IQR), the lower whisker extends from 25th percentile to the lowest value within 1.5 x IQR of the hinge. Any data points beyond the end of the whiskers were outliers. The ns sign indicates no statistical significance in the pairwise comparison between CRC and AN; p values were obtained using the One-sample t and Wilcoxon test ($p=0.0625$, GraphPad Prism 9).

2.8 Beta diversity analysis

Principal Coordinates Analysis (PCoA) was performed to explore the microbial species similarities and/or dissimilarities between tumor and AN samples. In the 3D PCoA plot, the close algorithm (distance matrix) between samples represented similar microbial species.

2.9 MA Plot analysis

MA plot represents the scattered plot of \log_2 fold-change (M on the y-axis) and means (A on the x-axis) between two samples. The base Mean was the average of normalized count values, dividing by size factors, taken for all samples. The

log2Fold Change is the size estimate.

2.10 OTU Venn Chart

A Venn diagram visually displays the number of common/unique OTUs in the tissue samples. The microbiomes of different samples are being obtained by combining with the OTUs representing the species. Based on the OTU abundance, OTU of each group was listed, and the Venn diagram was drawn using Bioinformatics & Evolutionary Genomics Venn Diagram online tool.

2.11 Heat Map and Hierarchical clustering

HeatMap differential abundance was constructed using DESeq2 data with significant OTUs ($p < 0.05$). The gathered data from DESeq2 were analyzed using NG-CHM GUI 2.20.2 BUILDER for building a Heat Map where the cluster analysis based on each OTU table to create dendrograms using Ward's clustering and Euclidean distance was carried out. The reason for this analysis was to recognize similarities in microorganisms within particular variables (Ryan et al. 2019). The joining of the Euclidean distance as the distance measure was adopted and Ward's method as a linkage rule. The color gradient key displays a linear scale of relative abundance.

2.12 Differential expression analysis

DESeq2 software was used to normalize by calculating the geometric mean across the 10 samples from 5 patients and determining the differential expression analysis using the fold-change expression of the bacterial genes between CRC and AN. (DESeq2 normalised data ($p < 0.05$) is shown in Supplementary table SI-3)

3. Results

3.1 Clinical findings

In this pilot study, demographic data such as age, gender, tumor site and size, tumor stage, and tumor pathology are described in detail in Table 2. Clinical symptoms of anorexia, nausea, vomiting, constipation/diarrhea, abdominal pain, weight loss, stool occult blood findings were noted. The male: female ratio was 4:1; tumor staging was determined according to the AJCC TNM system in CRC (Amin et al. 2017). Most of the patients belonged to stage II or higher, and 4 out of the 5 were obese.

Table 2
Demographic characteristics of CRC patients

Patients	Age/Sex	Stage	Stage grouping	Localisation	Histopathological findings	Diet	BMI	Obesity
1	48/m	IIb	pT3pN0pMx	Rectum	Moderately differentiated adenocarcinoma	Non-V	32.5	Obese
2	82/m	IIb	pT3pN0pMx	Colon	Moderately differentiated adenocarcinoma	Non-V	32.6	Obese
3	83/m	IV	pT4apN1pbMx	Colon	poorly differentiated adenocarcinoma	Veg	20.2	Normal
4	32/m	IIIb	pT3pN1pMx	Caecum	Moderately differentiated adenocarcinoma,	Non-V	26.8	overweight
5	52/f	IIIc	pT3pN2pMx	Rectum	Moderately differentiated adenocarcinoma,	Non-V	34.3	Obese

Patients were found to be anemic; mean \pm SD hemoglobin was lower than normal (8.7 ± 2.2 g/dl; normal range: 13-17 g/dL). The level of CEA was nearly twice (6.3 ± 3.7 ng/ml; normal range: 0- 3ng/ml), and 3/5 patients showed low levels of protein and albumin (6.2 ± 0.8 gm/dL and 3.1 ± 1.0 gm/dL respectively) as shown in Supplementary table SI-1.

3.2 Taxonomical composition in CRC and adjacent normal tissue

3.2.1 Ion Torrent Sequencing and pre-processing of reads

A total of 7.6 million raw reads were generated using the Ion Torrent™ platform. The sequencing indexed on an Ion 520 Chip provided 1.47 GB of data with a mode reads of 240 bp (Supplementary figure SI-1). Chips showed 93% loading with 100% enrichment. Pre-processing workflow consisting of a high-quality filtering approach resulted in 89% of the final library (Supplementary table SI-2).

3.2.2 Alpha Diversity

Shannon and Simpson's indexes were applied to determine the diversity and species evenness; Chao1 was employed to estimate the abundance of community richness in CRC and AN tissue samples. As shown in figure 1, the Shannon index and Simpson index showed fewer species-level diversity among CRC tissue samples than AN tissues. Chao1 index showed a similar observation of lower species richness in CRC than that of AN.

3.2.3 Beta Diversity

Beta diversity two-dimensional coordinate graph was used to measure the similarity or dissimilarity between the microbial Family of CRC and AN samples. Fig.1(d), shows close coordinates of 4 CRC samples (CRC P1,2,3,5) except for CRC-P4, whereas all AN samples were observed to be highly diverse in their representation of the microbial family. This PCoA graph showed a significant distinction in bacterial family composition between the CRC and AN tissue samples of the 5 patients.

Figure 1 A. Indices of microbial diversity. Box plot showing species-level microbial diversity of five different patient samples of CRC (red) and AN (blue). The alpha diversity was measured by diversity indices, including the Shannon index (a), Chao1 (b), and Simpson index (c) Alpha-diversity distances calculated using phylotype relative abundance measurements between AN and CRC tissues demonstrate that the microbial richness of AN tissues is higher than their CRC tissue, while the diversity has no statistical significance (ns) between the two groups. The indices in investigated

individuals were computed using the PAST software. **B.** Represents β diversity Principal-coordinate analysis (PCoA), Euclidean plot at Family level, in which each dot represents a sample, and each color represents a group: red for group CRC and blue for group AN. The plot describes the similarities and dissimilarities between samples of 5 patients (P1-P5). PC1 is the principal coordinate component causing the most considerable difference in samples, with an explanatory value of 64.16%, and PC2 and PC3, 22.82% and 6.65%, respectively.

3.3 Microbial Abundance and Distribution

Analysis of the V3-V4 region of filtered reads showed 17 phyla, 28 classes, 55 orders, 149 families, 222 genus, and 361 species (Supplementary figure SI-3). The 5 most abundant phyla were Firmicutes (42.9%), Proteobacteria (38.4%), Actinobacteria (7.8%), Bacteroidetes (8.3%), and Fusobacteria (1.4%) comprised of 98.9% (Fig. 2). Proteobacteria: Firmicutes ratio in CRC was 1:0.6 and in AN was 1: 1.1 (Supplementary figure SI-3). Detailed depictions using Krona graphs representing phylum, class, order, family, genus, and species are available in Supplementary figure SI-5. As per the metagenomics analysis, the percentage of each phyla abundance in CRC and AN is depicted in Supplementary figure SI-4. The taxonomic compositions of the remaining 1.02% were majorly represented by Acidobacteria, Chloroflexi, Deinococcus-Thermus, Elusimicrobia, Lentisphaerae, Nitrospinae, Nitrospirae, Spirochaetes, Synergistetes, Tenericutes, and Verrucomicrobia as shown in Supplementary figure SI-4. Cyanobacteria (0.14%) was abundant in CRC and Synergistetes (0.3%) in AN among these rare phyla.

3.4 Differentially enriched Operational Taxonomic Units in tissues of CRC and adjacent normal

The DESeq2 data based on the MA plot (Fig. 3) visualized illustrates the statistically differential gene expression pattern at species level between CRC and AN on a logarithmic scale of base 2 wherein each data point represents a bacterial species. The central (light shaded) clustering indicates a similar expression pattern across species between CRC and AN whereas expression of distinct species) or less similar species as dark/bright, shaded spots.

In order to find the relationship between and among CRC and AN at the species level, the data represented as Venn diagram (Fig. 4) depicts both common and unique OTUs at the species level. 210 species were common in both CRC and AN; 79 species were present only in CRC and not in AN, while 72 species were unique to AN. With a cut-off of >100 OTU counts, 4 species were significantly abundant (<0.05 from DEseq data) in CRC - *Bacteroides massiliensis*, *Alistipes onderdonkii*, *Bifidobacterium pseudocatenulatum* and *Corynebacterium appendicis*. Distinct abundance signature in AN showed significant presence of *Veillonella atypica*, from the Firmicutes phyla and *Paracoccus sphaerophysae*, *Campylobacter gracilis* and *Pasteurella pneumotropica* belonging from phyla proteobacteria (Table 3).

Table 3
Abundant species in CRC and AN

Genus/species in CRC	P value	Genus/species in AN	P value
<i>Corynebacterium appendicis</i>	1.28E-05	<i>Pasteurella pneumotropica</i>	0.00027
<i>Bacteroides massiliensis</i>	0.00024	<i>Campylobacter gracilis</i>	0.00052
<i>Bifidobacterium pseudocatenulatum</i>	0.00092	<i>Paracoccus sphaerophysae</i>	0.00032
<i>Alistipes onderdonkii</i>	0.00164	<i>Veillonella atypica</i>	0.00144

Further, the data comprising of significant ($p < 0.05$) OTU were analyzed using DESeq2 after normalized logarithmic transformation and represented by Heatmap (Fig. 5). The analysis showed differential species abundance in CRC and AN wherein enrichment of *Rumnicoccus callidus*, *Blautia producta*, *Micrococcus lylae*, *Butyrivibrio crossotus*, and

Corynebacterium appendicis ($p < 0.05$) was found in CRC, and *Streptococcus intermedius*, *Bacteriodes ovatus*, and *Bacteroides intestinalis* ($p < 0.05$) were found in AN.

4. Discussion

The gut microbiome has emerged as a central player in CRC pathogenesis, with multiple effects on the transformation process, tumor progression, and response to anticancer therapies (Helmink et al. 2019; Matson et al. 2018; Wong et al. 2017). Abundant taxa in CRC tumor tissues are the protagonists of tumor development by deregulating immune homeostasis, producing exoenzymes or toxins, and influencing the defense mechanism against pathogens. Thereafter, damaging the host tissue to spread deeper into the organ and/or body.

The majority of the literature shows microbiome studies conducted in fecal samples. This can pose a challenge as fecal matter does not capture all the gut microbes and, in particular, mucosal adherent microbes (Eckburg et al. 2005; Rivadeneyra et al. 2010). Pathogenic bacteria with invasive properties that can proliferate within the anaerobic tumor and its microenvironment most likely represent the tumor-associated signature (Zmora et al. 2018). Thus, studying the microbiome of tumor-tissue may help understand the progression of the disease. The present study attempted to decipher the CRC tumor-associated microbial signature.

In our study, the relative abundance of *Prevotella copri* and *Faecalibacterium prausnitzii* was found to be among the top 10 abundant species in both CRC and adjacent normal tissues. This correlates with studies by Dhakan et al. and Dubey et al. One of the likely reason could be the omnivorous diet of Indians that are rich in carbohydrates (Dhakan et al. 2019; Dubey et al. 2018).

CRC microbial signature

Using metagenomics (16s rRNA NGS), we identified the microbial composition and diversity in tumor/adjacent normal tissues pairs and uncovered microbial signatures inextricably linked to CRC. Our data observed an abundance of Bacteroidetes and Actinobacteria phylum in both CRC and AN. When the differences at each phylum level between the two groups were further studied, Proteobacteria were more abundant than Firmicutes in CRC. Literature has reported a higher abundance of Proteobacteria and Fusobacteria in tumor tissues and a lower abundance of Bacteroidetes, Firmicutes, and Actinobacteria (Kim et al. 2018; Ringel et al. 2015; Shah et al. 2018; Wang et al. 2020). Species of Proteobacteria phylum are involved in dysbiosis resulting in induction of beta-oxidation (Hughes et al. 2017). Expression of nitric oxide (NO_2) is inhibited by the activity of peroxisome proliferator-activated receptor-gamma (PPAR- γ) which in turn is activated by metabolite, butyrate, majorly produced by species belonging to Proteobacteria phyla (Tjalsma et al. 2012; Tomkovich et al. 2017).

Eighty percent (4 out of 5) of our CRC patients were obese. Proteobacteria have been shown to play a role in obesity. It induces the production of pro-inflammatory molecules such as lipopolysaccharides that aid in increasing host fat storage (Rizzatti et al. 2017). Obesity is one of the etiological factors associated with increased risk for CRC development.

A higher species abundance was observed among the CRC tissue samples than AN tissues. The richness and evenness of species (α -diversity) were similar in tumor tissues and their adjacent area. Further, CRC and AN tissues showed distinct bacterial family composition (β -diversity). He et al., found similar results of decreasing microbial diversity and abundance in CRC than that normal individual. Our 16s Sequencing data analysis showed distinct 79 species in CRC and 72 in AN (He et al. 2021). Based on cut-off (>100 OUT counts), 4 species significantly abundant (<0.005) were *Bacteroides massiliensis*, *Alistipes onderdonkii*, *Bifidobacterium pseudocatenulatum*, *Corynebacterium appendicis*. *Bacteroides massiliensis* was reported to be highly abundant in CRC patients (Ozawa et al. 1979) and Ulcerative colitis patients (Wu et al. 2018). Recently, Parker et al. reviewed the emerging role of different species of *Alistipes* genus. *Alistipes* species have contributed to tumor pathogenesis via the IL-6/STAT3 pathway (Moschen et al. 2016). The study by Goetz et al. found that the protein

Lipocalin 2 (LCN 2), produced by *Alistipes species*, binds to high-affinity iron-chelating compounds (siderophores), which causes reduction of iron availability (Goetz et al. 2002; Parker et al. 2020). In the 5 CRC patients in our study, haemoglobin level were extremely low. One of the reason could be significant high abundance of *Alistipes species*.

Adjacent normal microbial signature

A high abundance of species from Firmicutes and Protobacteria were observed in the AN region of the 5 CRC patients (*Bacillus sp*, *Veillonella atypica*, *Paracoccus sphaerophysae*, *Campylobacter gracilis*, *Pasteurella pneumotropica*, *Oribacterium sp*). *Bacillus sp*. belonging to Firmicute phyla are known to be involved in gut homeostasis by promoting the growth of other beneficial microbes and suppressing pathogen/pathogen-induced inflammatory response of intestinal mucosa (Khochamit et al. 2015; Nagal et al. 1996; Ozawa et al. 1979). Similarly, *Veillonella atypica* ferments lactate to propionate and acetate (Mashima and Nakazawa 2014) and stimulates the growth of a wide range of organisms through metabolic complementation (Zhou et al. 2015). *Oribacterium sp* belonging to Firmicutes phyla plays a significant role in repairing damaged mucous tissue, as shown by Zelante and team. The authors suggest that *Oribacterium species* binds to the host aryl-hydrocarbon receptor to maintain intestinal homeostasis via the production of short-chain fatty acids (Zelante et al. 2013).

5. Conclusion

The mortality rate of colorectal cancer is significantly high. The major limitation is its late identification, advanced stage of the disease with late diagnosis, and resistance to treatment. With evolving metagenomics technology, the genetic mutations of cancer and the expression pattern of the microbiome can play a critical role in CRC diagnosis and prediction. This preliminary study deciphered a panel of bacteria specific to CRC tumor tissue, which can enable the use of microbial biomarkers for early diagnosis of CRC. The study needs further validation in stratified groups that includes normal mucosa and adenomas.

Declarations

Funding

This study was supported by the grant provided by Research and Development programme (RDP grant number: 4.9.36 – 011), Sir Ganga Ram Hospital (SGRH), India.

Conflict of Interest

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

Ethics approval

This study was reviewed and approved by Sir Ganga Ram Hospital Ethics Committee (Ethics approval code: EC/06/16/1000). The patients/participants provided their informed consent to participate in this study.

Author Contributions

All authors contributed towards writing the manuscript and have given approval to the final version of the manuscript. RH performed collection of human samples/isolation/NGS assays and framing the manuscript. RR performed the metagenomics data analysis and prepared the figures. SR enrolled the patients as per inclusion criteria, provided the clinical samples and demographics of the patients. DB contributed in manuscript proof reading. PN & NKS assisted in

experimental standardization and NGS data acquisition and cleaning. SB guided, framed the manuscript, inferred the data and edited the manuscript. SC conceptualized, guided, framed and edited the manuscript.

Acknowledgments

The authors would like to thank Parul Chugh (SGRH) for statistical help and the co-workers at Department of Research, SGRH.

References

1. In. <http://rdp.cme.msu.edu>
2. Amin MB, Greene FL, Edge SB, et al. (2017) The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 67(2):93-99 doi:10.3322/caac.21388
3. Arnold M, Abnet CC, Neale RE, et al. (2020) Global Burden of 5 Major Types of Gastrointestinal Cancer. *Gastroenterology* 159(1):335-349 e15 doi:10.1053/j.gastro.2020.02.068
4. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F (2017) Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 66(4):683-691 doi:10.1136/gutjnl-2015-310912
5. Chen W, Liu F, Ling Z, Tong X, Xiang C (2012) Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One* 7(6):e39743 doi:10.1371/journal.pone.0039743
6. Das B, Ghosh TS, Kedia S, et al. (2018) Analysis of the Gut Microbiome of Rural and Urban Healthy Indians Living in Sea Level and High Altitude Areas. *Sci Rep* 8(1):10104 doi:10.1038/s41598-018-28550-3
7. Dhakan DB, Maji A, Sharma AK, et al. (2019) The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *Gigascience* 8(3) doi:10.1093/gigascience/giz004
8. Dubey AK, Uppadhyaya N, Nilawe P, et al. (2018) LogMPIE, pan-India profiling of the human gut microbiome using 16S rRNA sequencing. *Sci Data* 5:180232 doi:10.1038/sdata.2018.232
9. Eckburg PB, Bik EM, Bernstein CN, et al. (2005) Diversity of the human intestinal microbial flora. *Science* 308(5728):1635-8 doi:10.1126/science.1110591
10. Fidler MM, Soerjomataram I, Bray F (2016) A global view on cancer incidence and national levels of the human development index. *Int J Cancer* 139(11):2436-46 doi:10.1002/ijc.30382
11. Goetz DH, Holmes MA, Borregaard N, Bluhm ME, Raymond KN, Strong RK (2002) The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Mol Cell* 10(5):1033-43 doi:10.1016/s1097-2765(02)00708-6
12. Gupta A, Dhakan DB, Maji A, et al. (2019) Association of *Flavonifractor plautii*, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. *mSystems* 4(6) doi:10.1128/mSystems.00438-19
13. He T, Cheng X, Xing C (2021) The gut microbial diversity of colon cancer patients and the clinical significance. *Bioengineered* 12(1):7046-7060 doi:10.1080/21655979.2021.1972077
14. Helmkamp BA, Khan MAW, Hermann A, Gopalakrishnan V, Wargo JA (2019) The microbiome, cancer, and cancer therapy. *Nat Med* 25(3):377-388 doi:10.1038/s41591-019-0377-7
15. Hughes ER, Winter MG, Duerkop BA, et al. (2017) Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host Microbe* 21(2):208-219 doi:10.1016/j.chom.2017.01.005
16. Khochamit N, Siripornadulsil S, Sukon P, Siripornadulsil W (2015) Antibacterial activity and genotypic-phenotypic characteristics of bacteriocin-producing *Bacillus subtilis* KKU213: potential as a probiotic strain. *Microbiol Res* 170:36-

17. Kim K, Castro EJT, Shim H, Advincula JVG, Kim YW (2018) Differences Regarding the Molecular Features and Gut Microbiota Between Right and Left Colon Cancer. *Ann Coloproctol* 34(6):280-285 doi:10.3393/ac.2018.12.17
18. Mashima I, Nakazawa F (2014) The influence of oral *Veillonella species* on biofilms formed by *Streptococcus species*. *Anaerobe* 28:54-61 doi:10.1016/j.anaerobe.2014.05.003
19. Matson V, Fessler J, Bao R, et al. (2018) The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* 359(6371):104-108 doi:10.1126/science.aao3290
20. Milani C, Hevia A, Feroni E, et al. (2013) Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* 8(7):e68739 doi:10.1371/journal.pone.0068739
21. Moschen AR, Gerner RR, Wang J, et al. (2016) Lipocalin 2 Protects from Inflammation and Tumorigenesis Associated with Gut Microbiota Alterations. *Cell Host Microbe* 19(4):455-69 doi:10.1016/j.chom.2016.03.007
22. Nagai S, Okimura K, Kaizawa N, Ohki K, Kanatomo S (1996) Study on surfactin, a cyclic depsipeptide. II. Synthesis of surfactin B2 produced by *Bacillus natto* KMD 2311. *Chem Pharm Bull (Tokyo)* 44(1):5-10 doi:10.1248/cpb.44.5
23. Ozawa K, Yagu-Uchi K, Yamanaka K, Yamashita Y, Ueba K, Miwatani T (1979) Antagonistic effects of *Bacillus natto* and *Streptococcus faecalis* on growth of *Candida albicans*. *Microbiol Immunol* 23(12):1147-56 doi:10.1111/j.1348-0421.1979.tb00547.x
24. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A (2020) The Genus *Alistipes*: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. *Front Immunol* 11:906 doi:10.3389/fimmu.2020.00906
25. Raskov H, Burcharth J, Pommergaard HC (2017) Linking Gut Microbiota to Colorectal Cancer. *J Cancer* 8(17):3378-3395 doi:10.7150/jca.20497
26. Ringel Y, Maharshak N, Ringel-Kulka T, Wolber EA, Sartor RB, Carroll IM (2015) High throughput sequencing reveals distinct microbial populations within the mucosal and luminal niches in healthy individuals. *Gut Microbes* 6(3):173-81 doi:10.1080/19490976.2015.1044711
27. Rivadeneyra MA, Martin-Algarra A, Sanchez-Roman M, Sanchez-Navas A, Martin-Ramos JD (2010) Amorphous Calcium phosphate precursors for Ca-carbonate biominerals mediated by *Chromohalobacter marismortui*. *ISME J* 4(7):922-32 doi:10.1038/ismej.2010.17
28. Rizzatti G, Lopetuso LR, Gibiino G, Binda C, Gasbarrini A (2017) Proteobacteria: A Common Factor in Human Diseases. *Biomed Res Int* 2017:9351507 doi:10.1155/2017/9351507
29. Ryan MC, Stucky M, Wakefield C, et al. (2019) Interactive Clustered Heat Map Builder: An easy web-based tool for creating sophisticated clustered heat maps. *F1000Res* 8 doi:10.12688/f1000research.20590.2
30. Shah MS, DeSantis T, Yamal JM, et al. (2018) Re-purposing 16S rRNA gene sequence data from within case paired tumor biopsy and tumor-adjacent biopsy or fecal samples to identify microbial markers for colorectal cancer. *PLoS One* 13(11):e0207002 doi:10.1371/journal.pone.0207002
31. Siegel RL, Miller KD, Goding Sauer A, et al. (2020) Colorectal cancer statistics, 2020. *CA Cancer J Clin* 70(3):145-164 doi:10.3322/caac.21601
32. Sung H, Ferlay J, Siegel RL, et al. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 71(3):209-249 doi:10.3322/caac.21660
33. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE (2012) A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol* 10(8):575-82 doi:10.1038/nrmicro2819
34. Tomkovich S, Yang Y, Winglee K, et al. (2017) Locoregional Effects of Microbiota in a Preclinical Model of Colon Carcinogenesis. *Cancer Res* 77(10):2620-2632 doi:10.1158/0008-5472.CAN-16-3472
35. Ventura M, O'Flaherty S, Claesson MJ, et al. (2009) Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat Rev Microbiol* 7(1):61-71 doi:10.1038/nrmicro2047

36. Wang Q, Ye J, Fang D, et al. (2020) Multi-omic profiling reveals associations between the gut mucosal microbiome, the metabolome, and host DNA methylation associated gene expression in patients with colorectal cancer. *BMC Microbiol* 20(Suppl 1):83 doi:10.1186/s12866-020-01762-2
37. Wong SH, Zhao L, Zhang X, et al. (2017) Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice. *Gastroenterology* 153(6):1621-1633 e6 doi:10.1053/j.gastro.2017.08.022
38. Wu M, Wu Y, Li J, Bao Y, Guo Y, Yang W (2018) The Dynamic Changes of Gut Microbiota in Muc2 Deficient Mice. *Int J Mol Sci* 19(9) doi:10.3390/ijms19092809
39. Zelante T, Iannitti RG, Cunha C, et al. (2013) Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* 39(2):372-85 doi:10.1016/j.immuni.2013.08.003
40. Zhou P, Liu J, Merritt J, Qi F (2015) A YadA-like autotransporter, Hag1 in *Veillonella atypica* is a multivalent hemagglutinin involved in adherence to oral *streptococci*, *Porphyromonas gingivalis*, and human oral buccal cells. *Mol Oral Microbiol* 30(4):269-279 doi:10.1111/omi.12091
41. Zmora N, Zilberman-Schapira G, Suez J, et al. (2018) Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. *Cell* 174(6):1388-1405 e21 doi:10.1016/j.cell.2018.08.041

Figures

Figure 1

A. Indices of microbial diversity. Box plot showing species-level microbial diversity of five different patient samples of CRC (red) and AN (blue). The alpha diversity was measured by diversity indices, including the Shannon index (a), Chao1 (b), and Simpson index (c) Alpha-diversity distances calculated using phylotype relative abundance measurements between AN and CRC tissues demonstrate that the microbial richness of AN tissues is higher than their CRC tissue, while the diversity has no statistical significance (ns) between the two groups. The indices in investigated individuals were computed using the PAST software.

B. Represents β diversity Principal-coordinate analysis (PCoA), Euclidean plot at Family level, in which each dot represents a sample, and each color represents a group: red for group CRC and blue for group AN. The plot describes the similarities and dissimilarities between samples of 5 patients (P1-P5). PC1 is the principal coordinate component causing the most considerable difference in samples, with an explanatory value of 64.16%, and PC2 and PC3, 22.82% and 6.65%, respectively.

Figure 2

Relative abundance of major 5 bacterial phyla in CRC (1to5) and adjacent normal (1to5), respectively

Figure 3

MA-plot from base means and log2fold changes. The plot visualizes the intensity-dependent ratio of DESeq2 data. The brighter the spot, the more likely an observed difference between sample CRC and adjacent normal

Figure 4

A Venn diagram, representing the shared and unique OTUs of the microbiome at species level between CRC and adjacent normal

Figure 5

A heat map shows taxonomy assignment for each OTU from sample characterized by amplicon sequencing of 16S rRNA gene (V3-V4 region) within CRC and AN. The OTU heat map displays OTU counts per sample, where the counts are coloured based on the contribution of each OTU to the total OTU count present in the sample (green: contributes a low percentage of OTUs to sample; red: contributes a high percentage of OTUs). Vertical clustering (hierarchical clustering) indicates the similarity in the richness of different species among different samples. The closer the euclidean distance between two species, the shorter the branch length, indicating a more significant similarity in richness between the two species. The heat map was built using NG-CHM BUILDER. Scale on the right indicates the breakpoints associated with the colours

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SUPPLEMENTARYINFORMATION.docx](#)