

Tumour Microbiome-Based Subtypes of Colorectal Cancer Correlate with Clinical Variables

Barbora Zwinsová

Masarykova Univerzita

Vyacheslav Petrov

Masarykova Univerzita

Martina Hrivňáková

Masarykuv Onkologický Ústav

Stanislav Smatana

Masarykova Univerzita

Lenka Micenková

Masarykova Univerzita

Natálie Kazdová

Masarykova Univerzita

Vlad Popovici

Masarykova Univerzita

Roman Hrstka

Masarykuv Onkologický Ústav

Roman Šefr

Masarykuv Onkologický Ústav

Beatrix Bencsiková

Masarykuv Onkologický Ústav

Lenka Zdražilová-Dubská

Masarykuv Onkologický Ústav

Veronika Brychtová

Masarykuv Onkologický Ústav

Rudolf Nenutil

Masarykuv Onkologický Ústav

Petra Vídeňská

Masarykova Univerzita

Eva Budinská (✉ eva.budinska@mou.cz)

Masaryk Memorial Cancer Institute <https://orcid.org/0000-0002-9004-9187>

Research

Keywords: Colorectal cancer, 16S rRNA gene, tumour microbiome, microbial subtypes

Posted Date: August 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-64049/v1>

Abstract

Background

Long-term dysbiosis of the gut microbiome has a significant impact on the development, progression and the aggressiveness of colorectal cancer (CRC) and may explain part of the observed heterogeneity of the disease from phenotypic, prognostic and response to treatment perspectives. Although the shifts in gut microbiome in the normal-adenoma-carcinoma sequence have been described, the landscape of microbiome within CRC and its associations with clinical variables remain under-explored.

Results

We performed 16S rRNA gene sequencing of paired tumour tissue, adjacent visually-normal mucosa and stool swabs of N=186 patients with stage 0-IV CRC to describe the tumour microbiome and its association with clinical variable and to derive tumour microbial subtypes.

We identified new genera never previously associated with CRC tumour mucosa (*Flavonifractor*, *Haemophilus*, *Howardella*, *Pseudomonas*, *Sutterella*, *Treponema* 2) or CRC (*Actinobacillus*, *Aggregatibacter*, *Bergeyella*, *Phocaeiola*, *Defluviitaleaceae UCG-011*, *Massilia*, *Tyzzereella* 4). The bacteria residing on tumour-mucosa were dominated by genera belonging to (potential) oral pathogens. Based on tumour microbial profiles, we stratified CRC patients into three subtypes. The subtypes were significantly associated with prognostic factors such as tumor grade, primary tumour sidedness and TNM staging, with one subtype enriched in tumours with poor prognosis. Further, we inspected the associations of microbiome with clinical variables in a subtype-agnostic setting. The primary tumour-associated clinical variables predominantly correlated with tumour mucosal microbiome, while the presence of local and distant metastases was mostly associated with the stool microbiome.

Conclusions

Understanding the interactions of the bacteria residing on tumour mucosa within different CRC tumour microbiome subtypes will help to better understand the underlying biological background of the heterogeneity of this disease. Indeed, the tumour microbiome is a possible source of additional integrative markers of CRC patients' survival and prognosis. We found that CRC microbiome is strongly correlated with clinical variables, but these associations are dependent on the microbial environment (tumour mucosa, normal mucosa, stool). Our study thus identifies limitations of the usage of microbiome composition as marker of CRC progression, suggesting the need of combining several sampling sites (e.g. stool and tumour swabs).

Introduction

Colorectal cancer (CRC) is the third most frequent cancer worldwide, and the second leading cause of cancer mortality in Europe [1]. At the same time, it is a heterogeneous disease, both at the phenotypic level and from a prognostic and response to treatment perspective. The current standard treatments are limited and remain ineffective for a large group of CRC patients because of lack of adequate patient selection, resulting in unneeded toxicity and elevated cost due to the over-treating of patients that do not benefit [2,3]. Recent research shows that gut microbiota may have an important role in colorectal tumour initiation and progression [4–22].

Several studies showed that bacteria adherent to colorectal adenomas or carcinomas were different from bacteria adherent to healthy gut mucosa [8,11,12] due to the altered tumor microenvironment with decreased pH and modified metabolic conditions resulting from a hypoxia and onset of necrosis [23]. Gut microbiota can promote colon cancer

development or change the tumour invasion potential through (i) immunomodulation [10,24–26] or (ii) metabolic activity – via the production of specific toxins inducing DNA damage responses. Overall, the evidence of microbiome importance in colon cancer development is so overwhelming that a bacterial driver-passenger model for colorectal cancer development and progression has been suggested [27] as an alternative to the broadly accepted driver-passenger mutational adenoma-carcinoma model. Additionally, gut microbiota seems to play an important role also in response to anti-cancer therapy [28].

Despite the observation of several consistent patterns of gut microbial disruption between healthy and CRC cases, a full characterization of the microbial landscape in CRC is still missing. The published studies vary greatly in terms of techniques, specimen origin and sample size, thus hampering any integrative analysis. Most studies compared diseased and healthy subjects, and the few that try to characterize microbial composition within the CRC patients suffer from a small sample size. The specimens used in most studies are stool [4,6,7,15,17,18,20–22] or mucosa samples from colonoscopy biopsies [11,13,15] or after resection [6,12,16,19]. Stool microbiota sampling has the advantage of being non-invasive allowing its use for screening and follow-up studies. There have been efforts to incorporate information on the tumour-associated microbiome in order to improve the accuracy of the existing patient CRC prognostic scores [18] or to develop a new screening/prognostic model [29]. The results of two different meta-analyses showed that the accuracy of predicting diseased state was about 0.8, similar to occult blood test results, the main non-invasive clinical test for this type of cancer [30,31]. However, the microbial composition in stool only partially reflects the situation in tumour mucosa, a trend consistent across different nationalities of the subjects, techniques of sampling or sequencing methodology [32].

The microbiota adherent to the mucosal tissue differs from the faecal microbiota in its needs for oxygen and nutrient types [33,34], therefore the information derived from stool may be insufficient for capturing tumour-microbe interactions consistent with the disease prognosis. The relevance of the tumour mucosa microbiome assessment for screening purposes is dependent not only on the co-presence of the bacteria in both tumour mucosa and stool, but also on its association with relevant clinical parameters in both sample types. Additionally, studying the (dis)similarity of bacterial composition between tumour and visually-normal mucosa from the same individual may provide hints regarding the changes in microenvironment which have occurred favoring the growth of certain species and shed some light on the underlying tumour-immune-microbe interactions and metabolic pathways.

Recently, two studies provided a comparison of bacterial composition in both tumour tissue and visually-normal tissue, as well as the bacterial composition of stool samples from the same patients [30,31]. Liu et al. [30] showed that the bacterial communities in both tumour tissue and visually-normal tissue were similar, but the study was largely underpowered ($N = 8$ individuals) and did not answer the question of the clinical relevance of this similarity. Other studies associated microbiome on tumour or in stool with clinical variables [31,35,36] but had a similar disadvantage in terms of statistical power ($N = 25$, $N = 30$, $N = 53$, individuals, respectively).

The above results were mostly species-centric in the sense that they compared the abundance of microbial species between the groups of interest individually. However, a broader view is needed to account for lesser known species coupled with a larger sample size allowing for capturing enough of the inter-tumour heterogeneity thus better understanding the possible effects of bacteria on tumour growth, aggressiveness or response to therapy. In our study, we take a *microbial community-centric* approach to provide a comprehensive description of CRC tumour microbiome based on 16S rRNA sequencing. We analyse three sample types (tumour mucosa, visually-normal mucosa, stool) from $N = 186$ individuals with stage 0-IV colorectal cancer.

Our study has a dual nature, both exploratory and confirmatory. We explore and interpret the landscape of tumour mucosa associated microbiome with respect to clinical variables and microbial composition of paired adjacent visually-

normal mucosa and paired stool samples. Benefitting from a larger sample size, we advance the state-of-the-art knowledge by reporting previously unseen associations.

Materials And Methods

Patients and Specimens

All specimens were collected at Masaryk Memorial Cancer Institute (Brno, Czech Republic) during the years 2015–2019 from unselected patients newly diagnosed for CRC. The stool samples were collected from untreated patients before the scheduled surgery. Patients performed the collection at home, the morning of their hospitalization for the surgery and brought the samples to the hospital, where they were immediately frozen at -80 °C until further processing. Swabs from tumour and visually-normal mucosa were collected within 30 minutes of the tumour resection at the department of pathology. If possible, the normal tissue swab was taken at least 20 cm proximally to the tumour. The swabs were then stored immediately in a freezer at -20 °C and without unnecessary delay transferred to -80 °C until further processing. All samples, including stool, were collected using DNA free cotton swabs (Deltalab, Spain).

Overall, we analysed N = 505 samples from N = 186 patients with CRC. Of these, there were 133 triplets (all three sample types from the same patient) and 53 mucosa duplets (swab from tumour and visually-normal mucosa from the same patient).

The study was approved by the ethical committee of Masaryk Memorial Cancer Institute. All patients gave written informed consent in accordance with the Declaration of Helsinki prior to participating in the study.

DNA extraction, PCR amplification and sequencing of 16S rRNA gene

The DNA extraction was performed using DNeasy® PowerSoil® Isolation kit (QIAGEN, Germany) according to the manufacturer's instructions. Extracted DNA was used as a template in amplicon PCR to target the V4 hypervariable region of the bacterial 16S rRNA gene. The 16S metagenomics library was prepared according to the 16S Metagenomic Sequencing Library Preparation protocol (Illumina, USA) with some deviations described below. Each PCR was performed with HotStarTaq Master Mix Kit (QIAGEN, Germany) in triplicate, with the primer pair consisting of Illumina overhang nucleotide sequences, an inner tag and gene-specific sequences [37,38]. The Illumina overhang served to ligate the Illumina index and adapter. Each inner tag, i.e. a unique sequence of 7–9 bp, was designed to differentiate samples into groups. Primer sequences and PCR cycling conditions are summarized in table S1. After PCR amplification, triplicates were pooled, and the amplified PCR products were determined by gel electrophoresis. PCR clean-up was performed with Agencourt AMPure XP beads (Beckman Coulter Genomics). Samples with different inner tags were equimolarly pooled based on fluorometrically measured concentration using Qubit® dsDNA HS Assay Kit (Invitrogen™, USA) and microplate reader (Synergy Mx, BioTek, USA). Pools were used as a template for a second PCR with Nextera XT indexes (Illumina, USA). Differently indexed samples were quantified using the qPCR kit KAPA Library Quantification Complete Kit (Roche, USA) and LightCycler 480 Instrument (Roche, USA) and equimolarly pooled according to the measured concentration. The prepared libraries were checked with a 2100 Bioanalyzer Instrument using the High Sensitivity D5000 Screen tape (Agilent Technologies, USA) and concentration was measured with qPCR shortly prior sequencing. The final library was diluted to a concentration of 8 pM and 20% of PhiX DNA (Illumina, USA) was added. Sequencing was performed with the Miseq reagent kit V2 (500 cycles) using a MiSeq instrument according to the manufacturer's instructions (Illumina, USA).

Data analysis

Preprocessing and quality control

Forward and reverse pair-end reads were demultiplexed and barcodes and primers were trimmed. Denoising algorithm with DADA2 [39] was applied separately on forward and reverse reads that passed the quality and length filter and did not contain N's. Reads were merged using the fastq-join method [40]. In the next step chimeras were detected with the function removeBimeraDenovo in DADA2. Chimera sequences were subsequently excluded from the analysis and Amplicon Sequence Variant (ASV) table was created.

The number of reads after quality filtering and chimeras removing ranged from 2968 to 239116, with median of 44277 and mean of 52546 reads per sample. The number of reads did not differ between the sample types (paired Wilcoxon test, Figure S1).

Taxonomy assignment

Taxonomy was assigned to each ASV based on SILVA 123 reference database [41] using the algorithm UCLUST [42] in QIIME [43]. BLAST algorithm [44] was used to identify the species and all taxa with the maximum identity and minimum e-value were selected for each ASV. The observed species metric and the Chao1 and Shannon index were used to estimate alpha diversity for each sample in QIIME. Beta diversity was computed in QIIME using both weighted and unweighted UniFrac metrics [45].

Sample and taxa filtering

We filtered out the ASVs that were unassigned at phylum level and all the ASVs belonging to the phylum Cyanobacteria. Only the taxa present in at least 3 samples of the same sample type and at the same time represented by at least 9 reads were kept for further analysis to account for possible contaminations. The threshold of 9 reads represents 0.03% taxa abundance in the sample with the least number of reads (2968).

This filtering step resulted in discarding 42–58% of taxa at each taxa level (Table S2)

All comparisons between the three sample types were performed on triplet samples from 133 patients, totaling N = 399 samples for the analysis. For analysis of tumour - visually-normal mucosa pairs, we used paired tumour and visually-normal mucosa swabs from 186 patients (totaling 372 samples). For analyses within sample types, we used all the available samples (186 for tumour mucosa swabs, N = 186 for visually-normal tissue mucosa swabs and N = 133 for stool).

Statistical analysis and data mining

Data were analysed using composition appropriate corrections and approaches [46–49]. Zero multiplicative replacement [48] was applied prior to the centered log-ratio (clr) transformation.

Non-metric Multidimensional Scaling (R vegan package [50]) over Aitchinson distance matrices (R coda.base package [51]) was used to analyse tumour microbial heterogeneity and β -diversity. To estimate the contribution of clinical traits in microbiome, β -diversity permutational multivariate analysis of variance for distance matrices (R adonis function of vegan 2.5.4 package [50]) with 999 of permutations was used. To assess the differences between the sample types in alpha diversity we used a paired non-parametric two-way Mann-Whitney U test. We applied a non-parametric approach to identify differences in microbial composition between sample types and the associations between microbial relative abundance and clinical variables. For non-parametric analysis, Friedman test with paired Wilcoxon test and rank regression were used (R package Rfit [52]). Drop in dispersion test was used to produce overall p-values for rank regression models. Cochran Q test was used for analysis of differences in presence of genera across sample types (analysis of triplets). Benjamini-Hochberg correction for multiple hypothesis testing was applied [53]. Results were

considered significant at $FDR < 0.1$. The adjusted p-values are referred to as q-values. Visualization was performed with gplots 3.0.1.1, ggplot2, ComplexHeatmap 1.17.1, and circlize 0.4.8 packages [54–57].

When comparing the significance of discoveries in stool vs mucosa and stool vs visually-normal tissue, the dataset used was the one having patients with all 133 samples. However, for mucosa samples, the full set of 186 was exploited to maximize the statistical power.

For each clinical variable (or a combination of), we only tested genera present in at least 10 samples in one clinical group (or a combination of). We do emphasize that we approached this statistical testing from the point of view of a pilot discovery study.

Due to the known association between tumour grade and location on [58] (confirmed also in our data, $p < 0.001$, Fisher's exact test), we investigated the associations of microbiome with grade and tumour location in a model with the interaction between covariates compared to model without interaction. To ensure a more balanced design, we considered three categories for the location: right and transversum, left, rectosigmoid and rectum, respectively.

While we consider only associations with $FDR < 0.1$ to be statistically significant, we also report the unadjusted results $p < 0.05$ for the purposes of hypothesis confirmation by other studies.

Data Access

The data were uploaded to the European Nucleotide Archive under accession number PRJEB35990.

Results

In our effort to describe tumour microbial landscape, we explored the differences in microbiome abundance, diversity, the presence/absence of the species and the proportion of samples with the respective genera in different sample types across patient groups defined by clinical variables (Table 1).

Table 1

Table of clinical variables and their distribution in the full set of 186 patients and subset of 133 patients and CRC tumour microbial subtypes. (For categorical variable Fisher exact test was performed and for continuous data Kruskal-Wallis test was used.)

	Data subset comparison			Tumour microbiome subtypes			
	all tumours (N = 186)	triplets (N = 133)	p-value	TS1 (N = 54)	TS2 (N = 32)	TS3 (N = 100)	p-value
age at dg	mean (SD)	mean (SD)	0.769	mean (SD)	mean (SD)	mean (SD)	0.467
	67.03 (10.65)	66.68 (10.65)		67.24 (9.48)	64.94 (12.31)	67.59 (10.71)	
gender	N (%)	N (%)	1	N (%)	N (%)	N (%)	0.7463
male	104 (55.9)	74 (55.6)		30 (55.6)	16 (50.0)	58 (58.0)	
female	82 (44.1)	59 (44.4)		24 (44.4)	16 (50.0)	42 (42.0)	
tumour location	N (%)	N (%)	0.650	N (%)	N (%)	N (%)	< 0.001
right	66 (35.5)	50 (37.6)		32 (59.3)	3 (9.4)	31 (31.0)	
transversum	19 (10.2)	13 (9.8)		7 (13.0)	2 (6.2)	10 (10.0)	
left	47 (25.3)	38 (28.6)		5 (9.3)	9 (28.1)	33 (33.0)	
rectosigmoideum	32 (17.2)	23 (17.3)		5 (9.3)	8 (25.0)	19 (19.0)	
rectum	22 (11.8)	9 (6.8)		5 (9.3)	10 (31.2)	7 (7.0)	
grade	N (%)	N (%)	0.994	N (%)	N (%)	N (%)	< 0.001
NA, <i>in situ</i>	8 (4.3)	6 (4.5)		0 (0.0)	1 (3.1)	7 (7.0)	
1	17 (9.1)	11 (8.3)		0 (0.0)	2 (6.2)	15 (15.0)	
2	108 (58.1)	78 (58.6)		26 (48.1)	24 (75.0)	58 (58.0)	
3	53 (28.5)	38 (28.6)		28 (51.9)	5 (15.6)	20 (20.0)	
stage	N (%)	N (%)	0.963	N (%)	N (%)	N (%)	0.02702
0	8 (4.3)	6 (4.5)		0 (0.0)	1 (3.1)	7 (7.0)	
I	35 (18.8)	29 (21.8)		3 (5.6)	8 (25.0)	24 (24.0)	
II	69 (37.1)	48 (36.1)		27 (50.0)	12 (37.5)	30 (30.0)	
III	48 (25.8)	34 (25.6)		15 (27.8)	7 (21.9)	26 (26.0)	
IV	26 (14.0)	16 (12.0)		9 (16.7)	4 (12.5)	13 (13.0)	
tumour pathologic	N (%)	N (%)	0.987	N (%)	N (%)	N (%)	< 0.001
pTis	8 (4.3)	6 (4.5)		0 (0.0)	1 (3.1)	7 (7.0)	
pT1	11 (5.9)	10 (7.5)		0 (0.0)	3 (9.4)	8 (8.0)	
pT2	36 (19.4)	27 (20.3)		3 (5.6)	8 (25.0)	25 (25.0)	

	Data subset comparison			Tumour microbiome subtypes			
pT3	117 (62.9)	81 (60.9)		46 (85.2)	16 (50.0)	55 (55.0)	
pT4a	9 (4.8)	5 (3.8)		4 (7.4)	1 (3.1)	4 (4.0)	
pT4b	5 (2.7)	4 (3.0)		1 (1.9)	3 (9.4)	1 (1.0)	
node pathologic	N (%)	N (%)	0.827	N (%)	N (%)	N (%)	0.1382
pN0	114 (61.3)	84 (63.2)		30 (55.6)	21 (65.6)	63 (63.0)	
pN1a	25 (13.4)	20 (15.0)		7 (13.0)	1 (3.1)	17 (17.0)	
pN1b	21 (11.3)	16 (12.0)		5 (9.3)	5 (15.6)	11 (11.0)	
pN1c	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)	
pN2	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)	
pN2a	10 (5.4)	7 (5.3)		4 (7.4)	2 (6.2)	4 (4.0)	
pN2b	14 (7.5)	5 (3.8)		8 (14.8)	3 (9.4)	3 (3.0)	
pNX	2 (1.1)	1 (0.8)		0 (0.0)	0 (0.0)	2 (2.0)	
metastasis	N (%)	N (%)	0.737	N (%)	N (%)	N (%)	0.843
M0	160 (86.0)	117 (88.0)		45 (83.3)	28 (87.5)	87 (87.0)	
M1	26 (14.0)	16 (12.0)		9 (16.7)	4 (12.5)	13 (13.0)	

β -diversity analysis by NMDS performed on all sample types showed that tumor location was the factor with the highest influence on total microbiome composition for all sample types, while tumour histological grade affected only tumour and visually-normal samples, which had similar microbial profiles (Text S1, Figure S2, Figure S3).

Microbial categorization according to sample type

There was no significant difference between the read counts across different sample types (paired analysis of sample triplets, see Methods).

Overall, in all the 505 samples we identified 5553 ASVs, of these, 4920 ASVs in the 133 triplet samples. The QIIME assigned species only to 50 ASVs, hence we also performed manual BLAST search to the SILVA database (Table S3).

For further analysis, however, we mainly operated on higher taxonomic levels. After the taxa filtering step (Table S2) 14 phyla, 26 classes, 44 orders, 77 families and 268 genera were identified in the 133 triplets, most of which in all three sample types (Table S4). Inclusion of the additional 53 duplets (tumour mucosa and visually-normal mucosa swabs) resulted only in slight differences at the genus level – the identified taxa remained the same, what changed was their unique presence in some sample types (Text S1).

While most of the genera were found in all three sample types, their incidence and abundance across sample types varied greatly, mainly between mucosal samples and stool, both in overall and pairwise comparisons (Text S1). 14 genera (*Stomatobaculum*, *Pseudoramibacter*, *Pelomonas*, *Pasteurella*, *Mycoplasma*, *Kingella*, *Johnsonella*, *Helicobacter*, *Deinococcus*, *Centipeda*, *Bergeyella*, *Actinobacillus*, *Abiotrophia* and an unassigned genus from order *Comamonadaceae*) were detected only in mucosal (tumour and visually-normal) samples (Figure S4).

We further analysed the pairwise incidence of the 268 genera across sample types using Cochran's Q test and subsequent pairwise McNemar's tests and found that 128 genera varied significantly across sample types (analysis of 133 triplets, Text S1, Fig. 2A, Table S5).

To categorize the microbial genera based on their preferred environment we compared their abundance across sample types using Friedman rank sum test. Out of the 268 genera, 104 differed significantly in abundance across the sample types (Table 2, Fig. 1). Based on these results, we defined five microbial categories. The first is based solely on the results of tumour vs stool comparison: *tumour genera* (46 genera, more abundant in tumours compared to stool). Additionally, within the category of *tumour genera*, we defined *mucosa genera* (41 genera, enriched also in visually-normal mucosa compared to stool) and *tumour-specific genera* (18 genera of *tumour category*, additionally enriched in tumours compared to visually-normal mucosa). Fifty genera were significantly more abundant in stool compared to tumours and visually-normal mucosa form the group of *stool genera*. The fifth category was defined as the *no-difference genera* (164 genera, no difference across any of the sample types) (Text S1).

Table 2

Total counts of genera found significantly differentially abundant across the three sample types (133 triplets), divided into categories according to their enrichment in different sample types and top 10 significant genera for each category (see Table S6 for all genera). Abbreviations: "TtoS": tumour to stool, "VNtoS": visually-normal to stool, "TtoVN": tumour to visually-normal.

Category I	Freq	Category II	Freq	Category III	Freq	T to S	VN to S	T to VN	N	top 10 significant genera	
Tumour genera (enriched in tumours compared to stool)	46	Mucosa genera (enriched in both mucosal samples compared to stool)	41			↑	↑	↓	5	<i>G_Prevotella 1, G_Romboutsia, F_Peptostreptococcaceae, G_Unassigned, G_Faecalitalea, G_Massil</i>	
						↑	↑	-	23	<i>G_Haemophilus, G_Bacteroides, G_Veillonella, G_Pseudomonas, G_Escherichia-Shigella, G_Aggregatibacter, G_Actinobacillus, G_Neisseria, G_Coprobacter, G_Selenomonas 4</i>	
					Tumour-specific genera (enriched in tumours compared to both stool and normal mucosa)	18	↑	↑	↑	13	<i>G_Fusobacterium, G_Campylobacter, G_Gemella, G_Leptotrichia, G_Parvimonas, G_Granulicatella, G_Selenomonas 3, G_Eikenella, G_Selenomonas, G_Hungatella</i>
							↑	-	↑	4	<i>G_Howardella, G_Solobacterium, G_Streptococcus, G_Slackia</i>
							↑	↓	↑	1	<i>G_Defluviitaleaceae UCG-011</i>
No difference between stool and tumours	170			Enriched only in visually normal mucosa compared to stool	4	-	↑	↓	4	<i>O_Clostridiales, G_Odoribacter, G_Parabacteroides, G_Ruminococcaceae UCG-004</i>	
				No-difference genera between sample types	164	-	-	-	164		
				Depleted only in normal mucosa compared to stool		-	↓	-	1	<i>G_Methanobrevibacter</i>	
						-	↓	↑	1	<i>G_Cloacibacillus</i>	

Category I	Freq	Category II	Freq	Category III	Freq	T to S	VN to S	T to VN	N	top 10 significant genera
Depleted in tumours compared to stool	52								1	<i>G_Pseudobutyrvibrio</i>
				Depleted in tumours both compared to stool and normal mucosa	4	↓	↑	↓	1	<i>G_Coproccoccus 2</i>
		Stool genera (depleted in both mucosal samples compared to stool)	50			↓	↑	↓	3	<i>G_Ruminococcaceae UCG-013</i> , <i>G_Lachnospiraceae ND3007 group</i> , <i>G_Alistipes</i>
						↓	↓	↑	3	<i>G_Bifidobacterium</i> , <i>G_Acidaminococcus</i> , <i>G_Fusicatenibacter</i>
						↓	↓	-	44	<i>G_[Eubacterium] coprostanoligenes group</i> , <i>G_Ruminococcaceae UCG-002</i> , <i>F_Lachnospiraceae</i> ; <i>G_Unassigned</i> , <i>G_Christensenellaceae R-7 group</i> , <i>G_Ruminococcaceae NK4A214 group</i> , <i>G_Lachnospira</i> , <i>G_Ruminococcaceae UCG-005</i> , <i>G_Ruminococcus 2</i> , <i>G_Ruminococcaceae UCG-003</i> , <i>F_Clostridiales vadinBB60 group</i> ; <i>G_uncultured bacterium</i>
Overall	268								268	

The landscape of CRC tumour microbiome

For the description of tumour mucosa microbial heterogeneity without possible stool contaminants, we only considered species that were statistically significantly enriched in tumour mucosa compared to stool. We hence investigated the group of 46 *tumour genera* with special focus on the subgroup of 18 *tumour-specific genera* (*Bergeyella*, *Gemella*, *Granulicatella*, *Parvimonas*, *Hungatella*, *Peptoclostridium*, *Flavonifractor*, *Selenomonas*, *Selenomonas 3*, *Fusobacterium*, *Leptotrichia*, *Eikenella*, *Campylobacter*, *Slackia*, *Streptococcus*, *Howardella*, *Solobacterium*, *Defluviitaleaceae UCG-011*, Fig. 2A).

19 *tumour genera* (of which 14 *tumour-specific*), co-occurred in tumour mucosa significantly more often: *Eikenella*, *Campylobacter*, *Solobacterium*, *Gemella*, *Leptotrichia*, *Parvimonas*, *Bergeyella*, *Peptoclostridium*, *Howardella*, *Slackia*, *Selenomonas 3*, *Granulicatella*, *Selenomonas*, *Fusobacterium*, *Aggregatibacter*, *Alloprevotella*, *Treponema 2*, *Selenomonas 4* and *Veilonella* (Figure S5, Table S7). On the other hand, we observed significantly decreased co-occurrence of the following pairs of genera: *Selenomonas* and *Tyzzereella 4*, *Pseudomonas* and *Massillias*; *Selenomonas 3* and *Morganella*, *Coprobacter* and *Treponema 2* and *Coproccoccus* and *Faecalitalea*.

Tumour genera incidence ranged from 3.2–99.4% (median 34.4%) of tumours with the median abundance of the individual genera in the samples ranging from 0.015–28.8% (median 0.23%) (Fig. 2A). Overall, *tumour genera* constituted 10.8–96.7% (median 58.6%) while the *tumour-specific genera* constituted between 0.0%–63.7% (median 2.6%) of tumour mucosa microbiome (Fig. 2B).

We performed detailed literature search which revealed that *tumour genera* consisted predominantly of oral bacteria, many known as oral pathogens. The presence of some genera on tumor mucosa was never reported before nor associated with CRC (Table S8).

Microbiome and clinical variables

In the next step, we assessed the association of microbiome abundance with the clinical parameters and interpreted the results based on our microbial categorization. The results for each clinical variable are summarized in Table 3.

Table 3

Summary of rank regression results ($p < 0.05$) associating microbiome of the three different sample types with the clinical variables. **Bold text** denotes genera significant at FDR < 0.1 . Up and down arrows denote increase or decrease in abundance, respectively.

Regression covariate	Effect/Contrast	Tumour mucosa	Visually-normal mucosa	Stool	
grade	increasing grade	↑ <i>Fusobacterium</i> , <i>Campylobacter</i> , <i>Mobigacterium</i> , <i>Leptotrichia</i>			
location	right-sided/ transverse vs left- sided and rectum/rectosigmoid	↑ <i>Erysipelatoclostridium</i> , <i>Holdemania</i> , <i>Selenomonas 3</i> , <i>Selenomonas 4</i> , <i>Fretibacterium</i>	↑ <i>Fretibacterium</i> , <i>Selenomonas</i>		
	left-sided vs right- sided/ transverse	↑ <i>Fusicatenibacter</i> , <i>Christensenellaceae R-7</i> group, <i>Ruminococcaceae</i> <i>UCG - 013</i> , <i>Coprococcus 1</i> , Family <i>XIII AD3011</i> group ↓ <i>Campylobacter</i>	↑ <i>Bifidobacterium</i> , <i>Lachnospira</i>		
	rectosigmoid/rectum vs right- sided/transverse	↑ <i>Fusicatenibacter</i> , <i>Christensenellaceae R-7</i> group, <i>Ruminococcaceae</i> <i>UCG - 013</i> , <i>Bifidobacterium</i>	↑ <i>Bifidobacterium</i> , <i>Lachnospira</i> , [<i>Eubacterium</i>] ventriosum group		
interaction between grade and location	low-graded; left- sided	↑ <i>Lachnospira</i> , <i>Ruminoclostridium 6</i> , <i>Gemella</i> , [<i>Eubacterium</i>] ventriosum group, <i>Methanobrevibacter</i> , <i>an uncultured</i> <i>bacterium from</i> <i>Opitutae vadinBB60</i> <i>group family</i> , <i>Ruminococcaceae</i> <i>UCG-010</i> , <i>Victivallis</i> , an uncultured species and an Incertae Sedis genus from <i>Lachnospiraceae</i> family	↑ <i>Gemella</i>	↑ <i>Ruminoclostridium 6</i>	
	grade 2; right-sided			↑ <i>Ruminoclostridium 6</i>	
	grade 2; transverse, left-sided	↑ <i>Lachnospira</i>			
	grade 3; right-sided	↑ <i>Prevotella</i> , <i>Selenomonas</i>			
	grade3; rectosigmoid/rectum	↑ <i>Lachnospira</i>			↑ <i>Prevotella</i>

Regression covariate	Effect/Contrast	Tumour mucosa	Visually-normal mucosa	Stool
stage	III-IV vs 0-II	<p>↑ <i>Peptoclostridium</i>, <i>Fusobacterium</i></p> <p>↓ <i>Parabacteroides</i>, <i>Lachnospiraceae</i> <i>FCS020</i> group, <i>Tyzzereella</i> 4</p>	<p>↑ <i>Peptoclostridium</i></p>	<p>↑ <i>Streptococcus</i>, <i>Peptococcus</i>, <i>Akkermansia</i></p> <p>↓ <i>Dorea</i>, <i>Ruminiclostridium</i></p>
T	pT 3-4 vs pTis-2	<p>↑ <i>Fusobacterium</i>, <i>Campylobacter</i>, <i>Selenomonas</i>, <i>Holdemania</i>, <i>Anaerotruncus</i>, <i>Peptoclostridium</i>, <i>Mogibacterium</i>, <i>Parvimonas</i>, Family <i>XIII UCG-001</i>, one unassigned genera from <i>Ruminococcaceae</i> family</p> <p>↓ <i>Parabacteroides</i>, <i>Prevotellaceae</i> <i>NK3B31</i> group, <i>Coprococcus</i> 2, <i>Lachnospiraceae</i> <i>FCS020</i> group, <i>Tyzzereella</i> 4, [<i>Ruminococcus</i>] <i>gauvreauii</i> group, <i>Sutterella</i>, <i>Neisseria</i>, <i>Enterococcus</i>, <i>Actinobacillus</i>, <i>Staphylococcus</i>, <i>Peptostreptococcus</i></p>	<p>↑ <i>Peptoclostridium</i>, <i>Selenomonas</i>, <i>Gemella</i></p> <p>↓ <i>Corynebacterium</i>, <i>Coprobacter</i>, <i>Parabacteroides</i>, <i>Anaerostipes</i>, <i>Tyzzereella</i> 3, [<i>Ruminococcus</i>] <i>gauvreauii</i> group, <i>Lachnospiraceae</i> <i>FCS020</i> group, <i>Lachnospiraceae</i> <i>ND3007</i> group, <i>Lachnospiraceae</i> <i>UCG-001</i>, <i>Haemophilus</i></p>	<p>↑ <i>Fusobacterium</i>, <i>Anaerotruncus</i>, <i>Ruminococcaceae</i> <i>NK4A214</i> group</p> <p>↓ <i>Lachnospira</i>, <i>Coprococcus</i> 2, <i>Lachnospiraceae</i> <i>UCG-001</i>, <i>Hydrogenoanaerobacterium</i></p>
N	N1-2 vs N0	<p>↑ <i>Peptoclostridium</i></p> <p>↓ <i>Prevotellaceae</i> <i>UCG-001</i></p>	<p>↑ <i>Selenomonas</i></p> <p>↓ [<i>Eubacterium</i>] <i>hallii</i> group</p>	<p>↑ <i>Coprobacter</i>, <i>Peptococcus</i> <i>Campylobacter</i>, <i>Akkermansia</i>, <i>Selenomonas</i>, <i>Porphyromonas</i>, <i>Streptococcus</i></p> <p>↓ <i>Faecalibacterium</i>, <i>Ruminiclostridium</i>, <i>Dorea</i>, uncultured bacterium from <i>Bacteroidales</i> <i>S24-7</i> group family</p>
M	present vs absent	<p>↑ <i>Porphyromonas</i>, <i>Streptococcus</i>, <i>Ruminococcaceae</i> <i>UCG-005</i></p>	<p>↑ <i>Akkermansia</i></p> <p>↓ <i>Gelria</i>, [<i>Eubacterium</i>] <i>brachy</i> group, uncultured genera from <i>Christensenellaceae</i> family</p>	<p>↑ uncultured genus from <i>Erysipelotrichaceae</i> family, <i>Akkermansia</i>, <i>Coprococcus</i> 1, <i>Solobacterium</i></p> <p>↓ <i>Metanosphaera</i></p>

Regression covariate	Effect/Contrast	Tumour mucosa	Visually-normal mucosa	Stool
age	increasing age	↓ <i>Lachnospira</i> , <i>Pseudobutyvibrio</i> , Erysipelotrichaceae UCG-003	↑ <i>uncultured species from the Ruminococcaceae family</i> ↓ <i>Lachnospira</i> , <i>Erysipelotrichaceae</i> UCG - 003	↑ <i>Ruminiclostridium 5</i> ↓ <i>Pseudobutyvibrio</i>
gender	men vs women	↑ <i>Mogibacterium</i> , <i>Dorea</i> , <i>Roseburia</i> , <i>Bacteriodes</i> , <i>Paraprevotella</i> , <i>Eubacterium hallii</i> <i>group</i> , <i>Subdoligranulum</i> , <i>Sutterella</i> ↓ <i>Fretibacterium</i> , <i>[Eubacterium]</i> <i>nodatum group</i> , <i>Aeromonas</i>	↑ <i>Mogibacterium</i> , <i>Dorea</i> , <i>Roseburia</i> , <i>Collinsella</i> , <i>Howardella</i> , <i>Lachnospira</i> , <i>Ruminiclostridium</i> , <i>Acidaminococcus</i> , <i>Haemophilus</i> ↓ <i>Ruminiclostridium</i> 5	↑ <i>Acidaminococcus</i> , <i>Sutterella</i> , <i>Mogibacterium</i> ↓ <i>Actinomyces</i> , <i>Hungatella</i> , <i>Oscillospira</i> , <i>Ruminiclostridium 9</i> , <i>Erysipelatoclostridium</i> , <i>Tyzzereella</i> , <i>uncultured genus from family</i> <i>Erysipelotrichaceae</i> , <i>Eggerthella</i>

Tumour grade, location and microbiome

In the tumour mucosa samples, 35 genera were found significantly associated with grade or location, of which 18 with significant interaction (Table S9, Figure S6, Figure S7, Figure S8, Figure S9). Some of these associations were found also in visually-normal mucosa (analysis of 186 samples).

In tumour mucosa, three genera, *Fusobacterium*, *Campylobacter* and *Mogibacterium*, significantly increased with tumour grade regardless of primary tumour side. *Campylobacter* was also associated with lower abundance in the mucosa of the left sided tumours. Of these, in the visually-normal mucosa adjacent to these tumours, none remained associated with high tumour grade.

Leptotrichia was significantly increased in advanced grade tumours (2 or 3 compared to 1); while the interactional model was chosen according to the drop dispersion test, none of interactional coefficients were significant. *Prevotella* and *Selenomonas* both had significantly increased abundances in the grade 3 right sided tumours. *Prevotella* was also increased in the stool of patients with grade 3 rectosigmoid/rectum tumours. *Lachnoclostridium* was associated with grade 2 in all tumours except for rectosigmoid/rectum.

In contrast, *Lachnospira*, *Ruminiclostridium 6*, *Gemella*, *[Eubacterium] ventriosum group*, *Methanobrevibacter*, an uncultured bacterium from Opatutae vadinBB60 group family, *Ruminococcaceae UCG-010*, *Victivallis* and an uncultured species and an *Incertae Sedis genus from Lachnospiraceae family* were significantly enriched mainly in left-sided (for some including rectosigmoid/rectum) low-grade tumours. *Lachnospira* increased in abundance in grade 3 tumours of the rectosigmoid/rectum. In visually-normal mucosa, the same association was observed for *Gemella*, while *Lachnospira* was only significantly enriched in left, rectosigmoid and rectum location and *[Eubacterium] ventriosum group* only in rectum. *Ruminiclostridium 6* remained enriched in the stool of patients with low-grade left-sided tumours and grade 2 right-sided tumors.

Erysipelatoclostridium, *Holdemania*, *Selenomonas 3* and *Selenomonas 4* and *Fretibacterium* were increased in mucosa of the right-sided and transverse tumours. The same associations were found only for *Fretibacterium* and *Selenomonas* in the visually-normal mucosa. *Fusicatenibacter*, *Christensenellaceae R-7 group*, *Ruminococcaceae UCG - 013* were increased in mucosa of the left-sided, rectosigmoid and rectal tumours, *Coprococcus 1* and *Family XIII AD3011 group* in the mucosa of left-sided tumours and *Bifidobacterium* in the mucosa of rectosigmoid and rectal tumours. Similar associations were found for visually-normal mucosa for the *Bifidobacterium*.

For *Flavonifractor* and *Odoribacter*, although the model itself was significant in the drop dispersion test, none of the coefficients was significant alone.

Tumour stage and microbiome

When comparing early (0-II) and advanced (III-IV) stages, we identified increased abundance of *Peptoclostridium* and *Fusobacterium* associated with advanced stage and increased abundance of *Parabacteroides*, *Lachnospiraceae FCS020 group* and *Tyzzarella 4* associated with early stage tumours. Of these, only *Peptoclostridium* was associated with an advanced stage in visually-normal tissue. None of these genera were significant after the FDR correction ($p < 0.05$, $FDR > 0.1$). In stool, we found a different set of genera associated with advanced tumour stage: increased *Streptococcus*, *Peptococcus* and *Akkermansia* and decreased *Dorea* and *Ruminiclostridium* ($p < 0.05$, $FDR > 0.1$).

Patients with advanced T stages (pT 3–4) were characterized by significant increase in abundance of *Fusobacterium* and *Campylobacter* ($FDR < 0.1$) as well as *Selenomonas*, *Holdemania*, *Anaerotruncus*, *Peptoclostridium*, *Mogibacterium*, *Parvimonas*, *Family XIII UCG-001* and one unassigned genera from *Ruminococcaceae family* ($p < 0.05$, $FDR > 0.1$) in tumour mucosa, increased *Peptoclostridium*, *Selenomonas* and *Gemella* in visually-normal tissue, and increased *Fusobacterium*, *Anaerotruncus* and *Ruminococcaceae NK4A214 group* in stool.

Early T stage tumours (pTis-2) were associated with increase in *Parabacteroides*, *Prevotellaceae NK3B31 group*, *Coprococcus 2*, *Lachnospiraceae FCS020 group*, *Tyzzarella 4*, [*Ruminococcus*] *gouvreauii group*, *Sutterella*, *Neisseria*, *Enterococcus*, *Actinobacillus*, *Staphylococcus* and *Peptostreptococcus* ($p < 0.05$, $FDR > 0.1$) in tumour mucosa, increased *Corynebacterium*, *Coprobacter*, *Parabacteroides*, *Anaerostipes*, *Tyzzarella 3*, [*Ruminococcus*] *gouvreauii group*, *Lachnospiraceae FCS020 group*, *Lachnospiraceae ND3007 group*, *Lachnospiraceae UCG-001* and *Haemophilus* in visually-normal tissue and *Lachnospira*, *Coprococcus 2*, *Lachnospiraceae UCG-001*, or *Hydrogenoanaerobacterium*, but none of these associations was significant after FDR correction (Table S9, Figure S10, Figure S11).

The presence of metastases (local or distant) at the time of diagnosis was predominantly associated with changes in stool microbiome. Except for uncultured genus from the *Erysipelotrichaceae* family, none of these associations were significant after FDR correction.

Lymph-node metastases (N1-2) were associated with increased abundance of *Peptoclostridium* on tumour mucosa, *Selenomonas* on visually-normal mucosa and increased abundance of *Coprobacter*, *Peptococcus* *Campylobacter*, *Akkermansia*, *Selenomonas*, *Porphyromonas* and *Streptococcus* in stool ($p < 0.05$, $FDR > 0.1$). Tumours without lymph-node metastases showed increased abundance of *Prevotellaceae UCG-001* in tumour mucosa, [*Eubacterium*] *hallii group* on visually-normal mucosa and *Faecalibacterium*, *Ruminiclostridium*, *Dorea* and *uncultured bacterium* from *Bacteroidales S24-7 group family* in stool ($p < 0.05$, $FDR > 0.1$).

The presence of distant metastases was associated with increased abundance of *Porphyromonas*, *Streptococcus* and *Ruminococcaceae UCG-005* on tumour mucosa, *Akkermansia* in visually-normal tissue ($p < 0.05$, $FDR > 0.1$) and an uncultured genus from *Erysipelotrichaceae* family ($FDR < 0.1$), *Akkermansia*, *Coprococcus 1* and *Solobacterium* ($p < 0.05$, $FDR > 0.1$) in stool.

Patients without distant metastases had increased *Gelria*, [*Eubacterium*] *brachy* group and uncultured genera from *Christensenellaceae* family in visually-normal tissue and increased abundance of *Metanosphaera* in stool ($p < 0.05$, FDR > 0.1) (Table S9, Figure S10, Figure S12).

Gender, age and microbiome

We found three genera from tumour mucosa with abundances decreasing with increasing patients' age: *Lachnospira*, *Pseudobutyvibrio* ($p < 0.01$, FDR > 0.1) and *Erysipelotrichaceae* UCG-003 (FDR < 0.1). *Lachnospira* was decreasing with age also on visually-normal mucosa, together with *Erysipelotrichaceae* UCG - 003 ($p < 0.01$, FDR > 0.1). An uncultured species from the *Ruminococcaceae* family increased in abundance with increasing age in visually-normal mucosa ($p < 0.01$, FDR > 0.1). In stool, *Pseudobutyvibrio* decreased and *Ruminiclostridium 5* increased with age ($p < 0.01$, FDR > 0.1). (Table S9, Figure S10, Figure S13).

No significant (FDR < 0.1) associations between patients' gender and microorganisms' abundance were found for any sample type. Non-adjusted results ($p < 0.01$, FDR > 0.1), however, pointed to increased *Mogibacterium*, *Dorea*, *Roseburia* in both tumour and visually-normal mucosa in men, *Bacteriodes*, *Paraprevotella*, *Eubacterium hallii* group, *Subdoligranulum*, *Sutterella* in tumour mucosa of men, *Collinsella*, *Howardella*, *Lachnospira*, *Ruminiclostridium*, *Acidaminococcus* and *Haemophilus* in visually-normal mucosa of men and *Acidaminococcus*, *Sutterella* and *Mogibacterium* in stool of men with CRC. In contrast, women with CRC had increased *Fretibacterium*, [*Eubacterium*] *nodatum* group and *Aeromonas* in tumour mucosa, *Ruminiclostridium 5* in visually-normal mucosa and *Actinomyces*, *Hungatella*, *Oscillospira*, *Ruminiclostridium 9*, *Erysipelatoclostridium*, *Tyzzellerella*, uncultured genus from family *Erysipelotrichaceae* and *Eggerthella* in stool. (Table S9, Figure S10, Figure S14).

Tumour CRC microbial subtypes

We performed hierarchical clustering of patients based on the relative abundance of the 46 *tumour genera* in the tumour mucosa samples. As a distance measure between patients we used Aitchinson's distance, and Euclidean distance for the clustering of genera (see Methods).

Based on the tumour-mucosa microbial composition we observed three major subtypes of tumours (TS1-TS3), that could further be divided into seven groups (a-h) (Fig. 3). The bacteria clustered into six groups M1-M6 (Table S8). The seven minor groups a-h are reflecting profiles of certain individual species, such as *Sutterella*, *Peptoclostridium*, *Flavonifractor*, *Coprobacter*, *Aggregatibacter*, *Granulicatella*, *Hungatella*, *Alloprevotella* or *Slackia*. We associated these subtypes with clinical variables (Table 1).

The M1 group and M2 group are represented by typical gut microbiome members. The M1 group consists of five most common and most abundant genera *Fusobacterium*, *Lachnoclostridium*, *Bacteriodes*, *Escherichia-Shigella* and one uncultured genus from family *Lachnospiraceae*. All tumours contain at least three of these bacteria, most tumours (78,5%) all five. These bacteria have high co-occurrence across sample types (Fig. 2A, fourth panel), except for *Fusobacterium*. The M3, M4 and M5 groups contain predominantly oral microbiome genera. These genera have significantly different incidence across sample types, with 44.0-94.6% patients missing these genera in stool if present on tumour mucosa. Group M5 is more heterogeneous and with respect to the tumour microbial subtypes can be further divided into two subgroups: (i) *Alloprevotella*, *Howardella*, *Slackia* and (ii) *Treponema 2*, *Selenomonas*, *Selenomonas 3*, *Selenomonas 4*. Group M6 consists of sixteen less common species.

Tumour microbial subtype 1 (TS1) represents 29% (54) of tumours and is defined by presence of M1-3 and M5 microbial groups and overall contains most of the high-grade associated genera (*Fusobacterium*, *Campylobacter*, *Leptotrichia* and

Selenomonas). This subtype is enriched in right-sided tumours (59.3%), grade 3 tumours (51.9%), tumours with pT3 or pT4 stage (94.4%) and is depleted of stage 0 and stage I tumours (0% and 5.6%, respectively). Tumour microbial subtype 2 (TS2) represents 17% (32) of tumours and is defined mainly by the presence of M1-M3 bacteria. 84% of TS2 tumours are from the left side, rectosigmoid or rectum and mostly of grade 2 (75%). Finally, the largest subtype 3 (TS3) represents 54% (100) of tumours and is mostly missing the M3-5 bacterial groups as well as most of the high-grade related species, containing in median TS3 is thus characteristic by increased proportion of low grade tumours (15% grade 1). In TS3 microbial subtype, right-sided and left-sided tumours are equally represented.

Discussion

Carcinogenesis of colorectal cancer is a complex process with a unique set of somatic molecular changes that can be caused by different factors. Some studies correlated the dysbiosis of gut microbiome with the development of colorectal cancer in the healthy mucous-adenoma-carcinoma sequence or focused on elaborating the concrete role of selected bacterial species in gut (colorectal) pathogenesis progression [4–22]. In contrast, we characterized the heterogeneity of the microbiome in the ongoing disease through comprehensive description of microbial communities, as a first step for in-depth targeted studies.

The comparison of three sampling sites provided us with insights into the preferred environment of the observed species. The resulting microbial categories served for focusing further analyses and interpretation of our findings. We decided to base our characterization of colorectal cancer microbial landscape on bacteria with increased abundance in tumour mucosa compared to stool, specifically the 46 *tumour genera*, to filter out potential stool contaminants. With a median of 58.6%, the *tumour genera* represented an important fraction of total microbiome found on tumour mucosa. We consequently defined CRC tumour microbial subtypes and associated them with clinical variables. Naturally, we extended our search for clinical associations with microbiome to all available sample types.

For ten genera previously associated with stool of CRC patients we report for the first-time their presence on tumour mucosa (see below).

Both the unsupervised NMDS analysis and supervised comparison of microbial composition between sample types confirmed previously reported observations from 16S rRNA sequencing studies of much smaller sample size (31, 27 and 8 individuals) [8,11,30] or a qPCR of specific bacteria [16], that the tumour mucosal microbiome is dominated by mucosa-associated bacteria and that these species are at the same time associated also with non-cancerous (visually-normal) mucosa. It is, however, debatable whether the non-tumour, visually-normal tissue (however distant from the tumour) from the surgically removed segment can be considered healthy or is possibly already influenced by the bacteria initiating CRC development. Consistent with the bacterial driver-passenger model as proposed by Tjalsman et al. [27] our *mucosa genera* could be considered as candidates for potential drivers, while *tumour-specific* genera could be potential passengers.

In further agreement with the driver-passenger model, our analysis revealed that tumours harbour a diverse community of opportunistic pathogens mainly of oral origin (26 of 46 *tumour genera*). Increasing evidence suggests that oral bacteria can migrate to the colon and cause infections and inflammation [59,60].

Some of the *tumour genera* of (possible) oral origin identified in our study, were associated with tumour mucosa previously, namely *Fusobacterium*, *Campylobacter*, *Leptotrichia*, *Parvimonas*, *Gemella*, *Granulicatella*, *Eikenella*, *Selenomonas*, *Neisseria*, *Streptococcus*, *Alloprevotella*, *Veilonella*, *Haemophilus* [8,31,32,61–68]. In contrast, *Solobacterium* was reported increased [31], while *Slackia* and *Pseudomonas* were shown to be decreased [12,19] in fecal samples of CRC patients compared to healthy donors, and, in one epidemiological study, the presence of *Treponema* in

the oral cavity was associated with increased risk of CRC [29], but none of these were previously associated with tumour mucosa.

Amongst *tumour genera* of gut origin, *Bacteroides*, *Escherichia-Shigella*, *Coprobacter*, *Coprococcus 1*, *Peptoclostridium* (ex *Clostridium difficile*) [69], *Howardella*, *Prevotella 1*, *Romboutsia*, *Eggerthella*, *Morganella*, *Enterococcus* were previously associated with tumour mucosa [6,8,11,30,69–71], while *Lachnoclostridium*, *Flavonifractor* [72,73], *Sutterella* and *Hungatella* (ex-*Clostridium hathewayi*) [74] were only reported increased in stool of patients with CRC.

Most importantly, we newly identified genera of both oral and gut origin, not previously associated with colorectal cancer, with increased abundance in the tumour lesions: *Selenomonas 3*, *Selenomonas 4*, *Aggregatibacter*, *Actinobacillus*, *Bergeyella*, *Phocaeiola*, *Defluviitaleaceae UCG-011*, *Phocaeiola*, an uncultured species from *Veillonelaceae* family, and an uncultured species from *boneC3G7* at the family level (BLAST hit *Fusobacterium necrophorum*) (oral origin) and *Tyzzereella 4*, *Massilia*, and an unassigned genus from *Peptostreptococcaceae* family (gut origin). Some of these newly associated genera contain species that are known human pathogens causing infections of mucosal or other tissues. *Selenomonas*, *Phocaeiola* and newly established genera *Aggregatibacter* have been associated with periodontal disease [75–77]. *Bergeyella* and *Actinobacillus* genera comprises animal (zoonotic) pathogens causing infections in human through animal bites [78,79] and a few human pathogens causing endocarditis (*B. cardium*) or respiratory infections (*A. hominis*) [80,81]. In our study both genera were detected only in mucosal swabs of visually normal tissues or tumours. For other genera we newly associated with CRC, their potential involvement in CRC is not so obvious. The *tumour-specific* genera of *Defluviitaleaceae* might influence CRC through the metabolism of butyrate [82]. Butyrate is critical for intestinal microbial balance and colon health and the atrophy of butyrate metabolism was associated with the pathogenesis of various colonic diseases [83]. The association of *Tyzzereella 4* from the *Lachnospiraceae* family with CRC can be due to its increased occurrence in patients with higher cardiovascular risk (CVR) factor scores [84], that are associated also with CRC [85]. Presence of *Massilia* in CRC patients was also never described previously, but it was detected in patients with pancreatic cancer [86].

Indeed, some of the *tumour genera* were shown to be involved in the process of carcinogenesis, for some, a mechanism of action was described, but their role as bacterial drivers or passengers within the existing disease remains unknown. Correlating the tumour microbiome with clinical variables of tumour progression bears the promise of offering viable hypotheses on the role of bacteria in the progression of the disease. The identification of such associations within visually-normal mucosa or with the stool microbial composition has a significant potential in derivation of screening biomarkers. Currently, the association between clinical variables and gut microbial composition are understudied and only few efforts addressed the topic but on rather limited cohorts.

It is known that right and left-sided colorectal tumours are different in terms of genetic stability and prognosis [87] and, given the spatial variation in microbiome's taxonomic structure in different parts of gut [36], the effect of tumour localisation on grade-microbiota interaction should be estimated. So far, few studies have associated microbiome of CRC patients with the tumour grade using different approaches (real-time PCR, 16S rRNA sequencing) but without considering the interactions between tumour sidedness and grade. A study of *F. nucleatum's* DNA presence in formalin-fixed paraffin-embedded tumour slides showed that the proportion of this species increased with tumour histological grade [88]. Wu et al. [36] report 15 genera associated with tumour mucosa of moderately and poorly differentiated tumours compared to well differentiated tumours. In our study, the sample size allowed the inclusion of an interaction term, thus providing a finer estimation of differences in microbiome composition with respect to tumour grade. We reported 50 genera with association with grade and/or side for all tissue types studied, in total 66 of significant associations at FDR < 0.1. We confirmed previously reported high tumour grade associations of *Fusobacterium*, *Campylobacter* and *Mogibacterium*, in CRC tumour mucosa [36,88]. *Prevotella* and *Selenomonas* were associated with high grade (3), but only in the right-sided/transverse tumours.

Almost all these microorganisms belong to *tumour genera*, except for *Methanobrevibacter* (*stool genera*) and *Mogibacterium*, which showed no difference in abundance between the sample types. *Mogibacterium* is a known oral pathogen, but the mechanism of its action remains unknown [89].

In our analysis, we observed potentially beneficial effects of the increased abundance of 20 *stool genera* significantly associated with left location, namely decreasing tumour grade with increased abundance, e.g. *Bifidobacterium*, *Ruminococcaceae* UCG-010 and *Victivallis* in tumour mucosa; *Porphyromonas*, *Lachnospiraceae* UCG - 005 and *Gelria*, in stool. So, while these genera are significantly depleted in mucosal samples compared to stool, a significant difference in abundance between location remained. *Bifidobacterium* was previously shown to have anti-cancerogenic effects [73,90–93]. Similar association with left-sided and low-grade tumours showed 2 *tumour genera* (*Lachnospira*, *Gemella*) and *Methanobrevibacter*. Except for *Bifidobacterium*, none of these genera were previously reported to be associated with tumour grade.

We can only speculate whether the prolonged exposure of tumour mucosa to predominantly stool bacteria that is mechanistically related to tumours in distant part of the colon (left-sided or with onset in rectosigmoid and rectum) can have potentially harmful or beneficial effects or whether any associations are mostly due to the well-known molecular differences in the right vs left sided tumours [32,63,94].

In the study of Pu et al, *Fusobacterium*, *Corynebacterium*, *Enterococcus*, *Neisseria*, *Porphyromonas* and *Sclegella* were more abundant and *Oribacterium*, *Desulfovibrio*, *Clostridiales* and *Lactobacillus* were less abundant in the invasive cancer group [31]. As a result of our work we partially confirm these findings. In addition to *Fusobacterium*, we also identified *Campylobacter* to be increased in mucosa of advanced T stages tumours. On the contrary, we found *Corynebacterium*, *Enterococcus*, and *Neisseria* genera to be enriched in early stage tumours ($p < 0.05$, FDR > 0.1). This situation could be caused by a different T grouping strategy in our work and the absence of T4 tumours in the study of Pu et al. and by much larger sample size in our study (N = 186 compared to N = 25).

Early detection of local and distant metastases remains one of the most important tasks in cancer management. We confirmed a previously published increase of *Akkermansia* and *Porphyromonas* in stool of patients with local metastases ($p < 0.05$, FDR > 0.1) (82). Moreover, in our work, stool abundance of the uncultured genus from *Erysipelotrichaceae* family was directly associated with distant cancer metastases (FDR < 0.1), whereas in the study of Han et al. a close taxonomical relative of this genus, *Erysipelotrichaceae incertae sedis*, was linked with local metastasis presence. Family *Erysipelotrichaceae* is associated with the inflammation-related intestinal disorders, including inflammatory bowel diseases and Crohn's disease and has immunogenic properties [95]. In addition, abundance of this bacteria in stool was shown to be positively correlated with the blood level of TNF- α in human immunodeficiency virus infected individuals [96] and the TNF- α level itself is directly linked with rate of metastasis occurrence in colon cancer [97]. Based on this, we can speculate that *Erysipelotrichaceae*-associated inflammation in the gut of CRC patients could contribute to cancer metastasis rate.

Also, it is interesting to note that metastases occurrence is mainly associated with shifts in stool microbiome composition. On one side, it raises the possibility of potential microbiome-based non-invasive metastasis diagnostics in colorectal cancer or monitoring the patients at risk. On the other, it is a question still under discussion whether these changes are specific, or this alteration of the stool microbial community reflect the changes in overall health status in the presence of a metastasis and the cancer progression itself. As an example, it is known that non-colonic malignancies, for example breast cancer and lung cancer, are also accompanied by the shifts in gut microbiome [98–100].

While summarising the results, a common list of bacteria associated with tumour progression measured according to different criteria (histological grade, TNM stage, and T, M, N separately) was identified. Five bacterial genera including

Peptoclostridium, *Fusobacterium*, *Campylobacter*, *Streptococcus*, and *Akkermansia* were linked with three and more progression criteria at least for one specimen type, among them .

The screening potential of individual bacteria was assessed by pairwise analysis of incidence of all genera across sample types. The relevance of the observed differences is context dependent. On-tumour microbes with significant clinical associations and with no difference in incidence across samples types are perfect candidates for stool-based screening studies or stool-based prognostic and predictive classifiers. In our study these were 13 *tumour genera*, some of which previously associated with CRC like *Lachnoclostridium*, *Streptococcus*, *Bacteroides*, *Escherichia-Shigella*, *Coprobacter*, *Slackia* and *Sutterella*. Interestingly, *Fusobacterium*, the most studied bacterium in the context of CRC, the quantification of which was previously suggested for stool-based screening for advanced carcinoma [101] is only present in 63% of stool samples of patients with CRC. At the same time, in more than 31% patients with *Fusobacterium* present on tumour mucosa we failed to detect it in the stool. In fact, most of the *tumour-specific genera* if present on tumour mucosa, were not identified in stool of the same patients in more than 50% (e.g. *Haemophilus*, *Campylobacter*, *Gemella*, *Parvimonas*, *Leptotrichia*, *Solobacterium*, *Howardella*, *Hungatella*) or 80% (e.g. *Selenomonas*, *Selenomonas 3*, *Eigenella*, *Aggregatibacter*, *Massilia*, *Neisseria*) of cases. Given that these genera prefer the mucosal environment over stool, such associations are not entirely surprising. It needs to be emphasized, that although these genera are often significantly associated with clinical parameters, their stool screening potential remains poor and these genera are better candidates for colonoscopy biopsy sample screening.

The three tumour-mucosa based microbial subtypes we derived on patterns of similarity of abundance of the *tumour genera* were associated with tumour grade, location and stage. An interesting observation was that the tumour microbial subtypes differ not only in the type of the *tumour genera* they host, but also on the count of potentially pathogenic microbiome correlated with high grade and stage. Of the 13 high grade or high stage related genera, TS1 tumours had a median of 9 (69%), TS2 of 7 (54%) and TS3 of 5 (38%) differing this in what we could call “microbial pathogens burden” .

This subtyping could reflect differences in tumour biological properties linked with cancer progression: malignant tumours with active growth, cell and tissue atypia because of disruption of mucus layer and dysregulation of local immunity provides more comfortable conditions to aggressive microbial consortia expansion and unconventional (oral) species homing. Moreover, given the fact of bacteria-supported carcinogenesis, proofed in animal models [10,102], the pathogenic bacteria growth leads to additional dedifferentiation of tumour cells forming the pathogenetic loop. It is to be emphasized, that the proposed subtypes based on *tumour genera* are not comprehensive from the point of view of all the genera present on the tumour and other genera can still have important effects. As we discussed above, the presence of some *stool genera* on tumour mucosa of left-sided tumours was associated with low-grade.

It remains to be further investigated whether the subtypes could improve the prediction of patients’ survival and prognosis. We can speculate that high microbial pathogens burden could be worsening not only the tumour progression, but also potentially the patient’s condition after the surgical resection and during and after the chemotherapy treatment, since these genera are present also in visually-normal mucosa.

One possible complication, metachronous colorectal tumour, could potentially be associated with visually-normal mucosa microbiome alteration. Metachronous cancer, i.e. tumour occurring 6 months after the resection of primary tumour, is traditionally linked with hereditary conditions as familial adenomatous polyposis, but it also detected in patients above age 65 with sporadic cancer [103]. Given the fact that tumour-related genera reside also on visually-normal mucosa, they could initiate CRC tissue dysplastic changes and malignisation. There is limited evidence of linkage between mucosal microbiota and metachronous adenomas growing demonstrated by Liu et al [104]. On the other side, it is shown that the microbiome could interact and metabolise chemotherapeutic medicine which lead to

modulation of its activity and toxicity [105]. In the light of the above, modification of gut microbiome after colorectal cancer surgical removal might be considered as an additional step of treatment to prevent tumour recurrence and modulate chemotherapy effectiveness and toxicity.

Conclusion

In our study, by analysis of 505 samples from N = 186 patients, we extended the current characterization of colorectal cancer microbiome in several ways. Thanks to the large sample size, we identified bacterial genera that were not previously associated with CRC tumour mucosa, clinical variables or with colorectal carcinoma at all. These genera should be studied more in detail to describe their mechanism of interaction with the disease.

By focusing on the *microbial community* analysis, in contrast to classical *microbiome-centered* approaches we were able to identify co-occurring species and 3 major tumour-microbial subtypes that correlate with clinical variables, mainly grade, location and TNM staging. The subtypes also differ in what we describe as *microbial pathogens burden* – the number of pathogenic species correlated with increased grade and stage present on tumour mucosa, although the concept can be defined with respect to all three environments (tumour mucosa, visually-normal mucosa and stool).

It is well known that the gut microbial composition changes with dietary patterns and lifestyle, that could be region-based [106]. More studies of similar sample size or larger, from different geographical locations are needed to derive robust and generalizable patterns.

We make the full data available including clinical variables as a first step towards building a data corpus that could support such investigations. The technology chosen was high throughput, fitting the purpose of microbial community-based analysis. We did perform the sequence matching for the identified ASVs against the SILVA database, however, being aware of the limitations, we provide these results solely as supplementary information without discussing them here in detail.

Having sampled the microbiome at three different complementary sites allowed the study of several environments leading to the definition of novel microbial categories with multiple implications. Our study shows that the associations with clinical variables found for the tumour mucosal or adjacent visually-normal mucosa microbiome are rarely preserved in the microbial composition of stool and vice versa. While tumour histological grade, stage and location are reflected in corresponding mucosal microbiome, the presence of lymph node or distant metastases influences mainly stool microbiome. It seems that the mucosa and stool microbiome are complementary with respect to modulation of their effects on disease progression. Tumour-mucosa biopsies from colonoscopy might need to be coupled with stool sampling for efficient screening or diagnostic purposes.

Declarations

Ethics approval and consent to participate

The study was approved by the ethical committee of Masaryk Memorial Cancer Institute. All patients gave written informed consent in accordance with the Declaration of Helsinki prior to participating in the study.

Consent for publication

Not applicable.

Availability of data and material

Sequencing data were uploaded to the European Nucleotide Archive under accession number PRJEB35990.

Competing interests

The authors declare that they have no competing interests.

Funding

The work was supported by the Czech Ministry of Health (AZV 16-31966A).

Authors' contributions

EB, RS, RN and LD designed the study, BZ and SS performed bioinformatic analysis, BZ, ViaP, VlaP and EB performed statistical data analysis, MH, LM, NK and VB performed the sample processing, DNA isolation and sequencing; BZ, MH, ViaP, LM, EB, PV, LD and RH interpreted the results; BZ drafted the first version of the manuscript. All authors participated in writing or editing of the manuscript and approved the final manuscript.

Acknowledgements

We acknowledge the help of the volunteers enrolled in this study.

References

1. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer*. 2018;103:356–87.
2. Punt CJA, Koopman M, Vermeulen L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nature Reviews Clinical Oncology*. Nature Publishing Group; 2017;14:235–46.
3. Van der Jeught K, Xu H-C, Li Y-J, Lu X-B, Ji G. Drug resistance and new therapies in colorectal cancer. *World J Gastroenterol*. 2018;24:3834–48.
4. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst*. 2013;105:1907–11.
5. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science*. 2012;338:120–3.
6. Balamurugan R, Rajendiran E, George S, Samuel GV, Ramakrishna BS. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, *Desulfovibrio* and *Enterococcus faecalis* in the feces of patients with colorectal cancer. *J Gastroenterol Hepatol*. 2008;23:1298–303.
7. Chen J, Young SM, Allen C, Seeber A, Péli-Gulli M-P, Panchaud N, et al. Identification of a Small Molecule Yeast TORC1 Inhibitor with a Multiplex Screen Based on Flow Cytometry. *ACS Chem Biol*. 2012;7:715–22.
8. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE*. 2012;7:e39743.
9. Cipe G, Idiz UO, Firat D, Bektasoglu H. Relationship between intestinal microbiota and colorectal cancer. *World Journal of Gastrointestinal Oncology*. 2015;7:233–40.
10. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*. 2013;14:207–15.
11. Lu Y, Chen J, Zheng J, Hu G, Wang J, Huang C, et al. Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas. *Sci Rep*. 2016;6:26337.

12. Marchesi JR, Dutilh BE, Hall N, Peters WHM, Roelofs R, Boleij A, et al. Towards the human colorectal cancer microbiome. *PLoS ONE*. 2011;6:e20447.
13. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun*. 2015;6:8727.
14. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/ β -catenin signaling via its FadA adhesin. *Cell Host Microbe*. 2013;14:195–206.
15. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, et al. Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS ONE*. 2011;6:e16393.
16. Viljoen KS, Dakshinamurthy A, Goldberg P, Blackburn JM. Quantitative Profiling of Colorectal Cancer-Associated Bacteria Reveals Associations between *Fusobacterium* spp., Enterotoxigenic *Bacteroides fragilis* (ETBF) and Clinicopathological Features of Colorectal Cancer. *PLOS ONE*. 2015;10:e0119462.
17. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J*. 2012;6:320–9.
18. Wu N, Yang X, Zhang R, Li J, Xiao X, Hu Y, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb Ecol*. 2013;66:462–70.
19. Yang J, McDowell A, Kim EK, Seo H, Lee WH, Moon C-M, et al. Development of a colorectal cancer diagnostic model and dietary risk assessment through gut microbiome analysis. *Exp Mol Med* [Internet]. 2019 [cited 2020 Jan 28];51. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6802675/>
20. Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*. 2017;66:70–8.
21. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, et al. The Gut Microbiome Modulates Colon Tumorigenesis. *mBio* [Internet]. 2013 [cited 2020 Jan 28];4. Available from: <https://mbio.asm.org/content/4/6/e00692-13>
22. Zackular JP, Rogers MAM, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev Res (Phila)*. 2014;7:1112–21.
23. Vaupel P, Harrison L. Tumor hypoxia: causative factors, compensatory mechanisms, and cellular response. *Oncologist*. 2004;9 Suppl 5:4–9.
24. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol*. 2014;12:661–72.
25. Tlaskalova-Hogenova H, Vannucci L, Klimesova K, Stepankova R, Krizan J, Kverka M. Microbiome and colorectal carcinoma: insights from germ-free and conventional animal models. *Cancer J*. 2014;20:217–24.
26. Xiao Y, Freeman GJ. The microsatellite instable subset of colorectal cancer is a particularly good candidate for checkpoint blockade immunotherapy. *Cancer Discov*. 2015;5:16–8.
27. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol*. 2012;10:575–82.
28. Pennisi E. Cancer Therapies Use a Little Help From Microbial Friends. *Science*. 2013;342:921–921.
29. Yang Y, Cai Q, Shu X-O, Steinwandel MD, Blot WJ, Zheng W, et al. Prospective study of oral microbiome and colorectal cancer risk in low-income and African American populations. *Int J Cancer*. 2019;144:2381–9.
30. Liu C-J, Zhang Y-L, Shang Y, Wu B, Yang E, Luo Y-Y, et al. Intestinal bacteria detected in cancer and adjacent tissue from patients with colorectal cancer. *Oncol Lett*. 2019;17:1115–27.
31. Pu LZCT, Yamamoto K, Honda T, Nakamura M, Yamamura T, Hattori S, et al. Microbiota profile is different for early and invasive colorectal cancer and is consistent throughout the colon. *Journal of Gastroenterology and Hepatology*.

2020;35:433–7.

32. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut*. 2017;66:633–43.
33. Gao Z, Guo B, Gao R, Zhu Q, Qin H. Microbiota dysbiosis is associated with colorectal cancer. *Front Microbiol* [Internet]. 2015 [cited 2019 Mar 9];6. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313696/>
34. Li E, Hamm CM, Gulati AS, Sartor RB, Chen H, Wu X, et al. Inflammatory bowel diseases phenotype, *C. difficile* and NOD2 genotype are associated with shifts in human ileum associated microbial composition. *PLoS ONE*. 2012;7:e26284.
35. Han S, Wu W, Da M, Xu J, Zhuang J, Zhang L, et al. Adequate Lymph Node Assessments and Investigation of Gut Microorganisms and Microbial Metabolites in Colorectal Cancer. *OTT*. 2020;Volume 13:1893–906.
36. Wu Y, Shi L, Li Q, Wu J, Peng W, Li H, et al. Microbiota Diversity in Human Colorectal Cancer Tissues Is Associated with Clinicopathological Features. *Nutrition and Cancer*. Routledge; 2019;71:214–22.
37. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol*. 2015;75:129–37.
38. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*. 2011;108:4516–22.
39. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
40. Aronesty E. Comparison of Sequencing Utility Programs. *TOBIOIJ*. 2013;7:1–8.
41. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*. 2007;35:7188–96.
42. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
43. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. 2010;7:335.
44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology*. 1990;215:403–10.
45. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*. 2005;71:8228–35.
46. Aitchison J. *The statistical analysis of compositional data*. London; New York: Chapman and Hall; 1986.
47. Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It's all relative: analyzing microbiome data as compositions. *Annals of Epidemiology*. 2016;26:322–9.
48. Martín-Fernández J-A, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Statistical Modelling: An International Journal*. 2015;15:134–58.
49. Tsilimigras MCB, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of Epidemiology*. 2016;26:330–5.
50. Oaksen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. *vegan: Community Ecology Package*. R package. 2019.
51. Comas-Cufí M. *coda.base: A Basic Set of Functions for Compositional Data Analysis*. 2020.
52. Kloeke JD, McKean JW. Rfit: Rank-based Estimation for Linear Models. *The R journal*. 2012;4.

53. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995;57:289–300.
54. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2.
55. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9.
56. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, et al. *gplots: Various R Programming Tools for Plotting Data*. 2020.
57. Wickham H. *ggplot2: elegant graphics for data analysis*. Second edition. Cham: Springer; 2016.
58. Wray CM, Ziogas A, Hinojosa MW, Le H, Stamos MJ, Zell JA. Tumor Subsite Location Within the Colon Is Prognostic for Survival After Colon Cancer Diagnosis: *Diseases of the Colon & Rectum*. 2009;52:1359–66.
59. Han YW, Wang X. Mobile microbiome: oral bacteria in extra-oral infections and inflammation. *J Dent Res*. 2013;92:485–91.
60. Schmidt TS, Hayward MR, Coelho LP, Li SS, Costea PI, Voigt AY, et al. Extensive transmission of microbes along the gastrointestinal tract. *eLife*. 2019;8:e42693.
61. Allali I, Delgado S, Marron PI, Astudillo A, Yeh JJ, Ghazal H, et al. Gut microbiome compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain. *Gut Microbes*. 2015;6:161–72.
62. Drewes JL, White JR, Dejea CM, Fathi P, Iyadorai T, Vadivelu J, et al. High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes* [Internet]. 2017 [cited 2020 Apr 13];3. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5707393/>
63. Gao R, Kong C, Huang L, Li H, Qu X, Liu Z, et al. Mucosa-associated microbiota signature in colorectal cancer. *Eur J Clin Microbiol Infect Dis*. 2017;36:2073–83.
64. Ge W, Hu H, Cai W, Xu J, Hu W, Weng X, et al. High-risk Stage III colon cancer patients identified by a novel five-gene mutational signature are characterized by upregulation of IL-23A and gut bacterial translocation of the tumor microenvironment. *Int J Cancer*. 2019;
65. Hibberd AA, Lyra A, Ouwehand AC, Rolny P, Lindegren H, Cedgård L, et al. Intestinal microbiota is altered in patients with colon cancer and modified by probiotic intervention. *BMJ Open Gastroenterol* [Internet]. 2017 [cited 2020 Apr 13];4. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5609083/>
66. Kim K, Castro EJT, Shim H, Advincula JVG, Kim Y-W. Differences Regarding the Molecular Features and Gut Microbiota Between Right and Left Colon Cancer. *Ann Coloproctol*. 2018;34:280–5.
67. Leung PHM, Subramanya R, Mou Q, Lee KT, Islam F, Gopalan V, et al. Characterization of Mucosa-Associated Microbiota in Matched Cancer and Non-neoplastic Mucosa From Patients With Colorectal Cancer. *Front Microbiol*. 2019;10:1317.
68. Pasquereau-Kotula E, Martins M, Aymeric L, Dramsi S. Significance of *Streptococcus gallolyticus* subsp. *gallolyticus* Association With Colorectal Cancer. *Front Microbiol* [Internet]. 2018 [cited 2020 Apr 13];9. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5891635/>
69. Jahani-Sherafat S, Azimirad M, Alebouyeh M, Ahmadi Amoli H, Hosseini P, Ghasemian-Safaei H, et al. The rate and importance of *Clostridium difficile* in colorectal cancer patients. *Gastroenterol Hepatol Bed Bench*. 2019;12:358–63.
70. de Almeida CV, Taddei A, Amedei A. The controversial role of *Enterococcus faecalis* in colorectal cancer. *Therap Adv Gastroenterol*. 2018;11:175628481878360.
71. Soldevila Boixader L, Berbel D, Pujol M. Bacteriemia por *Eggerthella lenta* asociada a poliposis colónica y adenocarcinoma de colon. *Medicina Clínica*. 2017;149:418–9.

72. De Almeida C, Lulli M, di Pilato V, Schiavone N, Russo E, Nannini G, et al. Differential Responses of Colorectal Cancer Cell Lines to *Enterococcus faecalis*' Strains Isolated from Healthy Donors and Colorectal Cancer Patients. *JCM*. 2019;8:388.
73. Gupta A, Dhakan DB, Maji A, Saxena R, P.k VP, Mahajan S, et al. Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. *mSystems* [Internet]. American Society for Microbiology Journals; 2019 [cited 2020 Apr 13];4. Available from: <https://msystems.asm.org/content/4/6/e00438-19>
74. Ai D, Pan H, Li X, Gao Y, Liu G, Xia LC. Identifying Gut Microbiota Associated With Colorectal Cancer Using a Zero-Inflated Lognormal Model. *Front Microbiol* [Internet]. Frontiers; 2019 [cited 2020 Apr 13];10. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00826/full>
75. Gonçalves LFH, Fermiano D, Feres M, Figueiredo LC, Teles FRP, Mayer MPA, et al. Levels of *Selenomonas* species in generalized aggressive periodontitis. *J Periodontal Res*. 2012;47:711–8.
76. Scher JU, Ubeda C, Equinda M, Khanin R, Buischi Y, Viale A, et al. Periodontal disease and the oral microbiota in new-onset rheumatoid arthritis. *Arthritis & Rheumatism*. 2012;64:3083–94.
77. Herbert BA, Novince CM, Kirkwood KL. *Aggregatibacter actinomycetemcomitans*, a potent immunoregulator of the periodontal host defense system and alveolar bone homeostasis. *Mol Oral Microbiol*. 2016;31:207–27.
78. Lin W-R, Chen Y-S, Liu Y-C. Cellulitis and Bacteremia Caused by *Bergeyella zoohelcum*. *Journal of the Formosan Medical Association*. 2007;106:573–6.
79. Peel MM, Hornidge KA, Luppino M, Stacpoole AM, Weaver RE. *Actinobacillus* spp. and related bacteria in infected wounds of humans bitten by horses and sheep. *J Clin Microbiol*. 1991;29:2535–8.
80. Sohn KM, Huh K, Baek J-Y, Kim Y-S, Kang C-I, Peck KR, et al. A new causative bacteria of infective endocarditis, *Bergeyella cardium* sp. nov. *Diagn Microbiol Infect Dis*. 2015;81:213–6.
81. Friis-Møller A, Christensen JJ, Fussing V, Hesselbjerg A, Christiansen J, Bruun B. Clinical Significance and Taxonomy of *Actinobacillus hominis*. *J Clin Microbiol*. 2001;39:930–5.
82. Zha Z, Lv Y, Tang H, Li T, Miao Y, Cheng J, et al. An orally administered butyrate-releasing xylan derivative reduces inflammation in dextran sulphate sodium-induced murine colitis. *International Journal of Biological Macromolecules*. 2020;156:1217–33.
83. Leonel AJ, Alvarez-Leite JI. Butyrate: implications for intestinal function. *Curr Opin Clin Nutr Metab Care*. 2012;15:474–9.
84. Kelly TN, Bazzano LA, Ajami NJ, He H, Zhao J, Petrosino JF, et al. Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. *Circ Res*. 2016;119:956–64.
85. Niederseer D, Bracher I, Stadlmayr A, Huber-Schönauer U, Plöderl M, Obeid S, et al. Association between Cardiovascular Risk and Diabetes with Colorectal Neoplasia: A Site-Specific Analysis. *J Clin Med* [Internet]. 2018 [cited 2020 Aug 19];7. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6306807/>
86. Mei Q-X, Huang C-L, Luo S-Z, Zhang X-M, Zeng Y, Lu Y-Y. Characterization of the duodenal bacterial microbiota in patients with pancreatic head cancer vs. healthy controls. *Pancreatology*. 2018;18:438–45.
87. Ternes D, Karta J, Tsenkova M, Wilmes P, Haan S, Letellier E. Microbiome in Colorectal Cancer: How to Get from Meta-omics to Mechanism? *Trends in Microbiology*. 2020;28:401–23.
88. Ito M, Kanno S, Noshō K, Sukawa Y, Mitsuhashi K, Kurihara H, et al. Association of *Fusobacterium nucleatum* with clinical and molecular features in colorectal serrated pathway. *International Journal of Cancer*. 2015;137:1258–68.
89. Marchesan JT, Morelli T, Moss K, Barros SP, Ward M, Jenkins W, et al. Association of *Synergistetes* and Cyclodipeptides with Periodontitis. *J Dent Res*. 2015;94:1425–31.

90. Bahmani S, Azarpira N, Moazamian E. Anti-colon cancer activity of Bifidobacterium metabolites on colon cancer cell line SW742. *Turk J Gastroenterol.* 2019;30:835–42.
91. Mangifesta M, Mancabelli L, Milani C, Gaiani F, de'Angelis N, de'Angelis GL, et al. Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer. *Scientific Reports.* Nature Publishing Group; 2018;8:13974.
92. Parisa A, Roya G, Mahdi R, Shabnam R, Maryam E, Malihe T. Anti-cancer effects of Bifidobacterium species in colon cancer cells and a mouse model of carcinogenesis. *PLOS ONE.* Public Library of Science; 2020;15:e0232930.
93. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM, et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science.* 2015;350:1084–9.
94. Xi Y, Yuefen P, Wei W, Quan Q, Jing Z, Jiamin X, et al. Analysis of prognosis, genome, microbiome, and microbial metabolome in different sites of colorectal cancer. *J Transl Med [Internet].* 2019 [cited 2020 May 17];17. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6819376/>
95. Kaakoush NO. Insights into the Role of Erysipelotrichaceae in the Human Host. *Front Cell Infect Microbiol [Internet].* Frontiers; 2015 [cited 2020 Jun 22];5. Available from: <https://www.frontiersin.org/articles/10.3389/fcimb.2015.00084/full>
96. Dinh DM, Volpe GE, Duffalo C, Bhalchandra S, Tai AK, Kane AV, et al. Intestinal Microbiota, Microbial Translocation, and Systemic Inflammation in Chronic HIV Infection. *J Infect Dis.* Oxford Academic; 2015;211:19–27.
97. Zhao P, Zhang Z. TNF- α promotes colon cancer cell migration and invasion by upregulating TROP-2. *Oncology Letters.* Spandidos Publications; 2018;15:3820–7.
98. Luu TH, Michel C, Bard J-M, Dravet F, Nazih H, Bobin-Dubigeon C. Intestinal Proportion of *Blautia* sp. is Associated with Clinical Stage and Histoprognostic Grade in Patients with Early-Stage Breast Cancer. *Nutrition and Cancer.* 2017;69:267–75.
99. Wu AH, Tseng C, Vigen C, Yu Y, Cozen W, Garcia AA, et al. Gut microbiome associations with breast cancer risk factors and tumor characteristics: a pilot study. *Breast Cancer Res Treat.* 2020;182:451–63.
100. Zhuang H, Cheng L, Wang Y, Zhang Y-K, Zhao M-F, Liang G-D, et al. Dysbiosis of the Gut Microbiome in Lung Cancer. *Front Cell Infect Microbiol.* 2019;9:112.
101. Wong SH, Kwong TNY, Chow T-C, Luk AKC, Dai RZW, Nakatsu G, et al. Quantitation of faecal *Fusobacterium* improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut.* 2017;66:1441–8.
102. He Z, Gharaibeh RZ, Newsome RC, Pope JL, Dougherty MW, Tomkovich S, et al. *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut.* 2019;68:289–300.
103. Lindberg LJ, Ladelund S, Bernstein I, Therkildsen C, Nilbert M. Risk of Synchronous and Metachronous Colorectal Cancer: Population-Based Estimates in Denmark with Focus on Non-Hereditary Cases Diagnosed After Age 50. *Scand J Surg.* 2019;108:152–8.
104. Liu Y, Geng R, Liu L, Jin X, Yan W, Zhao F, et al. Gut Microbiota-Based Algorithms in the Prediction of Metachronous Adenoma in Colorectal Cancer Patients Following Surgery. *Front Microbiol.* 2020;11:1106.
105. Alexander JL, Wilson ID, Teare J, Marchesi JR, Nicholson JK, Kinross JM. Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat Rev Gastroenterol Hepatol.* 2017;14:356–65.
106. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature.* 2018;555:210–5.

Figures

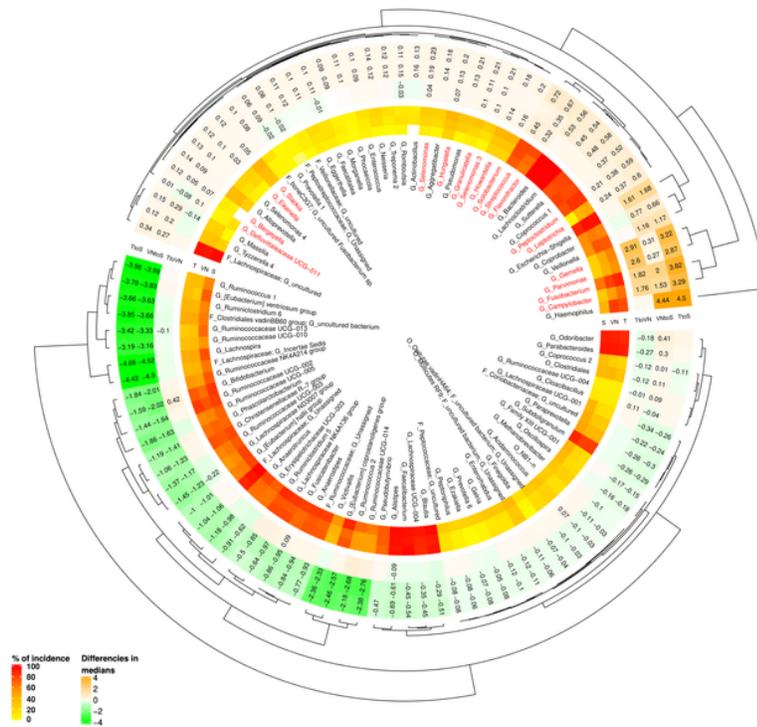


Figure 1

Differences in microbiome composition across the sample types. Circular heatmap summarizes 104 differentially abundant genera as identified in analysis between sample types on 133 triplets. Inner circle represents proportion of samples with the genus present in each sample type. The outer circle represents colour-coded difference of medians of clr-transformed abundance between sample types. The numbers (differences) are shown only if the difference was statistically significant (Friedman test $FDR < 0.1$, paired Wilcoxon test $FDR < 0.1$). Dendrograms represent the similarity in differences patterns of studied microorganisms. Top semiring represents tumour genera while tumour-specific genera names are colored red. T – incidence in tumor tissue, VN – incidence in visually-normal tissue, S – incidence in stool specimens. Numbers were shown only for genera significant after the post-hoc paired Wilcoxon test (q -value < 0.1).

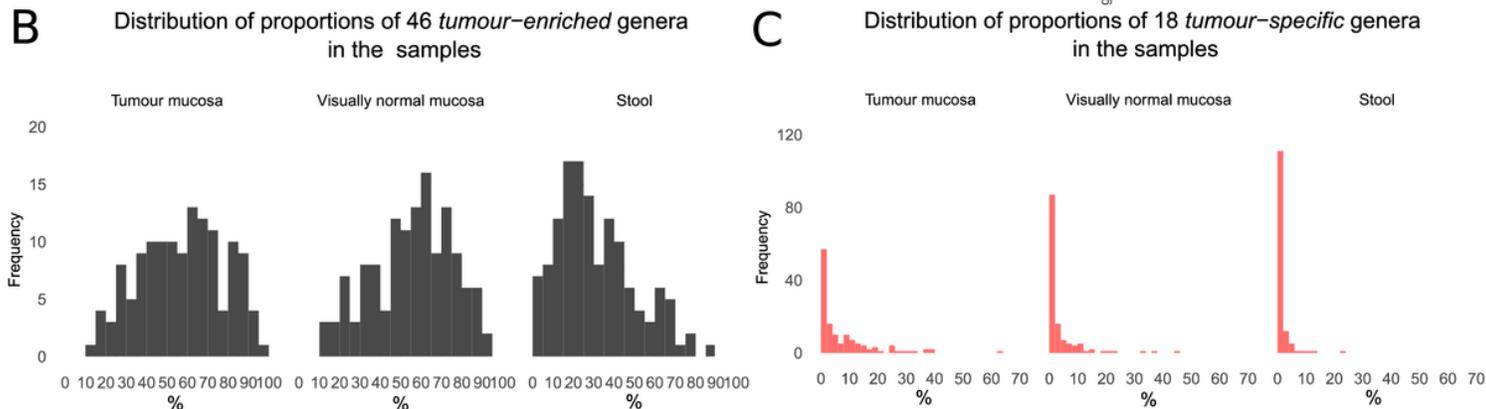
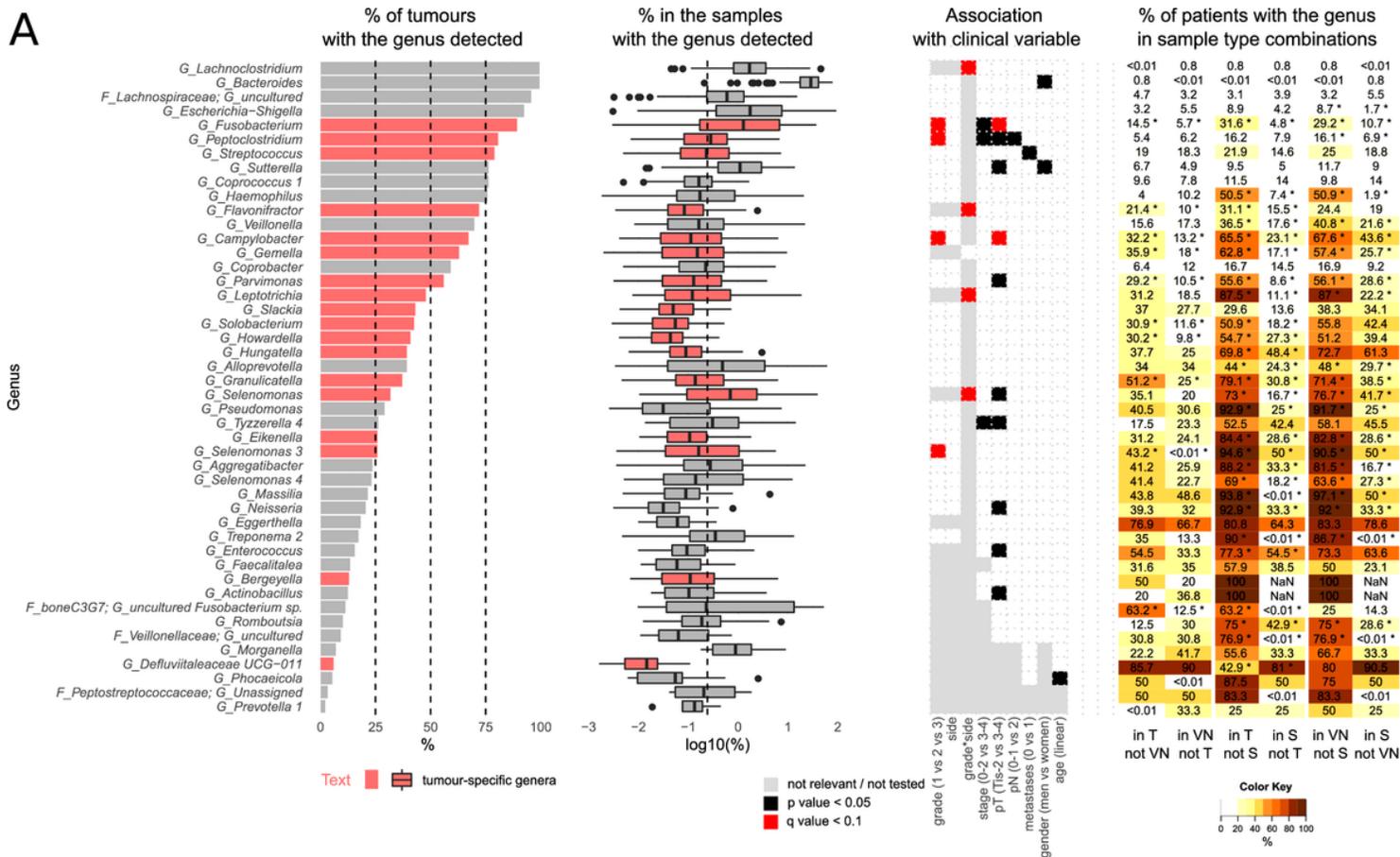


Figure 2

Tumour genera. A. (left) Proportion of tumours with the genera detected and distribution of their respective relative abundances (in %) in the samples where they were detected (middle) and association of the bacteria with clinical variables (right). Vertical dashed line represents median relative abundance of all 268 detected genera (med=0.24%). B. Overall proportion the 46 tumour genera in the three sample types (N=133) C. Overall proportion of the 18 tumour-specific genera in the three sample types (N=133).

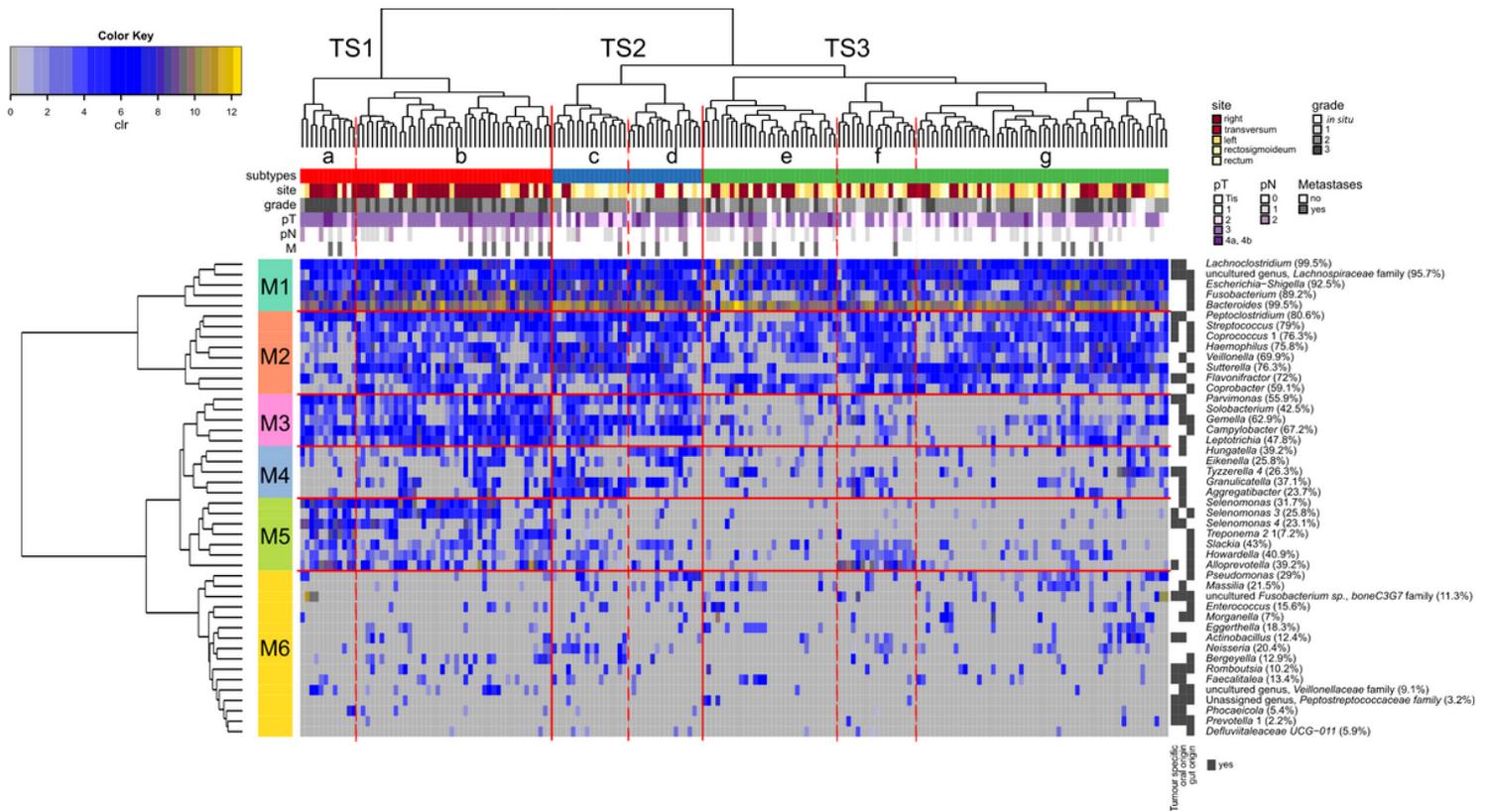


Figure 3

Tumour microbial subtypes. Hierarchical clustering of tumours (Aitchinson distance) and genera (Euclidean distance) based on the clr-transformed abundances of 46 tumour genera. Clinical variables of individual patients are shown. Proportions right to the genera name denote incidence of the genus in 186 tumour-mucosa samples.

Supplementary Files

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

- [TableS10.xlsx](#)
- [TableS9.xlsx](#)
- [TableS8.xlsx](#)
- [TableS7.xlsx](#)
- [TableS6.xlsx](#)
- [TableS5.xlsx](#)
- [TableS4.xlsx](#)
- [TableS3.xlsx](#)
- [TableS2.xlsx](#)
- [TableS1.xlsx](#)
- [FigS15.tif](#)
- [FigS14.tif](#)
- [FigS13.tif](#)

- FigS12.tif
- FigS11.tif
- FigS10.tif
- FigS9.tif
- FigS8.tif
- FigS7.tif
- FigS6.tif
- FigS5.tif
- FigS4.tif
- FigS3.tif
- FigS2.tif
- FigS1.tif
- TextS1.docx