

## **The local tumor microbiome is associated with survival in late-stage colorectal cancer patients**

Justine W. Debelius<sup>1,2,\*</sup>, Lars Engstrand<sup>1</sup>, Andreas Matussek<sup>3,4,5</sup>, Nele Brusselaers<sup>1,6,7</sup>, James T. Morton<sup>8</sup>, Margaretha Stenmarker<sup>9,10</sup>, Renate S. Olsen<sup>1,9,11\*</sup>

<sup>1</sup>Centre for Translational Microbiome Research, Department of Microbiology, Tumor, and Cell biology. Karolinska Institutet, Solna, Sweden

<sup>2</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health. Baltimore, MD, USA

<sup>3</sup>Laboratory Medicine, Jönköping Region County, Department of Clinical and Experimental Medicine, Linköping University, Jönköping, Sweden

<sup>4</sup>Division of Laboratory Medicine, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

<sup>5</sup>Department of Microbiology, Division of Laboratory Medicine, Oslo University Hospital, Oslo, Norway

<sup>6</sup>Global Health Institute, Antwerp University, Antwerp, Belgium

<sup>7</sup>Department of Head and Skin, Ghent University, Ghent, Belgium

<sup>8</sup> Biostatistics & Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

<sup>9</sup>Futurum/Department of Pediatrics, Region Jönköping County, and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

<sup>10</sup>The Institute of Clinical Sciences, Department of Paediatrics, the Sahlgrenska Academy at the University of Gothenburg, Sweden

<sup>11</sup>Department of Pathology, Division of Laboratory Medicine, Oslo University Hospital, Norwegian Radium Hospital, Oslo, Norway.

\*Corresponding Authors: Renate Slind Olsen, e-mail:resols@ous-hf.no; Justine Debelius email: justine.debelius@jhu.edu

Word Count: 3795

## ABSTRACT

Objective Colorectal cancer survival has been linked to the microbiome. Single organism analyses suggest *Fusobacterium nucleatum* as a marker of poor prognosis. However, *in situ* imaging of tumors demonstrate a polymicrobial tumor-associated community. To understand the role of these polymicrobial communities in survival, we performed an untargeted study of the microbiome in late-stage colorectal cancer patients.

Design We conducted a nested case-control study in late-stage cancer patients undergoing resection for primary adenocarcinoma. The microbiome of paired colorectal tumor and adjacent tissue samples was profiled using 16S rRNA sequencing; we used compositionally aware ordination and differential ranking to profile the microbial community.

Results. We found a consistent difference in the microbiome between paired tumor and adjacent tissue, despite strong individual microbial identities. Tumors had higher relative abundance of genus *Fusobacteria* and *Campylobacter* at the expense of members of families Lachnospriaceae and Rumminococeae. Furthermore, a larger difference between normal and tumor tissue was associated with prognosis: patients with shorter survival had a larger difference between normal and tumor tissue. We found the difference was specifically related to taxa previously associated with cancer. Within the tumor tissue, we identified a 39 member community statistic associated with survival; for every log<sub>2</sub> fold increase in this value, an individual's odds of survival increased by 20% (OR survival 1.20; 95% CI 1.04, 1.33).

Conclusion Our results suggest that a polymicrobial tumor-specific microbiome is associated with survival in late-stage colorectal cancer patients.

## 1 INTRODUCTION

2

3 Globally, colorectal cancer (CRC) is the second most common cause of cancer-related death and  
4 CRC-related mortality has been increasing since 2000 (1,2). One potentially modifiable area of  
5 interest in CRC survival is the gut microbiome. In a healthy gut, the intestinal microbiome  
6 contributes to homeostasis through epithelial cell renewal, maintaining gut barrier integrity, and  
7 immune modulation (3,4). However, CRC patients have demonstrated a consistently altered gut  
8 microbiome compared to healthy controls, including a higher relative abundance of organisms  
9 more commonly found in the oral cavity (5,6). Meta-analyses using targeted analyses show high  
10 levels of *Fusobacterium nucleatum* (*F. nucleatum*) in tumor tissue are detrimental to survival  
11 (7,8).

12

13 Fewer studies have explored the relationship between the gut microbiome and CRC prognosis  
14 using untargeted sequencing. Untargeted techniques can better characterize the bacterial  
15 community, and the ways in which potentially pathogenic organisms might interact with a host's  
16 unique, stable, microbiome (9–11). *In situ* microscopy shows that tumor tissue is colonized by a  
17 polymicrobial biofilm including Fusobacteria, Proteobacteria, Bacteroidetes, and  
18 Lachnospriaceae; monoculture biofilms have not been observed (12). Biofilms are also  
19 frequently localized to tumors, and paired normal tissue samples are rarely colonized, suggesting  
20 a localized effect and potential difference between tumor and adjacent tissue (12).

21

22 Previous untargeted studies of the gut microbiome and colorectal cancer survival have either  
23 focused exclusively on the tumor tissue (13) or have treated the tumor and adjacent tissue as

24 identical (14). Paired biopsy studies provide clues about whether local and global regulation of  
25 the microbiome drives tumorigenesis, although many paired studies have failed to account for  
26 survival (12,15–19), and in some cases, struggled to characterize the microbiome due to  
27 technical (19) or analytical (13–17) issues.

28

29 To address the gaps in knowledge, we followed 101 late-stage CRC patients recruited from a  
30 hospital in southern Sweden who underwent surgical resections of primary adenocarcinoma  
31 between 1997-2017. Patients were categorized into short- or long-term survivors based on their  
32 relapse free survival (less than two years or five or more years, respectively). We examined the  
33 relationship between the microbiome of colorectal tumors and adjacent normal tissue and  
34 survival, accounting for clinical covariates.

35

## 36 **METHODS**

37

### 38 **Study Population**

39 Patients were recruited from all consecutive CRC patients (n=540) who underwent surgical  
40 resection for primary colorectal adenocarcinoma at the Department of Surgery, Ryhov County  
41 Hospital, Region of Jönköping County, Jönköping, Sweden between 1997 and 2017. Patients  
42 with tumor-node-metastasis (TNM) stage III and IV cancer who had matched biopsies from  
43 normal and tumor tissue (n=116) were selected. Patient characteristics, including demographic,  
44 surgical, pathological information, and outcome were determined from a review of medical  
45 records.

46

47 The final study cohort included patients with paired high quality microbiome samples (n=101).  
48 Fifteen individuals were excluded due to insufficient sequencing depth in the tumor (n=8) or  
49 normal (n=7) tissue sample. There was no difference in the survival status in the samples with  
50 insufficient sequencing depth. Our final cohort included matched tumor- and normal tissue  
51 samples ( $\geq 10$  cm apart from tumor tissue) from 51 long- ( $\geq 5$  year survival) and 50 short-term ( $\leq 2$   
52 year survival) survivors.

53  
54 The study was approved by the Regional Ethical Review Board in Linköping, Linköping,  
55 Sweden (98113, 2013/271-31); a written informed consent was obtained from each patient.

56  
57 There was no patient or public involvement in this retrospective study.

### 58 59 **Statistical Analysis of patient characteristics**

60 A multivariable logistic regression was used to assess the predictive impact of the following  
61 patient, cancer and treatment related characteristics: age (categorized as  $<60$ , 60-69, 70-74,  $\geq 75$   
62 years), sex (female or male), American Society of Anesthesiologists physical status (ASA) score  
63 (I-healthy, II-mild, III-IV-severe; patients with V-VI were not eligible for surgery); localization  
64 of tumor (colon right, colon left, rectum), TNM stage (III or IV), grade of differentiation (from  
65 low differentiation to high differentiation, with the latter more closely resembling non-cancer  
66 histology); radical surgery (yes or no); period of surgery (1997-2005; 2006-2010; 2011-2017).  
67 All results are expressed as Odds Ratios (ORs) and 95% Confidence Intervals (CIs) and the  
68 calculations were conducted with Stata MP14 (Stata Corp, Texas).

69

## 70 **Microbiome Sequencing**

71 Paired tumor and normal tissue samples were collected during surgery. Tissue samples were  
72 frozen directly and stored at -80°C until use. Samples were processed as previously described  
73 (20). Briefly, DNA was extracted from tissue samples using physical and chemical lysis for  
74 extraction. The 16S rRNA amplicon library was amplified with 341F/805R primers  
75 (CCTACGGGNGGCWGCAG, GGACTACHVGGGTATCTAAT) using a program with 20  
76 cycles (21). The samples were sequenced with a 2x300 approach using an Illumina MiSeq (San  
77 Diego, CA, USA).

78  
79 The demultiplexed reads were denoised using the DADA2 algorithm (v1.13.1) in R (22). After  
80 reads were demultiplexed and primers were trimmed, forward reads were trimmed to 265  
81 nucleotides (nt) and reverse reads were trimmed to 225nt; the error rate model was trained on  
82 15% of the reads. Reads were joined with at least 30nt overlap, and anything shorter than 380nt  
83 after joining was discarded. Taxonomic assignment was performed using the naïve Bayesian  
84 classifier implemented in DADA2 against the Silva 128 database (23,24). The amplicon  
85 sequence variant (ASV) table from DADA2, taxonomy, and representative sequences were  
86 imported into QIIME 2 (v. 2020.11) for further processing (25). A phylogenetic tree was built  
87 using fragment insertion using the SEPP algorithm into the Silva 128 backbone with q2-  
88 fragment-insertion (24–26). The table and sequences were filtered to exclude any ASV without  
89 phylum level annotation or which could not be inserted into the phylogenetic tree.

90

## 91 **Microbiome Community Characterization**

92 *Between sample (beta) diversity*



93 For paired sample analysis, we calculated unweighted UniFrac (27), weighted UniFrac (28), and  
94 binary Jaccard (29) distances and Bray-Curtis dissimilarity (30) on a feature table rarified to  
95 2500 sequences/sample (31). Aitchison distance was calculated on unrarefied data with a  
96 pseudo-count of 1 (32,33). Beta diversity metrics were calculated using the q2-diversity plugin in  
97 QIIME 2 (25).

98

### 99 *Compositional Tensor Factorization ordination*

100 To account for subject-specific effects on ordinations, we used compositional tensor fraction  
101 (CTF) for paired samples using the Gemelli qiime2 plugin (0.7.0) (34). Features were filtered to  
102 exclude those present in fewer than 20 samples or less than 100 total counts. The distance in CTF  
103 subject space was calculated as the Euclidean distance between subject coordinates. The  
104 difference in intra-individual CTF space between normal and tumor tissue ( $\Delta PC$ ) were compared  
105 using the subject-state coordinates and feature-state coordinates, respectively.

106

### 107 *Robust Principal Components Analysis*

108 For each tissue type, we examined beta diversity using a Robust Principal Components Analysis  
109 (rPCA) using the DEICODE algorithm (v. 0.2.4) (35). For a given sample set, we filtered  
110 filtering features present in less than 10% of samples (n=10) or with fewer than 10 total counts.  
111 The auto-rPCA function was used to select the appropriate number of principle components  
112 (PCs) for the data. The PCs were divided into quartiles and dichotomized along the median  
113 value.

114

115 For tissue types where there was a significant association between a component and survival, we  
116 selected features which might be associated with outcome. Communality was calculated as the  
117 square root of the sum of squares across all PC. Features with a communality value of at least  
118 0.01 were selected as candidates for the additive log ratio (ALR) calculation (n=130). A  
119 pseudocount of 1 was added before the ALR calculation. The ALR was calculated as the log<sub>2</sub>  
120 ratio of features more extreme than the fourth quartile of samples over feature more extreme than  
121 the first quartile. Continuous ALR values or ALR divided into tertials were used for regression.

122

### 123 *Differential Ranking*

124 We performed hypothesis generating differential abundance testing between tumor and normal  
125 tissue using a modified differential ranking (DR) technique (36,37). We first filtered the table to  
126 remove any feature with a relative abundance of less than 1/1000 in fewer than 10% of samples,  
127 leaving 243 features for testing. We then used a modified Bayesian method for DR testing. ASV  
128 counts were modeled through a negative binomial process. We started with naive priors of a  
129 between 0-and 5-fold change in a ASV and fit the model using 4000 iterations. The data was fit  
130 to a linear mixed effects model using subject as a random intercept, modeling either for tissue or  
131 for the intersection between tissue and survival. Modeling was done with pystan (v 3.4.0) within  
132 the QIIME 2 2021.11 conda environment (38,39).

133

134 We used the ranks to identify “extreme” features. Starting from the feature with the strongest  
135 signal associated with each possible value for a variable (e.g. normal vs tumor, short vs long  
136 survival), we added features until every sample contained at least one of the extreme features. A

137 pooled ALR was calculated as the sum of all normal-tissue associated features over the tumor-  
138 associated features.

139

## 140 **Statistical Analysis**

141 Paired distances were extracted as the distance between an individual's tumor and adjacent  
142 normal tissue. Interindividual distance was compared to the interindividual distance to samples  
143 of the same tissue type, anatomical location, and survival group with a permutative two sample t-  
144 test with 999 permutations.

145

146 Associations with per-subject CTF coordinates were checked by calculating the Euclidean  
147 distance between samples and applying a permanova test with 999 permutations in scikit-bio (v.  
148 0.5.6) (40). The change between tissue types in CTF coordinate space were modeled with a  
149 paired sample t-test was used to determine if there was a global difference between tumor and  
150 normal tissue along either PC; the effect of change on survival was compared using a  
151 permutative Welch's t-test looking at the difference between groups with 999 permutations. ALR  
152 interactions were evaluated using a linear mixed effects model with individual as the grouping  
153 factor.

154

155 Survival was modeled using logistic regression. Models were fit using a crude (unadjusted)  
156 model and a model adjusted for age, sex, ASA score, tumor location, surgery period, TNM stage,  
157 radical surgery, and differentiation grade.

158

159 For all analyses, a p-value of 0.05 was considered significant.

160

161 Modeling was performed using statsmodels (v 0.11.1), scipy (v 1.4.1), and numpy (v 1.18.5) in  
162 python (v. 3.6) (41–43). Figures were plotted using with matplotlib (3.2.2) and seaborn (0.10.1)  
163 The dendrogram was plotted using Empress (q2-empress v.0.0.1-dev, commit b705358) (44);  
164 three dimensional ordinations were rendered using Emperor (v. 1.0.3) (45). Taxonomic colors  
165 come from the microshades colorblind friendly palette (46). Figures were assembled in Adobe  
166 Illustrator 2021 (Adobe Inc, San Jose, CA, USA).

167

168 **RESULTS**

169

170 In our nested case-control study of late-stage colorectal cancer patients, the 51 long-term  
171 survivors were more likely to be younger, male, and healthier compared to the 50 short-term  
172 survivors (Table S1). The short-term survivors presented with metastatic tumors and lower  
173 differentiation than long-term survivors, and fewer received radical surgery. We found that age,  
174 TNM stage and tumor differentiation were strong predictors of long-term survival (Table 1).  
175 Individuals aged between 70-74 years were 14 times more likely to be short-term survivors (OR  
176 14.24; 95% CI 1.21, 167.40) than those younger than 60. TNM-stage IV was associated with an  
177 almost 50 times higher risk of being a short-term survivor (OR 49.32; 95% CI 5.86, 415.12)  
178 compared to TNM stage III (Table 1).

179

180 Following sequencing, quality filtering, and denoising to ASVs, we retained 202 paired tumor  
181 and adjacent normal tissue samples. The broad pattern in the overserved microbiome reflect  
182 those seen in previous studies of Swedish adults (Figure S1) (47). We found the patient was the  
183 strongest predictor of microbiome composition, and that an individual's paired samples were  
184 more similar to each other than tissue samples from patients with similar characteristics (Figure  
185 S2), reflecting what appears to be a common pattern in CRC patients and beyond (10,18,48).

186

**Table 1. Risk factors for short-term survival**

	Crude risk		Adjusted risk	
		OR (95% CI)	Model 1 <sup>1</sup> OR (95% CI)	Model 2 <sup>2</sup> OR (95% CI)
<b><u>Patient characteristics at time of surgery</u></b>				
<b>Age, years</b>				
	<60	1.00 (ref)	1.00 (ref)	1.00 (ref)
	60-69	0.87 (0.24-3.09)	2.45 (0.26-22.72)	2.59 (0.28-24.38)
	70-74	2.40 (0.65-8.81)	12.55 (1.06-149.26)	14.24 (1.21-167.40)
	≥75	1.96 (0.56-6.91)	8.68 (0.79-95.19)	10.55 (0.99-112.75)
<b>Sex</b>				
	Female	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Male	0.76 (0.35-1.67)	0.47 (0.14-1.56)	0.44 (0.13-1.41)
<b>ASA score</b>				
	I (healthy)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	II (mild)	0.80 (0.32-2.02)	2.29 (0.45-11.78)	2.69 (0.56-12.96)
	III-IV (severe or worse)	2.01 (0.65-6.19)	4.10 (0.60-27.92)	4.99 (0.79-31.45)
<b>Pre-operative treatment</b>				
	None	1.00 (ref)	1.00 (ref)	-
	Radiotherapy	1.17 (0.74-1.84)	0.71 (0.12-4.15)	-
<b><u>Tumor characteristics</u></b>				
<b>Localization</b>				
	Colon right	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Colon left	0.47 (0.16-1.32)	0.78 (0.16-3.82)	0.76 (0.16-3.63)
	Rectum	0.72 (0.28-1.84)	2.03 (0.33-12.63)	1.61 (0.36-7.21)
<b>Mucinous cancer</b>				
	no	1.00 (ref)	1.00 (ref)	-
	yes	0.83 (0.24-2.93)	0.50 (0.05-5.39)	-
<b>TNM stage</b>				
	III	1.00 (ref)	1.00 (ref)	1.00 (ref)
	IV	10.8 (3.68-31.72)	44.67 (5.53-360.63)	49.32 (5.86-415.12)
<b>Grade of differentiation</b>				
	Low	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Medium	0.20 (0.07-0.54)	0.23 (0.05-0.98)	0.24 (0.06-1.00)
	High	0.21 (0.05-0.97)	0.09 (0.01-1.24)	0.10 (0.01-1.27)
<b><u>Surgical characteristics</u></b>				
<b>Period of surgery</b>				
	1997-2005	1.00 (ref)	1.00 (ref)	1.00 (ref)
	2006-2010	0.54 (0.21-1.37)	0.44 (0.10-1.92)	0.44 (0.10-1.89)
	2011-2017	0.59 (0.21-1.65)	1.19 (0.22-6.47)	1.08 (0.22-5.36)
<b>Radical operation</b>				
	no	1.00 (ref)	1.00 (ref)	1.00 (ref)
	yes	0.05 (0.01-0.41)	0.13 (0.01-1.51)	0.12 (0.01-1.34)

<sup>1</sup> Adjusted for all variables

<sup>2</sup> Adjusted for all other variables except for preoperative treatment and mucinous cancer

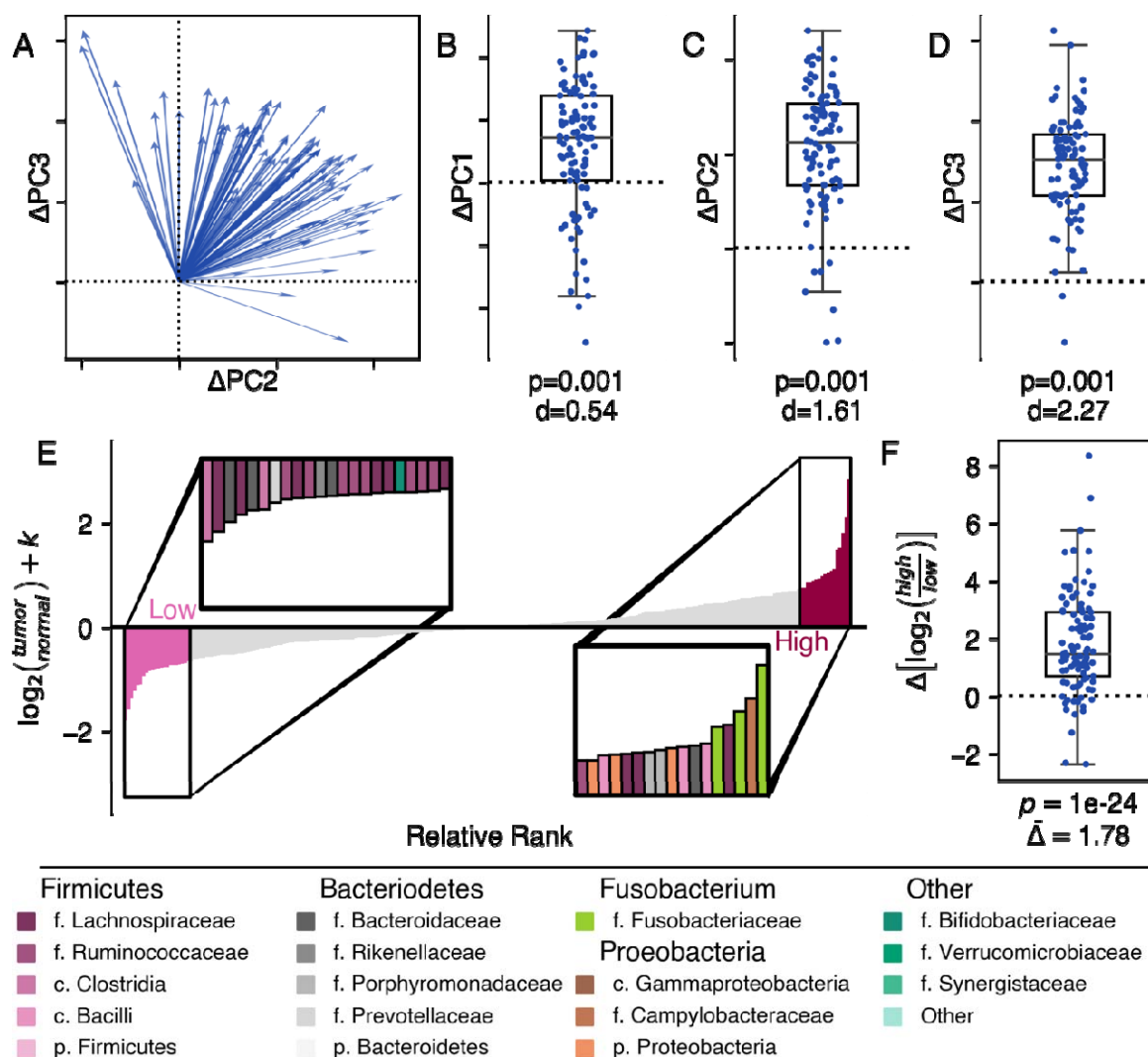
187 **The microbiome of tumor and normal tissue differ**

188 To address individual microbial identities, we applied a subject-aware compositional tensor  
189 factorization (CTF) technique (34). The approach integrates feature-based information from each  
190 sample to build both a subject-specific profile, and to describe changes within subjects and  
191 features across a gradient. We did not find a statistically significant association between a  
192 sample's position in CTF space and survival (unadjusted permanova  $R^2=0.012$ ;  $p=0.296$ , 999  
193 permutation, Figure S3, Table S2). However, we found differences between normal and tumor  
194 tissue, with a consistent shift between paired samples in CTF space, primarily along principal  
195 component (PC) 2 and PC 3 (Figure 1A-D).

196

197 Given evidence of consistent, community-level changes in the microbiome between the tissue  
198 types, we looked for features, which might be driving these differences. We used an individual-  
199 aware differential ranking technique, which first ranked the features with the greatest differences  
200 associated with tissue type, and then we selected a subset of these features to build an additive  
201 log ratio (ALR), a summary of taxa which likely describe the difference (Figure 1E, Table S3,  
202 Files S1). We found tumor tissue was associated with a higher relative abundance of  
203 *Fusobacteria*, *Porphyromonas*, *Granulicatella*, and *Campylobacter* at the expense of members  
204 of genus *Blautia*, and *Ruminococcus*. Tumor tissue had a 1.78 (95% CI 1.50, 2.18,  $p < 1 \times 10^{-12}$ )  
205 log<sub>2</sub>-fold increase in the features selected by DR compared to normal tissue, suggesting a tissue-  
206 specific signature (Figure 1F).

207





## Differences between normal- and tumor-associated microbiome are associated with survival

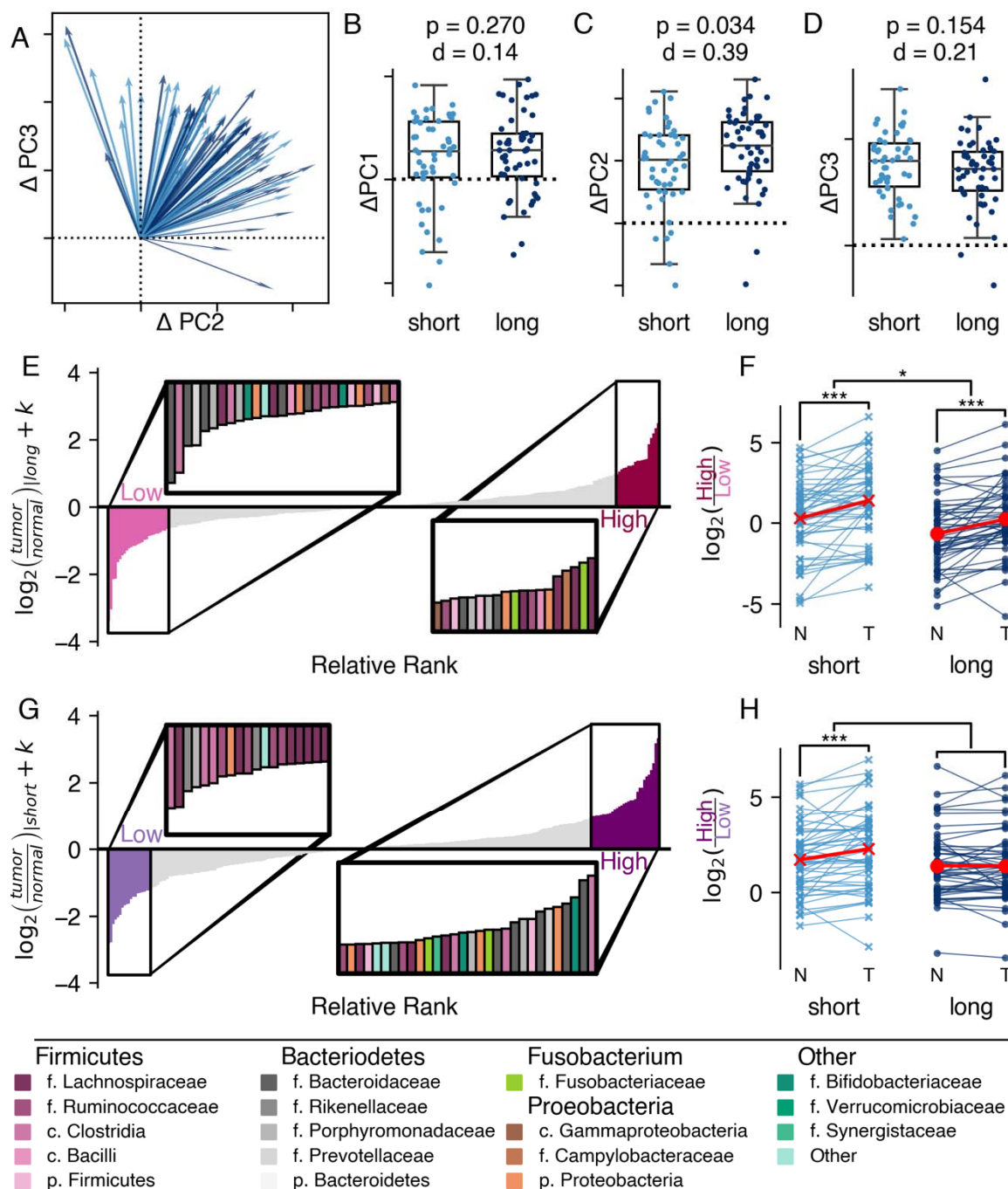
208

209 Since we saw consistent differences between tumor and normal tissue, we wondered if there  
210 might be a relationship between the magnitude of the difference and survival. We found that  
211 tumor and normal tissue were more similar in long-term survivors than short-term, a difference  
212 primarily driven by changes in abundance (Table S4). Additionally, long-term survivors showed  
213 a larger change along PC 2 compared to short-term survivors (Cohen's  $d$  0.40,  $p=0.016$ , 999  
214 permutations; Figure 2). This suggested enough of a community-level change in the microbiome  
215 to motivate looking for features which might explain the differences.

216

217 Therefore, we applied a subject-aware differential ranking technique looking at the interaction  
218 between tissue type and survival to further refine the features (Figure 2E-G). We used an  
219 interaction model to identify features that changed in tumor tissue based on survival group.  
220 Based on the tissue associated taxa associated with long-term survival, we defined an ALR  
221 where tumor tissue was associated with ASVs from genus *Fusobacterium*, *Campylobacter*, and  
222 *Escherichia/Shigella*. We found members of genus *Butyricoccus*, *Roseburia*, and *Streptococcus*  
223 associated with both normal and tumor tissue (Table S5, File S2). There was a higher relative  
224 abundance of the tumor-associated organisms in both survival groups, and the overall relative  
225 abundance was higher in short-term survivors (Figure 2F). However, the magnitude of the  
226 change did not differ between the two groups.

227



**Figure 2. The magnitude of the difference between tumor and normal tissue is associated with survival**

There is a difference in the magnitude of change between long-term and short-term survivors. (A) In two dimensions, the change along PC2 and PC3 is visualized as a vector from normal tissue to tumor issue. Short-term survivors ( $< 2$  year) are shown in light blue. Long term survivors ( $\geq 5$  years) are shown in dark blue. The corresponding relationships can be visualized along the individual components: (B) PC 1, (C) PC 2, and (D) PC 3. Ticks along PC 2 (C) and PC 3 (D) match the two-dimensional axes in (A). All boxplots are shown with a Cohen's d effect size, and p-value from a permutative Welch's t-test with 999 permutations, comparing the two survival groups. (E-H) A differential ranking model was fit to consider

the interaction between survival and tissue. The ranks associated with (E,F) tumor tissue in long-term survivors and (G,H) tumor tissue in short survivors (interaction) are show. (E,F) The relative associated with the model, insets highlight ASVs associated with the extremes of each group. Taxonomic assignments are provided in the legend. (F,H) The additive long ratio associated with the ranks. Paired differences are connected by a line between normal (N) and tumor (T) tissue. The effect was modeled using a linear mixed effects model, treating the individual as random. \* $p < 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ .

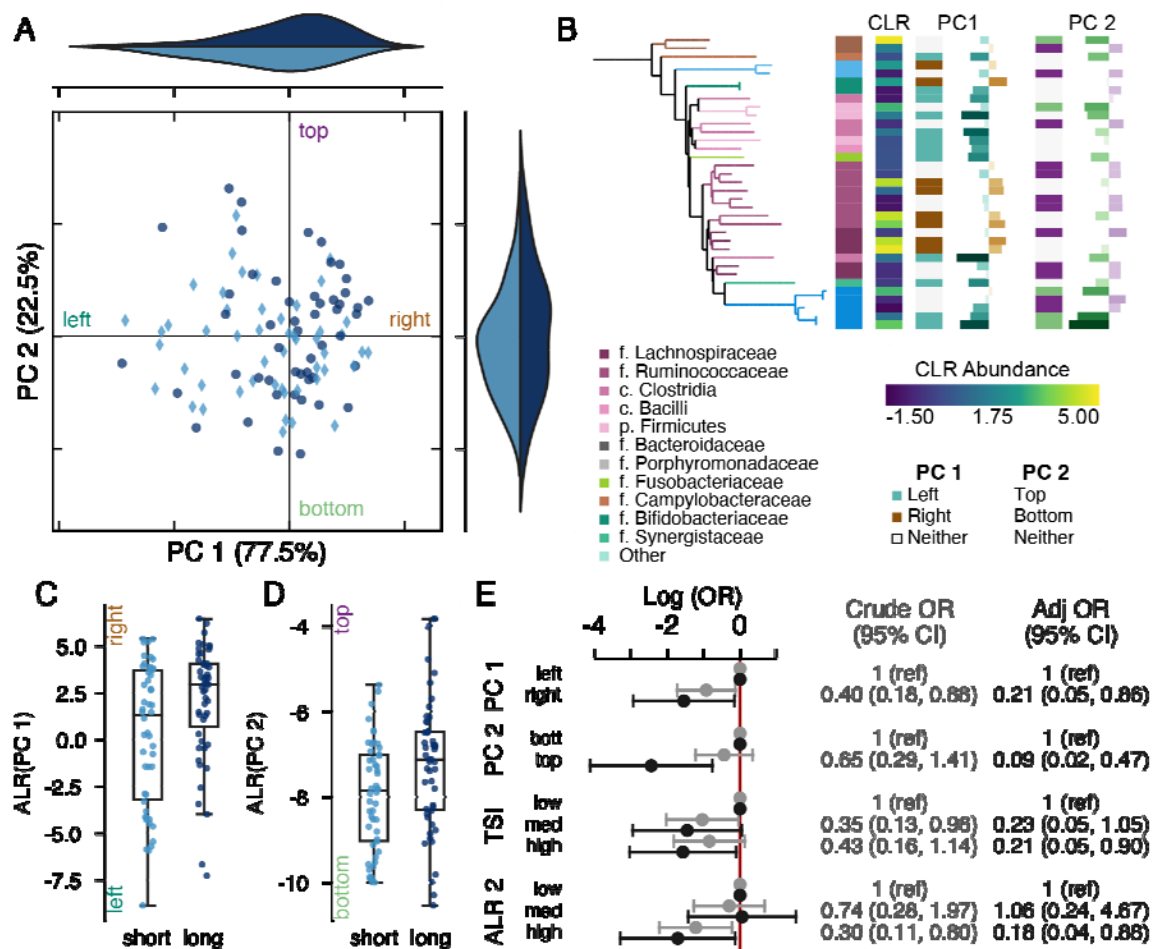
228 In contrast, the interaction term identified a set of taxa, which were significantly different  
229 between the tissue types in short-term survivors but not among long-term survivors (Figure 2G,  
230 Table S6, File S2). Once again, we found tumor tissue in short-term survivors to be strongly  
231 associated with an ASV from *Fusobacteria* and as well as a few members of family  
232 Veillonellaceae, although again, there were not clear taxonomic patterns in other families, such  
233 as family Lachnospiraceae or Rumminococaceae. These results indicate the survival-associated  
234 changes in the microbiome may be largest in tumor tissue and help to identify specific set of  
235 organisms responsible for these changes.

236

### 237 **The tumor microbiome is associated with survival**

238 Based on our observation that differences in tissue types were more pronounced in short-term  
239 survivors, and since past work focused on tumor tissues, we also chose to further interrogate the  
240 tumor-specific microbiome. A rPCA approach showed separation in the microbiomes between  
241 short- and long-term survivors (Figure 3A, D) (34). After adjustment for confounders and both  
242 PCs, patients with larger values for PC 1 had 3.5 lower (OR 0.29; 95% CI 0.08, 0.97) odds of  
243 short-term survival, while those with higher values for PC 2 were five times less likely to be  
244 short-term survivors (OR 0.19; 95% CI 0.05, 0.80). Individuals in the quadrant defined by these  
245 two extremes in the data were at least 7.5 times more likely to survive than any other group in  
246 the ordination (Figure S4).

247



248 We found 37 features associated with separation along PC 1. To the left of PC 1, we found  
249 members of genus *Fusobacterium*, *Parvimonas* and *Porphyromonas*, and other common oral  
250 genera like *Gemella* and *Dialster* (Figure 3B). In contrast, higher values along PC 1 (to the right)  
251 were correlated to more common gut taxa, including members of families Lachnospiraceae and  
252 Rumminococceae. We defined the log<sub>2</sub> fold ratio between the organisms separating PC 1 as a  
253 tumor-survival index (Table S7, File S3). For every 2-fold increase in this index in tumor tissue,  
254 the odds of survival increased by 20% (adjusted OR 0.80; 95% CI 0.67, 0.96). There were no  
255 clear patterns in the taxa separating along PC 2, beyond the association between  
256 *Escherichia/Shigella* and short-term survival (Table S8, File S3), although there was a significant  
257 relationship between these selected taxa and survival (OR 0.64; 95% CI 0.41, 0.98 for every log<sub>2</sub>  
258 increase).

259

## 260 **DISCUSSION**

261

262 Our results show a clear and consistent difference between normal and tumor tissue once we had  
263 accounted for individual microbiome effects. Across all patients, tumors carried a higher  
264 proportion of ASVs mapped to genus *Fusobacterium*, *Gemella*, *Dialster* and *Campylobacter* at  
265 the expense of genera like *Blautia* and *Allistipes*. The tumor-associated features reflect  
266 organisms found more commonly in CRC patients compared to healthy controls, whereas the  
267 organisms associated with normal tissue belong to clades commonly associated with short chain  
268 fatty acids and widely believed to be beneficial (5,6,49–51). Further, we are among the first to  
269 show that the magnitude of the difference between the two tissue types can be associated with  
270 prognosis. Our differential ranking analysis identified a set of 38 ASVs, which changed between

271 the tumor and normal tissue in short-term survivors, but not long-term survivors. This suggests  
272 survival may be associated with localized changes in the microbiome.

273  
274 We are among the first to report differences a different between tumor and normal tissue in  
275 paired samples, let alone an association between the degree of dissimilarity and survival. Drewes  
276 et al (12) demonstrated clear difference between paired tumor and normal tissue samples using  
277 microscopy, although their 16S analysis did not explicitly test paired samples. These results  
278 seemingly conflict with much of the existing literature (15–18,48). Several previous studies  
279 reported no difference in the microbiome between the two tissue types, let alone an intra-  
280 individual difference associated with survival. Like the past studies, we observed and described a  
281 strong intraindividual similarity. A personal microbial signature is a normal feature of the  
282 microbiome seen in a variety of settings including population-based studies (10), dietary  
283 interventions (52), and among CRC patients (15–18,48). However, unlike prior work, the  
284 statistical models we selected accounted for this strong intra-individual similarity. Re-analysis of  
285 prior publications using subject aware methods may identify the same patterns we found: strong  
286 individual microbial signatures with a difference between the tissue types. Our results indicate  
287 that the tumor-specific microenvironment, rather than the overall microbiome, is important for  
288 understanding CRC pathology. At a minimum, future sequencing survey studies will need to  
289 account for tissue-specific effects in their analysis, and studies treating tumor and non-tumor  
290 biopsy samples as identical may need to check for biases.

291  
292 Based on the difference in the microbiome between tissue types, we specifically focused on the  
293 relationship between the tumor microbiome and survival. Two previous studies have explored

294 the relationship between the tumor microbiome and survival using untargeted sequencing. In that  
295 study of 67 Irish patients, Flemer and colleagues defined microbiome groups using a non-  
296 compositional abundance-based clustering approach (14). They found a higher relative  
297 abundance of a cluster defined by members of genus *Bacteroidetes*, *Blautia*, *Roseburia*,  
298 *Rumminococcus*, and an unclassified member of family Lachnospiraceae was associated with  
299 shorter survival, while higher abundance of a cluster characterized by *Streptococcus*,  
300 *Fusobacterium*, and unclassified family Enterobacteraceae was associated with longer survival.  
301 These groupings are contradictory to the features associated with survival in our tumor tissue  
302 results. In contrast, our tumor survival index, defined by an ALR of features along PC 1, showed  
303 a decrease in the relative abundance in *Fusobacterium* among our long-term survivors, who were  
304 characterized by a higher relative abundance of *Blautia*, *Roseburia*, among others. It is likely this  
305 disagreement is due to differences in methods used for differential abundance (36,53). Our  
306 results are more in line with results from a Chinese cohort (13). In that study, a higher  
307 untransformed relative abundance of genus *Fusobacterium*, or higher relative abundance of reads  
308 mapped to *Bacteroidetes fragilis* were associated with an increased hazard of death, while a  
309 higher relative abundance of genus *Faecalibacterium* was protective. We find similar trends in  
310 our tumor survival index, where short-term survival was associated with ASVs mapped to genus  
311 *Fusobacterium* and a *Bacteroidetes* ASV, while longer survival was associated with  
312 *Faecalibacterium*. Our results and those of the Chinese cohort suggest that a more normal (gut-  
313 like) microbiome is associated with long-term survival, while a more disrupted (oral)  
314 microbiome led to a poor prognosis.

315



316 Our conclusions are supported by our nested case-control design, which helps establishing  
317 temporality: changes in the local tumor microbiome at the time of surgery are associated with  
318 future outcomes, increasing the probability that the observation is a real phenomenon, rather than  
319 a change in the microbiome in response to disease state. Our analysis used statistically  
320 appropriate methods, which accounted for analytical challenges in describing the microbiome,  
321 decreasing the possibility of false positives, especially among the identified taxa (36,54). Our  
322 analysis has also addressed confounders, which may affect the microbiome and survival,  
323 including the strong individual microbiome signature.

324

325 However, our study has some limitations. First, our results focus on late-stage cancer patients in  
326 northern Europe, and therefore may not be broadly generalizable. There are reports of  
327 differences in the tumor microbiome between early and late stage CRC patients (55), and  
328 differences in the healthy microbiome between countries (56). However, past work has suggested  
329 that CRC is characterized by a set of organisms similar to the ones we identified, and our work  
330 overlaps with the results of a Chinese cohort, despite methodological differences (5,6,13).

331 Additionally, we did not characterize our specific taxonomic profiles in a validation cohort,  
332 meaning the taxa separating tissue types and the tumor survival index may be specific features of  
333 our cohort, rather than able to predict survival in a broader population of late-stage CRC patients.

334 Finally, we profiled the microbiome using 16S rRNA sequencing, with all the assumptions,  
335 benefits, and limitations of the technique. Our work is predicated on the assumption that  
336 phylogenetic similarity correlates to genetic and niche similarity. Without robust functional  
337 prediction and the ability to assemble genome units, we are limited in our mechanistic insight.

338 However, our 16S sequencing is, in many cases, able to capture species or sub-species level

339 resolution as the amplicon sequence variant ID, even if the name cannot be inferred accurately  
340 (57,58).

341

## 342 **CONCLUSION**

343 We performed a nested case-control of the role of the microbiome in relapse free survival  
344 following primary resection in late-stage CRC patients. We identified clear differences in the  
345 microbiome between normal and tumor tissue and that a larger difference between tissue types  
346 was associated with poor prognosis. We found the tumor microbiome was associated with  
347 survival. This suggests a need to focus microbiome-based interventions at the tumor-specific  
348 community, rather trying to modify prognosis by changing the gut microbiome overall.

349

## 350 **DATA AVAILABILITY**

351 Due to patient privacy concerns, sequences are available upon reasonable request. Patient data  
352 will not be released.

353

## 354 **ACKNOWLEDGEMENTS**

355 The authors wish to thank the Department of Surgery, County Hospital Ryhov, for the collection  
356 of tissue biopsies. Thanks to the lab core at Centre for Translational Microbiome Research for  
357 support in extracting, processing, and sequencing the samples. We are also grateful to Cameron  
358 Martino, Marcus Fedarko, and Kalen Cantrell for their rapid responses to bug reports and feature  
359 requests for the gemelli and empress qiime2 plugins.

360

361

362 **FUNDING**

363 This work was funded by Futurum-Academy for Healthcare, Region Jönköping County, Sweden

364 (grant FUTURUM-933436 and FUTURUM-809281), as well as centre grant from Ferring

365 Pharmaceuticals for the establishment of the Centre for Translational Microbiome Research.

366 JTM was funded by the intramural research program of the Eunice Kennedy Shriver National

367 Institute of Child Health and Human Development (NICHD).

368 The funders were not involved in the development, analysis, or interpretation of the study.

369

370 **AUTHOR CONTRIBUTIONS**

371 RSO, MS, LE, and AM designed the study. RSO collected the tissue samples and reviewed

372 medical records. JWD prepared the data. JWD performed the bioinformatic analysis; JWD and

373 NB analyzed the data with advice from JTM. JWD drafted the manuscript. All authors reviewed

374 and approved the final manuscript.

375

376 **CONFLICTS OF INTEREST**

377 The authors do not have any conflicts of interest.

378

## 379 REFERENCES

- 380 1. Mattiuzzi C, Sanchis-Gomar F, Lippi G. Concise update on colorectal cancer epidemiology.  
381 Ann Transl Med [Internet]. 2019 Nov [cited 2021 May 7];7(21). Available from:  
382 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7011596/>
- 383 2. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality,  
384 survival, and risk factors. Prz Gastroenterol. 2019;14(2):89–103.
- 385 3. Hooper LV, Gordon JI. Commensal host-bacterial relationships in the gut. Science. 2001  
386 May 11;292(5519):1115–8.
- 387 4. Hooper LV, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient  
388 environment of the mammalian intestine. Annu Rev Nutr. 2002;22:283–307.
- 389 5. Dai Z, Coker OO, Nakatsu G, Wu WKK, Zhao L, Chen Z, et al. Multi-cohort analysis of  
390 colorectal cancer metagenome identified altered bacteria across populations and universal  
391 bacterial markers. Microbiome. 2018 Apr 11;6(1):70.
- 392 6. Young C, Wood HM, Seshadri RA, Van Nang P, Vaccaro C, Melendez LC, et al. The  
393 colorectal cancer-associated faecal microbiome of developing countries resembles that of  
394 developed countries. Genome Medicine. 2021 Feb 16;13(1):27.
- 395 7. Lauka L, Reitano E, Carra MC, Gaiani F, Gavriilidis P, Brunetti F, et al. Role of the  
396 intestinal microbiome in colorectal cancer surgery outcomes. World Journal of Surgical  
397 Oncology. 2019 Dec 2;17(1):204.
- 398 8. Gethings-Behncke C, Coleman HG, Jordao HWT, Longley DB, Crawford N, Murray LJ, et  
399 al. *Fusobacterium nucleatum* in the Colorectum and Its Association with Cancer Risk and  
400 Survival: A Systematic Review and Meta-analysis. Cancer Epidemiol Biomarkers Prev.  
401 2020 Mar 1;29(3):539–48.
- 402 9. Byrd AL, Segre JA. Adapting Koch’s postulates. Science. 2016 Jan 15;351(6270):224–6.
- 403 10. Chen L, Wang D, Garmaeva S, Vila AV, Gacesa R, Sinha T, et al. The long-term genetic  
404 stability and individual specificity of the human gut microbiome. Cell. 2021 Apr  
405 29;184(9):2302–2315.e12.
- 406 11. Gibson TE, Carey V, Bashan A, Hohmann EL, Weiss ST, Liu YY. On the Stability  
407 Landscape of the Human Gut Microbiome: Implications for Microbiome-based Therapies  
408 [Internet]. 2017 Aug [cited 2021 Nov 25] p. 176941. Available from:  
409 <https://www.biorxiv.org/content/10.1101/176941v1>
- 410 12. Drewes JL, White JR, Dejea CM, Fathi P, Iyadorai T, Vadivelu J, et al. High-resolution  
411 bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal  
412 cancer consortia. npj Biofilms Microbiomes. 2017 Nov 29;3(1):1–12.

- 413 13. Wei Z, Cao S, Liu S, Yao Z, Sun T, Li Y, et al. Could gut microbiota serve as prognostic  
414 biomarker associated with colorectal cancer patients' survival? A pilot study on relevant  
415 mechanism. *Oncotarget*. 2016 Jun 15;7(29):46158–72.
- 416 14. Flemer B, Herlihy M, O'Riordain M, Shanahan F, O'Toole PW. Tumour-associated and  
417 non-tumour-associated microbiota: Addendum. *Gut Microbes*. 2018 Jul 4;9(4):369–73.
- 418 15. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human Intestinal Lumen and Mucosa-  
419 Associated Microbiota in Patients with Colorectal Cancer. *PLOS ONE*. 2012 Jun  
420 28;7(6):e39743.
- 421 16. Wirth U, Garzetti D, Jochum LM, Spriewald S, Kühn F, Ilmer M, et al. Microbiome  
422 Analysis from Paired Mucosal and Fecal Samples of a Colorectal Cancer Biobank. *Cancers*.  
423 2020 Dec;12(12):3702.
- 424 17. Leung PHM, Subramanya R, Mou Q, Lee KT wei, Islam F, Gopalan V, et al.  
425 Characterization of Mucosa-Associated Microbiota in Matched Cancer and Non-neoplastic  
426 Mucosa From Patients With Colorectal Cancer. *Frontiers in Microbiology* [Internet]. 2019  
427 [cited 2021 Oct 12];10. Available from:  
428 <https://www.ncbi.nlm.nih.gov/labs/pmc/articles/PMC6581718/>
- 429 18. Sheng QS, He KX, Li JJ, Zhong ZF, Wang FX, Pan LL, et al. Comparison of Gut  
430 Microbiome in Human Colorectal Cancer in Paired Tumor and Adjacent Normal Tissues.  
431 *Onco Targets Ther*. 2020 Jan 21;13:635–46.
- 432 19. Debesa-Tur G, Pérez-Brocal V, Ruiz-Ruiz S, Castillejo A, Latorre A, Soto JL, et al.  
433 Metagenomic analysis of formalin-fixed paraffin-embedded tumor and normal mucosa  
434 reveals differences in the microbiome of colorectal cancer patients. *Scientific Reports*. 2021  
435 Jan 11;11(1):391.
- 436 20. Hugerth LW, Seifert M, Pennhag A a. L, Du J, Hamsten MC, Schuppe-Koistinen I, et al. A  
437 comprehensive automated pipeline for human microbiome sampling, 16S rRNA gene  
438 sequencing and bioinformatics processing. *bioRxiv*. 2018 Mar 21;286526.
- 439 21. Hugerth LW, Wefer HA, Lundin S, Jakobsson HE, Lindberg M, Rodin S, et al. DegePrime,  
440 a Program for Degenerate Primer Design for Broad-Taxonomic-Range PCR in Microbial  
441 Ecology Studies. *Appl Environ Microbiol*. 2014 Aug 15;80(16):5116–23.
- 442 22. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2:  
443 High-resolution sample inference from Illumina amplicon data. *Nature Methods*. 2016  
444 Jul;13(7):581–3.
- 445 23. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment  
446 of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol*. 2007 Aug  
447 15;73(16):5261–7.

- 448 24. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal  
449 RNA gene database project: improved data processing and web-based tools. *Nucleic Acids*  
450 *Res.* 2013 Jan;41(Database issue):D590-596.
- 451 25. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al.  
452 Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2.  
453 *Nature Biotechnology.* 2019 Aug;37(8):852–7.
- 454 26. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, et al.  
455 Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical  
456 Information. *mSystems* [Internet]. 2018 Jun 26 [cited 2021 Feb 8];3(3). Available from:  
457 <https://msystems.asm.org/content/3/3/e00021-18>
- 458 27. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial  
459 communities. *Appl Environ Microbiol.* 2005 Dec;71(12):8228–35.
- 460 28. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity  
461 measures lead to different insights into factors that structure microbial communities. *Appl*  
462 *Environ Microbiol.* 2007 Mar;73(5):1576–85.
- 463 29. Jaccard P. The Distribution of the Flora in the Alpine Zone.1. *New Phytologist.*  
464 1912;11(2):37–50.
- 465 30. Sørensen TJ. A method of establishing groups of equal amplitude in plant sociology based  
466 on similarity of species content and its application to analyses of the vegetation on Danish  
467 commons. København: I kommission hos E. Munksgaard; 1948.
- 468 31. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and  
469 microbial differential abundance strategies depend upon data characteristics. *Microbiome.*  
470 2017 Mar 3;5(1):27.
- 471 32. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are  
472 Compositional: And This Is Not Optional. *Front Microbiol.* 2017;8:2224.
- 473 33. Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. Logratio  
474 Analysis and Compositional Distance. *Mathematical Geology.* 2000 Apr 1;32(3):271–5.
- 475 34. Martino C, Shenhav L, Marotz CA, Armstrong G, McDonald D, Vázquez-Baeza Y, et al.  
476 Context-aware dimensionality reduction deconvolutes gut microbial community dynamics.  
477 *Nature Biotechnology.* 2021 Feb;39(2):165–8.
- 478 35. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A Novel  
479 Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems* [Internet].  
480 2019 Feb 26 [cited 2020 May 4];4(1). Available from:  
481 <https://msystems.asm.org/content/4/1/e00016-19>

- 482 36. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, et al.  
483 Establishing microbial composition measurement standards with reference frames. *Nat*  
484 *Commun.* 2019 20;10(1):2719.
- 485 37. Morton JT, Jin DM, Mills RH, Shao Y, Rahman G, Berding K, et al. Multi-omic analysis  
486 along the gut-brain axis points to a functional architecture of autism [Internet]. *bioRxiv*;  
487 2022 [cited 2022 Mar 2]. p. 2022.02.25.482050. Available from:  
488 <https://www.biorxiv.org/content/10.1101/2022.02.25.482050v1>
- 489 38. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A  
490 Probabilistic Programming Language. *Journal of Statistical Software.* 2017 Jan 11;76:1–32.
- 491 39. Sorensen T, Hohenstein S, Vasishth S. Bayesian linear mixed models using Stan: A tutorial  
492 for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for*  
493 *Psychology.* 2016;12(3):175–200.
- 494 40. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral*  
495 *Ecology.* 2001;26(1):32–46.
- 496 41. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python.  
497 *Proceedings of the 9th Python in Science Conference.* 2010 Jan 1;2010.
- 498 42. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy  
499 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods.* 2020  
500 Mar;17(3):261–72.
- 501 43. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al.  
502 *Array programming with NumPy.* *Nature.* 2020 Sep;585(7825):357–62.
- 503 44. Cantrell K, Fedarko MW, Rahman G, McDonald D, Yang Y, Zaw T, et al. EMPress  
504 Enables Tree-Guided, Interactive, and Exploratory Analyses of Multi-omic Data Sets.  
505 *mSystems.* 6(2):e01216-20.
- 506 45. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-  
507 throughput microbial community data. *GigaScience* [Internet]. 2013 Dec 1 [cited 2021 Feb  
508 8];2(2047-217X-2-16). Available from: <https://doi.org/10.1186/2047-217X-2-16>
- 509 46. Dahl E, Neer E, Karstens L. A custom color palette for improving data visualization  
510 [Internet]. [cited 2022 Jan 12]. Available from: <https://karstenslab.github.io/microshades/>
- 511 47. Hugerth LW, Andreasson A, Talley NJ, Forsberg AM, Kjellström L, Schmidt PT, et al. No  
512 distinct microbiome signature of irritable bowel syndrome found in a Swedish random  
513 population. *Gut.* 2020 Jun;69(6):1076–84.
- 514 48. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-  
515 associated and non-tumour-associated microbiota in colorectal cancer. *Gut.* 2017  
516 Apr;66(4):633–43.

- 517 49. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis M. The  
518 Controversial Role of Human Gut Lachnospiraceae. *Microorganisms*. 2020 Apr;8(4):573.
- 519 50. Barcenilla A, Pryde SE, Martin JC, Duncan SH, Stewart CS, Henderson C, et al.  
520 Phylogenetic Relationships of Butyrate-Producing Bacteria from the Human Gut. *Applied*  
521 *and Environmental Microbiology*. 2000 Apr;66(4):1654–61.
- 522 51. Hajjar R, Richard CS, Santos MM. The role of butyrate in surgical and oncological  
523 outcomes in colorectal cancer. *American Journal of Physiology-Gastrointestinal and Liver*  
524 *Physiology*. 2021 Apr;320(4):G601–8.
- 525 52. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-  
526 term dietary patterns with gut microbial enterotypes. *Science*. 2011 Oct 7;334(6052):105–8.
- 527 53. Nearing JT, Douglas GM, Hayes M, MacDonald J, Desai D, Allward N, et al. Microbiome  
528 differential abundance methods produce disturbingly different results across 38 datasets.  
529 2021 May 10;2021.05.10.443486.
- 530 54. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are  
531 Compositional: And This Is Not Optional. *Frontiers in Microbiology*. 2017;8:2224.
- 532 55. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic  
533 and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in  
534 colorectal cancer. *Nat Med*. 2019 Jun;25(6):968–76.
- 535 56. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al.  
536 *American Gut: an Open Platform for Citizen Science Microbiome Research*. *mSystems*.  
537 2018 Jun;3(3).
- 538 57. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational  
539 taxonomic units in marker-gene data analysis. *ISME J*. 2017 Dec;11(12):2639–43.
- 540 58. Eren AM, Borisy GG, Huse SM, Mark Welch JL. Oligotyping analysis of the human oral  
541 microbiome. *Proceedings of the National Academy of Sciences*. 2014 Jul  
542 15;111(28):E2875–84.

543



## **Supplemental tables**

Table S1. Patient characteristics in the cohort

Table S2. Per-subject predictors of the microbiome in CTF ordination space

Table S3. ASVs separating normal and tumor tissue based on differential ranking

Table S4. ASVs associated with the survival term in the interaction differential ranking model

Table S5. ASVs associated with the tissue term in the interaction differential ranking model

Table S6. ASVs associated with the interaction term in the interaction differential ranking model

Table S7. ASVs used for the tumor rPCA ALR calculation

## **Supplemental Figures**

Figure S1. A family level view of the microbiome

Figure S2. There is a strong individual effect on the microbiome

Figure S3. Metadata predictors of the microbiome in Compositional Tensor Factorization (CTF) space

Figure S4. Survival separates individuals by tumor-associated PCA quadrant

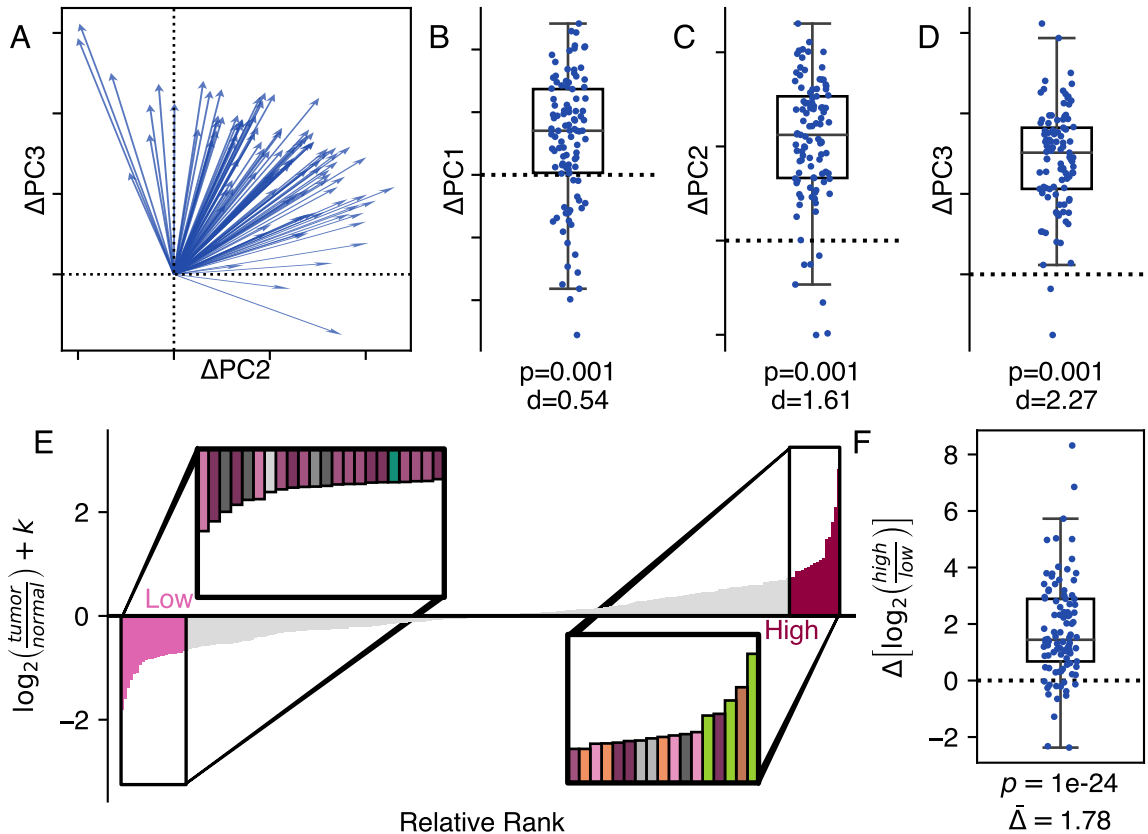
## **Supplemental Files**

File S1. Representative sequences separating tumor tissue

File S2. Representative sequences identified in the interaction model

File S3. Representative sequences for ASVs associated with PC1 in rPCA space

File S4. STORMS checklist



### Firmicutes

- f. Lachnospiraceae
- f. Ruminococcaceae
- c. Clostridia
- c. Bacilli
- p. Firmicutes

### Bacteroidetes

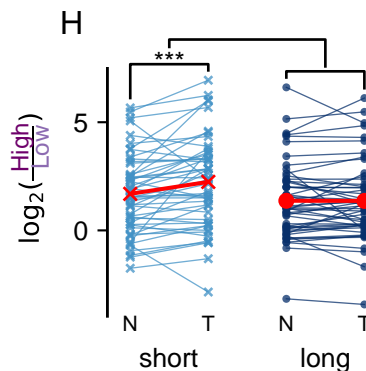
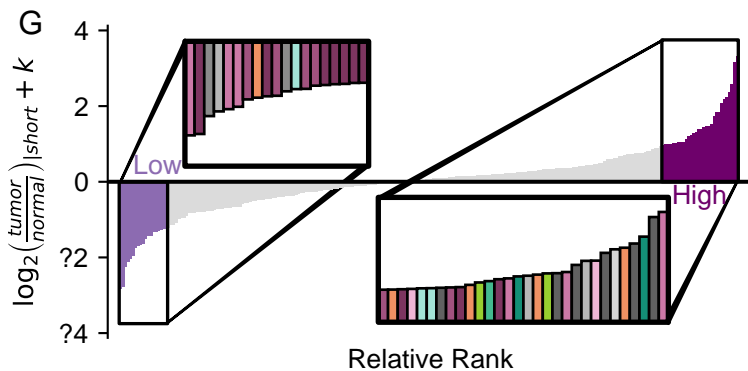
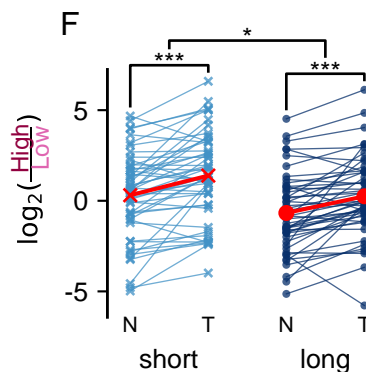
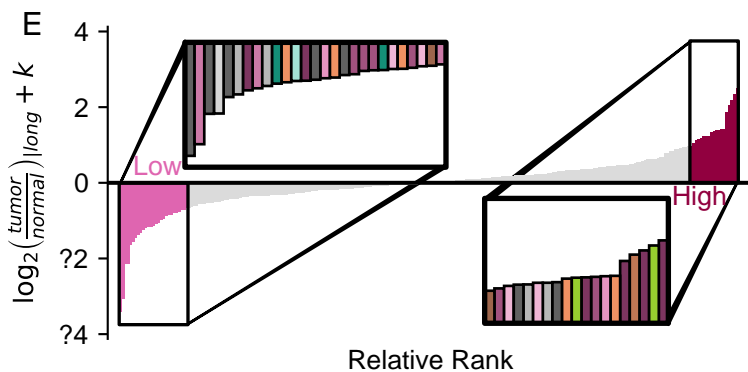
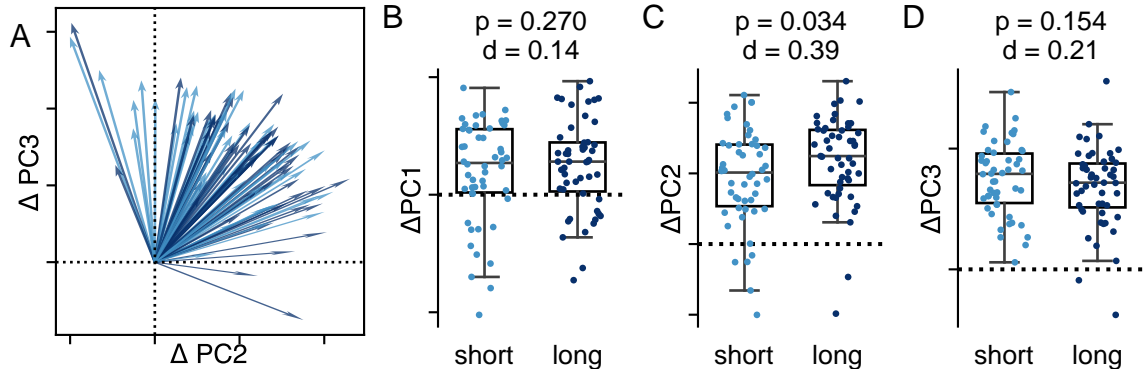
- f. Bacteroidaceae
- f. Rikenellaceae
- f. Porphyromonadaceae
- f. Prevotellaceae
- p. Bacteroidetes

### Fusobacterium

- f. Fusobacteriaceae
- Proeobacteria**
- c. Gammaproteobacteria
- f. Campylobacteraceae
- p. Proteobacteria

### Other

- f. Bifidobacteriaceae
- f. Verrucomicrobiaceae
- f. Synergistaceae
- Other



**Firmicutes**

- f. Lachnospiraceae
- f. Ruminococcaceae
- c. Clostridia
- c. Bacilli
- p. Firmicutes

**Bacteroidetes**

- f. Bacteroidaceae
- f. Rikenellaceae
- f. Porphyromonadaceae
- f. Prevotellaceae
- p. Bacteroidetes

**Fusobacterium**

- f. Fusobacteriaceae
- Proeobacteria**
- c. Gammaproteobacteria
- f. Campylobacteraceae
- p. Proteobacteria

**Other**

- f. Bifidobacteriaceae
- f. Verrucomicrobiaceae
- f. Synergistaceae
- Other

