

# Spatial variation of the native colon microbiota in healthy adults

Running title: Spatial variation of native colon microbiota

Kaitlin J. Flynn<sup>1</sup>, Charles C. Koumpouras<sup>1</sup>, Mack T. Ruffin IV<sup>2</sup>, D. Kim Turgeon<sup>3†</sup>, and Patrick  
D. Schloss<sup>1†</sup>

† Corresponding authors: kturgeon@umich.edu and pschloss@umich.edu

1. Department of Microbiology and Immunology, University of Michigan Medical School, Ann  
Arbor, Michigan 48109

2. Department of Family and Community Medicine, College of Medicine, Pennsylvania State  
University, Hershey, Pennsylvania 17033

3. Department of Internal Medicine, Division of Gastroenterology, University of Michigan Medical  
School, Ann Arbor, Michigan

Grant support: Rose and Lawrence C. Page Foundation

Disclosures: nothing to disclose

Word count: 3,953/5,000

Figures: 6

## Abstract

The microbiome has been implicated in the development of colorectal cancer (CRC) and inflammatory bowel diseases (IBD). The specific traits of these diseases vary along the axis of the digestive tract. Further, variation in the structure of the gut microbiota has been associated with both diseases. Here we profiled the microbiota of the healthy proximal and distal mucosa and lumen to better understand how bacterial populations vary along the colon. We used a two-colonoscopy approach to sample proximal and distal mucosal and luminal contents from the colons of 20 healthy subjects that had not undergone any bowel preparation procedure. The biopsies and home-collected stool were subjected to 16S rRNA gene sequencing and Random Forest classification models were built using taxa abundance and location to identify microbiota specific to each site. The right mucosa and lumen had the most similar community structures of the five sites we considered from each subject. The distal mucosa had higher relative abundance of *Finegoldia*, *Murdochiella*, *Peptoniphilus*, *Porphyromonas* and *Anaerococcus*. The proximal mucosa had more of the genera *Enterobacteriaceae*, *Bacteroides* and *Pseudomonas*. The classification model performed well when classifying mucosal samples into proximal or distal sides (AUC=0.850). Separating proximal and distal luminal samples proved more challenging (AUC=0.580) and specific microbiota that differentiated the two were hard to identify. By sampling the unprepped colon, we identified distinct bacterial populations native to the proximal and distal sides. Further investigation of these bacteria may elucidate if and how these groups contribute to different disease processes on their respective sides of the colon.

Words: 259/260

Keywords: microbiome; colon cancer; proximal and distal colon

## 39 Introduction

40 The human colon is an ecosystem comprised of numerous microenvironments that select for  
41 different microbiota. Concentrations of oxygen, water, and anti-microbial peptides change along  
42 the gut axis and influence which microbiota reside in each location. Microenvironments differ not  
43 only longitudinally along the colon, but also radially from the epithelium to mucosa to intestinal  
44 lumen, offering several sites for different microbial communities to flourish. The identity of these  
45 specific microbiota and communities are important for understanding the etiology of complex  
46 diseases such as Colorectal Cancer (CRC) and Inflammatory Bowel Disease (IBD). CRC and IBD  
47 can be preceded or accelerated by perturbations the structure of the gut microbiota (1–3). The  
48 manifestations of these diseases are known to vary based upon the location in which they occur.  
49 For instance, CRC that arises in the distal (left) colon are of hindgut origin and tend to have large  
50 chromosomal alterations indicative of chromosomal instability (1). In contrast, CRC arising in  
51 the proximal (right) colon tumors are of midgut origin and tend to be sessile and microsatellite  
52 instable (MSI with BRAF and KRAS mutations) (1). In addition to the environmental gradients  
53 within the colon, the distal and proximal sides of the colon differ in the amount of inflammation  
54 present and the genomic instability of precancerous cells, respectively (1,4,5). In IBD patients,  
55 disease occurring in the distal colon extending proximally is usually indicative of ulcerative colitis  
56 (UC), whereas Crohn's disease (CD) can occur anywhere along the GI tract, most commonly in  
57 the ileum and the cecum (2). UC presents as continuous disease with only mucosal involvement,  
58 where as CD has skip lesions and full thickness involvement that may cause abscesses, strictures  
59 and fistulas (2). Thus, given the varied physiology of the proximal-distal axis of the colon and  
60 known differences in disease patterns at these sites, symbiotic microbiota and their metabolites  
61 likely vary as well, and may influence the heterogeneous disease prognoses of IBD and CRC.  
62 Because CRC can be a long-term complication of IBD, the distribution of microbiota is important  
63 to understanding the pathophysiology of both diseases.

64 Several recent findings have shown that development and progression of IBD or CRC can be

65 attributed to specific molecular events as a result of interactions between the gut microbiota  
66 and human host (1,3,6). For instance, comparison of the bacteria present on CRC tumors with  
67 those found on nearby healthy tissue has identified specific species that are tumor-associated  
68 (7). Specific bacteria have also been identified in fecal samples of patients with varying stages of  
69 colon tumorigenesis (8,9). These species include the oral pathogens *Fusobacterium nucleatum*  
70 and *Porphyromonas asacharolytica*. *F. nucleatum* has also been found to be elevated in the stool  
71 and biopsies of patients with IBD as compared to healthy controls (10,11). Furthermore, studies  
72 of *F. nucleatum* isolated from mucosal biopsies showed that more invasive *F. nucleatum* positively  
73 correlates with IBD disease level (10). Like many intestinal pathogens, the bacteria appear to  
74 have a high-impact despite being lowly-abundant in the community (2). The physiology of these  
75 rare taxa may contribute to the colonic disease state. These studies often examined only shed  
76 human stool or the small intestine, preventing fine-resolution analysis of paired samples from the  
77 proximal and distal sides of the colon. Similarly, comparisons of on- or off-tumor/lesion bacteria  
78 rarely have matched tissue from the other side of the colon from the same, disease-bearing patient,  
79 limiting what conclusions can be drawn about the colonic microbiome overall, let alone at that  
80 specific site (12). Due to these limitations, the contribution of the gut microbiota to CRC and  
81 IBD disease location in the colon is largely undefined. Characterizing these communities in healthy  
82 individuals could provide needed insight into disease etiology, including how the disruption of the  
83 healthy community could promote the initiation or proliferation of the distinct proximal and distal  
84 CRC tumors or IBD flares.

85 The few existing profiles of the microbial spatial variation of the colon have been limited by sample  
86 collection methods. The majority of human gut microbiome studies have been performed on whole  
87 shed feces or on samples collected during colonoscopy or surgery (5). While the latter method  
88 allows investigators to acquire samples from inside the human colon, typically these procedures  
89 are preceded by the use of bowel preparation methods such as the consumption of laxatives to  
90 cleanse the bowel. Bowel preparation is essential for detecting cancerous or precancerous lesions  
91 in the colon, but complicates microbiome profiling as the chemicals strip the bowel of contents

and disrupt the mucosal layer (13,14). As such, what little information we do have about the spatial distribution of the microbiota in the proximal and distal colon is confounded by the bowel preparation procedure.

Here we address the limitations of previous studies and identify the microbes specific to the lumen and mucosa of the proximal and distal healthy human colon. We used an unprepared colonoscopy technique to sample the natural community of each location of the gut without prior disruption of the native bacteria in 20 healthy volunteers. To address the inherent inter-individual variation in microbiota, we used a machine-learning classification algorithm trained on curated 16S rRNA sequencing reads to identify the microbiota that were specific to each location. We found that our classification models were able to separate mucosal and luminal samples as well as differentiate between sides of the colon based on populations of particular microbiota. By identifying the distinguishing microbiota we are poised to ask if and how the presence or disruption of the microbiota at each site contribute to the development of the tumor subtypes of CRC in the proximal and distal human colon.

## **Methods**

### **Human subjects**

The procedures in this study and consent were approved by the Institutional Review Board at the University of Michigan Health System with protocol number HUM00082721. Subjects were recruited using the online recruitment platform and were pre-screened prior to enrollment in the study. Exclusion criteria included: use of aspirin or NSAIDs within 7 days, use of antibiotics within 3 months, current use of anticoagulants, known allergies to Fentanyl or Benadryl, prior history of colon disease, diabetes, abdominal surgery, respiratory, liver, kidney or brain impairments, undergoing current chemotherapy or radiation treatment and subjects that were pregnant or trying to conceive. 20 subjects that met the criteria were selected and provided signed informed consent prior to the procedure. There were 13 female and 7 male subjects ranging in age from 25 to 64.

## **Sample collection**

At a baseline visit, subjects gave consent and were given a home collection stool kit (Zymo). At least one week prior to the scheduled colonoscopy, subjects collected whole stool at home and shipped the samples to a research coordinator on ice. Notably, subjects did not undergo any bowel preparation method prior to sampling. On the procedure day, subjects reported to the Michigan Clinical Research Unit at the University of Michigan Health System. Subjects were consciously sedated using Fentanyl, Versed and/or Benadryl as appropriate. A flexible sigmoidoscope was first inserted about 25cm into the colon and jumbo biopsy forceps used to collect the luminal contents. Two luminal samples were collected and the contents immediately deposited into RNeasy (Fisher) and flash-frozen in liquid nitrogen. The forceps were withdrawn and new biopsy forceps were used to collect mucosal biopsies on sections of the colon that were pink and free of stool matter. Three mucosal biopsies were collected and flash-frozen in RNeasy. These samples comprised the distal colon samples. The sigmoidoscope was then withdrawn and a pediatric colonoscope was inserted to reach the proximal colon. Samples were then collected in the same manner as was done in the distal colon and the colonoscope withdrawn. All samples were stored at -80°C.

## **Sample processing, sequencing and analysis**

DNA extraction was performed using the PowerMicrobiome DNA/RNA Isolation Kit (MO BIO Laboratories). For tissue biopsies, Bond-Breaker TCEP solution (Fisher) and 2.8mm ceramic beads (MO BIO Laboratories) were added to the bead beating step to enhance DNA recovery from mucosal samples. The resulting DNA was used as template for amplification of the V4 region of the 16S rRNA gene and fragments were sequenced on an Illumina MiSeq as previously described (15). Sequences were curated using the mothur software as described previously (16). The sequences were assigned a taxonomic classification using a naive Bayesian classifier trained using a 16S rRNA gene training set from the Ribosomal Database Project (RDP) (17) and clustered into operational taxonomic units (OTUs) based on a 97% similarity cutoff. Sequencing and analysis of a mock community revealed the error rate to be 0.018%. Samples were rarefied to 4231 sequences

per sample in order to reduce the effects of uneven sampling bias.

Diversity analysis was performed using the Simpson diversity calculator and  $\theta_{YC}$  calculator metrics in mothur version 1.39.5 (16).  $\theta_{YC}$  distances were calculated to determine the dissimilarity between two samples. Random Forest classification models were built using the AUCRF R package using a leave-one-subject out approach (18). For each model the data were split into a 19-subject training set and a 1-subject test set. The model was built and cross-validated using AUCRFcv on the training set. The model was then tested on the left-out patient. This process was repeated iteratively for all subjects and results plotted as Receiver Operator Characteristic curves using the pROC R package (19). Resultant models were used to identify the OTUs that were most important for classifying each location. Species-level information for sequences of interest was obtained by aligning the sequences to the GenBank nucleotide database using blastn. The species name was only used if the identity score was  $\geq 99\%$  over the full-length of the contig and matched a single reference.

## Statistical analysis

Differences in community membership at the phyla level were tested using the analysis of molecular variance (AMOVA) metric in mothur. Differences in  $\theta_{YC}$  distances by location were tested using the Wilcoxon rank-sum test adjusted for multiple comparisons using the Benjamini-Hochberg procedure.

## Data availability

16S rRNA gene sequence reads and experiment metadata are available on the NCBI Sequence Read Archive (SRA) with accession number XXXX. A reproducible data analysis pipeline can be found at [https://github.com/SchlossLab/Flynn\\_LRColon\\_CancPrevRes\\_2017](https://github.com/SchlossLab/Flynn_LRColon_CancPrevRes_2017).

## Results

### Microbial membership and diversity of the proximal and distal colon

Luminal and mucosal samples were collected from the proximal and distal colon of 20 healthy individuals who had not undergone bowel preparation (Fig. 1). Subjects also collected stool at home one week prior to the procedure. To characterize the bacterial communities present at these sites, 16S rRNA gene sequencing was performed on DNA extracted from each sample. As expected, each site was primarily dominated by *Firmicutes* and *Bacteroidetes* (Fig. 2A) (20). Samples had varying levels of diversity at each site, irrespective of the individual (Fig. 2B). For example, the proximal mucosa was more diverse than the distal for some individuals while the opposite was true for others. Therefore we could not identify a clear pattern of changes in microbial diversity along the gut axis.

To compare similarity between the proximal and distal sides and within the lumen and mucosa, we compared the community structure of these sites based on the relative abundances of individual Operational Taxonomic Units (OTUs). Across all subjects we observed wide variation when comparing sample locations (Fig. 3A). Those ranges did not follow a clear pattern on an individual basis. However, when comparing median dissimilarity between the communities found in the proximal lumen and mucosa, the proximal and distal lumen, the proximal and distal mucosa, and the distal lumen and mucosa, we found that the proximal lumen and mucosa were most similar to each other than to the other samples ( $P < 0.005$ , Wilcoxon, BH adjustment).

### Fecal samples resemble luminal samples from the distal colon

Next, we compared the luminal and mucosal samples to the fecal sample of each subject. Amidst the large inter-subject variation, we did identify significantly less dissimilarity between the distal luminal sample and the feces (Fig. 3B,  $P < 0.05$ , Wilcoxon, BH adjustment). Furthermore, there was an even larger difference in the communities found in the distal mucosa compared to the fecal communities, indicating that the mucosa is as different from the stool as compared to lumen ( $P$



190 < 0.0005, Wilcoxon, BH adjustment). These results suggest that the contents of the distal lumen  
191 were most representative of the subjects' feces, and the mucosal microbiota are distinct from the  
192 fecal and luminal communities.

### 193 **Interpersonal community variation is greater than the variation between sites**

194 To determine what factors may have driven the differences seen among the samples, we compared  
195 the community dissimilarity between samples from all subjects (interpersonal) versus samples from  
196 within one subject (intrapersonal). We found that samples from one individual were far more  
197 similar to each other than to matched samples from the other subjects (Fig. 3C); this is consistent  
198 with previous human microbiome studies that have sampled multiple sites of the human colon  
199 (21–23). Thus interpersonal variation drove the differences between samples more than whether  
200 the sample came from the proximal or distal side of the colon or from the lumen or mucosa.

### 201 **Random Forest classification models identify important OTUs on each side**

202 To identify OTUs that were distinct at each site, we constructed several Random Forest models  
203 trained using OTU relative abundances. We built the first model to classify the luminal versus  
204 mucosal samples for the proximal and distal sides, independently (Fig. 4A). The models performed  
205 well when classifying these samples (proximal AUC = 0.764, distal AUC = 0.908). The OTUs that  
206 were most predictive of each site were identified by their greatest mean decrease in accuracy when  
207 removed from the model. For distinguishing the proximal lumen and mucosa, OTUs affiliated with  
208 the *Bacteriodes*, *Actinomyces*, *Psuedomonas* and *Enterobacteraceae* were included in the best  
209 model (Fig. 5A). The model to differentiate between the distal lumen and mucosa included OTUs  
210 affiliated with the *Turicibacter*, *Finegoldia*, *Peptoniphilus* and *Anaerococcus* (Fig. 5B). These  
211 results indicated that there were fine differences between the different sites of the colon, and that  
212 these could be traced to specific OTUs on each side.

213 Next, we built a Random Forest model to differentiate the proximal and distal luminal samples.

The model performed best when distinguishing the proximal versus distal mucosa (Fig. 4B, AUC = 0.850) whereas the model to differentiate between the proximal versus distal lumen performed poorly (AUC = 0.580). The OTUs included in the model differentiating the distal and proximal mucosa included members of the *Porphyromonas*, *Murdochiella*, *Finegoldia*, *Anaerococcus* and *Peptoniphilus* (Fig. 6A). The model that attempted to separate the the proximal and distal lumen included OTUs affiliated with the *Bacteroides*, *Clostridium IV* and *Oscillibacter* (Fig. 6B). Interestingly, *Anaerococcus* and *Finegoldia* were distinct between the mucosa and lumen and also helped to differentiate between the proximal and distal sides.

### **Bacterial OTUs associated with CRC and IBD are found in healthy individuals**

Given that specific bacterial species have been associated with colorectal cancer and IBD, we probed our sample set for these OTUs. Among our 100 samples, the most frequent sequence associated with the *Fusobacterium* genus was OTU179, which aligned via blastn to *Fusobacterium nucleatum subsp animalis* (100% over full length). This is the only species of *Fusobacterium* known to have oncogenic properties and be found on the surfaces of colorectal cancer tumors (24). There were 14 samples from 8 subjects with the *F. nucleatum subsp. animalis* sequences. Of the samples with the highest relative abundance of *F. nucleatum subsp. animalis*, four of the samples were from the proximal mucosa and three from the distal mucosa (Supplementary Fig. S1A). The second most frequent *Fusobacterium* sequence was OTU472, which aligned with 99% identity to *F. varium*. In addition to *F. nucleatum*, *F. varium* has been associated with IBD (25). Four subjects harbored *F. varium* and the samples were split evenly between the proximal and distal mucosa (Supplementary Fig. S1B). OTU152 was similar to the members of the *Porphyromonas* genus and the most frequent sequence in that OTU aligned to *Porphyromonas asacharolytica* (99% over full length), another bacterium commonly detected and isolated from colorectal tumors. OTU152 was only detected on the distal mucosa, and in fact was one of the OTUs the classification model identified as separating distal and proximal sides (Supplementary Fig. S1C). Among the 11 distal mucosa samples that were positive for *P. asacharolytica*, the relative abundances for this OTU

240 ranged from 0.01% to 16%. Thus, disease-associated OTUs could be found in our sample set of  
241 20 healthy individuals.

## 242 Discussion

243 We identified bacterial taxa that were specific to the lumen and mucosa of the proximal and distal  
244 sides of the human colon using samples collected during an unprepared colonoscopy of healthy  
245 subjects. We found that all locations contained a range of phylum relative abundances and a  
246 range of diversity, but that there was a wide variability between subjects. Pairwise comparisons  
247 of each of the sites revealed that the proximal mucosa and lumen were most similar to each  
248 other. Further, comparison of colonoscopy-collected samples with fecal samples demonstrated  
249 that the distal lumen was most similar to feces. Random Forest models built using OTU relative  
250 abundances from each sample identified microbiota that were particular to each location of the  
251 colon. Finally, we were able to detect some bacterial OTUs associated with colonic disease in our  
252 healthy cohort. Using unprepped colonoscopies and machine learning, we have identified bacterial  
253 taxa specific to the healthy proximal and distal human colon.

254 When examining the relative abundance of the dominant phyla at each site (i.e. *Bacteriodes* and  
255 *Firmicutes*), there was a wide amount of variation. This likely reflects not only the variability  
256 between human subjects, caused by differences in age, sex, diet, but also spatial “patchiness”  
257 in the gut microbiome. One study noted that the bacteria recoverable from the same mucosal  
258 sample location can be vastly different when the samples are taken just 1 cm away from each other  
259 (26). Similar patchiness was also observed in luminal contents and fecal samples themselves; there  
260 was separation of different interacting microbes along the length of a stool sample, for instance  
261 (27). That said, across our samples, the mucosal samples harbor more *Proteobacteria*, consistent  
262 with previous studies comparing mucosal swabs to luminal content in humans (4). Hence, the  
263 conclusions we were able to draw from phyla analysis may have been impacted by inter-subject  
264 patchiness.

265 To get around the noisiness from a diverse set of samples, we built Random Forest classification  
266 models to identify the microbiota that were specific to each side and in the lumen and mucosa. For  
267 each comparison we identified the top five OTUs that were strongly predictive of one site or another.  
268 Generally, OTUs identified in each location were consistent with known physiological gradients along  
269 the gut axis (5). For instance, the proximal mucosa contains the highest oxygen concentrations  
270 of the colon and harbored mucosa-associated facultative anaerobes such as *Actinomyces* and  
271 *Enterobacteraceae* and aerobic *Psuedomonas*. The distal mucosa was far more likely to host  
272 strictly anaerobic species such as *Porphyromonas*, *Anaerococcus*, *Finegoldia* and *Peptoniphilus*.  
273 Thus the gut microenvironment of each location likely enriches for these specific microbiota.

274 In addition to identifying features that are specific to each side of the gut, the ability of the  
275 Random Forest to classify samples can serve as a proxy for similarity. That is, a higher AUC value  
276 indicates the samples are more efficiently classified (and thus more different) than a model with a  
277 lower AUC value. For instance, the model separating the proximal and distal mucosa had an AUC  
278 of 0.850 whereas the model for classifying the proximal and distal lumen had a much lower AUC  
279 of 0.580. Further, the latter model required 44 OTUs to best separate the samples whereas the  
280 models separating the mucosa only needed 10 OTUs. The much lower AUC and need for a high  
281 number of features compared to other models suggest these locations are the most similar of the  
282 comparisons tested. We speculate that the model was less effective at classifying the proximal and  
283 distal luminal contents because the mucosal microenvironments have variable selective pressure  
284 along the colon than the luminal microenvironments.

285 We detected *F. nucleatum* and *P. asacharolytica* in 8 and 5 of our subjects, respectively. These  
286 bacteria have been shown to be predictive of colorectal cancer in humans (9) and have oncogenic  
287 properties in cell culture and in mice (28). Interestingly, while *F. nucleatum* was found on both  
288 sides of the colon, *P. asacharolytica* was only detected in the distal mucosa. Not much is known  
289 about the distribution of *P. asacharolytica* along the colon, but given its anaerobic lifestyle and  
290 asacharolytic metabolism, it is not surprising that it resides in the less-oxygen-rich and protein-rich

291 distal mucosa (4). In studies examining bacteria on colorectal cancer tumors, *F. nucleatum* was  
292 more commonly detected on proximal-sided tumors, and distribution of *F. nucleatum* decreased  
293 along the colon to rectum (29). In another study, *Fusobacterium* was associated with MSI with  
294 BRAF and KRAS mutations, molecular features of proximal CRC (30). Of the 8 (40%) individuals  
295 positive for *F. nucleatum* in the present study, the bacterium was spread across the proximal  
296 mucosa, distal lumen and distal mucosa. Data examining bacterial biofilms on the mucosa of  
297 CRC tumors suggests that *Fusobacteria* species are more commonly found on proximal tumors  
298 and in biofilms, indicating that it is not only the presence of the bacterium but the structure  
299 of the tumor community that contributes to *Fusobacterium*'s role in tumorigenesis (7). Finally,  
300 *Fusobacterium* and *Porphyromonas* populations not only co-occur on CRC tumors but also to  
301 synergize to promote tumorigenesis in an oral cancer model (31,32). Further analysis of the  
302 distribution and activities of these pathogens along the colon is needed to elucidate a mechanism  
303 for development of CRC or IBD subtypes in the proximal or distal colon.

304 The *Fusobacterium* species *nucleatum* and *varium* have been commonly isolated from mucosal  
305 biopsies of patients with IBD (25). Laboratory experiments with these isolates have shown that  
306 disease-isolated *F. nucleatum* are more invasive and stimulate more TNF- $\alpha$  production than strains  
307 from healthy individuals (10), suggesting the bacteria may increase inflammation in the gut as  
308 well (33). *F. varium* isolated from UC patients caused colonic ulcers in an experimental mouse  
309 model (34). *F. varium* was only detected in three subjects and two of those samples were isolated  
310 from the proximal mucosa (Supplementary Fig. S1B). *F. varium* is most commonly isolated from  
311 UC patient biopsies from the ileum or cecum (35), suggesting this species may exhibit preference  
312 for the different environmental conditions of these gastrointestinal sites. Further work will assess  
313 how gut environment may select for species which may then cause localized disease.

314 Specific comparisons of our findings to previously published studies of spatial variation are  
315 confounded by the use of bowel preparation methods. A rare report of a matched-colonoscopy  
316 study sampled 18 patient's colonic mucosa and luminal contents prior to and after bowel cleansing

(36). This study found that mucosal and luminal samples were distinguishable prior to bowel cleansing, but that bowel preparation resulted in an increase in shared OTUs between each site (36). After seven days, bowel cleansing not only made the samples more difficult to distinguish, but it also decreased the diversity observed across sites. Bowel preparation clearly biases the representation of microbiota recovered from sampling the lumen or mucosa.

By revealing specific differences in microbial populations at each location in the gut via sampling an unprepared bowel, we can begin to form hypotheses about how specific host-microbe interactions can affect disease progression of proximal and distal CRC and IBD subtypes. A better understanding of microbial activities in the gut can enhance microbiome-based screening and treatment modalities for these colon diseases.

## **Acknowledgments**

We thank all the individuals who volunteered for the study. This work was supported by the Rose and Lawrence C. Page Foundation (DKT). We would also like to thank Brian Kleiner, Chelsea Crofoot, and Kirk Herman for their roles in study coordination, subject recruitment, sample collection and sample processing.

## References

1. Yamauchi M, Lochhead P, Morikawa T, Huttenhower C, Chan AT, Giovannucci E, et al. Colorectal cancer: A tale of two sides or a continuum?: Figure 1. *Gut* [Internet]. *BMJ*; 2012;61:794–7. Available from: <https://doi.org/10.1136/gutjnl-2012-302014>
2. Forbes JD, Domselaar GV, Bernstein CN. The gut microbiota in immune-mediated inflammatory diseases. *Frontiers in Microbiology* [Internet]. *Frontiers Media SA*; 2016;7. Available from: <https://doi.org/10.3389/fmicb.2016.01081>
3. Halfvarson J, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature Microbiology* [Internet]. *Springer Nature*; 2017;2:17004. Available from: <https://doi.org/10.1038/nmicrobiol.2017.4>
4. Albenberg L, Esipova TV, Judge CP, Bittinger K, Chen J, Laughlin A, et al. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology* [Internet]. *Elsevier BV*; 2014;147:1055–1063.e8. Available from: <https://doi.org/10.1053/j.gastro.2014.07.020>
5. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology* [Internet]. *Springer Nature*; 2015;14:20–32. Available from: <https://doi.org/10.1038/nrmicro3552>
6. 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* [Internet]. *Elsevier BV*; 2013;14:207–15. Available from: <https://doi.org/10.1016%2Fj.chom.2013.07.007>
7. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Welch JLM, Rossetti BJ, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. *Proceedings of the National Academy of Sciences* [Internet]. *Proceedings of the National Academy of Sciences*; 2014;111:18321–

- 357 6. Available from: <https://doi.org/10.1073/pnas.1406199111>
- 358 8. McCoy AN, Araújo-Pérez F, Azcárate-Peril A, Yeh JJ, Sandler RS, Keku TO. Fusobacterium  
359 is associated with colorectal adenomas. Goel A, editor. PLoS ONE [Internet]. Public Library  
360 of Science (PLoS); 2013;8:e53653. Available from: [https://doi.org/10.1371/journal.pone.](https://doi.org/10.1371/journal.pone.0053653)  
361 0053653
- 362 9. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD. Microbiota-based model improves the  
363 sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Medicine [Internet].  
364 Springer Nature; 2016;8. Available from: <https://doi.org/10.1186/s13073-016-0290-3>
- 365 10. Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, DeVinney R, et al. Invasive  
366 potential of gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of  
367 the host. Inflammatory Bowel Diseases [Internet]. Ovid Technologies (Wolters Kluwer Health);  
368 2011;17:1971–8. Available from: <https://doi.org/10.1002/ibd.21606>
- 369 11. Brennan CA, Garrett WS. Gut microbiota, inflammation, and colorectal cancer. Annual  
370 Review of Microbiology [Internet]. Annual Reviews; 2016;70:395–411. Available from: <https://doi.org/10.1146/annurev-micro-102215-095513>  
371
- 372 12. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated  
373 and non-tumour-associated microbiota in colorectal cancer. Gut [Internet]. BMJ Publishing Group;  
374 2017;66:633–43. Available from: <http://gut.bmj.com/content/66/4/633>
- 375 13. Jalanka J, Salonen A, Salojärvi J, Ritari J, Immonen O, Marciani L, et al. Effects of bowel  
376 cleansing on the intestinal microbiota. Gut [Internet]. BMJ; 2014;64:1562–8. Available from:  
377 <https://doi.org/10.1136/gutjnl-2014-307240>
- 378 14. Harrell L, Wang Y, Antonopoulos D, Young V, Lichtenstein L, Huang Y, et al. Standard  
379 colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. Singh  
380 SR, editor. PLoS ONE [Internet]. Public Library of Science (PLoS); 2012;7:e32545. Available



from: <https://doi.org/10.1371/journal.pone.0032545>

15. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq illumina sequencing platform. *Applied and Environmental Microbiology* [Internet]. American Society for Microbiology; 2013;79:5112–20. Available from: <https://doi.org/10.1128/aem.01043-13>

16. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* [Internet]. American Society for Microbiology; 2009;75:7537–41. Available from: <https://doi.org/10.1128/aem.01541-09>

17. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* [Internet]. American Society for Microbiology; 2007;73:5261–7. Available from: <https://doi.org/10.1128/aem.00062-07>

18. Calle ML, Urrea V, Boulesteix A-L, Malats N. AUC-RF: A new strategy for genomic profiling with random forest. *Human Heredity* [Internet]. S. Karger AG; 2011;72:121–32. Available from: <https://doi.org/10.1159/000330778>

19. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: An open-source package for r and s to analyze and compare ROC curves. *BMC Bioinformatics* [Internet]. Springer Nature; 2011;12:77. Available from: <https://doi.org/10.1186/1471-2105-12-77>

20. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Medicine* [Internet]. Springer Nature; 2016;8. Available from: <https://doi.org/10.1186/s13073-016-0307-y>

21. Eckburg PB. Diversity of the human intestinal microbial flora. *Science* [Internet]. American Association for the Advancement of Science (AAAS); 2005;308:1635–8. Available from: <https://doi.org/10.1126/science.1110591>

405 //doi.org/10.1126/science.1110591

406 22. Cárcer DA de, Cuív PÓ, Wang T, Kang S, Worthley D, Whitehall V, et al. Numerical ecology  
407 validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria  
408 along the human colon. *The ISME Journal* [Internet]. Springer Nature; 2010;5:801–9. Available  
409 from: <https://doi.org/10.1038/ismej.2010.177>

410 23. Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, et al. Spatial heterogeneity and co-occurrence  
411 patterns of human mucosal-associated intestinal microbiota. *The ISME Journal* [Internet]. Springer  
412 Nature; 2013;8:881–93. Available from: <https://doi.org/10.1038/ismej.2013.185>

413 24. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. *Fu-*  
414 *sobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Re-*  
415 *search* [Internet]. Cold Spring Harbor Laboratory Press; 2011;22:299–306. Available from:  
416 <https://doi.org/10.1101/gr.126516.111>

417 25. Lee Y, Eun CS, Lee AR, Park CH, Han DS. *Fusobacterium* Isolates recovered from colonic  
418 biopsies of inflammatory bowel disease patients in Korea. *Annals of Laboratory Medicine* [Internet].  
419 Korean Society for Laboratory Medicine (KAMJE); 2016;36:387. Available from: [https://doi.org/](https://doi.org/10.3343/alm.2016.36.4.387)  
420 [10.3343/alm.2016.36.4.387](https://doi.org/10.3343/alm.2016.36.4.387)

421 26. Hong P-Y, Croix JA, Greenberg E, Gaskins HR, Mackie RI. Pyrosequencing-based analysis  
422 of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-  
423 heterogeneity. Ahmed N, editor. *PLoS ONE* [Internet]. Public Library of Science (PLoS);  
424 2011;6:e25042. Available from: <https://doi.org/10.1371/journal.pone.0025042>

425 27. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, et  
426 al. Bacterial biogeography of the human digestive tract. *Scientific Reports* [Internet]. Springer  
427 Nature; 2011;1. Available from: <https://doi.org/10.1038/srep00170>

428 28. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. *Cell Host & Microbe* [Internet].

- Elsevier BV; 2014;15:317–28. Available from: <https://doi.org/10.1016/j.chom.2014.02.007>
29. Mima K, Cao Y, Chan AT, Qian ZR, Nowak JA, Masugi Y, et al. *Fusobacterium nucleatum* in colorectal carcinoma tissue according to tumor location. *Clinical and Translational Gastroenterology* [Internet]. Springer Nature; 2016;7:e200. Available from: <https://doi.org/10.1038/ctg.2016.53>
30. Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, et al. *Fusobacterium* in colonic flora and molecular features of colorectal carcinoma. *Cancer Research* [Internet]. American Association for Cancer Research (AACR); 2014;74:1311–8. Available from: <https://doi.org/10.1158%2F0008-5472.can-13-1865>
31. Whitmore SE, Lamont RJ. Oral bacteria and cancer. Goldman WE, editor. *PLoS Pathogens* [Internet]. Public Library of Science (PLOS); 2014;10:e1003933. Available from: <https://doi.org/10.1371/journal.ppat.1003933>
32. Flynn KJ, Baxter NT, Schloss PD. Metabolic and community synergy of oral bacteria in colorectal cancer. McMahon K, editor. *mSphere* [Internet]. American Society for Microbiology; 2016;1:e00102–16. Available from: <https://doi.org/10.1128/msphere.00102-16>
33. Dharmani P, Strauss J, Ambrose C, Allen-Vercoe E, Chadee K. *Fusobacterium nucleatum* infection of colonic cells stimulates MUC2 mucin and tumor necrosis factor alpha. *Infection and Immunity* [Internet]. American Society for Microbiology; 2011;79:2597–607. Available from: <https://doi.org/10.1128/iai.05118-11>
34. Ohkusa T. Induction of experimental ulcerative colitis by *fusobacterium varium* isolated from colonic mucosa of patients with ulcerative colitis. *Gut* [Internet]. BMJ; 2003;52:79–83. Available from: <https://doi.org/10.1136/gut.52.1.79>
35. Ohkusa T, Sato N, Ogihara T, Morita K, Ogawa M, Okayasu I. *Fusobacterium varium* localized in the colonic mucosa of patients with ulcerative colitis stimulates species-specific antibody. *Journal of Gastroenterology and Hepatology* [Internet]. Wiley-Blackwell; 2002;17:849–53. Available from:

453 <https://doi.org/10.1046/j.1440-1746.2002.02834.x>

454 36. Shobar RM, Velineni S, Keshavarzian A, Swanson G, DeMeo MT, Melson JE, et al. The effects  
455 of bowel preparation on microbiota-related metrics differ in health and in inflammatory bowel  
456 disease and for the mucosal and luminal microbiota compartments. *Clinical and Translational*  
457 *Gastroenterology* [Internet]. Springer Nature; 2016;7:e143. Available from: [https://doi.org/10.](https://doi.org/10.1038/ctg.2015.54)  
458 [1038/ctg.2015.54](https://doi.org/10.1038/ctg.2015.54)

**Figures**

**Figure 1**

Sampling strategy. A flexible sigmoidoscope was used to sample the distal colonic luminal contents and mucosa. The scope was inserted ~ 25cm into the subject and biopsy forceps were used to sample the luminal contents (D, inset). A separate set of biopsy forceps was used to sample the distal mucosa (D, inset). The sigmoidoscope was removed. A pediatric colonoscope was inserted and used to access the proximal colon (P, inset). Biopsies were taken of the proximal luminal contents and mucosa as described. One week prior to the procedure stool was collected at home and sent into the laboratory. Representative images from one individual are shown.

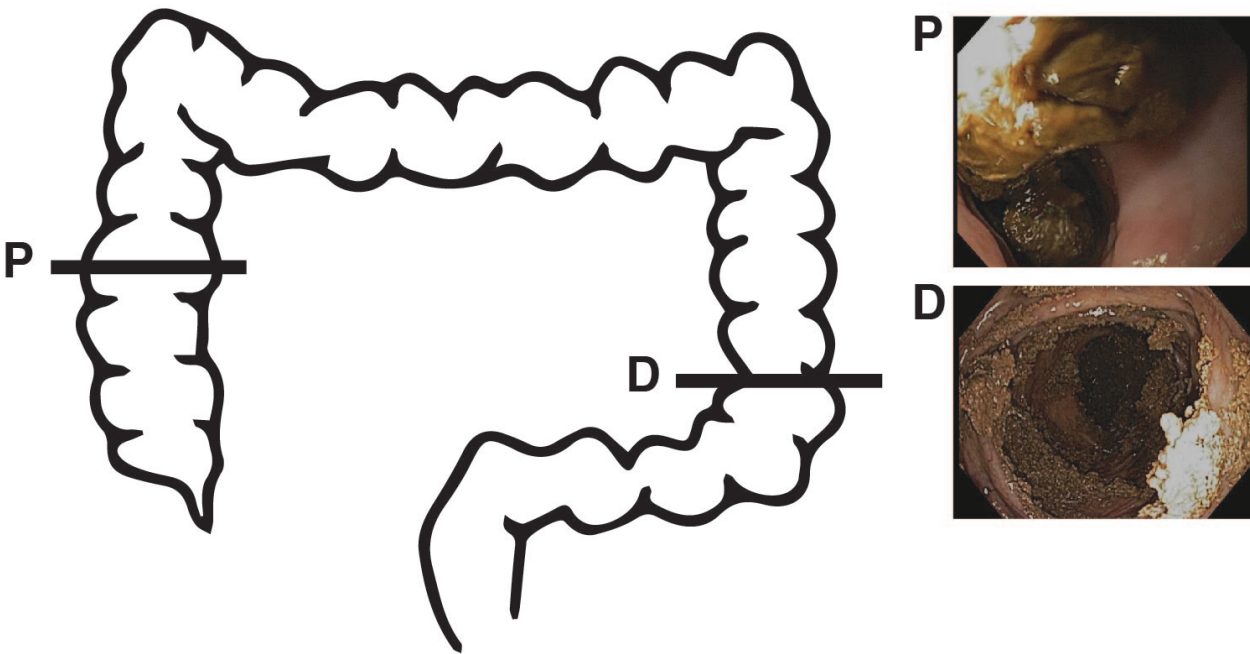


Figure 1: Fig 1

**Figure 2**

Phylum-level relative abundance and diversity in the proximal and distal human colon. A) Relative abundance of the top five bacterial phyla in each sampling site. Each box represents the median and interquartile range. B) Simpson diversity of the microbial communities at each location. The horizontal lines represent the median values.

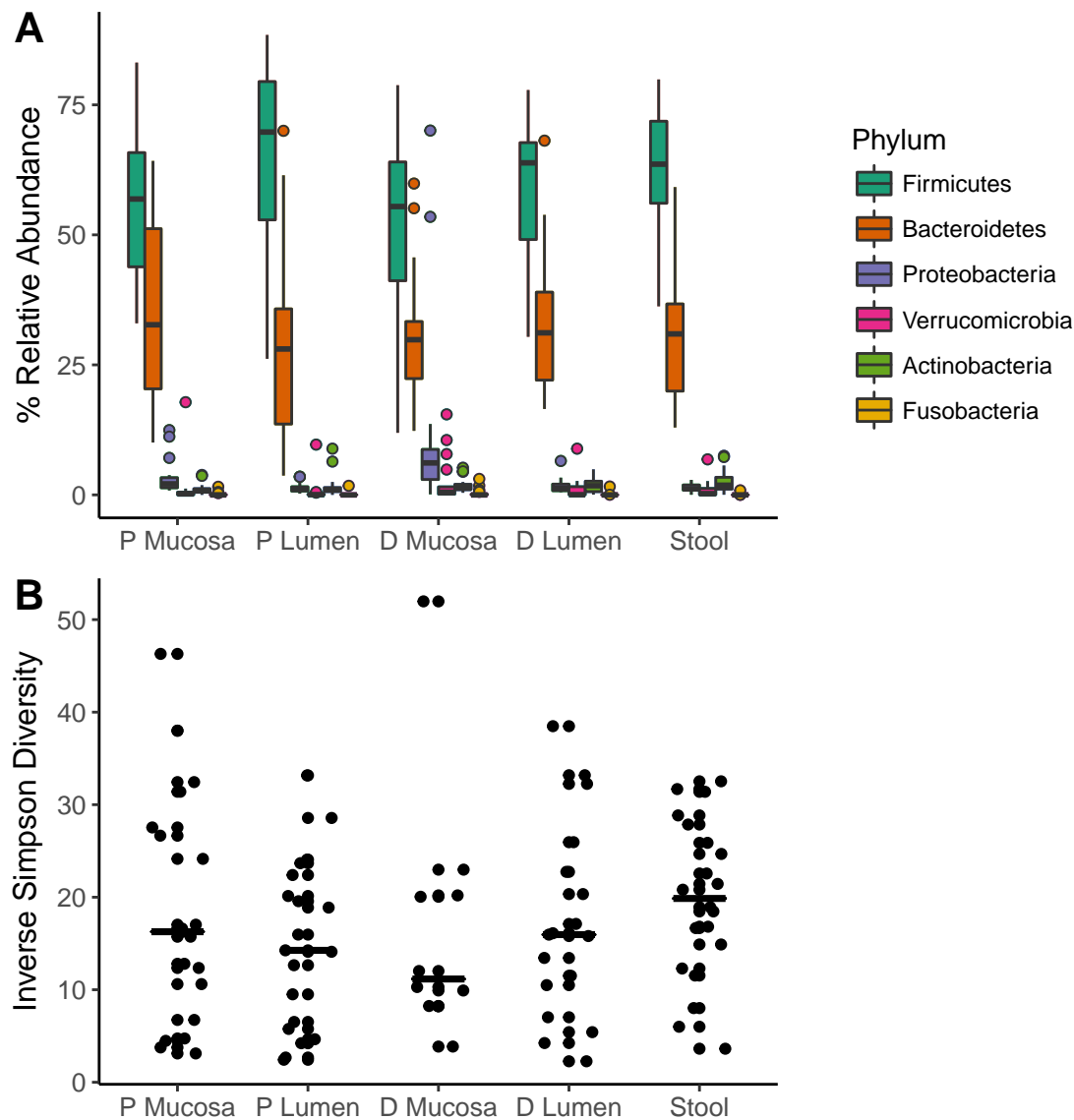


Figure 2: Fig 2

**Figure 3**

Comparison of microbial community structure between sites of the gut.  $\theta_{YC}$  distances are shown to indicate the interpersonal dissimilarities between two sites – each point represents one individual. In (A), comparisons of the proximal and distal mucosal and lumen are shown. In (B), comparisons of each site to the exit stool are shown. In (C), comparisons of samples from all subjects to each other (interpersonal) or within one subject (intrapersonal) are shown.

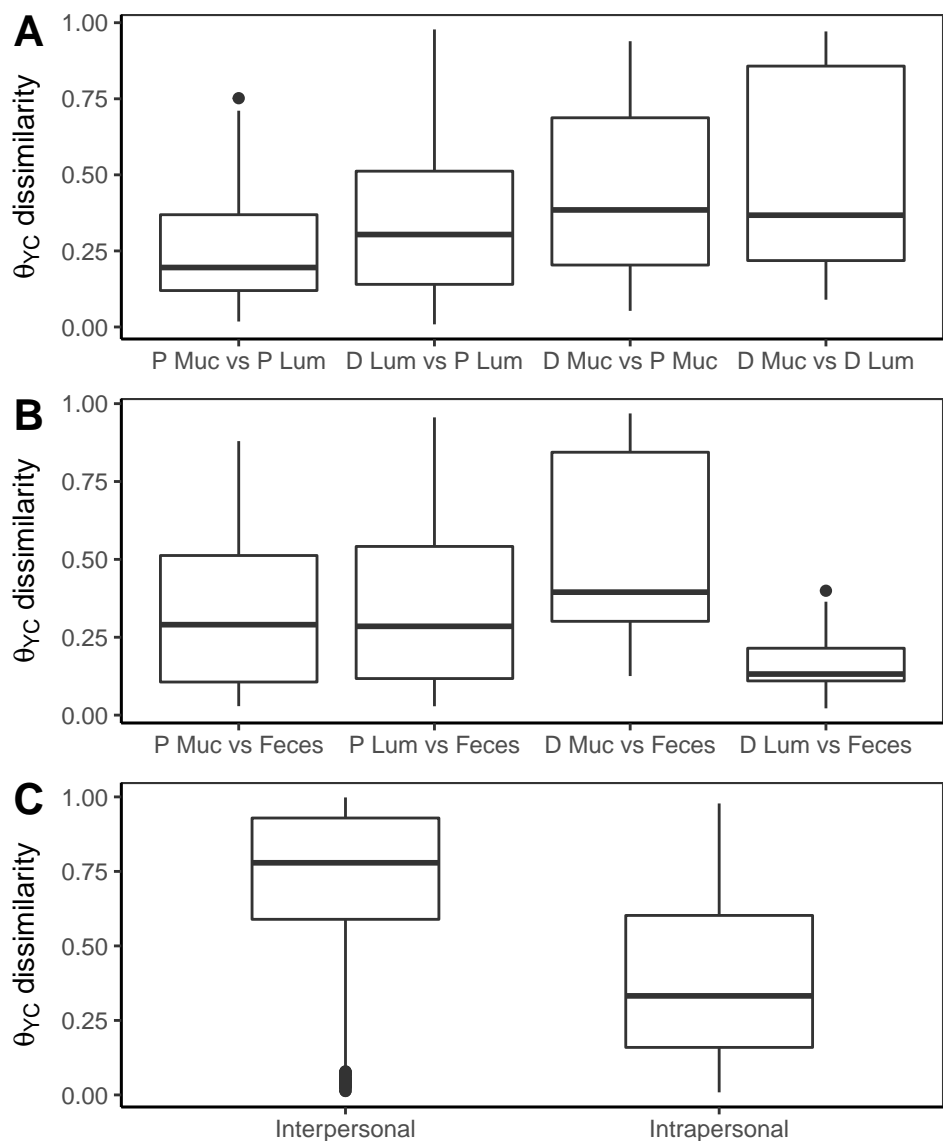


Figure 3: Fig 3

479 **Figure 4**

480 Random Forest classifies locations in the colon. A) Receiver Operator Characteristic curves are  
 481 shown for the Random Forest model classifying lumen and mucosal samples for the distal (red)  
 482 and proximal (blue) sides of the colon. (B) Receiver Operator Characteristic curves are shown for  
 483 the 10-fold cross validation of the Random Forest model classifying distal mucosa vs proximal  
 484 mucosa (green) and distal lumen versus proximal lumen (purple).

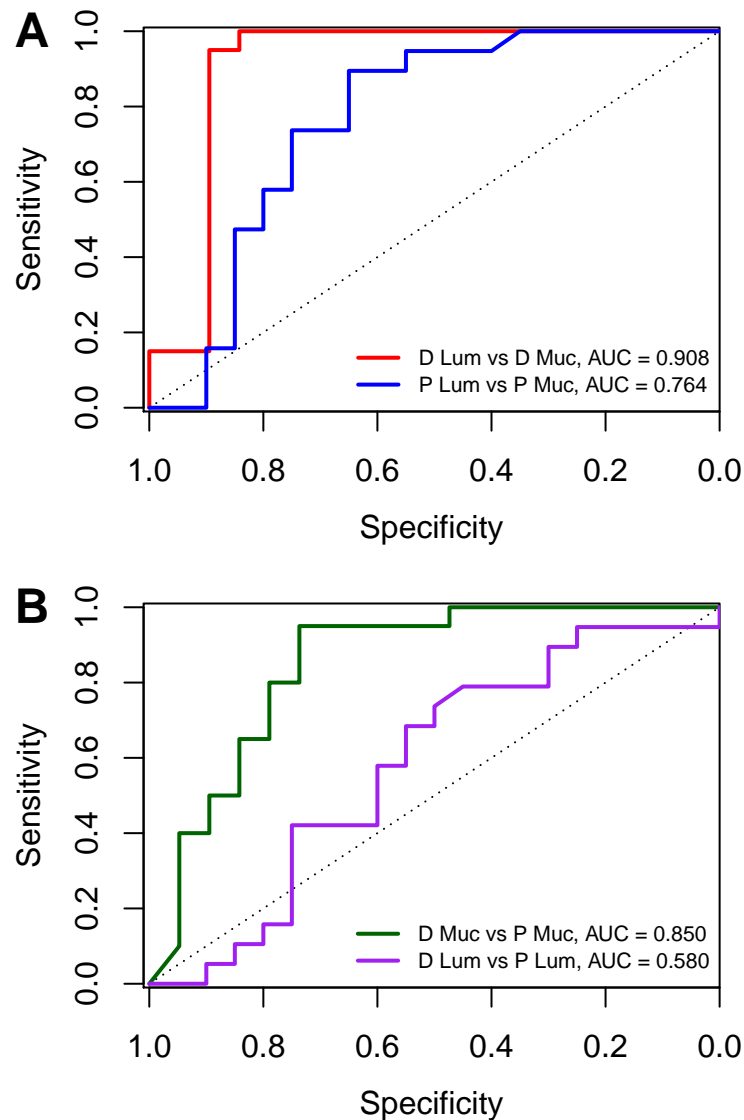


Figure 4: Fig 4



**Figure 5**

Taxa specific to the distal and proximal sides of the colon. Top five OTUs that are most important for the classification model for the distal mucosa and lumen (A) and the proximal mucosa and lumen (B). The vertical lines represent the median values for each OTU.

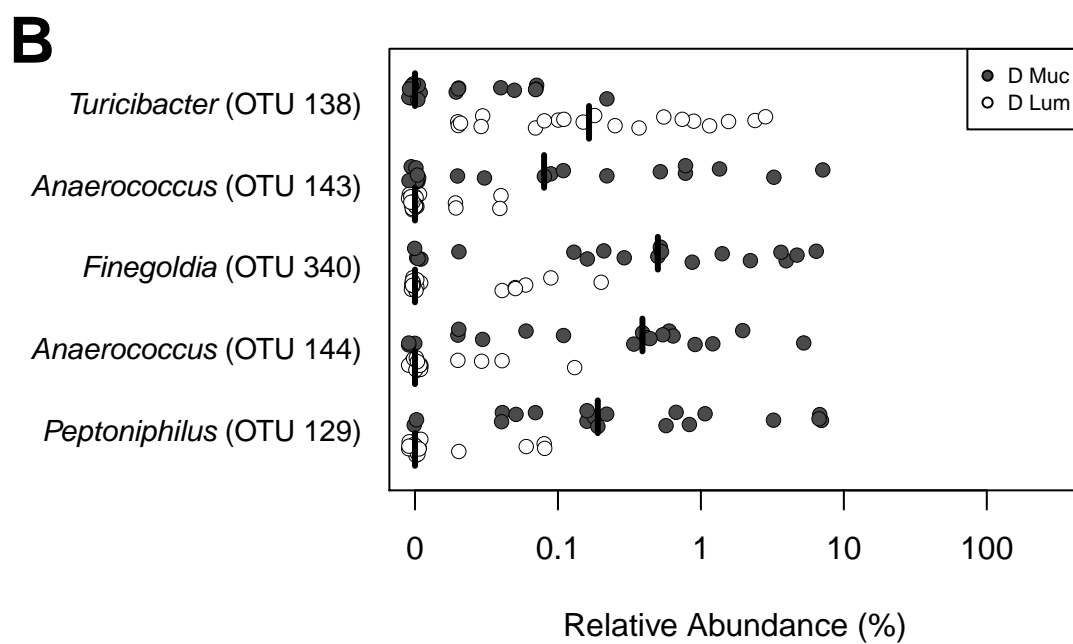
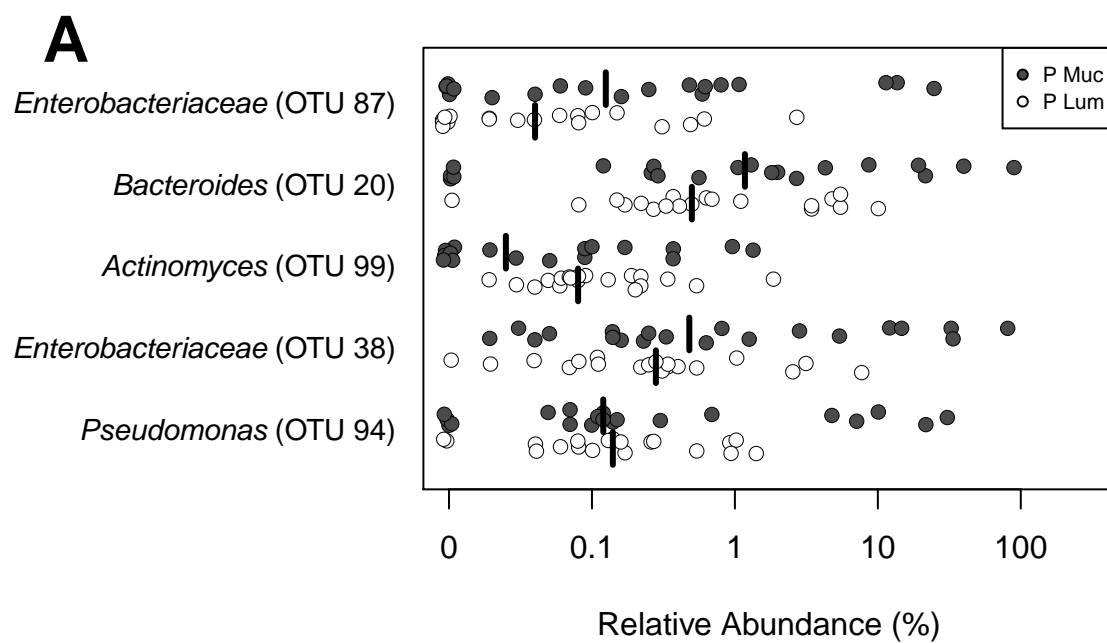


Figure 5: Fig 5

**Figure 6**

Taxa specific to the distal and proximal mucosa and lumen. The five OTUs that were most important differentiating the distal and proximal mucosa (A) and the distal and proximal lumen (B). The vertical lines represent the median values for each OTU.

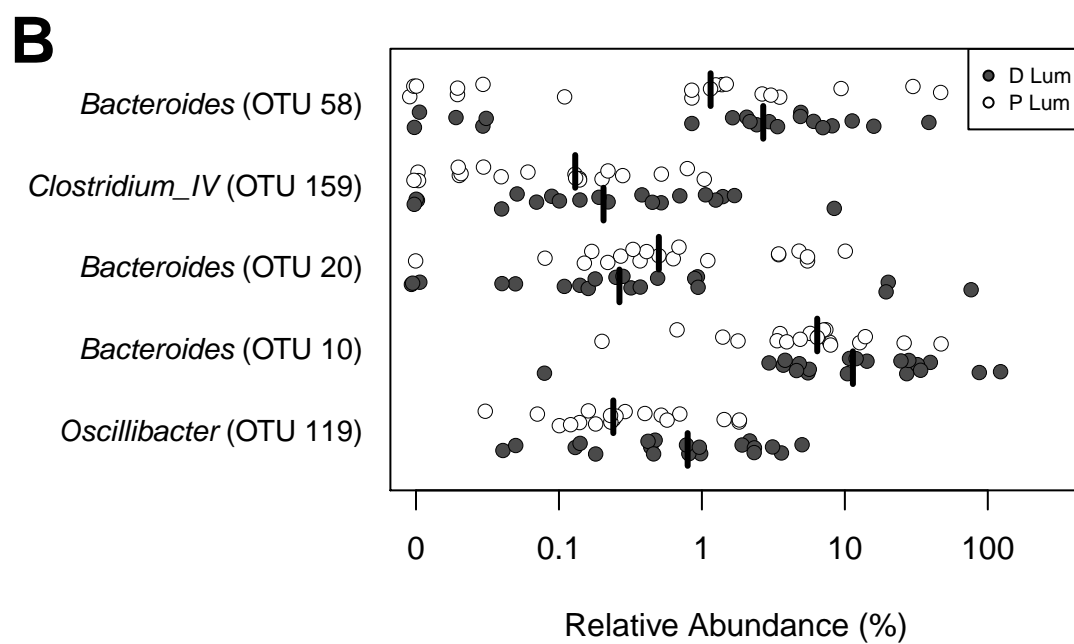
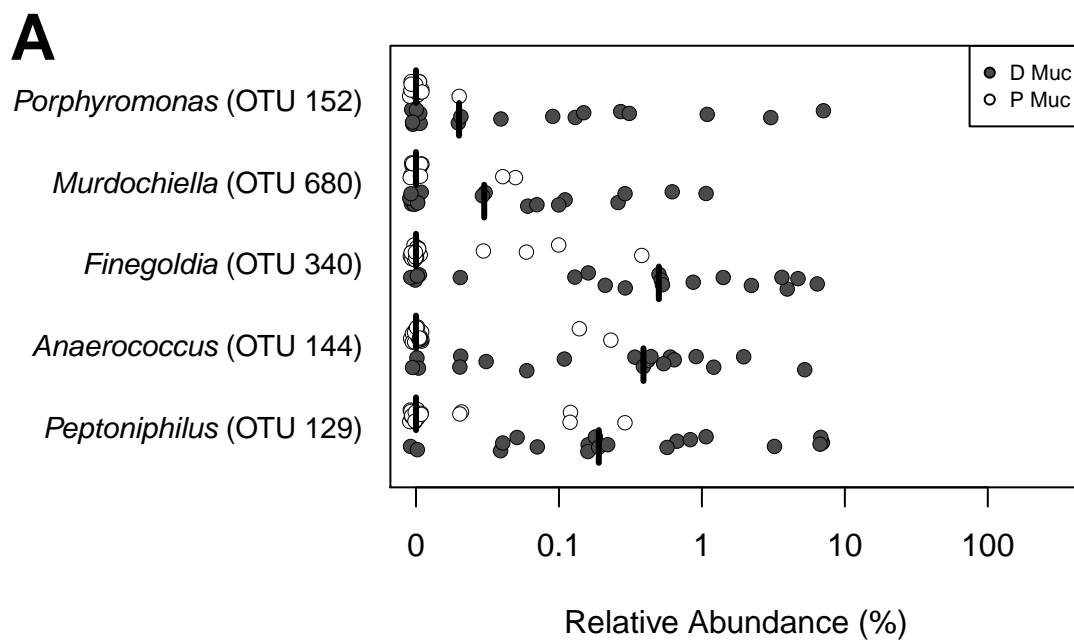


Figure 6: Fig 6

**Figure S1**

Location and relative abundance of cancer-associated OTUs. Relative abundance was calculated and plotted by sample site for each OTU of interest: (A) *Fusobacterium nucleatum* subsp. *animalis* (B) *Fusobacterium varium* and (C) *Porphyromonas asacharolytica*

