Title: Distinct Colon Mucosa Microbiomes associated with Tubular Adenomas and Serrated Polyps

Authors: Julio Avelar-Barragan*, Lauren DeDecker**, Zachary Lu**, Bretton Coppedge*, William E. Karnes**, Katrine L. Whiteson*

* School of Biological Sciences, University of California, Irvine

** School of Medicine, University of California, Irvine.

Corresponding author:

Julio Avelar-Barragan – javelarb@uci.edu

Keywords: Colorectal cancer, polyps, human, gut, microbiome, serrated, tubular, adenoma

Abbreviations:

CRC	Colorectal cancer
APC	Adenomatous polyposis coli
HPP	Hyperplastic polyp
TSA	Traditional serrated adenoma
SSP	Sessile serrated polyp
TA	Tubular adenoma
OTU	Operational taxonomic unit
ASV	Amplicon sequence variant
ORF	Open reading frame
LME	Linear mixed effects model
GH	Glycoside hydrolase

Abstract:

***Background*:** Colorectal cancer is the second most deadly and third most common cancer in the world. Its development is heterogenous, with multiple mechanisms of carcinogenesis. Two distinct mechanisms include the adenoma-carcinoma sequence and the serrated pathway. The gut microbiome has been identified as a key player in the adenoma-carcinoma sequence, but its role in serrated carcinogenesis is unclear. In this study, we characterized the gut microbiome of 140 polyp-free and polyp-bearing individuals using colon mucosa and fecal samples to determine if microbiome composition was associated with each of the two key pathways.

***Results*:** We discovered significant differences between colon mucosa and fecal samples, explaining 14% of the variation observed in the microbiome. Multiple mucosal samples were collected from each individual to investigate the gut microbiome for differences between polyp and healthy intestinal tissue, but no such differences were found. Colon mucosa sampling revealed that the microbiomes of individuals with tubular adenomas and serrated polyps were significantly different from each other and polyp-free individuals, explaining 2-10% of the variance in the microbiome. Further analysis revealed differential abundances of *Eggerthella lenta*, *Clostridium scindens*, and three microbial genes across tubular adenoma, serrated polyp, and polyp-free cases.

***Conclusion*:** By directly sampling the colon mucosa and distinguishing between the different developmental pathways of colorectal cancer, this study helps characterize potential mechanistic targets and diagnostic biomarkers for serrated carcinogenesis. This research also provides insight into multiple microbiome sampling strategies by assessing each method's practicality and effect on microbial community composition.

Introduction:

Colorectal cancer (CRC) is the second most deadly and third most common cancer globally, accounting for over 900,000 deaths in 2020.[1] The etiologies of CRC are multifactorial, with only 5-10% of cases being attributable to hereditary germline mutations.[2] Significant risk factors include diets high in red meat and low in fiber, obesity, physical inactivity, drug and alcohol usage, and chronic bowel inflammation.[3-6] Each of these factors is associated with compositional and functional changes in the collective community of bacteria, fungi, archaea, and viruses that inhabit the colon.[7-10] Commonly referred to as the gut microbiome, this community of microorganisms has been identified as a potential regulator of CRC initiation and progression.

Colorectal polyp formation precedes cancer development and is influenced by various environmental factors and host genetics. Polyps most commonly progress into malignancy through the adenoma-carcinoma sequence.[11] This pathway is characterized by chromosomal instability and mutations in the adenomatous polyposis coli (APC) gene, KRAS oncogene, and TP53 tumor suppressor gene.[12] Alternatively, 15 to 30% of CRCs develop through the serrated pathway.[13] This pathway is characterized by the epigenetic hypermethylation of gene promoters to produce a CpG island methylator phenotype.[13] In addition to the epigenetic inactivation of tumor suppressor genes, BRAF or KRAS mutations are also common.[13] The serrated pathway often results in the production of hyperplastic polyps (HPPs), traditional serrated adenomas (TSAs), and sessile serrated polyps (SSPs). Premalignant polyps from both pathways can be screened for and removed during colonoscopy to prevent CRC formation, but incomplete polyp resection or escaped detection can result in the development of interval cancers. Compared to other colorectal polyps, SSPs are disproportionately responsible for interval cancers, as their flat morphology makes them

difficult to detect.[14] Therefore, additional detection methods, such as SSP-specific biomarkers, would assist with CRC prevention.

One potential avenue for polyp-specific biomarker discovery is the gut microbiota. SSPs often overexpress mucin forming genes, like MUC6, MUC5aC, MUC17, and MUC2, producing a mucus cap, which may harbor mucin-degrading bacteria and other microbes.[15] Finding microbiome alterations in patients consistent with the presence of SSPs would enable gastroenterologists to personalize their technique and screening frequency for these higher risk patients. Additionally, elucidating the microbiome alterations specific to the adenoma-carcinoma sequence and the serrated pathway would help better understand the mechanisms of how particular microbes, their metabolites, and dysbiosis may contribute to CRC carcinogenesis.

Studies comparing the microbiomes of these two pathways with healthy controls have yet to discover differences between healthy individuals and those with serrated polyps.[16-18] One limitation of existing CRC research is the dominant use of stool samples for assessing the microbiome, which has been shown to reflect the bacteria found in the lumen of the colon.[19] Mucosal samples, on the other hand, reflect the bacteria that are adherent to the epithelium and thus may be more clinically relevant in CRC pathogenesis.[20] To investigate this, and the role of the microbiome in the adenoma-carcinoma sequence and serrated pathway, we used multiple sampling techniques to obtain microbiome samples from colorectal polyps. Samples were collected during and after colonoscopy from healthy individuals or those with tubular adenomas (TAs), HPPs, or SSPs. When possible, samples from the same individual were collected from polyps and the healthy colon epithelium opposite from these polyps. Stool samples were also collected 4-6 weeks after colonoscopy. We used a combination of amplicon (16S and ITS) and shotgun sequencing to characterize the microbial communities of samples. The purpose of our

work was to: 1) investigate the feasibility and utility of using two novel methods of microbiome sampling during colonoscopy; 2) elucidate hyperlocal differences in the microbiome of the colon to find polyp specific biomarkers; and 3) elucidate changes in the microbiome specific to CRC precursors in the adenoma-carcinoma sequence versus the serrated pathway. Our key hypothesis was that there would be distinct differences between the microbiomes of individuals with serrated polyps versus TAs.

Methods and Materials:

Subject Recruitment and Criteria:

Individuals who presented for colonoscopy with indications of screening or a prior history of colorectal polyps were asked to participate in the study. Subjects who were pregnant, had taken antibiotics within 6 weeks of colonoscopy, or with known inflammatory bowel diseases, were excluded. In total, 140 individuals were recruited for this study.

Colonoscopy Preparation, Procedure, and Sample Collection:

Before colonoscopy, subjects were asked to adhere to a clear liquid diet for 24 hours. Bowel cleansing was done using Miralax, or polyethylene glycol with electrolytes administered as a split dose, typically 12 and 5 hours before the procedure. Sample collection focused on two direct and two indirect microbiome sampling methods. The first direct sampling method involved brushing the mucosa of colon tissue during colonoscopy. Since mucosal brushes can potentially damage or agitate the intestine, we also employed a novel method of direct microbiome sampling in which colonoscopy washing fluid was sprayed directly on to the target mucosa and immediately re-suctioned into a storage vial. Participants with polyps had additional mucosal washing aspirates taken near the polyp, as well as brushings of the polyp and opposite wall of the polyp. When more than one polyp was found, the largest polyp was targeted for mucosal brushing and aspirate

sampling. The first indirect sampling method involved collecting an aliquot of the post-colonoscopy lavage fluid. This lavage fluid was from the entirety of the procedure, thus reflecting the microbiome throughout the colon. Sampling during colonoscopy resulted in a total of 1,685 samples, which were collected in sterile cryogenic tubes and placed on ice until the colonoscopy procedure was finished. Afterwards, the samples were stored at -80°C. Additional data collected included indication for procedure, age, sex, ethnicity, BMI, family history, and findings, including the size, shape, location, and pathology of all polyps found or removed.

Collection of Fecal Samples:

For the second indirect sampling method, subjects were encouraged to send follow-up fecal samples four to six weeks post-colonoscopy. Subjects who complied were compensated $20 USD. Subjects were provided with a fecal collection kit, which contained collection equipment, prepaid shipping labels, and Zymo DNA/RNA shield preservation buffer (R1101). Samples were returned via the United States Postal Service. After arrival, samples were stored at -80°C. A total of 38 fecal samples were returned.

Polyp and Subject Type classification:

Polyp biopsies collected during colonoscopy were sent to a pathologist for classification. This information was then recorded for the corresponding mucosal brush aspirate samples. Polyp pathology reports were also used to broadly categorize samples by subject type. The three subject types were polyp-free subjects, TA-bearing subjects, and serrated polyp-bearing subjects, which included HPPs and SSPs. For example, if a sample was taken from healthy intestinal tissue of an individual who was found to have a TA, that sample and all others from the same individual would be included in the TA-bearing subject type. Three individuals had both a TA and an SSP and were classified as serrated polyp-bearing subjects.

DNA extraction:

Colonoscopy and fecal samples were thawed on ice for DNA extraction. For mucosal aspirates and lavage aliquot samples, 250 uL of fluid were taken from each sample and then extracted using ZymoBiomics DNA Miniprep Kit (D4300) according to the manufacturer's protocol. For mucosal brushes, 750 uL of ZymoBIOMICS™ Lysis Solution was mixed with the brushes in their original sterile cryogenic tube and vortexed for 5 minutes to suspend the contents of the brush into solution. The solution was then transferred and extracted according to the manufacturer's protocol. Fecal samples stored in Zymo DNA/RNA shield were thawed, mixed by vortexing, and 750 uL of fecal slurry was extracted according to the manufacturer's protocol.

Amplicon Library preparation and sequencing:

To amplify the V4 region of the bacterial 16S rRNA gene, 0.5 ng of DNA template was combined with 9.5 uL of PCR grade water, 12.5 uL of 1x AccustartII PCR tough mix (QuantaBio), 1 ug of BSA, 0.2 uM of 926R reverse 16S primer, and 0.2 uM of barcoded, Illumina adapter sequence-tagged forward primer 515F in 25 uL reactions. Each sample was amplified for 30 cycles (94°C for 3 min; 94°C for 45 sec, 55°C for 30 sec, 72°C for 20 sec; repeat steps 2-4 30 times; 72°C for 10 min).

The ITS2 region of the fungal 18S rRNA gene was amplified using the methods described by Looby et al.[51] Briefly, 0.5 ng of DNA template was combined with 9.5 uL of PCR grade water, 12.5 uL of 1x AccustartII PCR tough mix (QuantaBio), 1 ug of BSA, 0.3 uM of ITS9, and 0.3 uM of a staggered, barcoded reverse ITS4 primer for a total reaction volume of 25 uL. Samples were PCR amplified at 94°C for 5 min, 35 cycles of 95°C for 45 sec, 50°C for 1 min, 72°C for 90 sec, and a final extension step of 72°C for 10 min.

Both purified amplicon libraries were quantified using the Qubit dsDNA HS Assay Kit and pooled at equimolar concentrations. The pooled amplicon library was cleaned and concentrated using Agencourt AMPure XP beads (Beckman-Coulter, A63880) according to the manufacturer's protocol. Equimolar PhiX was added at 10% final volume to the amplicon library and sequenced on the Illumina MiSeq platform, yielding 300bp paired-end sequences. An average of 41,578 +/- 35,920 reads per sample were obtained for 16S amplicons, while an average of 22,252 +/- 17,000 reads per sample were obtained for ITS amplicons.

Shotgun Library preparation and sequencing:

Shotgun sequencing libraries were prepared using the Illumina Nextera Flex kit, using an adapted protocol as described by Bruinsma et al.[52] Here, a maximum of 5 uL or 50 ng (whichever was reached first) of DNA from each sample was tagmented by Nextera bead-linked transposomes for 15 min at 55°C. Next, 1.25 uL of 1 uM forward and 1.25 uL of 1 uM reverse barcodes were added to each sample and annealed via PCR using KAPA HiFi DNA Polymerase (Roche Life Science). Samples were then pooled and cleaned of smaller fragments using the included sample purification beads according to the manufacturer's protocol. The pooled sample libraries were quantified using Quanti-iT PicoGreen dsDNA (P7589) and ran on a bioanalyzer to check fragment size. Lastly, libraries were packaged on dry ice and shipped overnight to Novogene Corporation Inc. (Sacramento, CA) to be sequenced using Illumina's Hiseq 4000 for 150 bp paired-end sequencing. An average of 1,267,359 +/- 690,384 reads per sample were obtained.

OTU Table generation:

For amplicon data, sequences were processed using Qiime2-2019.1.[53] Index sequences were extracted from reads and demultiplexed in the presence of quality filtering using DADA2 in QIIME2.[54] Reads were filtered to remove chimeric sequences, and clustered into amplicon

sequence variants (ASVs). Taxonomy was assigned with the Greengenes database (May 2013) for bacteria.[55]

For sequences produced by shotgun metagenomic sequencing, we first removed sequencing adapters, barcodes, and sequencing artifacts, followed by demultiplexing using BBMap.[56] After demultiplexing, sequences were quality filtered so that the minimum quality score did not fall below 28 using PRINSEQ++.[57] Removal of human-derived reads was performed in Bowtie2 by aligning samples to a reference human genome, hg38.[58] Next, we used IGGSearch on the lenient preset to characterize the taxonomy of our quality-filtered sequences to generate an operational taxonomic unit (OTU) table.[59]

OTU Table analysis:

Analysis of OTUs was performed using R 3.6.3. A sequencing control was included for each library to remove potential contaminants before analysis (ZymoBIOMICS #D6305). OTUs that appeared in the control, but which were not part of the microbial community standard were removed from the OTU table. To account for read-depth variability, 16S and ITS OTU counts were rarefied to 3,000 and 1,000 reads, respectively.[60] For shotgun sequence data, 1,500 marker gene reads from each sample were randomly subsampled ten times and then averaged.

Alpha diversity for each OTU table was calculated using the diversity and specnumber functions in the Vegan v2.5-6 package.[61] Significance of alpha diversity scores was tested using a linear-mixed effect model, adjusted for repeated measurements and sample plate effects using the nlme package, v3.1-148.[62] The adonis function in Vegan was used to generate Bray-Curtis matrices and perform PERMANOVA analyses from OTU tables. Non-metric multidimensional scaling was performed using the metaMDS function in Vegan. SparCC was used for correlation analysis of OTUs using the fastspar and MakeBootstraps.py functions.[63] Pseudo p_1 values were

197 calculated for sparCC correlations using the PseudoPvals.py script. Plotting for all analyses was
198 done using ggplot2 v3.3.0.[64]

199 To identify differentially abundant microbes, we utilized Random Forest and ANCOM
200 v2.1.[65] Random forests were performed using the rfPermute v2.1.81 package in R. Random Forest
201 analysis was performed to identify the top variables of importance for healthy to TA-bearing,
202 healthy to serrated polyp-bearing, and TA to serrated polyp-bearing host comparisons.
203 Significance testing for differentially abundant microbes was performed using Dunn's Kruskal-
204 Wallis on the top variables of importance. Microbial abundances were averaged for each subject
205 to avoid repeated measurements, and p-values were adjusted for false discovery rate. Random
206 Forest out-of-box accuracy was assessed by plotting Random Forest proximity and confidence
207 scores, using the proximityPlot and plotConfMat function within rfPermute. ANCOM v2.1 was
208 performed on OTU tables to identify differentially abundant microbes between sample types while
209 accounting for multiple samples per individual.

210 ***Functional Microbiome analysis:***

211 Functional annotation of shotgun sequences was performed by cross-assembling samples into
212 contigs using MEGAHIT v1.1.1.[66] Contigs larger than 2,500 bp had open reading frames (ORFs)
213 called by Prodigal v2.6.3.[67] The resulting ORFs were functionally annotated using eggNOG
214 mapper v2, using the eggNOG v5.0 database.[68] Individual samples were aligned to annotated
215 ORFs using bowtie2 v2.3.5.1 to obtain per-sample ORF abundances. Per sample ORF abundances
216 were compiled into a single ORF abundance table using the pileup.sh script from BBMap v38.79.
217 Principal coordinate analysis was performed by normalizing gene abundances using
218 MicrobeCensus and the cmdscale function from vegan.[69] KEGG BRITE annotations were used to
219 visualize the relative abundances of functional hierarchies. DESeq2 v1.2.6 was used to identify

differentially abundant microbial genes using the DESeq function in R.[70] Proximal and distal mucosal aspirates were processed through DESeq2 separately to avoid repeated measurements. Differentially abundant genes that were not present in at least 10% of subjects were removed from the analysis. The taxonomy of differentially abundant genes was resolved by aligning their protein sequences to the NCBI nr database using blastp.

Results:

Microbiomes of Mucosal and Lavage Samples are similar to each other but different from those in Feces:

To determine whether microbiomes varied between sample types, mucosal brushes, mucosal aspirates, and lavage aliquots were collected from 140 individuals during colonoscopy; 38 of these individuals provided a fecal sample 4-6 weeks after their procedure (Figure 1). Using 16S amplicon sequencing on a subset of individuals, we observed no significant differences in Shannon diversity or richness across mucosal brushes, mucosal aspirates, and lavage aliquots (linear mixed effects model (LME): $p > 0.05$, Figure 2A). PERMANOVA analysis of Bray-Curtis dissimilarities revealed that the individual explained the greatest amount of variation in microbiome composition ($R^2 = 0.56$, $p = 0.001$; Supplemental table 1). This analysis found no significant differences in the microbiomes associated with mucosal brushes, mucosal aspirates, and lavage aliquots from within the same individual ($R^2 = 0.12$, $p = 0.49$; Supplemental table 1). The lack of significance was consistent with no discernable clusters based on sample type (Figure 2B). We were unable to identify any microbes whose abundances significantly differed across the three sampling methods, except for three ASVs; one from the *Gemellaceae* family and two *Streptococcus spp.*, whose abundances were all higher in mucosal aspirates compared to mucosal brushes (ANCOM2: FDR < 0.05; Supplemental figure 1).

ITS2 sequencing was also performed on the same subset of samples to investigate the effect of sampling method on the fungal microbiome. We observed no differences in Shannon diversity or richness across mucosal brushes, mucosal aspirates, and lavage aliquots (LME: $p > 0.05$, Supplemental figure 2). Like 16S amplicon data, PERMANOVA analysis of Bray-Curtis dissimilarities showed that the individual explained the greatest amount of variation in fungal community composition ($R^2 = 0.28$, $p < 0.001$), with no significant associations with fungal community composition and our three sampling methods ($R^2 = 0.38$, $p = 0.122$; Supplemental table 3).

Following the collection of fecal samples, we performed shotgun sequencing on a second set of samples, representing 105 individuals. Mucosal brushes were excluded from the second sample set because a pilot shotgun sequencing run revealed these samples contained an large percentage of human-derived reads (Supplemental figure 3). Based on estimates of Shannon diversity and species richness, the microbiomes in fecal samples were significantly more diverse than those in the mucosal aspirates (LME: $p = 0.007$ and $p = 0.002$, respectively) and marginally more diverse than those in lavage aliquots (LME: $p = 0.053$ and $p = 0.047$, respectively; Figure 2C). With respect to microbial beta diversity, the individual explained the greatest amount of variation in microbiome composition (PERMANOVA: $R^2 = 0.75$, $p < 0.001$; Supplemental table 4). In addition, fecal samples had microbial communities that were distinct when compared to mucosal aspirates and lavage aliquots from the same individual, with sampling method explaining 14% of variation in the microbiome (PERMANOVA: $p = 0.001$; Figure 2D). Fecal samples had a mean relative abundance of 63% for Firmicutes, 27% for Bacteroides, 3.5% for Actinobacteria, and 4.5% for Proteobacteria. Mucosal aspirates and lavage aliquots were more similar and had a mean relative abundance of 73% and 75% for Firmicutes, 15% and 11% for Bacteroides, 4.7%

and 5.2% for Actinobacteria, and 4.0% and 6.6% for Proteobacteria, respectively (Supplemental figure 4). Differential abundance analysis of microbes revealed 44 OTUs that were significantly different between fecal samples and mucosal aspirates (ANCOM2: FDR < 0.05; Supplemental table 5). Six OTUs were differentially abundant between fecal samples and lavage aliquots (Supplemental table 6), and no OTUs were significantly different between mucosal aspirates and lavage aliquots (ANCOM2; FDR > 0.05).

The Microbiomes of Polyps and Opposite Wall Healthy Tissue are similar within Individuals:

To identify any polyp-specific microbial biomarkers, 14 mucosal brush samples were collected from polyps and opposite wall healthy tissue and sequenced as part of the first sample set (Figure 3A). Based on 16S sequencing, we observed no significant differences in Shannon diversity or richness between polyp and opposite wall healthy tissue from within the same individual (Figure 3B). Similarly, there were no significant differences in beta-diversity across polyp and opposite wall healthy tissue pairs (PERMANOVA: $R^2 = 0.18$, $p = 0.62$; Figure 3C; Supplemental table 7). We were unable to identify any differentially abundant microbes between polyp and opposite wall tissue brushes. Microbiomes were mostly individualistic, with subject origin explaining 53% of the variance in microbiome composition (PERMANOVA: $p = 0.005$; Figure 3D; Supplemental table 7). We detected significant associations between microbiome composition and colon side (right/proximal versus left/distal), with 16% of the variance in microbiome composition explained by the colon side (PERMANOVA: $p = 0.003$; Supplemental table 7). Significant associations within the microbiome were observed when both polyp and opposite wall tissue pairs were categorized by subject type, explaining approximately 10% of variance (PERMANOVA: $p = 0.03$; Supplemental table 7).

Tubular Adenoma-bearing, Serrated Polyp-bearing, and Healthy Individuals have distinct Microbiomes:

We next analyzed all samples from the first and second sample sets to ask whether the type of polyp an individual had (subject type) was significantly associated with microbial diversity and composition. In both 16S and shotgun data, we observed no significant differences between polyp-free, TA-bearing, and serrated polyp-bearing samples based on Shannon diversity or richness estimates (LME: $p > 0.05$; Supplemental figure 5). In ITS data, we observed significantly increased Shannon diversity, but not richness, in samples from polyp-free individuals when compared to those from TA-bearing individuals (LME: $p = 0.03$; Supplemental figure 5). Beta diversity analysis of 16S data from the first sample set demonstrated that subject type explained 5% and 2% of the variance associated with the microbiome, respectively (16S PERMANOVA: $p = 0.001$; Supplemental table 1 and ITS PERMANOVA: $p = 0.09$; Supplemental table 3). A similar result was observed in second sample set mucosal aspirates. We found significant associations between the microbiome and an individual's polyp type, explaining 2% of the variance observed (PERMANOVA: $p = 0.001$; Supplemental table 4). This association between microbiome composition and polyp type was only observed in samples obtained directly from the mucosa and not in lavage aliquots (PERMANOVA: $p > 0.05$; Supplemental table 8) or fecal samples (PERMANOVA: $p > 0.05$; Supplement table 9).

Plotting the microbial relative abundances from shotgun mucosal aspirates revealed elevations of *Lachnospiraceae* and depletions of *Bacteroidaceae* in TA-bearing individuals (Figure 4A). Random Forest classification of shotgun sequences from healthy and TA-associated mucosal aspirates resulted in an out-of-box accuracy of approximately 86%, suggesting distinct microbial compositions between the two groups (Figure 4B). The top OTUs of importance for this

Random Forest classification included *Ruthenibacterium sp.*, *Tyzzerella* HGM12567, *Ruminococcus gnavus*, *Coprococcus comes*, and *Clostridium scindens* (Figure 4E). Differential abundance analysis revealed an increased abundance of *C. scindens* in TA mucosal aspirates when compared to serrated polyps (Dunn's Kruskal-Wallis: p-adjusted = 0.005) and healthy controls (Dunn's Kruskal-Wallis: p-adjusted = 0.016). *Tyzzerella* HGM12567 was also significantly enriched in TA mucosal aspirates when compared to healthy aspirates (Dunn's Kruskal-Wallis: p-adjusted = 0.01), but not serrated polyp aspirates (Dunn's Kruskal-Wallis: p-adjusted = 0.07).

Despite more similar microbial compositions, Random Forest classification of healthy and serrated polyp mucosal aspirates resulted in an out-of-box accuracy of approximately 87% (Figure 4C). OTUs of importance for this Random Forest classification included *Eggerthella lenta*, *Dorea longicatena*, *Anaerostipes hadrus*, *Ruminococcus obeum*, and *DTU089 HGM12731* (Figure 4F). Differential abundance analysis revealed a significant depletion of *E. lenta* in shotgun mucosal aspirates from serrated polyp-bearing individuals compared to healthy (Dunn's Kruskal-Wallis: q = 0.004) and TA-bearing individuals (Dunn's Kruskal-Wallis: q = 0.004). Lower abundances of *E. lenta* were also observed in serrated polyp mucosal aspirates from the first sample set, although the differences were not statistically significant (Supplemental figure 6).

To determine if the microbiome could be used to distinguish mucosal aspirates from TAs and serrated polyps, another Random Forest classification was performed, which produced an out-of-box accuracy of approximately 85% (Figure 4D). The top five variables of importance were *E. lenta*, *C. scindens*, *Ruthenibacterium sp.*, *Sellimonas sp.*, and *UBA7182 HGM12585*. SparCC correlation of *E. lenta* and *C. scindens* demonstrated that they significantly most positively correlated with one another (SparCC: pseudo-$p_1 < 0.05$; Figure G and H).

Microbiome Functional Potential is distinct across Sampling Methods and Subject Types:

The functional characteristics of the shotgun sequence data were next explored using KEGG and eggNOG annotations. Based on KEGG BRITE hierarchies, we observed a high degree of functional conservation across metagenomes, spanning both sample and subject types (Supplemental figure 7). A principal coordinate analysis of individual gene abundances revealed a distinct cluster of fecal samples when compared to mucosal aspirates and lavage aliquots. PERMANOVA analysis of gene abundances confirmed an association between functional microbiome capacity and sampling methods, significantly explaining 10.9% of the observed variance (PERMANOVA: $p = 0.001$; Supplemental table 10). In contrast, the individual of origin explained approximately 75% of the observed variance in the functional microbiome (PERMANOVA: $p = 0.001$; Supplemental table 10).

Subject type was also significantly associated with functional gene composition, explaining 1.2% of the observed variance (PERMANOVA: $p = 0.001$; Supplemental table 10). Differential abundance analysis revealed three microbial genes that varied across polyp-bearing and polyp-free controls. *SdaA*, a gene which encodes an L-serine dehydratase, was depleted in serrated mucosal aspirates when compared to polyp-free controls (Dunn's Kruskal-Wallis: p-adjusted < 0.001) and TA mucosal aspirates (Dunn's Kruskal-Wallis: p-adjusted < 0.001, Figure 5B). Taxonomic characterization of *sdaA* showed that it belonged to *E. lenta* (Blastp: E-value = 0.0, per. identity = 99.81%). Next, an alpha-galactosidase in the glycoside hydrolase (GH) family 31 was found to be depleted in serrated mucosal aspirates when compared to polyp-free mucosal aspirates (Dunn's Kruskal-Wallis: p-adjusted = 0.035), but not TA mucosal aspirates (Dunn's Kruskal-Wallis: p-adjusted = 0.22, Figure 5C). Lastly, the transporter-encoding *tonB* was enriched in polyp-free mucosal aspirates when compared to TA mucosal aspirates (Dunn's Kruskal-Wallis: p-adjusted = 0.038), but not serrated mucosal aspirates (Dunn's Kruskal-Wallis: p-adjusted =

0.065; Figure 5D). Both GH31 (Blastp: E-value = 0.0, per. identity = 99.92%) and *tonB* (Blastp: E-value = 0.0, per. identity = 100%) belonged to *Phocaeicola massiliensis.*

Discussion:

Sampling Method and Microbiome Characterization:

In this study we used direct and indirect methods to sample the colon to characterize the microbiomes of healthy and colorectal polyp-bearing individuals. Using amplicon sequencing, we did not observe significant differences in the diversity or composition of microbiomes in samples obtained directly from mucosal brushes and mucosal aspirates. In contrast, fecal samples were significantly more diverse and compositionally distinct when compared to mucosal aspirates.

Due to their ease of collection, fecal samples are commonly used to study the human microbiome in the context of CRC. Fecal samples poorly represent the microbiota adherent to the colon mucosa, and instead capture those found in the intestinal lumen.[19,20] Additionally, they do not provide information on the biogeography of the microbiome within the gastrointestinal tract.[21,22] These limitations suggest stool samples are not ideal for studying the microbiota adherent to colorectal polyps, which have less robust microbial signatures of dysbiosis than adenocarcinomas and carcinomas, and therefore would benefit from more direct sampling methods.

Using stool samples, Peters *et al.* found significant associations between the microbiota and distal conventional adenoma cases, but not proximal, hypothesizing that stool samples were a poor proxy for measuring the proximal microbiota.[16] Through direct sampling methods, we observed significant associations between the microbial composition and both proximal and distal TA cases using mucosal aspirates, supporting this hypothesis. Peters *et al.* also did not find significant differences in microbial diversity or composition between healthy controls and serrated

polyps, which predominantly develop in the proximal colon.[16] Using mucosal aspirates, but not fecal samples, we found compositional differences between serrated polyp cases and healthy controls. Together, these data suggest that mucosal brushes or aspirates, but not fecal samples, are sensitive enough to study the microbiome of colorectal polyps found within the proximal colon. These results contradict a study published by Yoon *et al.*, who did not find significant compositional differences the in mucosa-associated gut microbiota among polyp-free, conventional adenoma, SSP, and CRC bearing individuals.[17] The authors did note, however, that this result was likely driven by the small sample size of the study, with only 6 samples per group, and 24 samples total. In contrast, we managed to find significant compositional differences in the mucosa-associated microbiome which could be used to distinguish individuals with and without polyps using more samples.

Compared to mucosal brushes, mucosal aspirates had a lower risk of damaging the host epithelium, provided larger collection volumes for downstream sample processing, and resulted in a lower proportion of human derived reads with shotgun sequencing. Because of these advantages, we recommend using mucosal aspirates rather than mucosal brushing for characterizing the microbiomes of colorectal polyps.

Hyperlocal Microbiome Comparisons:

Although the sample size was limited, direct sampling of polyp mucosa with brushes revealed no differences in the hyperlocal microbiome of polyp tissue versus opposite wall healthy tissue. As a result, we were unable to identify microbial biomarkers specific to SSPs or other polyp types. One factor which could have disrupted the potential hyperlocal differences in the gut microbiota is the colonoscopy preparation and lavage. As part of the preparation, individuals were advised to adhere to a low fiber, clear liquid diet 24 hours prior to colonoscopy. Dietary fiber is

403 important in maintaining the longitudinal and lateral organization of the gut microbiota within the
404 colon, as mice on a low fiber diet show disrupted microbial organization.[19] Changes in diet can
405 rapidly shift the composition of the gut microbiome, often within 24 hours, in both humans and
406 mice.[7,23,24] A laxative-based cleansing and colonoscopy rinse was also performed, potentially
407 obscuring the hyperlocal organization further. Nevertheless, significant compositional differences
408 between the microbiomes of samples taken from the proximal and distal colon were observed,
409 suggesting that broad microbial organization remained present in the gut after colonoscopy
410 preparation and lavage.

411 ***Microbiome Signatures of the CRC Carcinogenesis Pathways:***

412 Compositional differences were observed in the microbiome across TA-bearing, serrated
413 polyp-bearing, and polyp-free individuals using mucosal aspirate sampling, but not fecal sampling.
414 Based on mucosal aspirates, the microbiomes of serrated polyp-bearing individuals more closely
415 resembled those of polyp-free controls rather than those from TA cases. Despite this, we still found
416 significant differences in the microbial composition of serrated and polyp-free mucosal aspirates.
417 These findings may suggest that the gut microbiome functions differently across the adenoma-
418 carcinoma sequence and the serrated pathway. In the adenoma-carcinoma sequence, the gut
419 microbiome exists in, and potentially contributes to, an inflammatory environment known to
420 promote CRC development. For example, enterotoxigenic *Bacteroides fragilis* produces a
421 metalloprotease that causes oxidative DNA damage and cleaves the tumor suppressor protein, E-
422 cadherin.[25-27] Another CRC-associated microbe, pks+ *E. coli*, synthesizes colibactin that induces
423 double stranded DNA breaks.[28.29]

424 Although we did not find any significant increases in the relative abundances of these CRC-
425 associated microbes in our data, we did see an increased abundance of *C. scindens and Tyzzerella*

426 using mucosal aspirates in TA-bearing individuals. No dietary information was collected in this
427 study, but it is possible that the increased abundance of *C. scindens* was influenced by a high fat
428 diet. High fat diets stimulate increased primary bile acid concentrations above normal
429 physiological concentrations.[30] Gut microbes, like *C. scindens*, can 7α-dehydroxylate excess
430 primary bile acids not absorbed by the small intestine into carcinogenic secondary bile acids.[31,32]
431 High concentrations of bile acids can cause oxidative stress, nitrosative stress, DNA damage,
432 apoptosis, and mutations.[30] Secondary bile acids also act as farnesoid X receptor antagonists,
433 resulting in enhanced *wnt* signaling during the adenoma-carcinoma sequence.[33] With respect to
434 *Tyzzerella*, this microbe has been found to be elevated in APC $^{min+/-}$ mice, which develop colitis-
435 associated CRC, when compared to wild-type mice.[34] We also observed an increasing trend of *R.*
436 *gnavus* in TA-associated mucosal aspirates, which has been previously associated with CRC and
437 inflammatory bowel disease.[35,36]

438 Another CRC-associated microbe is *Fusobacterium nucleatum*, which can activate Wnt
439 signaling by binding to host E-cadherin using its FadA adhesin to facilitate CRC development in
440 the adenoma-carcinoma sequence.[37] As reviewed in DeDecker et al, *F. nucleatum* has also been
441 implicated in the serrated pathway through its association with serrated pathway lesions and
442 features, such as mismatch repair deficiency, MLH1 methylation, CpG island methylator
443 phenotype, and high microsatellite instability.[38] A study by Rezasoltani et al. quantified the fecal
444 abundance of *F. nucleatum* and seven other CRC-associated microbes across TA,
445 villous/tubulovillous, HPP, SSP, and polyp-free individuals using qPCR.[18] Elevated levels of *F.*
446 *nucleatum, E. faecalis, S. bovis,* enterotoxigenic *B. fragilis*, and *Porphyromonas* spp. were
447 observed in TA and villous/tubulovillous groups, but not polyp-free, HPP, and SSP cases.[18] Here,
448 we did not find differences in *F. nucleatum* abundances across HPPs, SSPs, TAs, and polyp-free

449 controls. Instead, we found that *E. lenta* was depleted in mucosal aspirates from serrated polyp
450 bearing individuals.

451 Within the colon, *E. lenta* metabolizes inert plant lignans into bioactive enterolignans, such
452 as enterolactone and enterodiol.[39] These enterolignans have anti-proliferative and anti-
453 inflammatory effects, and help modulate estrogen signaling, lipid metabolism, and bile acid
454 regulation.[40] They have also been associated with reduced cancer risk.[40,41] Interestingly, the *sdaA*
455 gene, which was found to be depleted in serrated mucosal aspirates, was identified as belonging to
456 *E. lenta*. This gene catalyzes the conversion of L-serine into pyruvate and ammonia during
457 gluconeogenesis. It is not clear whether this enzyme is involved in enterolignan production in *E.*
458 *lenta*, but *sdaA* in *Campylobacter jejuni* is required for avian gut colonization.[42] Further
459 investigation is needed to determine *sdaA* is necessary for *E. lenta* colonization in the human gut.

460 Another microbial gene that was depleted in serrated mucosal aspirates was an alpha-
461 galactosidase in the glycoside hydrolase family 31 from *Phocaeicola massiliensis,* formerly named
462 *Bacteroides massilliensis*. These enzymes are carried by microbes to digest the glycosidic linkages
463 that join plant fibers. *P. massiliensis* has been shown to utilize starch and porcine mucin O-linked
464 glycans as sole carbon sources.[43] Diets rich in plant fiber, like starch, have been associated with
465 decreased CRC risk.[6,44] Fiber is fermented by the intestinal microbiota to produce short chain fatty
466 acids, including acetate, butyrate, and propionate. Butyrate is the primary energy source for
467 colonocytes and has anti-inflammatory and anti-tumor properties.[45-47] Butyrate also is involved in
468 the epigenetic expression of genes as a histone deacetylase inhibitor.[48] In the serrated pathway, the
469 gene SLC5A8, which mediates short chain fatty acid uptake into colonic epithelial cells, is
470 frequently inhibited via promoter methylation, suggesting that dietary fiber may be required for
471 proper cellular epigenetic regulation.[49] Taken together, we hypothesize that decreased abundances

of *E. lenta*, *sdaA*, and GH31 in serrated polyp samples result from low-dietary fiber consumption, which leads to subsequent epigenetic modifications within colonocytes to promote serrated carcinogenesis. Nevertheless, more research is needed to elucidate the role of the gut microbiome during serrated carcinogenesis.

TonB from *P. massiliensis* is another gene which we observed to be increased in polyp-free controls. *TonB* transporters allow microbes to actively transport essential micronutrients, such as iron. Diets high in red meat consumption are rich in iron and have been associated with increased CRC risk.[6] It is possible that our polyp-free controls consumed less red meat, thus necessitating more *tonB* transporters. This contradicts a recent metaproteomic analysis of stool samples from healthy individuals and those with CRC, which found an increase of *tonB* in the CRC group.[50]

***Conclusion*:**

The complex and individualistic nature of the human gut microbiome has made it difficult to mechanistically link the microbiome with CRC carcinogenesis. By describing the association between the gut microbiota and serrated polyp development, our study aims to elucidate mechanistic targets for the epigenetic-based serrated pathway to CRC. In addition, our data underscores the importance of distinguishing between different pathways of colorectal carcinogenesis when investigating the gut microbiome. Finally, transitioning future microbiome studies to direct sampling methods will enable the discovery of previously unassociated microbes and novel mechanistic targets as demonstrated here.

Declarations:

Author contributions:

Katrine L. Whiteson and William E. Karnes devised the study design with support from Lauren DeDecker. Subject recruitment was performed by William E. Karnes and Zachary Lu. Sample collection was performed by William E. Karnes, Lauren DeDecker, and Zachary Lu with guidance from Katrine L. Whiteson. Julio Avelar-Barragan, Bretton Coppedge, and Zachary Lu processed samples for data acquisition. Julio Avelar-Barragan performed the data analysis and wrote the manuscript with guidance from Katrine L. Whiteson.

Ethics approval:

This study was approved by the Institutional Review Board (IRB) of the University of California, Irvine (HS# 2017-3869).

Funding details:

This study was funded by institutional research grant #IRG-16-187-13 from the American Cancer Society.

Disclosure of interest:

The authors declare no competing or conflicts of interest.

Data availability statement:

Bash and R code for data processing and analysis is available on GitHub at: https://github.com/Javelarb/ACS_polyp_study. Additional data and materials are available upon reasonable request.

Data deposition:

Sequencing data is available on the Sequence Read Archive under the BioProject ID, PRJNA745329.

Acknowledgements:

We would like to thank Claudia Weihe and Jennifer B.H Martiny for allowing us to borrow laboratory equipment and giving insightful feedback, Andrew Oliver and Jason A. Rothman for their bioinformatic expertise, Clark Hendrickson for his assistance with sample preparation, and Heather Maughan for her helpful edits and suggestions.

519 **REFERENCES:**

520 1. Sung, Hyuna, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle
521 Soerjomataram, Ahmedin Jemal, and Freddie Bray. "Global Cancer Statistics 2020:
522 GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185
523 Countries." *CA: A Cancer Journal for Clinicians* 71, no. 3 (May 2021): 209–49.
524 https://doi.org/10.3322/caac.21660.

525 2. Stoffel, Elena M., Pamela B. Mangu, Stephen B. Gruber, Stanley R. Hamilton, Matthew F.
526 Kalady, Michelle Wan Yee Lau, Karen H. Lu, Nancy Roach, and Paul J. Limburg. "Hereditary
527 Colorectal Cancer Syndromes: American Society of Clinical Oncology Clinical Practice
528 Guideline Endorsement of the Familial Risk–Colorectal Cancer: European Society for Medical
529 Oncology Clinical Practice Guidelines." *Journal of Clinical Oncology* 33, no. 2 (January 10,
530 2015): 209–17. https://doi.org/10.1200/JCO.2014.58.1322.

531 3. Collins, S.M., E. Denou, E.F. Verdu, and P. Bercik. "The Putative Role of the Intestinal
532 Microbiota in the Irritable Bowel Syndrome." *Digestive and Liver Disease* 41, no. 12
533 (December 2009): 850–53. https://doi.org/10.1016/j.dld.2009.07.023.

534 4. Verdam, Froukje J., Susana Fuentes, Charlotte de Jonge, Erwin G. Zoetendal, Runi Erbil, Jan
535 Willem Greve, Wim A. Buurman, Willem M. de Vos, and Sander S. Rensen. "Human
536 Intestinal Microbiota Composition Is Associated with Local and Systemic Inflammation in
537 Obesity: Obese Gut Microbiota and Inflammation." *Obesity* 21, no. 12 (December 2013):
538 E607–15. https://doi.org/10.1002/oby.20466.

539 5. Song, Mingyang, Wendy S. Garrett, and Andrew T. Chan. "Nutrients, Foods, and Colorectal
540 Cancer Prevention." *Gastroenterology* 148, no. 6 (May 2015): 1244-1260.e16.
541 https://doi.org/10.1053/j.gastro.2014.12.035.

6. "Diet, Nutrition, Physical Activity, and Colorectal Cancer." World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report, 2018. dietandcancerreport.org.

7. David, Lawrence A., Corinne F. Maurice, Rachel N. Carmody, David B. Gootenberg, Julie E. Button, Benjamin E. Wolfe, Alisha V. Ling, et al. "Diet Rapidly and Reproducibly Alters the Human Gut Microbiome." *Nature* 505, no. 7484 (January 2014): 559–63. https://doi.org/10.1038/nature12820.

8. Engen, Phillip A., Stefan J. Green, Robin M. Voigt, Christopher B. Forsyth, and Ali Keshavarzian. "The Gastrointestinal Microbiome: Alcohol Effects on the Composition of Intestinal Microbiota." *Alcohol Research: Current Reviews* 37, no. 2 (2015): 223–36.

9. Ley, Ruth E. "Obesity and the Human Microbiome:" *Current Opinion in Gastroenterology* 26, no. 1 (January 2010): 5–11. https://doi.org/10.1097/MOG.0b013e328333d751.

10. Mailing, Lucy J., Jacob M. Allen, Thomas W. Buford, Christopher J. Fields, and Jeffrey A. Woods. "Exercise and the Gut Microbiome: A Review of the Evidence, Potential Mechanisms, and Implications for Human Health." *Exercise and Sport Sciences Reviews* 47, no. 2 (April 2019): 75–85. https://doi.org/10.1249/JES.0000000000000183.

11. Nakanishi, Yuki, Maria T. Diaz-Meco, and Jorge Moscat. "Serrated Colorectal Cancer: The Road Less Travelled?" *Trends in Cancer* 5, no. 11 (November 2019): 742–54. https://doi.org/10.1016/j.trecan.2019.09.004.

12. Pino, Maria S., and Daniel C. Chung. "The Chromosomal Instability Pathway in Colon Cancer." *Gastroenterology* 138, no. 6 (May 2010): 2059–72. https://doi.org/10.1053/j.gastro.2009.12.065.

13. De Palma, Fatima, Valeria D'Argenio, Jonathan Pol, Guido Kroemer, Maria Maiuri, and Francesco Salvatore. "The Molecular Hallmarks of the Serrated Pathway in Colorectal Cancer." *Cancers* 11, no. 7 (July 20, 2019): 1017. https://doi.org/10.3390/cancers11071017.

14. Kahi, Charles J. "Screening Relevance of Sessile Serrated Polyps." *Clinical Endoscopy* 52, no. 3 (May 31, 2019): 235–38. https://doi.org/10.5946/ce.2018.112.

15. Delker, Don A., Brett M. McGettigan, Priyanka Kanth, Stelian Pop, Deborah W. Neklason, Mary P. Bronner, Randall W. Burt, and Curt H. Hagedorn. "RNA Sequencing of Sessile Serrated Colon Polyps Identifies Differentially Expressed Genes and Immunohistochemical Markers." Edited by Frank T. Kolligs. *PLoS ONE* 9, no. 2 (February 12, 2014): e88367. https://doi.org/10.1371/journal.pone.0088367.

16. Peters, Brandilyn A., Christine Dominianni, Jean A. Shapiro, Timothy R. Church, Jing Wu, George Miller, Elizabeth Yuen, et al. "The Gut Microbiota in Conventional and Serrated Precursors of Colorectal Cancer." *Microbiome* 4, no. 1 (December 2016): 69. https://doi.org/10.1186/s40168-016-0218-6.

17. Yoon, Hyuk, Nayoung Kim, Ji Hyun Park, Yong Sung Kim, Jongchan Lee, Hyoung Woo Kim, Yoon Jin Choi, et al. "Comparisons of Gut Microbiota Among Healthy Control, Patients With Conventional Adenoma, Sessile Serrated Adenoma, and Colorectal Cancer." *Journal of Cancer Prevention* 22, no. 2 (June 30, 2017): 108–14. https://doi.org/10.15430/JCP.2017.22.2.108.

18. Rezasoltani, Sama, Hamid Asadzadeh Aghdaei, Hossein Dabiri, Abbas Akhavan Sepahi, Mohammad Hossein Modarressi, and Ehsan Nazemalhosseini Mojarad. "The Association between Fecal Microbiota and Different Types of Colorectal Polyp as Precursors of Colorectal

586 Cancer." *Microbial Pathogenesis* 124 (November 2018): 244–49.
587 https://doi.org/10.1016/j.micpath.2018.08.035.

588 19. Riva, Alessandra, Orest Kuzyk, Erica Forsberg, Gary Siuzdak, Carina Pfann, Craig Herbold,
589 Holger Daims, Alexander Loy, Benedikt Warth, and David Berry. "A Fiber-Deprived Diet
590 Disturbs the Fine-Scale Spatial Architecture of the Murine Colon Microbiome." *Nature*
591 *Communications* 10, no. 1 (December 2019): 4366. https://doi.org/10.1038/s41467-019-
592 12413-0.

593 20. Chen, Weiguang, Fanlong Liu, Zongxin Ling, Xiaojuan Tong, and Charlie Xiang. "Human
594 Intestinal Lumen and Mucosa-Associated Microbiota in Patients with Colorectal Cancer."
595 Edited by Antonio Moschetta. *PLoS ONE* 7, no. 6 (June 28, 2012): e39743.
596 https://doi.org/10.1371/journal.pone.0039743.

597 21. Eckburg, P. B. "Diversity of the Human Intestinal Microbial Flora." *Science* 308, no. 5728
598 (June 10, 2005): 1635–38. https://doi.org/10.1126/science.1110591.

599 22. Tropini, Carolina, Kristen A. Earle, Kerwyn Casey Huang, and Justin L. Sonnenburg. "The
600 Gut Microbiome: Connecting Spatial Organization to Function." *Cell Host & Microbe* 21, no.
601 4 (April 2017): 433–42. https://doi.org/10.1016/j.chom.2017.03.010.

602 23. Wu, G. D., J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S. A. Keilbaugh, M. Bewtra, et
603 al. "Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes." *Science* 334, no.
604 6052 (October 7, 2011): 105–8. https://doi.org/10.1126/science.1208344.

605 24. Turnbaugh, P. J., V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight, and J. I. Gordon. "The Effect
606 of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic
607 Mice." *Science Translational Medicine* 1, no. 6 (November 11, 2009): 6ra14-6ra14.
608 https://doi.org/10.1126/scitranslmed.3000322.

609 25. Haghi, Fakhri, Elshan Goli, Bahman Mirzaei, and Habib Zeighami. "The Association between
610 Fecal Enterotoxigenic B. Fragilis with Colorectal Cancer." *BMC Cancer* 19, no. 1 (December
611 2019): 879. https://doi.org/10.1186/s12885-019-6115-1.
612 26. Ulger Toprak, N., A. Yagci, B.M. Gulluoglu, M.L. Akin, P. Demirkalem, T. Celenk, and G.
613 Soyletir. "A Possible Role of Bacteroides Fragilis Enterotoxin in the Aetiology of Colorectal
614 Cancer." *Clinical Microbiology and Infection* 12, no. 8 (August 2006): 782–86.
615 https://doi.org/10.1111/j.1469-0691.2006.01494.x.
616 27. Cheng, Wai Teng, Haresh Kumar Kantilal, and Fabian Davamani. "The Mechanism of
617 Bacteroides Fragilis Toxin Contributes to Colon Cancer Formation." *The Malaysian Journal*
618 *of Medical Sciences: MJMS* 27, no. 4 (July 2020): 9–21.
619 https://doi.org/10.21315/mjms2020.27.4.2.
620 28. Nougayrède, Jean-Philippe, Stefan Homburg, Frédéric Taieb, Michèle Boury, Elzbieta
621 Brzuszkiewicz, Gerhard Gottschalk, Carmen Buchrieser, Jörg Hacker, Ulrich Dobrindt, and
622 Eric Oswald. "Escherichia Coli Induces DNA Double-Strand Breaks in Eukaryotic Cells."
623 *Science (New York, N.Y.)* 313, no. 5788 (August 11, 2006): 848–51.
624 https://doi.org/10.1126/science.1127059.
625 29. Pleguezuelos-Manzano, Cayetano, Jens Puschhof, Axel Rosendahl Huber, Arne van Hoeck,
626 Henry M. Wood, Jason Nomburg, Carino Gurjao, et al. "Mutational Signature in Colorectal
627 Cancer Caused by Genotoxic Pks+ E. Coli." *Nature* 580, no. 7802 (April 2020): 269–73.
628 https://doi.org/10.1038/s41586-020-2080-8.
629 30. Ajouz, Hana, Deborah Mukherji, and Ali Shamseddine. "Secondary Bile Acids: An
630 Underrecognized Cause of Colon Cancer." *World Journal of Surgical Oncology* 12, no. 1
631 (2014): 164. https://doi.org/10.1186/1477-7819-12-164.

632 31. Ridlon, Jason M., and Phillip B. Hylemon. “Identification and Characterization of Two Bile
633 Acid Coenzyme A Transferases from Clostridium Scindens, a Bile Acid 7α-Dehydroxylating
634 Intestinal Bacterium.” *Journal of Lipid Research* 53, no. 1 (January 2012): 66–76.
635 https://doi.org/10.1194/jlr.M020313.
636 32. Marion, Solenne, Nicolas Studer, Lyne Desharnais, Laure Menin, Stéphane Escrig, Anders
637 Meibom, Siegfried Hapfelmeier, and Rizlan Bernier-Latmani. “In Vitro and in Vivo
638 Characterization of Clostridium Scindens Bile Acid Transformations.” *Gut Microbes* 10, no.
639 4 (July 4, 2019): 481–503. https://doi.org/10.1080/19490976.2018.1549420.
640 33. Ocvirk, Soeren, and Stephen J.D. O’Keefe. “Dietary Fat, Bile Acid Metabolism and Colorectal
641 Cancer.” *Seminars in Cancer Biology*, October 2020, S1044579X2030208X.
642 https://doi.org/10.1016/j.semcancer.2020.10.003.
643 34. Zhang, Yong, Xiao-lan Wang, Min Zhou, Chao Kang, He-dong Lang, Meng-ting Chen, Suo-
644 cheng Hui, Bin Wang, and Man-tian Mi. “Crosstalk between Gut Microbiota and Sirtuin-3 in
645 Colonic Inflammation and Tumorigenesis.” *Experimental & Molecular Medicine* 50, no. 4
646 (April 2018): 1–11. https://doi.org/10.1038/s12276-017-0002-0.
647 35. Hall, Andrew Brantley, Moran Yassour, Jenny Sauk, Ashley Garner, Xiaofang Jiang, Timothy
648 Arthur, Georgia K. Lagoudas, et al. “A Novel Ruminococcus Gnavus Clade Enriched in
649 Inflammatory Bowel Disease Patients.” *Genome Medicine* 9, no. 1 (November 28, 2017): 103.
650 https://doi.org/10.1186/s13073-017-0490-5.
651 36. Burns, Michael B., Emmanuel Montassier, Juan Abrahante, Sambhawa Priya, David E.
652 Niccum, Alexander Khoruts, Timothy K. Starr, Dan Knights, and Ran Blekhman. “Colorectal
653 Cancer Mutational Profiles Correlate with Defined Microbial Communities in the Tumor

654 Microenvironment." *PLoS Genetics* 14, no. 6 (June 2018): e1007376.
655 https://doi.org/10.1371/journal.pgen.1007376.
656 37. Gholizadeh, Pourya, Hosein Eslami, and Hossein Samadi Kafil. "Carcinogenesis Mechanisms
657 of Fusobacterium Nucleatum." *Biomedicine & Pharmacotherapy* 89 (May 2017): 918–25.
658 https://doi.org/10.1016/j.biopha.2017.02.102.
659 38. DeDecker, Lauren, Bretton Coppedge, Julio Avelar-Barragan, William Karnes, and Katrine
660 Whiteson. "Microbiome Distinctions between the CRC Carcinogenic Pathways." Gut
661 Microbes, January 15, 2021, 1–12. https://doi.org/10.1080/19490976.2020.1854641.
662 39. Bess, Elizabeth N., Jordan E. Bisanz, Fauna Yarza, Annamarie Bustion, Barry E. Rich,
663 Xingnan Li, Seiya Kitamura, et al. "Genetic Basis for the Cooperative Bioactivation of Plant
664 Lignans by Eggerthella Lenta and Other Human Gut Bacteria." *Nature Microbiology* 5, no. 1
665 (January 2020): 56–66. https://doi.org/10.1038/s41564-019-0596-1.
666 40. Webb, Amy L., and Marjorie L. McCullough. "Dietary Lignans: Potential Role in Cancer
667 Prevention." *Nutrition and Cancer* 51, no. 2 (March 2005): 117–31.
668 https://doi.org/10.1207/s15327914nc5102_1.
669 41. Adlercreutz, Herman. "Lignans and Human Health." *Critical Reviews in Clinical Laboratory*
670 *Sciences* 44, no. 5–6 (January 2007): 483–525. https://doi.org/10.1080/10408360701612942.
671 42. Velayudhan, Jyoti, Michael A. Jones, Paul A. Barrow, and David J. Kelly. "Serine Catabolism
672 via an Oxygen-Labile l-Serine Dehydratase Is Essential for Colonization of the Avian Gut by
673 Campylobacter Jejuni." *Infection and Immunity* 72, no. 1 (January 2004): 260–68.
674 https://doi.org/10.1128/IAI.72.1.260-268.2004.
675 43. Pudlo, Nicholas A., Karthik Urs, Supriya Suresh Kumar, J. Bruce German, David A. Mills,
676 and Eric C. Martens. "Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for

677 Host Mucosal Glycans." Edited by Vanessa Sperandio. *MBio* 6, no. 6 (December 31, 2015).
678 https://doi.org/10.1128/mBio.01282-15.
679 44. Aune, D., D. S. M. Chan, R. Lau, R. Vieira, D. C. Greenwood, E. Kampman, and T. Norat.
680 "Dietary Fibre, Whole Grains, and Risk of Colorectal Cancer: Systematic Review and Dose-
681 Response Meta-Analysis of Prospective Studies." *BMJ* 343, no. nov10 1 (November 11, 2011):
682 d6617–d6617. https://doi.org/10.1136/bmj.d6617.
683 45. Donohoe, Dallas R., Nikhil Garge, Xinxin Zhang, Wei Sun, Thomas M. O'Connell, Maureen
684 K. Bunger, and Scott J. Bultman. "The Microbiome and Butyrate Regulate Energy Metabolism
685 and Autophagy in the Mammalian Colon." *Cell Metabolism* 13, no. 5 (May 4, 2011): 517–26.
686 https://doi.org/10.1016/j.cmet.2011.02.018.
687 46. Hague, Angela, Douglas J. E. Elder, Diane J. Hicks, and Christos Paraskeva. "Apoptosis in
688 Colorectal Tumour Cells: Induction by the Short Chain Fatty Acids Butyrate, Propionate and
689 Acetate and by the Bile Salt Deoxycholate." *International Journal of Cancer* 60, no. 3 (January
690 27, 1995): 400–406. https://doi.org/10.1002/ijc.2910600322.
691 47. Hamer, H. M., D. Jonkers, K. Venema, S. Vanhoutvin, F. J. Troost, and R.-J. Brummer.
692 "Review Article: The Role of Butyrate on Colonic Function." *Alimentary Pharmacology &*
693 *Therapeutics* 27, no. 2 (October 26, 2007): 104–19. https://doi.org/10.1111/j.1365-
694 2036.2007.03562.x.
695 48. Davie, James R. "Inhibition of Histone Deacetylase Activity by Butyrate." *The Journal of*
696 *Nutrition* 133, no. 7 (July 1, 2003): 2485S-2493S. https://doi.org/10.1093/jn/133.7.2485S.
697 49. Goldstein, Neal S. "Serrated Pathway and APC (Conventional)-Type Colorectal Polyps:
698 Molecular-Morphologic Correlations, Genetic Pathways, and Implications for Classification."

699 *American Journal of Clinical Pathology* 125, no. 1 (January 2006): 146–53.
700 https://doi.org/10.1309/87BD0C6UCGUG236J.

701 50. Long, Shuping, Yi Yang, Chengpin Shen, Yiwen Wang, Anmei Deng, Qin Qin, and Liang
702 Qiao. "Metaproteomics Characterizes Human Gut Microbiome Function in Colorectal
703 Cancer." *Npj Biofilms and Microbiomes* 6, no. 1 (December 2020): 14.
704 https://doi.org/10.1038/s41522-020-0123-4.

705 51. Looby, Caitlin I., Mia R. Maltz, and Kathleen K. Treseder. "Belowground Responses to
706 Elevation in a Changing Cloud Forest." *Ecology and Evolution* 6, no. 7 (April 2016): 1996–
707 2009. https://doi.org/10.1002/ece3.2025.

708 52. Bruinsma, Stephen, Joshua Burgess, Daniel Schlingman, Agata Czyz, Natalie Morrell,
709 Catherine Ballenger, Heather Meinholz, et al. "Bead-Linked Transposomes Enable a
710 Normalization-Free Workflow for NGS Library Preparation." *BMC Genomics* 19, no. 1
711 (December 2018): 722. https://doi.org/10.1186/s12864-018-5096-9.

712 53. Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet,
713 Gabriel A. Al-Ghalith, Harriet Alexander, et al. "Reproducible, Interactive, Scalable and
714 Extensible Microbiome Data Science Using QIIME 2." *Nature Biotechnology* 37, no. 8
715 (August 2019): 852–57. https://doi.org/10.1038/s41587-019-0209-9.

716 54. Callahan, Benjamin J, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson,
717 and Susan P Holmes. "DADA2: High-Resolution Sample Inference from Illumina Amplicon
718 Data." *Nature Methods* 13, no. 7 (July 2016): 581–83. https://doi.org/10.1038/nmeth.3869.

719 55. McDonald, Daniel, Morgan N Price, Julia Goodrich, Eric P Nawrocki, Todd Z DeSantis,
720 Alexander Probst, Gary L Andersen, Rob Knight, and Philip Hugenholtz. "An Improved
721 Greengenes Taxonomy with Explicit Ranks for Ecological and Evolutionary Analyses of

722 Bacteria and Archaea." *The ISME Journal* 6, no. 3 (March 2012): 610–18.
723 https://doi.org/10.1038/ismej.2011.139.

724 56. Bushnell, Brian. "BBMap: A Fast, Accurate, Splice-Aware Aligner." Lawrence Berkeley
725 National Lab.(LBNL), 2014.

726 57. Cantu, Vito Adrian, Jeffrey Sadural, and Robert Edwards. "PRINSEQ++, a Multi-Threaded
727 Tool for Fast and Efficient Quality Control and Preprocessing of Sequencing Datasets."
728 Preprint. PeerJ Preprints, February 27, 2019. https://doi.org/10.7287/peerj.preprints.27553v1.

729 58. Langmead, Ben, and Steven L Salzberg. "Fast Gapped-Read Alignment with Bowtie 2."
730 *Nature Methods* 9, no. 4 (April 2012): 357–59. https://doi.org/10.1038/nmeth.1923.

731 59. Nayfach, Stephen, Zhou Jason Shi, Rekha Seshadri, Katherine S. Pollard, and Nikos C.
732 Kyrpides. "New Insights from Uncultivated Genomes of the Global Human Gut Microbiome."
733 *Nature* 568, no. 7753 (April 2019): 505–10. https://doi.org/10.1038/s41586-019-1058-x.

734 60. Salazar, Guillem. *EcolUtils: Utilities for Community Ecology Analysis.* (version 0.1).
735 https://github.com/GuillemSalazar/EcolUtils, 2019.

736 61. Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan
737 McGlinn, Peter R. Minchin, et al. *Vegan: Community Ecology Package.* (version R package
738 version 2.5-6), 2019. https://CRAN.R-project.org/package=vegan.

739 62. Pinheiro, J, D Bates, S DebRoy, D Sarkar, and R Core Team. *Nlme: Linear and Nonlinear*
740 *Mixed Effects Models.* (version R package version 3.1-148), 2020. https://CRAN.R-
741 project.org/package=nlme.

742 63. Friedman, Jonathan, and Eric J. Alm. "Inferring Correlation Networks from Genomic Survey
743 Data." Edited by Christian von Mering. *PLoS Computational Biology* 8, no. 9 (September 20,
744 2012): e1002687. https://doi.org/10.1371/journal.pcbi.1002687.

745 64. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis.* (version 3.3.0). Springer-Verlag
746 New York, 2016.
747 65. Mandal, Siddhartha, Will Van Treuren, Richard A. White, Merete Eggesbø, Rob Knight, and
748 Shyamal D. Peddada. "Analysis of Composition of Microbiomes: A Novel Method for
749 Studying Microbial Composition." *Microbial Ecology in Health & Disease* 26, no. 0 (May 29,
750 2015). https://doi.org/10.3402/mehd.v26.27663.
751 66. Li, Dinghua, Ruibang Luo, Chi-Man Liu, Chi-Ming Leung, Hing-Fung Ting, Kunihiko
752 Sadakane, Hiroshi Yamashita, and Tak-Wah Lam. "MEGAHIT v1.0: A Fast and Scalable
753 Metagenome Assembler Driven by Advanced Methodologies and Community Practices."
754 *Methods* 102 (June 2016): 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
755 67. Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and
756 Loren J Hauser. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site
757 Identification." *BMC Bioinformatics* 11, no. 1 (December 2010): 119.
758 https://doi.org/10.1186/1471-2105-11-119.
759 68. Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K
760 Forslund, Helen Cook, Daniel R Mende, et al. "EggNOG 5.0: A Hierarchical, Functionally and
761 Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502
762 Viruses." *Nucleic Acids Research* 47, no. D1 (January 8, 2019): D309–14.
763 https://doi.org/10.1093/nar/gky1085.
764 69. Nayfach, Stephen, and Katherine S Pollard. "Average Genome Size Estimation Improves
765 Comparative Metagenomics and Sheds Light on the Functional Ecology of the Human
766 Microbiome." *Genome Biology* 16, no. 1 (December 2015): 51.
767 https://doi.org/10.1186/s13059-015-0611-7.

768 70. Love, Michael I, Wolfgang Huber, and Simon Anders. “Moderated Estimation of Fold Change
769 and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15, no. 12 (December
770 2014): 550. https://doi.org/10.1186/s13059-014-0550-8.

Figure legends:

Figure 1: Study design.

A total of 140 different individuals were recruited for this study, including 50 healthy/polyp-free individuals, 45 with tubular adenomas, and 33 with serrated polyps (HPP, TSA, or SSP). For the remaining 12 individuals, 11 had unknown pathology and one had an adenocarcinoma. Multiple samples were taken from each subject during colonoscopy. This included mucosal brushes (orange), mucosal aspirates (yellow), and lavage aliquots (purple). Fecal samples (brown) were collected from participants four to six weeks post-colonoscopy. DNA extraction and sequencing produced two sample sets. The first sample set had a median age of 60, a median BMI of 25, and was 57% male and 43% female. The second sample set had a median age of 65, a median BMI of 25, and was 50% male, 40% female, and 10% unknown/other. Some individuals appeared in both sample sets.

Figure 2: Microbiomes of Mucosal and Lavage Samples are similar to each other but different from those in Feces.

A and C) Shannon diversity and richness estimates across mucosal aspirates (yellow), mucosal brushes (orange), lavage aliquots (purple), and fecal samples (brown). The first sample set was sequenced using 16S sequencing **(A)**, while the second sample set was sequenced using shotgun sequencing **(C)**. **B and D)** Non-metric multidimensional scaling of Bray-Curtis dissimilarities in the first **(B)** and second **(D)** sample sets. Each point corresponds to a single sample with multiple samples per individual. The individual of origin is denoted numerically within each point. The number of samples per sample type and subject category are annotated parenthetically. Significant comparisons ($p < 0.05$) are denoted by an asterisk (*).

Figure 3: The Microbiomes of Polyps and Opposite Wall Healthy Tissue are similar within Individuals.

A) An illustration of the sampling strategy used to characterize the hyperlocal differences in the microbial community of polyps (red) and opposite wall healthy tissue (green). Brush samples were sequenced as part of the first sample set using 16S sequencing. **B)** Shannon diversity and richness estimates of mucosal brushes from polyp and opposite wall healthy tissue. **C)** Non-metric multidimensional scaling of Bray-Curtis dissimilarities of polyp and opposite wall healthy tissue brushes. The individual of origin is denoted numerically within each point. The shape of each point denotes the right (proximal) and left (distal) side of the colon. **D)** The relative abundance of the top ten microbial genera across all samples. Samples are grouped by each individual and labeled by polyp type, where tubular adenoma = TA, hyperplastic polyp = HPP, and sessile serrated polyp = SSP.

Figure 4: Tubular Adenoma-bearing, Serrated Polyp-bearing, and Healthy Individuals have distinct Microbiomes.

A) The top seven most abundant microbial families across all samples from the second sample set. The number of samples per sampling method and subject type is annotated parenthetically. **B-D)** Non-metric multi-dimensional scaling of random forest proximity scores per sample using only mucosal aspirates from the second sample set. Central points are colored by the known subject type. The outer circles represent predicted subject type as determined by the Random Forest. The out-of-box accuracy percentage for each Random Forest classification is embedded within each graph. **E-F)** The relative abundance of the top five variables of importance in second sample set mucosal aspirates for healthy vs. TA **(E)** and healthy vs. serrated **(F)** Random Forest comparisons. Significant comparisons (p-adjusted < 0.05) are denoted by an asterisk (*). **G-H)** The top ten most significantly correlated microbes for *C. scindens* **(G)** and *E. lenta* **(H)** as determined by sparCC correlation across all second sample set samples (pseudo-$p_1 < 0.05$).

Figure 5: Microbiome Functional Potential is distinct across Sampling Methods and Subject Types.

A) Principal coordinate analysis of gene Bray-Curtis dissimilarities across second sample set mucosal aspirates, lavage aliquots, and fecal samples. **B-D)** The abundance of DESeq2 normalized reads across each subject type in second sample set mucosal aspirates for the genes *sdaA* **(B)**, GH31 **(C)**, and *tonB* **(D)**. The number of samples per sampling method and subject type are annotated parenthetically. Significant comparisons (p-adjusted < 0.05) are denoted by an asterisk (*).

827 **Main figures:**

828 ***Figure 1: Study design***

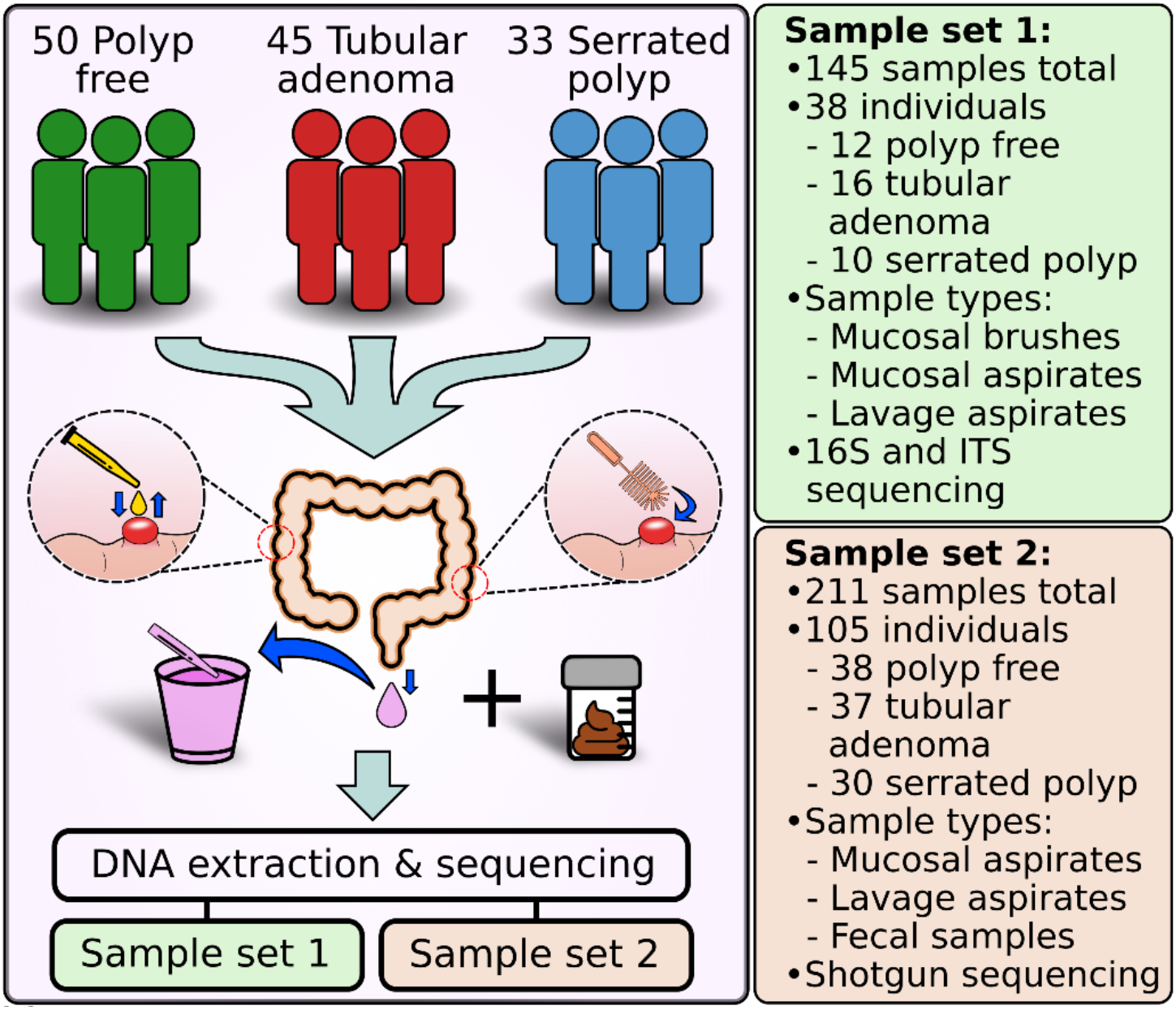
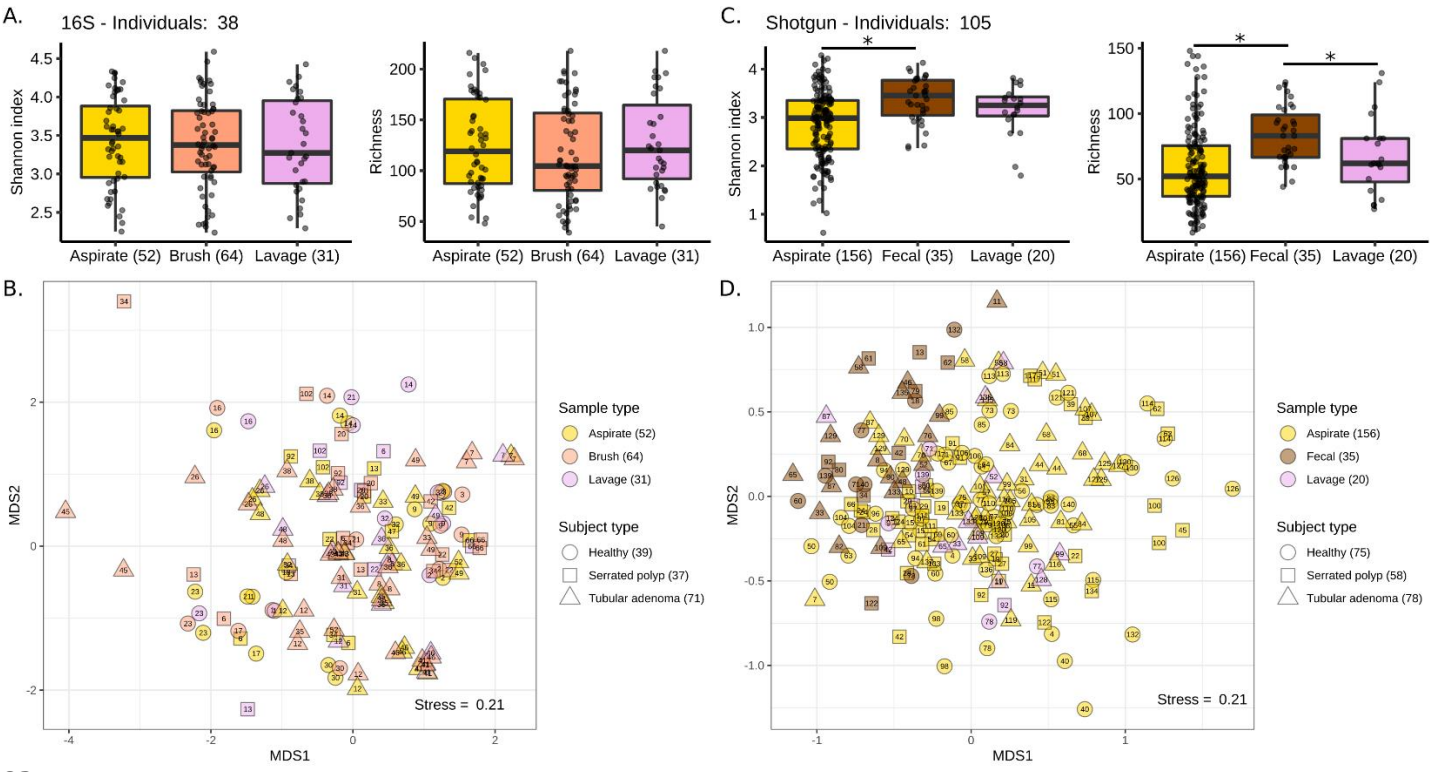
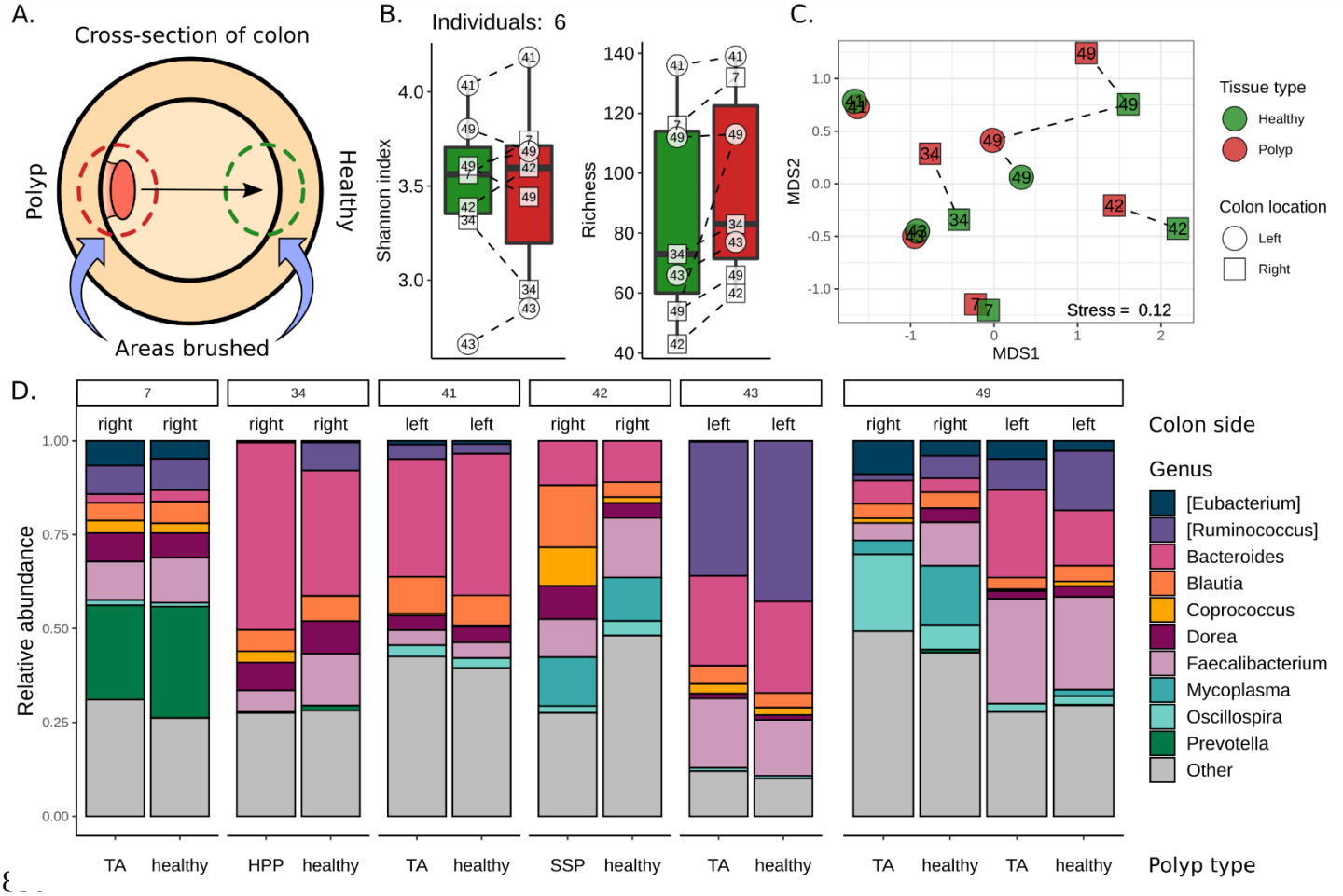


Figure 2: Microbiomes of Mucosal and Lavage Samples are similar to each other but different from those in Feces.



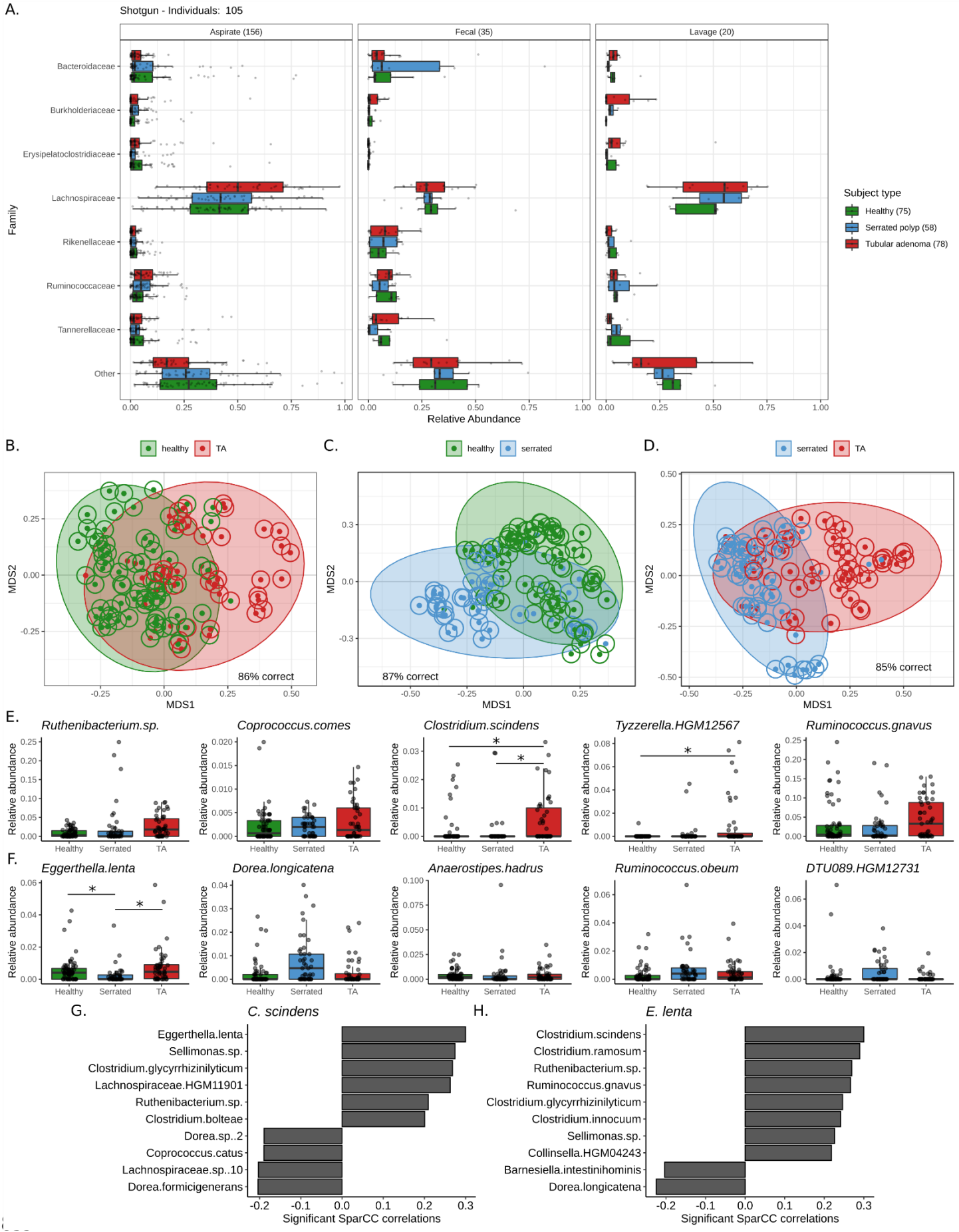
834 ***Figure 3: The Microbiomes of Polyps and Opposite Wall Healthy Tissue are similar within***

835 ***Individuals.***



837 ***Figure 4: Tubular Adenoma-bearing, Serrated Polyp-bearing, and Healthy Individuals have***

838 ***distinct Microbiomes.***



840 ***Figure 5: Microbiome Functional Potential is distinct across Sampling Methods and Subject***

841 ***Types.***