Revisiting Short-Chain Fatty Acids and the Microbiota in Colorectal Cancer

Running title: SCFAs and Colorectal Cancer

Marc A Sze[1], Nicholas A Lesniak[1], Mack T Ruffin IV[2], Patrick D. Schloss[1][†]

† To whom correspondence should be addressed: pschloss@umich.edu

1 Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109

2 Department of Family Medicine and Community Medicine, Penn State Hershey Medical Center, Hershey, PA

Abstract

Background. Certain bacterial community members have been shown to increase colon tumorigenesis while other taxa have been associated with normal colons. The some of the taxa associated with normal colons can use fiber to produce short-chain fatty acids which have been shown to inhibit tumor growth in model systems. However, the data supporting the importance of SCFAs in humans is less certain. Here we test the hypothesis that short-chain fatty acid concentration changes in individuals with colorectal tumors.

Methods. We analyzed a cross-sectional (n=490) and pre- and post-treatment (n=67) group for their concentrations of acetate, butyrate, and propionate. Analysis also included tumor classification models using Random Forest, imputed gene relative abundance with PICRUSt, and metagenomic sequencing on a subset (n=85) of the total cross-sectional group.

Results. No difference in short-chain fatty acid concentration was found between normal individuals and patients with adenoma or carcinoma (P-value > 0.15). There was no difference in the ability of classification models with or without short-chain fatty acids added to them in predicting adenoma or carcinoma (P-value > 0.05). Using metagenomic sequencing, there was also no difference in genes involved with short-chain fatty acid synthesis (P-value > 0.70).

Conclusions. Although our data does not support the hypothesis that short-chain fatty acids decrease in individuals that have colorectal tumors, there may be specific scenarios where short-chain fatty acids may still be beneficial for treatment of colorectal cancer. Alternatively, there may be other metabolites that have not been thoroughly investigated that are more important to human colorectal cancer.

Introduction

Colorectal cancer (CRC) is currently the third leading cancer-related cause of death within the US (1, 2). The envrionment has been attributed as an important risk factor for CRC (3). This has lead to the investigation of how the microbiota may exacerbate or cause tumorigensis (4–6) and whether the bacterial community is altered (7, 8). Multiple reports in case/control studies on the microbiota and CRC have identified taxa commonly associated with the normal colon to be decreased in patients with carcinoma tumors (9–11). Many of these taxa actively produce short-chain fatty acids (SCFAs) from fiber that is part of our general diet (12). The most extensively studied of these SCFAs are acetate, butyrate, and propionate (13). Overall, CRC is a growing health problem and the microbiota through SCFAs have shown promise as a viable treatment option.

Within model systems SCFAs, in particular butyrate, have shown positive results in CRC treatment (14). In particular, butyrate has been shown to inhibit cancer cell growth in *in vitro* systems (15). Additionally, supplementation with food sources that bacteria use to create these SCFAs may also be able to confer beneficial effects. Fiber supplementation in mouse models of CRC caused an overall reduction in tumor burden while also increasing SCFA concentrations (16). Although these model systems provide important preliminary evidence towards the ability of SCFAs to reduce and treat tumors, the studies reporting benefit in humans has been less convincing.

There is a lack of evidence on the benefit of increasing SCFA concentrations to protect against CRC in human populations. The initial case/control studies that investigated SCFA concentrations in CRC have found that patients with carcinomas had lower concentrations of acetate, butyrate, and propionate versus patients with adenomas and individuals with normal colons (17). Although this would argue that increasing SCFA concentrations could be protective against tumorigenesis, fiber supplementation in randomized controlled trials have consistently failed to protect against tumor recurrence (18). These findings would argue against the utility of treatments that aim to use SCFAs to reduce or protect against tumorigenesis. Given the lack of clear evidence in human studies of the benefit of SCFAs in CRC there is a need for more investigation into this area.

Our current study will fill some of the current gaps in the literature that relate to the study of SCFAs

and CRC in human populations. Specifically, it will attempt to replicate previous case/control findings on SCFA concentrations in individuals with and without tumors. Additionally, by increasing the adenoma tumor cohort it we will be able to identify if the observation of a continuous gradual reduction in SCFA concentrations occurs as tumor severity increases. It also will build upon these observations and assess the utility of using SCFAs and Operational Taxonomic Units (OTUs) as a potential risk stratification tool of colorectal tumors. This study will provide important information on the replicability of previous findings in humans by extensively studying how SCFAs could be associated with colorectal tumors.

To accomplish this task we will directly measure the concentration of acetate, butyrate, and propionate within fecal samples for two different groups. The first group will be at a cross sectional point in time while the second group will be before (pre-) and after (post-)treatment. Additionally, we also will assess whether adding SCFA concentrations to OTU data improves the classification of adenoma or carcinoma, use PICRUSt and metagenomic sequencing to assess the genes involved in SCFA synthesis, and analyze how well 16S rRNA gene sequencing can predict SCFA concentrations. This extensive investigation will provide additional information as to whether the taxa that are decreased in carcionma use SCFAs to help protect against tumorigenesis.

Results

Decreased SCFA concentrations are not associated with tumors. We used frozen fecal samples from 490 individuals and HPLC to measure acetate, butyrate, and propionate concentrations at a cross-sectional point in time. There was no difference between individuals with normal colons (n=172) and patients with either an adenoma (n=198) or carcinoma (n=120) for any of the SCFAs measured after multiple comparison correction (P-value > 0.15) [Figure 1A - 1C]. We next measured the concentration of SCFAs in 67 patients with an adenoma (n=41) or carcinoma (n=26) in which we had pre- and post-treatment fecal samples. Although there was a general trend for increasing acetate, butyrate, and propionate concentrations after treatment for tumors, there was no significant difference pre- and post-treatment for patients with adenoma (P-value > 0.20) or carcinoma (P-value > 0.80) [Figure 1D - 1F]. Even though there was no change in SCFA concentrations between individuals with normal colons and those with tumors, this information could still be important to help classify disease.

Random Forest models with SCFA concentrations do not classify tumors better. Using the Random Forest algorithm we built models to classify normal versus adenoma and normal versus carcinoma with OTUs or OTUs and SCFA concentrations. For both adenoma and carcinoma models, there was no difference between the median AUC of models with or without SCFA concentrations (P-value > 0.05) [Figure 2]. Although including SCFA concentrations did not add extra information for classification of disease using Random Forest models, it is possible that the genes for enzymes involved in SCFA synthesis may change due to disease status; where a smaller number of microbes are responsible for the observed SCFA concentrations.

Changes in genes for enzymes involved in SCFA synthesis are not associated with tumors. Using a list of specific genes that are important for the synthesis of SCFAs [Table S1], we looked for differences in gene abundance based on individuals having normal colons, adenomas, or carcinomas. First, using imputed gene relative abundance based on 16S rRNA gene sequencing we found no difference in any of the genes inolved with acetate, butyrate, or propionate synthesis (P-value > 0.90) [Table S2]. This similarity between groups is highlighted by visualizing genes important in butyrate synthesis [Figure 3A]. Using a Wilcoxson rank-sum test, there also was no

93 difference in imputed gene relative abundance between pre- and post-treatment samples for any

94 genes involved with SCFA synthesis (P-value > 0.70) [Table S3]. Next, we took a subset of these

95 490 fecal samples (n=85) and performed metagenomic sequencing to confirm these results. Like

96 the imputed gene results, metagenomic analysis found that there was no difference in any of the

97 genes involved in SCFA synthesis between individuals with normal colons (n=29) and patients

98 with adenoma (n=28) or carcinoma (n=28) (P-value > 0.70) [Table S4]. This lack of difference

99 is highlighted when we visualize the results for butyrate kinase [Figure 3B]. These observations

100 provide evidence that the gene content does not change due to tumors. Although these results

101 suggest that SCFA concentrations do not change due to tumor, there may be some errors present

102 within our results that may account for the lack of associations.

103 **Expected taxa are associated with higher SCFA concentrations regardless of tumor status.**

104 Using OTU data we built Random Forest models to classify higher than median and lower than

105 median SCFA concentrations. Overall, OTU data had a reasonable ability to classify high and low

106 SCFA concentrations [Figure S1A]. However, these models tended to be overfit, suggesting that

107 rarer taxa may be important for this classification [Figure S1A]. The most important OTUs to these

108 models (assessed with mean decrease in accuracy (MDA)) were to taxa that are normally associated

109 with SCFA production [Figure S1B]. These results highlight that there are no immediately apparent

110 errors in our data because SCFA concentrations are associated with taxa known to produce acetate,

111 butyrate, and propionate. Additionally, OTUs associated with these taxa are the most important to

112 models that can classify high and low SCFA concentrations. Overall, our results do not support the

113 hypothesis that lower SCFA concentrations are associated with colorectal cancer.

6

Discussion

The observations from this study do not support the hypothesis that SCFA concentrations are lower in individuals with tumors. Whether we directly measured the SCFA concentration or investigate genes associated with their production no difference could be identified [Figure 1 & 3]. Further, no benefit could be found in using the concentrations to help classify individuals with and without tumors [Figure 2]. These observations are in stark contrast to some of the previous literature.

Much of the previous literature on SCFA benefit to human colorectal cancer has been illustrated in model systems (14). Many SCFAs are produced through the breakdown of fiber (12). A recent study in mice found that fiber supplementation increased SCFA concentration and decreased tumor burden (16). Additionally, SCFAs such as butyrate have been found to inhibit tumor growth in *in vitro* experiments (15). Yet, observations in humans has been mixed. Previous case/control studies have suggested that SCFAs are lower in individuals with carcinoma tumors (17). However, individual studies and meta-analysis on fiber supplementation to prevent tumor recurrence have found no benefit (18, 19). Our results align with what has been reported in randomized-controlled trials that SCFAs do not protect against all types of colorectal tumors. It is possible though that there are specific instances where SCFAs may be beneficial.

One limitation of ours and others research into the effect of SCFAs in colorectal cancer has been that all tumors are treated as the same type. However, there are known differences in the types of mutations that occur (20) and treating all tumors as equal may actually hide any benefit that could be found in specific subsets of individuals. Similar to the idea of specific immunotherapy for specific tumors (21), SCFAs may have beneficial effects for specific types of colorectal cancers. Future research will need to test if this is a valid hypothesis. Although there are limitations with the current research on SCFAs and colorectal tumors, evidence is building that these specific metabolites may not be protective. Yet, taxa that are associated with SCFA production are consistently higher in normal colons than those with carcinomas (7, 8, 22).

The potential protection against colorectal cancers may not be from SCFAs even though taxa associated with their production are higher in individuals with normal colons versus patients

7

141 with carcinomas (7, 8, 22). Protection could be via a different pathway and other metabolites

142 that have not been extensively studied. Alternatively, protection may not occur via a metabolite

143 but instead occur through niche exclusion of mouth-associated microbes (e.g. *Fusobacterium*,

144 *Porphyromonas*, *Parvimonas*, *Peptostreptococcus* (9, 10)). The idea of niche exclusion is similar

145 to how the community protects against *Clostridium difficile* infection with chronic inflammation

146 replacing the role of antibiotics. Although we did not find lower concentrations of SCFAs associated

147 with colorectal tumors, we think that there are many exciting new avenues to explore because of

148 these observations.

Conclusions

Our observations found no difference in SCFA concentration, their utility as a classification tool, and for genes for enzymes involved in SCFA synthesis between individuals with normal colons and patients with either adenoma or carcinoma tumors. Although these results are different than other reports in the literature, they do align with the randomized controlled trials that have tested fiber use in preventing colorectal tumor recurrence. Overall, these results suggest that the SCFAs produced by resident microbes do not protect against tumor. By focusing on other types of metabolites, the identification of more promising therapeutic options for use in treating colorectal cancer may be found.

Materials and Methods

Study design and sampling. The overall protocol has been described in detail previously (23, 24). In brief, this study used fecal samples obtained at either a single cross-sectional time point (n=490) or from before (pre-) and after (post-) treatment for their tumor (n=67). For patients undergoing treatment for their tumor the length of time between their initial and follow up sample ranged from 188 - 546 days. Our use of treatment has been previously defined as encompassing removal of a tumor with or without chemotherapy and radiation (23). Diagnosis of tumor was made by colonoscopic examination and histopathological review of biopsies obtained (23, 24). The University of Michigan Institutional Review Board approved the study and informed consent was obtained from all participants in accordance to the guidelines set out by the Helsinki Decleration.

Measuring specific SCFAs. Our protocol for the measurement of acetate, butyrate, and propionate followed a previously published protocol (25). The following changes to this protocol included the use of frozen fecal samples suspended in 1ml of PBS instead of fecal suspensions in DNA Genotek OmniGut tubes, and the use of the acutal weight of fecal samples instead of the average weight for SCFA concentration normalizations. These changes did not affect the overall median concentrations of these SCFAs between the two studies (see Table 1 (25) and Figure 1 in this report).

16s rRNA gene sequencing. The workflow and processing have been described previously (23, 26, 27). The major differences from these previous reports include: the use of version 1.39.5 of the mothur software package and clustering Operational Taxonomic Units (OTUs) at 97% similarity used the OptClust algorithm (28).

Generating imputed metagenomes. The use of PICRUSt version 1.1.2 with the recommended standard operating protocol (29) was used. Briefly, the mothur shared file and metadata was converted into a biom formated table using the biom convert function, the subsequent biom file was processed with the normalize_by_copy_number.py function, and subsequent imputed metagenomes created using the predict_metagenomes.py function.

Obtaining OPFs from metagenomes. A subset of the cross-sectional group (n=490) containing a total of 85 individuals (normal n=29 normal, adenoma n=28, and carcinoma n=28) was shotgun

10

sequenced on an Illumina HiSeq using 125 bp paired end reads and a previously described method (30). Briefly, the sequences were quality filtered and sequences aligning to the human genome were removed prior to contig assembly with MEGAHIT (31). Open Reading Frames (ORFs) were identified using Prodigal (32), counts generated using Diamond (33), subsequent clustering into Operational Protein Families (OPFs) used mmseq2 (34), and OPF alignment used the KEGG database (35).

Pulling genes involved with SCFA synthesis. Specific genes located near the end of the pathways involved in the synthesis of acetate, butyrate, and propionate were analyzed for any differences between individuals with normal colons and those with tumors. These genes were based on pathways from KEGG as well as previous research (35, 36) and a list can be found in the supplemental material [Table S1].

Random Forest Models. The model was first trained on 80% of the data and then tested on the held out 20% (80/20 split) using the Random Forest algorithm for both classification and regression models (37). This was repeated on 100 differen 80/20 splits of the data to generate a reasonable range for the AUC of the model. The reported AUCs, unless otherwise specified, are for the test sets. Classification models were built to group normal versus adenoma and normal versus carcinoma or high versus low SCFA concentrations.

Statistical analysis workflow. All analysis was performed using the statistical language R (38). Generally, differences between the different disease groups used a Kruskal-Walis rank sum test with a Dunn's post-hoc test. Where appropriate Benjamini-Hochberg was used to correct for multiple comparisons (39). First, we assessed differences in SCFA concentrations measured by HPLC between individuals with normal colons and patients with tumors (adenoma or carcinoma). We then analyzed whether SCFA concentrations changed in patients with an adenoma or carcinoma pre- versus post-treatment. Next, we assessed whether OTUs alone or OTUs and SCFAs were better able to classify individuals with and without tumor using Random Forest models. Next, the imputed gene counts of important mediators of SCFA creation were tested. Finally, the counts generated for OPFs that matched important genes involved with SCFA creation were analyzed. Finally, models to classify high or low SCFA concentration based on 16S rRNA gene sequencing data were created

11

213 using the Random Forest algorithm.

Acknowledgements

13

References

1. **Haggar F**, **Boushey R**. 2009. Colorectal cancer epidemiology: Incidence, mortality, survival, and risk factors. Clinics in Colon and Rectal Surgery **22**:191–197. doi:10.1055/s-0029-1242458.

2. **Siegel RL**, **Miller KD**, **Jemal A**. 2016. Cancer statistics, 2016. CA: A Cancer Journal for Clinicians **66**:7–30. doi:10.3322/caac.21332.

3. **Lichtenstein P**, **Holm NV**, **Verkasalo PK**, **Iliadou A**, **Kaprio J**, **Koskenvuo M**, **Pukkala E**, **Skytthe A**, **Hemminki K**. 2000. Environmental and heritable factors in the causation of cancer analyses of cohorts of twins from sweden, denmark, and finland. New England Journal of Medicine **343**:78–85. doi:10.1056/nejm200007133430201.

4. **Zackular JP**, **Baxter NT**, **Iverson KD**, **Sadler WD**, **Petrosino JF**, **Chen GY**, **Schloss PD**. 2013. The gut microbiome modulates colon tumorigenesis. mBio **4**:e00692–13–e00692–13. doi:10.1128/mbio.00692-13.

5. **Baxter NT**, **Zackular JP**, **Chen GY**, **Schloss PD**. 2014. Structure of the gut microbiome following colonization with human feces determines colonic tumor burden. Microbiome **2**:20. doi:10.1186/2049-2618-2-20.

6. **Zackular JP**, **Baxter NT**, **Chen GY**, **Schloss PD**. 2015. Manipulation of the gut microbiota reveals role in colon tumorigenesis. mSphere **1**:e00001–15. doi:10.1128/msphere.00001-15.

7. **Shah MS**, **DeSantis TZ**, **Weinmaier T**, **McMurdie PJ**, **Cope JL**, **Altrichter A**, **Yamal J-M**, **Hollister EB**. 2017. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. Gut **67**:882–891. doi:10.1136/gutjnl-2016-313189.

8. **Sze MA**, **Schloss PD**. 2018. Leveraging existing 16S rRNA gene surveys to identify reproducible biomarkers in individuals with colorectal tumors. doi:10.1101/285486.

9. **Zeller G**, **Tap J**, **Voigt AY**, **Sunagawa S**, **Kultima JR**, **Costea PI**, **Amiot A**, **Bohm J**, **Brunetti F**, **Habermann N**, **Hercog R**, **Koch M**, **Luciani A**, **Mende DR**, **Schneider MA**, **Schrotz-King P**, **Tournigand C**, **Nhieu JTV**, **Yamada T**, **Zimmermann J**, **Benes V**, **Kloor M**, **Ulrich CM**, **Knebel**

14

246 **Doeberitz M von**, **Sobhani I**, **Bork P**. 2014. Potential of fecal microbiota for early-stage detection

247 of colorectal cancer. Molecular Systems Biology **10**:766–766. doi:10.15252/msb.20145645.

248 10. **Baxter NT**, **Ruffin MT**, **Rogers MAM**, **Schloss PD**. 2016. Microbiota-based model improves

249 the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Medicine **8**.

250 doi:10.1186/s13073-016-0290-3.

251 11. **Kostic AD**, **Gevers D**, **Pedamallu CS**, **Michaud M**, **Duke F**, **Earl AM**, **Ojesina AI**, **Jung J**,

252 **Bass AJ**, **Tabernero J**, **Baselga J**, **Liu C**, **Shivdasani RA**, **Ogino S**, **Birren BW**, **Huttenhower**

253 **C**, **Garrett WS**, **Meyerson M**. 2011. Genomic analysis identifies association of fusobacterium with

254 colorectal carcinoma. Genome Research **22**:292–298. doi:10.1101/gr.126573.111.

255 12. **Holscher HD**. 2017. Dietary fiber and prebiotics and the gastrointestinal microbiota. Gut

256 Microbes **8**:172–184. doi:10.1080/19490976.2017.1290756.

257 13. **Louis P**, **Flint HJ**. 2016. Formation of propionate and butyrate by the human colonic microbiota.

258 Environmental Microbiology **19**:29–41. doi:10.1111/1462-2920.13589.

259 14. **O'Keefe SJD**. 2016. Diet, microorganisms and their metabolites and colon cancer. Nature

260 Reviews Gastroenterology & Hepatology **13**:691–706. doi:10.1038/nrgastro.2016.165.

261 15. **Encarnação JC**, **Pires AS**, **Amaral RA**, **Gonçalves TJ**, **Laranjo M**, **Casalta-Lopes JE**,

262 **Gonçalves AC**, **Sarmento-Ribeiro AB**, **Abrantes AM**, **Botelho MF**. 2018. Butyrate, a dietary

263 fiber derivative that improves irinotecan effect in colon cancer cells. The Journal of Nutritional

264 Biochemistry **56**:183–192. doi:10.1016/j.jnutbio.2018.02.018.

265 16. **Bishehsari F**, **Engen P**, **Preite N**, **Tuncil Y**, **Naqib A**, **Shaikh M**, **Rossi M**, **Wilber S**, **Green**

266 **S**, **Hamaker B**, **Khazaie K**, **Voigt R**, **Forsyth C**, **Keshavarzian A**. 2018. Dietary fiber treatment

267 corrects the composition of gut microbiota, promotes SCFA production, and suppresses colon

268 carcinogenesis. Genes **9**:102. doi:10.3390/genes9020102.

269 17. **Ohigashi S**, **Sudo K**, **Kobayashi D**, **Takahashi O**, **Takahashi T**, **Asahara T**, **Nomoto**

270 **K**, **Onodera H**. 2013. Changes of the intestinal microbiota, short chain fatty acids, and

271 fecal pH in patients with colorectal cancer. Digestive Diseases and Sciences **58**:1717–1726.

272 doi:10.1007/s10620-012-2526-4.

273 18. **Yao Y**, **Suo T**, **Andersson R**, **Cao Y**, **Wang C**, **Lu J**, **Chui E**. 2017. Dietary fibre for the
274 prevention of recurrent colorectal adenomas and carcinomas. Cochrane Database of Systematic
275 Reviews. doi:10.1002/14651858.cd003430.pub2.

276 19. **Schatzkin A**, **Lanza E**, **Corle D**, **Lance P**, **Iber F**, **Caan B**, **Shike M**, **Weissfeld J**, **Burt R**,
277 **Cooper MR**, **Kikendall JW**, **Cahill J**, **Freedman L**, **Marshall J**, **Schoen RE**, **Slattery M**. 2000.
278 Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. New England
279 Journal of Medicine **342**:1149–1155. doi:10.1056/nejm200004203421601.

280 20. **Fearon ER**, **Vogelstein B**. 1990. A genetic model for colorectal tumorigenesis. Cell
281 **61**:759–767. doi:10.1016/0092-8674(90)90186-i.

282 21. **Thomas X**, **Heiblig M**. 2016. The development of agents targeting the BCR-ABL tyrosine
283 kinase as philadelphia chromosome-positive acute lymphoblastic leukemia treatment. Expert
284 Opinion on Drug Discovery **11**:1061–1070. doi:10.1080/17460441.2016.1227318.

285 22. **Sze MA**, **Baxter NT**, **Ruffin MT**, **Rogers MAM**, **Schloss PD**. 2017. Normalization of the
286 microbiota in patients after treatment for colonic lesions. Microbiome **5**. doi:10.1186/s40168-017-0366-3.

287 23. **Sze MA**, **Baxter NT**, **Ruffin MT**, **Rogers MAM**, **Schloss PD**. 2017. Normalization of the
288 microbiota in patients after treatment for colonic lesions. Microbiome **5**. doi:10.1186/s40168-017-0366-3.

289 24. **Baxter NT**, **Ruffin MT**, **Rogers MAM**, **Schloss PD**. 2016. Microbiota-based model improves
290 the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Medicine **8**.
291 doi:10.1186/s13073-016-0290-3.

292 25. **Venkataraman A**, **Sieber JR**, **Schmidt AW**, **Waldron C**, **Theis KR**, **Schmidt TM**. 2016.
293 Variable responses of human microbiomes to dietary supplementation with resistant starch.
294 Microbiome **4**. doi:10.1186/s40168-016-0178-x.

295 26. **Schloss PD**, **Westcott SL**, **Ryabin T**, **Hall JR**, **Hartmann M**, **Hollister EB**, **Lesniewski RA**,
296 **Oakley BB**, **Parks DH**, **Robinson CJ**, **Sahl JW**, **Stres B**, **Thallinger GG**, **Horn DJV**, **Weber CF**.

297 2009. Introducing mothur: Open-source, platform-independent, community-supported software

298 for describing and comparing microbial communities. Applied and Environmental Microbiology

299 **75**:7537–7541. doi:10.1128/aem.01541-09.

300 27. **Kozich JJ**, **Westcott SL**, **Baxter NT**, **Highlander SK**, **Schloss PD**. 2013. Development of a

301 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on

302 the MiSeq illumina sequencing platform. Applied and Environmental Microbiology **79**:5112–5120.

303 doi:10.1128/aem.01043-13.

304 28.   **Westcott SL**, **Schloss PD**. 2017.   OptiClust, an improved method for assigning

305 amplicon-based sequence data to operational taxonomic units.   mSphere **2**:e00073–17.

306 doi:10.1128/mspheredirect.00073-17.

307 29. **Langille MGI**, **Zaneveld J**, **Caporaso JG**, **McDonald D**, **Knights D**, **Reyes JA**, **Clemente JC**,

308 **Burkepile DE**, **Thurber RLV**, **Knight R**, **Beiko RG**, **Huttenhower C**. 2013. Predictive functional

309 profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology

310 **31**:814–821. doi:10.1038/nbt.2676.

311 30. **Hannigan GD**, **Duhaime MB**, **Ruffin MT**, **Koumpouras CC**, **Schloss PD**. 2017. Diagnostic

312 potential & the interactive dynamics of the colorectal cancer virome. doi:10.1101/152868.

313 31. **Li D**, **Liu C-M**, **Luo R**, **Sadakane K**, **Lam T-W**. 2015. MEGAHIT: An ultra-fast single-node

314 solution for large and complex metagenomics assembly via succinct de bruijn graph. Bioinformatics

315 **31**:1674–1676. doi:10.1093/bioinformatics/btv033.

316 32. **Hyatt D**, **Chen G-L**, **LoCascio PF**, **Land ML**, **Larimer FW**, **Hauser LJ**. 2010. Prodigal:

317 Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119.

318 doi:10.1186/1471-2105-11-119.

319 33. **Buchfink B**, **Xie C**, **Huson DH**. 2014. Fast and sensitive protein alignment using DIAMOND.

320 Nature Methods **12**:59–60. doi:10.1038/nmeth.3176.

321 34. **Steinegger M**, **Söding J**. 2017. MMseqs2 enables sensitive protein sequence searching for

322 the analysis of massive data sets. Nature Biotechnology. doi:10.1038/nbt.3988.

323 35. **Kanehisa M**, **Sato Y**, **Kawashima M**, **Furumichi M**, **Tanabe M**. 2015. KEGG as a
324 reference resource for gene and protein annotation. Nucleic Acids Research **44**:D457–D462.
325 doi:10.1093/nar/gkv1070.

326 36. **Baxter NT**, **Zackular JP**, **Chen GY**, **Schloss PD**. 2014. Structure of the gut microbiome
327 following colonization with human feces determines colonic tumor burden. Microbiome **2**:20.
328 doi:10.1186/2049-2618-2-20.

329 37. **Liaw A**, **Wiener M**. 2002. Classification and regression by randomForest. R News **2**:18–22.

330 38. **R Core Team**. 2017. R: A language and environment for statistical computing. R Foundation
331 for Statistical Computing, Vienna, Austria.

332 39. **Benjamini Y**, **Hochberg Y**. 1995. Controlling the false discovery rate: A practical and powerful
333 approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological)
334 **57**:289–300.

Figure 1. Using HPLC no change in SCFA measurements was observed between normal, adenoma, and carcinoma individuals. Acetate concentrations in fecal samples of individuals with normal colons, adenomas, and carcinomas (A). Butyrate concentrations in fecal samples of individuals with normal colons, adenomas, and carcinomas (B). Propionate concentrations in fecal samples of individuals with normal colons, adenomas, and carcinomas (C). The black links indicate the median SCFA concentration. Acetate concentrations in fecal samples before and after treatment for adenoma (yellow) and carcinoma (red) (D). Butyrate concentrations in fecal samples before and after treatment for adenoma (yellow) and carcinoma (red) (E). Propionate concentrations in fecal samples before and after treatment for adenoma (yellow) and carcinoma (red) (F). The black dots and lines represent the median change in SCFA concentration.

Figure 2. SCFAs do not improve OTU-based Random Forest models. Difference between the area under the curve of 100 different 80/20 split OTU-based normal versus adenoma 10-fold CV models with and without SCFAs (A). Difference between the area under the curve of 100 different 80/20 OTU-based normal versus carcinoma 10-fold CV models with and without SCFAs (B). The black linke represents the median AUC. The dotted line highlights an AUC of 0.5.

Figure 3. No change in butyrate producing genes identified between normal, adenoma, and carcinoma individuals. Imputed gene relative abundance of important butyrate pathway genes using PICRUSt (A). Counts per million (corrected for size and number of contigs in an OPF) for the Butyrate Kinase gene (B). The other genes from the PICRUSt analysis did not align to any of the OPFs in the metagenome analysis.

19

Figure S1. **OTU-based Random Forest models of SCFA concentrations.** Classification Random Forest train and tests of 100 different 80/20 OTU-based models with 10-fold CV based on higher or lower than the medain SCFA concentration (A). The top 10 OTUs based on mean decrease in accuracy (MDA) for each model, colored by their lowest taxonomic identification (B). Regression Random Forest train and tests of 100 different 80/20 OTU-based models with 10-fold CV based on correlation to actual SCFA concentration (C). The top 10 OTUs based on mean decrease in accuracy (MDA) for each model, colored by their lowest taxonomic identification (D).