

# Revisiting the Relationship between Short-Chain Fatty Acids, the Microbiota, and Colorectal Tumors

Fecal short chain fatty acids are not predictive of colorectal cancer status and cannot be predicted based on bacterial community structure

Marc A. Sze<sup>1</sup>, Begüm Topçuoğlu<sup>1</sup>, Nicholas A. Lesniak<sup>1</sup>, Mack T. Ruffin IV<sup>2</sup>, Patrick D. Schloss<sup>1†</sup>

† To whom correspondence should be addressed: [pschloss@umich.edu](mailto:pschloss@umich.edu)

<sup>1</sup> Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109

<sup>2</sup> Department of Family Medicine and Community Medicine, Penn State Hershey Medical Center, Hershey, PA

1 **Abstract (250 words)**

2 Something clever

3 **Importance (150 words)**

4 Something clever

Colorectal cancer is the third leading cancer-related cause of death within the United States (1, 2). Less than 10% of cases can be attributed to genetic risk factors (3). This leaves a significant role for environmental, behavioral, and other factors such as smoking and diet (4, 5). Colorectal cancer is thought to be initiated by a series of mutations that accumulate as mutated cells proliferate leading to adenomatous lesions, which are succeeded by carcinomas (3). Throughout this progression, there are ample opportunities for bacterial populations to create mutations, induce inflammation, and accelerate tumorigenesis (6). Numerous studies in murine models have supported this model (7–11). Additional cross sectional studies in humans have identified microbiome-based biomarkers of disease (12–20). These studies suggest that in some cases, it is the loss of bacterial populations that produce short-chain fatty acids (SCFAs) that results in increased inflammation and tumorigenesis.

SCFAs have anti-inflammatory and anti-proliferative activities (21–24). Furthermore, manipulation of SCFAs in mouse models of colorectal cancer by direct supplementation or feeding of fiber caused an overall reduction in tumor burden (25, 26). These results suggest that supplementation with substrates that bacteria can ferment to produce SCFAs may confer beneficial effects against colorectal cancer. Regardless, there is a lack of evidence that increasing SCFA concentrations can protect against colorectal cancer in humans. Based on similar observations, many microbiome studies use the concentrations of SCFAs and the presence of 16S rRNA gene sequences from organisms and the genes involved in producing them as a biomarker of a healthy microbiota (27–30). Case-control studies that have investigated SCFA concentrations in colorectal cancer found that patients with carcinomas had lower concentrations of SCFAs versus patients with adenomas or individuals without colon tumors (31). Although this would argue that increasing SCFA concentrations could be protective against tumorigenesis, in randomized controlled trials fiber supplementation has been inconsistently associated with protection against tumor formation and recurrence (32–36). These findings temper enthusiasm for treatments that aim to use SCFAs as biomarkers or protection against tumorigenesis.

**SCFA concentrations do not meaningfully vary with diagnosis or treatment.** To quantify the associations between colorectal cancer, the microbiome, and SCFAs, we quantified the concentration of acetate, propionate, isobutyrate, and butyrate in feces of previously characterized

individuals with normal colons (N=172) and those with colonic adenomas (N=198) or carcinomas (N=120) (14). The only SCFA that had a significantly different concentration across the diagnoses was isobutyrate ( $P=0.0091$ ; Figure 1A). The median concentration of isobutyrate was 3.30 mmol/kg in people with normal colons and it was 3.00 and 3.84 mmol/kg in people with colonic adenomas or carcinomas. The difference in isobutyrate concentration between people with adenomas and carcinomas was significantly different ( $P=0.0065$ ); however, the differences between people with normal colons and those with adenomas or carcinomas was not significant ( $P=0.19$  and  $P=0.11$ ). The median concentration of isobutyrate in people with normal colons (3.30 mmol/kg) was between those with adenomas (3.00 mmol/kg) and carcinomas (3.84 mmol/kg). Among the subjects with adenomas and carcinomas, a subset ( $N_{\text{adenoma}}=41$ ,  $N_{\text{carcinoma}}=26$ ) were treated and sampled a year later (37). The only SCFA that changed following treatment was isobutyrate, which decreased by 0.99 mmol/kg ( $P=0.002$ ; Figure 1B). For both the pre-treatment cross-sectional data and the pre/post treatment data, we pooled the SCFA concentrations on a per molecule of carbon basis and again failed to see any significant differences ( $P>0.15$ ). The low concentration of isobutyrate relative to the other SCFAs, inconsistent concentrations, and unexpected decrease in concentration with treatment makes it difficult to ascribe much biological relevance to this observation.

**Combining SCFA and microbiome data does not improve the ability to diagnose individual as having adenomas or carcinomas.** We previously found that binning 16S rRNA gene sequence data into operational taxonomic units based on 97% similarity or into genera enabled us to classify individuals as having adenomas or carcinomas using Random Forest machine learning models (13, 14). We repeated that analysis but added the concentration of the individual or pooled SCFAs as possible features to train a model. Models trained using SCFAs to classify individuals as having adenomas or carcinomas rather than normal colons had median areas under the receiver operator characteristic curve (AUROC) that were significantly greater than 0.5 ( $P_{\text{adenoma}}<0.001$  and  $P_{\text{carcinoma}}<0.001$ ); however, the AUROC values to detect the presence of adenomas or carcinomas were only 0.54 and 0.55, respectively (Figure 2A). When we trained the models with the SCFAs concentrations and OTU or genus-level relative abundances the AUROC values were not significantly different from the models trained without the SCFA concentrations ( $P>0.21$ ; Figure 2A). We also trained models using a smaller dataset that generated shotgun metagenomic

sequencing data from a subset of our cohort ( $N_{\text{normal}}=27$ ,  $N_{\text{adenoma}}=25$ , and  $N_{\text{cancer}}=26$ ) (16). We binned genes extracted from the assembled metagenomes into operational protein families (OPFs) or KEGG categories. Again, the performance of the models trained with the metagenomic data did not improve when the SCFA concentrations were added as possible features when training the model ( $P>0.24$ ; Figure 2A). These data demonstrate that knowledge of the SCFA profile from a patient's fecal sample does not improve the ability to diagnose a colonic lesion.

**Knowledge of microbial community structure does not predict SCFA concentrations.**

Regardless of a person's diagnosis, we next asked whether the fecal community structure was predictive of fecal SCFA concentrations. We trained Random Forest regression models using 16S rRNA gene sequence data binned into OTUs and genera to predict the concentration of the individual and pooled SCFAs. Regardless of the binning method or SCFA, the largest amount of variation that the model could explain was **0.XXXX** (Figure 2B). Next, we trained Random Forest regression models using metagenomic sequence data binned into OPFs and KEGG categories to predict the concentration of the individual and pooled SCFAs. Similar to the analysis using 16S rRNA gene sequence data, the metagenomic data was not predictive of SCFA concentration; the largest amount of variation that the models could explain was **0.XXXX** (Figure 2B). Because of the limited number of samples that we were able to generate metagenomic sequence data from, we used our 16S rRNA gene sequence data to impute metagenomes that were binned into metabolic pathways, enzyme commission numbers, or KEGG categories using picrust2. Again, SCFA concentrations could not be predicted based on the imputed metagenomic data; the largest amount of variation that the models could explain was **0.XXXX** (Figure 2B). The inability to model SCFA concentrations from microbiome data indicates that the knowledge of the abundance of organisms and genes is insufficient to predict SCFA concentrations.

**Conclusion.** Our data indicate that fecal SCFA concentrations are not associated with the presence of adenomas or carcinomas and that they provide weak predictive power to improve the ability to diagnose someone with one of these lesions. Furthermore, knowledge of the taxonomic and genetic structure of gut microbiota was not predictive of individual or pooled SCFA concentrations. These results complement existing literature that suggest that fiber consumption and the production of SCFAs are unable to prevent the risk of developing colonic tumors. It is important to note

92 that our analysis concerned fecal SCFA concentrations and microbiome characterization and  
93 that observations along the mucosa near the site of lesions may provide a stronger association.  
94 Regardless, given the growing literature in this area, it is unlikely that SCFAs are the primary  
95 mechanism that limits tumorigenesis. This may be a cautionary result to temper enthusiasm for  
96 SCFAs as a biomarker of gut health more generally. Going forward it is critical to develop additional  
97 hypotheses for how the microbiome and host interact to drive tumorigenesis to better understand  
98 the disease process and identify biomarkers that will allow early detection of tumors.

## Materials and Methods

**Study design and sampling.** The overall protocol has been described in detail previously (14, 37). In brief, this study used fecal samples obtained at either a single cross-sectional time point ( $n = r$  length(meta sample)\*\*) or from before (pre-) and after (post-) treatment of a patient's tumor (adenoma = \*\*r length(filter(meta f, dx == "adenoma") time) - r max(meta f time) days. Our use of treatment has been previously defined as encompassing removal of a tumor (surgery or colonoscopy) with or without chemotherapy and radiation (37). Diagnosis of tumor was made by colonoscopic examination and histopathological review of biopsies obtained (14, 37). 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 Declaration.

**Measuring specific SCFAs.** Our protocol for the measurement of acetate, butyrate, and propionate followed a previously published protocol that used a High-Performance Liquid Chromatography (HPLC) machine (38). 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 actual weight of fecal samples instead of the average weight for SCFA concentration normalizations. These methodological changes did not affect the overall median concentrations of these SCFAs between the two studies (see Table 1 (38) and Figure 1 here).

**16S rRNA gene sequencing.** The workflow and processing have been previously described (37, 39, 40). In brief, sequences were quality filtered and contigs created from the paired end reads. Any sequences with ambiguous base calls were discarded. Contigs were then checked for matches to the V4 region of the 16S rRNA gene using the SILVA database (41). Chimeras were identified and removed using UCHIME and OTUs clustered at 97% similarity (42). 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 using the OptiClust algorithm (43).

**Generating imputed metagenomes.** The use of PICRUSt version 1.1.2 with the recommended standard operating protocol (44) was used. Briefly, the mothur shared file and metadata was converted into a biom formatted 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 Operational Protein Families from metagenomes.** A subset of the cross-sectional group ( $n = \text{length}(\text{metasample}) * *) \text{containing a total of } * * \text{length}(\text{geof\_samplesX1})$  individuals (normal  $n = \text{filter}(\text{geof\_samples}, \text{X30} == \text{"Healthy"}) \%>\% \text{pull}(\text{X30}) \%>\% \text{length}()$ , adenoma  $n = \text{filter}(\text{geof\_samples}, \text{X30} == \text{"Adenoma"}) \%>\% \text{pull}(\text{X30}) \%>\% \text{length}()$ , and carcinoma  $n = \text{filter}(\text{geof\_samples}, \text{X30} == \text{"Cancer"}) \%>\% \text{pull}(\text{X30}) \%>\% \text{length}()$ ) was shotgun sequenced on an Illumina HiSeq using 125 bp paired end reads and a previously described method (16). Briefly, the sequences were quality filtered and sequences aligning to the human genome were removed prior to contig assembly with MEGAHIT (45). Open Reading Frames (ORFs) were identified using Prodigal (46), counts generated using Diamond (47), subsequent clustering into Operational Protein Families (OPFs) used mmseq2 (48), and OPF alignment used the KEGG database (49).

**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 classification and regression models via the caret package (50, 51). This was repeated on 100 different 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. The classification models were built to group normal versus adenoma, normal versus carcinoma, and high versus low SCFA concentrations. The regression models were built to classify the SCFA concentrations of acetate, butyrate, and propionate regardless of disease status.

**Statistical analysis workflow.** All analysis was performed using the statistical language R (52). Generally, a Kruskal-Wallis rank sum test with a Dunn's post-hoc test was used to assess differences between individuals without colon tumors, patients with adenomas, and patients with carcinomas. Where appropriate Benjamini-Hochberg was used to correct for multiple comparisons [XXXXXX]. 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, the imputed gene counts of genes encoding enzymes involved in SCFA synthesis was tested.



154 Additionally, metagenomic sequencing counts for important genes involved with SCFA production  
155 were analyzed. From here we analyzed the number of significant positive and negative correlations  
156 between OTU relative abundance and SCFA concentrations in individuals without colon tumors  
157 and patients with adenomas or carcinomas using Spearman's rho. Next, we assessed whether  
158 OTUs alone or OTUs and SCFAs were better able to classify individuals with and without tumors  
159 using Random Forest models. Finally, models to classify the actual SCFA concentration or high/low  
160 SCFA concentration based on the median of each SCFA using 16S rRNA gene sequencing data  
161 was created using the Random Forest algorithm. For all Random Forest models, the assessment  
162 of the most important variables was based on the top 10 features (OTUs or SCFAs) using the mean  
163 decrease in accuracy.

## **Acknowledgements**

The authors thank the Great Lakes-New England Early Detection Research Network for providing the fecal samples that were used in this study. We would thank the University of Michigan Center for Microbial Systems for enabling our short-chain fatty acid analysis. Support for MAS came from the Canadian Institute of Health Research and the National Institutes of Health (UL1TR002240). Support for PDS came from the National Institutes of Health (P30DK034933 and R01CA215574).

## References

1. **Haggard 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. **Fearon ER, Vogelstein B.** 1990. A genetic model for colorectal tumorigenesis. *Cell* **61**:759–767. doi:10.1016/0092-8674(90)90186-i.
4. **Fliss-Isakov N, Zelber-Sagi S, Webb M, Halpern Z, Kariv R.** 2017. Smoking habits are strongly associated with colorectal polyps in a population-based case-control study. *Journal of Clinical Gastroenterology* **1**. doi:10.1097/mcg.0000000000000935.
5. **Lee J, Jeon JY, Meyerhardt JA.** 2015. Diet and lifestyle in survivors of colorectal cancer. *Hematology/Oncology Clinics of North America* **29**:1–27. doi:10.1016/j.hoc.2014.09.005.
6. **Flynn KJ, Baxter NT, Schloss PD.** 2016. Metabolic and community synergy of oral bacteria in colorectal cancer. *mSphere* **1**. doi:10.1128/msphere.00102-16.
7. **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.
8. **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.
9. **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.
10. **Shields CED, Meerbeke SWV, Housseau F, Wang H, Huso DL, Casero RA, O'Hagan HM, Sears CL.** 2016. Reduction of murine colon tumorigenesis driven by

Enterotoxigenic *Bacteroides fragilis* Using cefoxitin treatment. *Journal of Infectious Diseases* **214**:122–129. doi:10.1093/infdis/jiw069.

11. **Tomkovich S, Yang Y, Winglee K, Gauthier J, Mühlbauer M, Sun X, Mohamadzadeh M, Liu X, Martin P, Wang GP, Oswald E, Fodor AA, Jobin C.** 2017. Locoregional effects of microbiota in a preclinical model of colon carcinogenesis. *Cancer Research* **77**:2620–2632. doi:10.1158/0008-5472.can-16-3472.

12. **Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, Baselga J, Liu C, Shivdasani RA, Ogino S, Birren BW, Huttenhower C, Garrett WS, Meyerson M.** 2011. Genomic analysis identifies association of fusobacterium with colorectal carcinoma. *Genome Research* **22**:292–298. doi:10.1101/gr.126573.111.

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

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

15. **Zackular JP, Rogers MAM, Ruffin MT, Schloss PD.** 2014. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prevention Research* **7**:1112–1121. doi:10.1158/1940-6207.capr-14-0129.

16. **Hannigan GD, Duhaime MB, Ruffin MT, Koumpouras CC, Schloss PD.** 2017. Diagnostic potential & the interactive dynamics of the colorectal cancer virome. doi:10.1101/152868.

17. **Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, Herczeg 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 Doeberitz M von, Sobhani I, Bork P.** 2014. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology* **10**:766–766. doi:10.15252/msb.20145645.

18. **Shah MS, DeSantis TZ, Weinmaier T, McMurdie PJ, Cope JL, Altrichter A, Yamal J-M,**

- 220 **Hollister EB**. 2017. Leveraging sequence-based faecal microbial community survey data to identify  
221 a composite biomarker for colorectal cancer. *Gut* **67**:882–891. doi:10.1136/gutjnl-2016-313189.
- 222 19. **Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, Wu X, Shields CED,**  
223 **Hechenbleikner EM, Huso DL, Anders RA, Giardiello FM, Wick EC, Wang H, Wu S, Pardoll**  
224 **DM, Housseau F, Sears CL**. 2018. Patients with familial adenomatous polyposis harbor colonic  
225 biofilms containing tumorigenic bacteria. *Science* **359**:592–597. doi:10.1126/science.aah3648.
- 226 20. **Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell**  
227 **BJ, Abujamel T, Dogan B, Rogers AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku**  
228 **TO, Fodor AA, Jobin C**. 2012. Intestinal inflammation targets cancer-inducing activity of the  
229 microbiota. *Science* **338**:120–123. doi:10.1126/science.1224820.
- 230 21. **Encarnação JC, Pires AS, Amaral RA, Gonçalves TJ, Laranjo M, Casalta-Lopes JE,**  
231 **Gonçalves AC, Sarmiento-Ribeiro AB, Abrantes AM, Botelho MF**. 2018. Butyrate, a dietary  
232 fiber derivative that improves irinotecan effect in colon cancer cells. *The Journal of Nutritional*  
233 *Biochemistry* **56**:183–192. doi:10.1016/j.jnutbio.2018.02.018.
- 234 22. **Verma MS, Fink MJ, Salmon GL, Fornelos N, Ohara TE, Ryu SH, Vlamakis H, Xavier RJ,**  
235 **Stappenbeck TS, Whitesides GM**. 2018. A common mechanism links activities of butyrate in the  
236 colon. *ACS Chemical Biology* **13**:1291–1298. doi:10.1021/acscchembio.8b00073.
- 237 23. **Zheng L, Kelly CJ, Battista KD, Schaefer R, Lanis JM, Alexeev EE, Wang RX, Onyiah**  
238 **JC, Kominsky DJ, Colgan SP**. 2017. Microbial-derived butyrate promotes epithelial barrier  
239 function through IL-10 receptorDependent repression of claudin-2. *The Journal of Immunology*  
240 **199**:2976–2984. doi:10.4049/jimmunol.1700105.
- 241 24. **O’Keefe SJD**. 2016. Diet, microorganisms and their metabolites and colon cancer. *Nature*  
242 *Reviews Gastroenterology & Hepatology* **13**:691–706. doi:10.1038/nrgastro.2016.165.
- 243 25. **Bishehsari F, Engen P, Preite N, Tuncil Y, Naqib A, Shaikh M, Rossi M, Wilber S, Green**  
244 **S, Hamaker B, Khazaie K, Voigt R, Forsyth C, Keshavarzian A**. 2018. Dietary fiber treatment  
245 corrects the composition of gut microbiota, promotes SCFA production, and suppresses colon

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

26. **Tian Y, Xu Q, Sun L, Ye Y, Ji G.** 2018. Short-chain fatty acids administration is protective in colitis-associated colorectal cancer development. *The Journal of Nutritional Biochemistry* **57**:103–109. doi:10.1016/j.jnutbio.2018.03.007.

27. **Vital M, Howe AC, Tiedje JM.** 2014. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5**:e00889–14–e00889–14. doi:10.1128/mbio.00889-14.

28. **Sanna S, Zuydam NR van, Mahajan A, Kurilshikov A, Vila AV, Vösa U, Mujagic Z, Masclee AAM, Jonkers DMAE, Oosting M, Joosten LAB, Netea MG, Franke L, Zhernakova A, Fu J, Wijmenga C, McCarthy MI.** 2019. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature Genetics*. doi:10.1038/s41588-019-0350-x.

29. **Liu S, Li E, Sun Z, Fu D, Duan G, Jiang M, Yu Y, Mei L, Yang P, Tang Y, Zheng P.** 2019. Altered gut microbiota and short chain fatty acids in chinese children with autism spectrum disorder. *Scientific Reports* **9**. doi:10.1038/s41598-018-36430-z.

30. **Meisel M, Mayassi T, Fehlner-Peach H, Koval JC, O'Brien SL, Hinterleitner R, Lesko K, Kim S, Bouziat R, Chen L, Weber CR, Mazmanian SK, Jabri B, Antonopoulos DA.** 2016. Interleukin-15 promotes intestinal dysbiosis with butyrate deficiency associated with increased susceptibility to colitis. *The ISME Journal* **11**:15–30. doi:10.1038/ismej.2016.114.

31. **Ohigashi S, Sudo K, Kobayashi D, Takahashi O, Takahashi T, Asahara T, Nomoto K, Onodera H.** 2013. Changes of the intestinal microbiota, short chain fatty acids, and fecal pH in patients with colorectal cancer. *Digestive Diseases and Sciences* **58**:1717–1726. doi:10.1007/s10620-012-2526-4.

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

33. **Yao Y, Suo T, Andersson R, Cao Y, Wang C, Lu J, Chui E.** 2017. Dietary fibre for the

prevention of recurrent colorectal adenomas and carcinomas. Cochrane Database of Systematic Reviews. doi:10.1002/14651858.cd003430.pub2.

34. **Kunzmann AT, Coleman HG, Huang W-Y, Kitahara CM, Cantwell MM, Berndt SI.** 2015. Dietary fiber intake and risk of colorectal cancer and incident and recurrent adenoma in the prostate, lung, colorectal, and ovarian cancer screening trial. *The American Journal of Clinical Nutrition* **102**:881–890. doi:10.3945/ajcn.115.113282.

35. **Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G, Dahm CC, Overvad K, Olsen A, Tjønneland A, Clavel-Chapelon F, Boutron-Ruault MC, Racine A, Kaaks R, Teucher B, Boeing H, Bergmann MM, Trichopoulou A, Trichopoulos D, Lagiou P, Palli D, Pala V, Panico S, Tumino R, Vineis P, Siersema P, Duijnhoven F van, Peeters PHM, Hjartaker A, Engeset D, González CA, Sánchez M-J, Dorronsoro M, Navarro C, Ardanaz E, Quirós JR, Sonestedt E, Ericson U, Nilsson L, Palmqvist R, Khaw K-T, Wareham N, Key TJ, Crowe FL, Fedirko V, Wark PA, Chuang S-C, Riboli E.** 2012. Dietary fibre intake and risks of cancers of the colon and rectum in the european prospective investigation into cancer and nutrition (EPIC). *PLoS ONE* **7**:e39361. doi:10.1371/journal.pone.0039361.

36. **Gianfredi V, Salvatori T, Villarini M, Moretti M, Nucci D, Realdon S.** 2018. Is dietary fibre truly protective against colon cancer? A systematic review and meta-analysis. *International Journal of Food Sciences and Nutrition* **69**:904–915. doi:10.1080/09637486.2018.1446917.

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

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

39. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF.** 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*

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

300 **40. 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 **41. Quast C, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO**.  
305 2012. The SILVA ribosomal RNA gene database project: Improved data processing and web-based  
306 tools. *Nucleic Acids Research* **41**:D590–D596. doi:10.1093/nar/gks1219.

307 **42. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R**. 2011. UCHIME improves sensitivity  
308 and speed of chimera detection. *Bioinformatics* **27**:2194–2200. doi:10.1093/bioinformatics/btr381.

309 **43. Westcott SL, Schloss PD**. 2017. OptiClust, an improved method for assigning  
310 amplicon-based sequence data to operational taxonomic units. *mSphere* **2**:e00073–17.  
311 doi:10.1128/mspheredirect.00073-17.

312 **44. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC,**  
313 **Burkepile DE, Thurber RLV, Knight R, Beiko RG, Huttenhower C**. 2013. Predictive functional  
314 profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*  
315 **31**:814–821. doi:10.1038/nbt.2676.

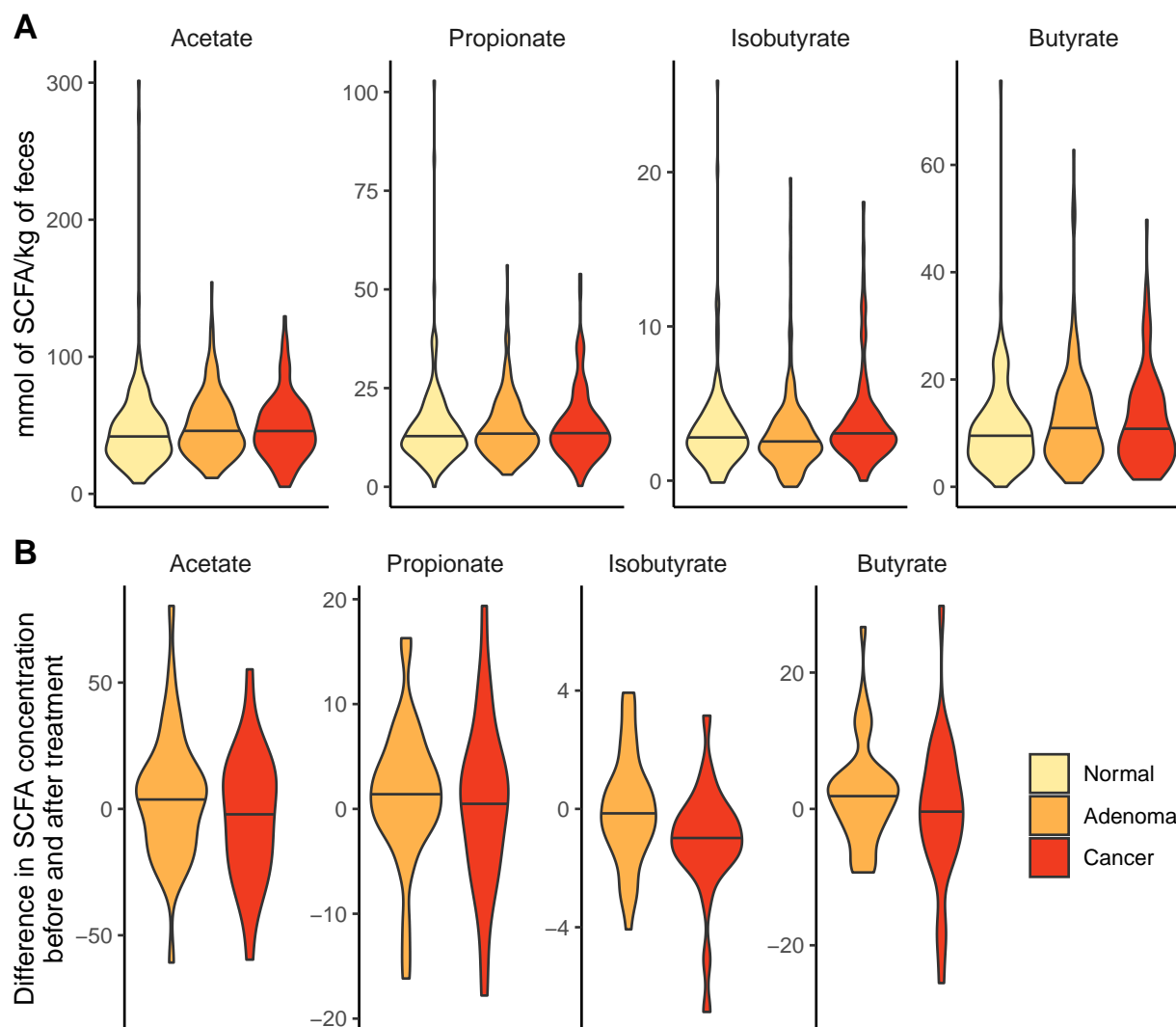
316 **45. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W**. 2015. MEGAHIT: An ultra-fast single-node  
317 solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics*  
318 **31**:1674–1676. doi:10.1093/bioinformatics/btv033.

319 **46. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ**. 2010. Prodigal:  
320 Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119.  
321 doi:10.1186/1471-2105-11-119.

322 **47. Buchfink B, Xie C, Huson DH**. 2014. Fast and sensitive protein alignment using DIAMOND.  
323 *Nature Methods* **12**:59–60. doi:10.1038/nmeth.3176.



- 324 48. **Steinegger M, Söding J.** 2017. MMseqs2 enables sensitive protein sequence searching for  
325 the analysis of massive data sets. *Nature Biotechnology*. doi:10.1038/nbt.3988.
- 326 49. **Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M.** 2015. KEGG as a  
327 reference resource for gene and protein annotation. *Nucleic Acids Research* **44**:D457–D462.  
328 doi:10.1093/nar/gkv1070.
- 329 50. **Liaw A, Wiener M.** 2002. Classification and regression by randomForest. *R News* **2**:18–22.
- 330 51. **Kuhn M.** 2017. Caret: Classification and regression training.
- 331 52. **R Core Team.** 2017. R: A language and environment for statistical computing. R Foundation  
332 for Statistical Computing, Vienna, Austria.



**Figure 1. SCFA concentrations did not vary meaningfully with diagnosis of colonic lesions or with treatment for adenomas or carcinomas.** (A) We measured the concentration of fecal SCFAs from individuals with normal colons (N=172) or those with adenoma (N=198) or carcinomas (N=120) was for isobutyrate. (B) A subset of individuals diagnosed with adenomas (N=41) or carcinomas (N=26) who underwent treatment were resampled a year after the initial sampling; one extreme propionate value (124.4 mmol/kg) was included in the adenoma analysis but censored from the visualization for clarity.

341 **Figure 2. SCFA concentrations do not improve models for diagnosing the presence of**  
342 **adenomas, carcinomas, or all lesions. 16S rRNA gene and metagenomic sequence data do**  
343 **not predict SCFAs concentrations.**

344 **Figure S1. Comparison of training and testing results.**