We appreciate the comments provided by the two reviewers and are grateful for their suggestions and overall positive view of the manuscript. We have done our best to address their comments below. Where appropriate, we have referenced line numbers in the marked up version of the manuscript.

**Reviewer #1 (Comments for the Author):**

**This paper investigates associations between colon cancer and fecal microbiota/SCFAs. This is an important study because intersections between each of these variables are current active areas of research in the field. Overall, I found this paper concise and easy to read. I also found the analyses clear. My primary comments (described below) center on the interpretation of those analyses.**

**Major points:**

**This manuscript deals with the tricky case of how to report non-significant (i.e. p>0.05) results. The approach the authors take is to argue for a lack of association between the measured variables (e.g. fecal SCFAs & tumor status). I have two concerns about this interpretation. First, statisticians are increasingly clarifying that the lack of a significant statistical result does not mean there is no association between two variables (see Amrhein et al., Nature 2019). Second, p-values reflect the odds that a given statistical result was arrived at by chance, assuming the null hypothesis is true. Those odds, in my opinion, are not fundamentally different between 0.05 and say 0.058 or even 0.077 (L66; L70). Additionally, the authors even report an instance where a model associating SCFAs and tumor status has a p-value below the conventional cutoff (p < 0.0001; L79,80).**

**Still, I don't mean to encourage alternative problematic interpretations such as trumpeting significant associations when p-values are marginal. I think instead what might be merited here is nuance. Perhaps rather than rejecting an association, the authors could consider simply reporting the p-values, as well as a measure of effect (e.g. difference of means or odds ratios). Those effect sizes, I suspect, will support the conclusion that even if associations exist, they are weak (which is how the random forest models in L72-85 are presented).**

> We agree that relying heavily on an arbitrary threshold for establishing significance or lack of significance is problematic. This was why we also showed the data in Figure 1. These box plots hopefully make it clear that although there may be a couple of nearly significant P-values, the effect sizes are small and are inconsistent. We have added a sentence to emphasize this point at L75-77. We have considered different ways of reporting the P-values and effect sizes and have concluded that the revised approach is the most suitable given the word length restrictions on an

Observation format manuscript.

For the reviewer's benefit in the paragraph from L63 to L78...

- The "all P>0.148" set of P-values for the cross sectional data included values of ***0.148***, 0.222, and 0.655 for acetate, butyrate, and propionate, respectively. The largest effect sizes for these three comparisons were 2.45, 1.72, and 0.352 mmol/kg
- The "all P>0.058" set of P-values for the follow-up data included values of 0.508 [cancer] and 0.350 [adenoma], 0.916 [cancer] and 0.101 [adenoma], 0.853 [cancer] and ***0.058*** [adenoma] for acetate, butyrate, and propionate, respectively. The corresponding differences between the before and after samples were 3.00 [cancer] and -3.84 [adenoma], 0.312 [cancer] and -1.75 [adenoma], -0.396 [cancer] and -1.68 [adenoma] for acetate, butyrate, and propionate, respectively.
- The "all P>0.16" set of P-values for the relative concentrations of SCFA included values of 0.571, 0.765, and 0.245 for acetate, butyrate, and propionate, respectively from the cross sectional data. They were ***0.164*** [cancer] and 0.322 [adenoma], 0.692 [cancer] and 0.432 [adenoma], and 0.210 [cancer] and 0.365 [adenoma] for acetate, butyrate, and propionate follow-up data, respectively. The largest effect sizes for the cross-sectional data were 0.006, 0.004, and 0.011 for acetate, butyrate, and propionate, respectively. The differences in relative SCFA concentrations in the follow-up data were 0.029 [cancer] and 0.015 [adenoma], -0.006 [cancer] and -0.007 [adenoma], and -0.020 [cancer] and -0.011 [adenoma] for acetate, butyrate, and propionate, respectively.
- The "all P>0.077" set of P-values for the total concentration of SCFA included values of 0.189 for the cross sectional data and 0.833 for the follow-up cancer data and 0.150 for the follow-up adenoma data. In addition the set included values for the per-molecule of carbon basis which were 0.216 for the cross sectional data and 0.916 for the follow-up cancer data and ***0.077*** for the follow-up adenoma data.

Again, aside from cluttering the figure with P-values or the text with extra effect sizes we think it is best to leave it to the reader to see that the effect sizes in Figure 1 are quite small.

Finally, the reviewer points out that "a model associating SCFAs and tumor status has a p-value below the conventional cutoff (p < 0.0001; L79,80)". We point out at that location in the manuscript that the comparison is to a random classification and that the actual AUROC was only 0.54 and 0.55, which was unimpressive. We feel that this was sufficient to explain the P-value.

**I encourage the authors to consider tempering their language regarding the association between fiber intake, SCFAs and protection tumor risk (e.g. L18). I am aware that butyrate and other SCFAs may be over-hyped as a panacea for GI disease. Still, if protection is mediated by SCFA uptake (to me the most reasonable model) then excreted (i.e. fecal SCFAs) should not necessarily be correlated with disease. Additionally, the authors are examining a single point measurement of SCFAs, without consideration of how noisy they may be. It's possible that more frequent longitudinal measurements would reveal a stronger association.**

> We appreciate the reviewer's concern that we are over interpreting negative results. In various places where we provide background to the problem, including L18, we are pointing to epidemiological data which provide a confused picture of the value of fiber and SCFAs in preventing colorectal cancer (CRC). We have gone back through the reporting of our results to emphasize the use of fecal SCFAs, a single time point, and the goal of classification. The most significant edits were made at L31-32, L114, and L118. For example, at L31-33, we rewrote, *"These data indicate that there is no conclusive link between the gut microbiome, SCFAs, and tumor burden."* as *"Our results indicate that the association between fecal SCFAs, the gut microbiome, and tumor burden is weak."*
>
> It should be noted that our manuscript is in the context of diagnosis, not necessarily identifying mechanisms of tumorigenesis that are mediated by microorganisms. We are limited in the amount of discussion we can provide because of word limitations, but it is important to note that it is technically challenging/impossible to obtain mucosal SCFA concentrations since the typical bowel preparation strips away much of the mucosa limiting the ability to sample the SCFAs at the interface. Furthermore, considering our goal is to develop a diagnostic to detect colonic lesions, longitudinal sampling is unlikely to be possible.

**Minor points:**

**L38: This line could be interpreted as saying any bacterium could cause mutations, inflammations, etc. But, to date, only select bacteria are known to cause such events.**

> We have revised the text from "...role as bacteria are known to..." to "...role as some bacteria are known to..."

**L45: "SCFA have ... anti-proliferative activities." This is true but only contextually. Butyrate in particular has proliferative activity in colonocytes. (See Donohoe et al. Mol Cell. The Warburg Effect**

**Dictates the Mehcanism of Butyrate Mediated Histone Acetylation and Cell Proliferation. 2012)**

This sentence has been edited to highlight the contextual nature of this observation.

**L55: Insuring should be ‘ensuring’**

This as been corrected as suggested.

**L65: Is it really surprising that there are no effects 1 year after treatment? Also, what exactly were the treatments?**

We have added text to the sentence to make this more clear at L70 - “None of the *individuals showed signs of recurrence and yet none of the* SCFAs exhibited a significant change with treatment (all P>0.058; Figure 1B)”. We previously showed that the microbiome of those individuals diagnosed with carcinomas returned to a more normal state following treatment. If SCFAs were associated with disease, then the prediction would be that the concentrations would change with a return to health. That we did not see such a change further supports our assertion that there is a tenuous link between fecal SCFA concentrations and health status.

Because of the limits to the number of words available in the *mBio* Observation Format, we do not have the space to recap the treatments. Patients underwent surgical resection and/or chemotherapy. This is more extensively covered in the referenced paper by Sze et al. (2017) [citation 17].

**L92: To be consistent with other statistical tests reported in this manuscript, can the authors provide some measure of how often an R2 of 0.14 would be achieved by chance?**

The R2 values that we report are the median observed values for our 80/20 split hold out permutation test based on 100 permutations. The distribution of observed values are provided in Supplemental Figure 2. If the association were truly random, the R2 value would be near zero. Our point is that the observed R2 values are quite low (e.g. at most 0.14) and although they are significantly different from zero, the models do a poor job of predicting SCFA concentrations. Because we are not comparing models to each other (e.g. an OTU-based model to an OPF-based model), there is no P-value to compare across R2 values.

**L135: The authors use an aqueous extraction and aqueous injection method for measurement of SCFA by HPLC, and do not acidify the aqueous solution to a pH of ~3 before filtration, as has**

**previously been described as essential for accurate quantitation by GC (see Zhao, Nyman, Jönsson. Biomedical Chromatorgraphy (2006)), and remains a standard practice in GC (see Teixeira et al, British Journal of Nutrition (2013)). Why would acidification of the sample to ensure protonation of the SCFA not be necessary for efficient extraction, as the anionic form of the conjugate base is extracted less efficiently from the stool matrix? The literature often does not provide a rationale for the lack of a need for acidification for quantitation of SCFA from feces. If such a rationale has been published, can the authors please provide this reference?**

> We are unaware of a specific reference as to why SCFAs are extracted differently when using GC or HPLC. The protocol we used is based on that used by Venkataraman et al. (2016)[citation 19], which developed their methods based on the existing literature. The approach is effective and has worked well for a variety of fecal samples.

**L173: Can the authors define 'mtry'? Some readers might not be familiar with this term.**

> We have added a sentence at L183-185 to clarify the meaning of this parameter, *"The mtry parameter represents the number of features randomly sampled from the available features at a question point in the classification tree (i.e. called splits of nodes) that, when answered, lead to the greatest improvement in classification"*.

**Fig. 1B: Are the units here the same as in Fig 1A? Can these be specified?**

> They are the same units and we've inserted them with the y-axis label for Figure 1B.

**Reviewer #2 (Comments for the Author):**

**This manuscript by Sze et al., carries out testing on fecal SCFAs, fecal 16S/metagenomics, in order to assess the value of SCFAs to a colorectal cancer classifier. In this aspect, it is a fine contribution and of interest to the readers of mBio. On that basis I like it, with one major caveat:**

**The language of the claims is problematic. It is very hard to prove that something cannot be done. I would have less of a problem with these statements if there were multiple methods or an attempt to quantify the overall amount of information embedded in SCFAs to try to find an upper bound for how predictive they could be. Instead, the authors try one or two tests in each case and declare that SCFAs cannot be used for CRC classification. In other words, the scientific diligence is not commensurate with the broad nature of the claims.**

This reviewer shares the concerns of the first reviewer, which we have hopefully addressed. We agree that it is impossible to prove that something cannot be done - scientifically we can't *prove* anything. We can only propose hypotheses and either reject them or find support for the hypothesis.

We disagree with the reviewer's comment that we haven't pursued multiple methods. First of all, we used a large collection of samples from a cross sectional study and a collection of samples from people with a significant colonic lesion who underwent treatment. Second, our analysis compared SCFA concentrations across diagnosis groups with straightforward statistical tests and we used those data to see if they would improve the ability to diagnose individuals using 16S rRNA gene sequence data, metagenomic data, and imputed metagenomic sequence data. Finally, we attempted to use the 16S rRNA gene sequence data, metagenomic data, and imputed metagenomic sequence data to predict the fecal SCFA concentrations. For these comparisons we tested for an association using the individual concentrations of each SCFA, the relative concentration of each SCFA, the total concentration of SCFAs, and the total amount of SCFA on a per carbon molecule basis. Although, the only classification modeling approach we used was random forest; however, this method and our testing framework are considered to be less sensitive to overfitting, able to incorporate non-linearities in the data, and treat the data in a context-dependent manner. Other methods are available (e.g. logistic regression, SVM, decision tree, etc.), but at least in our hands with similar data do not perform as well as random forest.

**More specifically, I would suggest the following edits to reign this in:**

**1. "Combining SCFA and microbiome data does not improve the ability to diagnose individual as having adenomas or carcinomas" should be changed to "Combining SCFA and microbiome data does not improve the ability to diagnose individual as having adenomas or carcinomas in a random forest machine learning model."**

**2. "Knowledge of microbial community structure does not predict SCFA concentrations in a random forest regression model."**

Both of these section headings have been edited as suggested by the reviewer.