# Response to Reviewers

## Reviewer #1

**Main points**

* Post-treatment samples were collected 188-546 days after lesion treatment, I wonder if that is not too long to evaluate the effect of the treatment. Many other factors may interfere such as an adaptation in their nutritional habits or changes in their medical chronic treatments, for instance.
* I think "Treatment" is not accurate enough, a treatment is a "fixed" factor in a design, almost any intervention can be consider as a treatment. It can be nutritional, pharmacological, surgical, .. I would specify it more, maybe something like "anti-cancer treatment".
* The consumption of antibiotics is not reported in the study. Cancer patients received antibiotics for several reasons, pending the clinical situation. It is hard to decipher what is due to the antibiotic treatment versus the surgical intervention and the chemotherapy. At least, antibiotic regimen should be reported. Statement in lines 220-222 is not enough. Antibiotics can have long-term effects on the gut microbiota composition by completely eradicating sensitive microorganisms.

We appreciate the reviewer's concern as articulated in these three points. It is clearly impossible to control for the myriad differences in the specific details of each patient's treatment regimen and behaviors. Alas, this is human subjects research and the reviewer's definition of a treatment as a "fixed factor" is too stringent. When we refer to "treatment" in the manuscript, we are describing the overall effects of removing the lesion in fairly general terms (e.g. surgery, chemotherapy, radiation). We have added text at L262-271 to reflect this. We would consider antibiotic usage, bowel prep, etc. to be part of these therapies as they are clinically proscribed. It should be noted that, at least in healthy individuals, the microbiota generally recovers within several weeks from these perturbations. Removal of a lesion, surgery, chemotherapy, and radiation would be expected to have the most profound impact on a patient's microbiota.

* If the point is actually to search for microbial biomarkers of recurrence and survival (as stated in the conclusion of the abstract and the manuscript), it would be relevant to provide recurrence and survival data, at least for the 26 patients with carcinoma.

The point of this manuscript was to characterize how the communities change following treatment (L4-6 and L71-74). Our long term goal is to identify biomarkers of recurrence. We are clearly not powered to do that analysis here. Furthermore, because of the difficulties of following up with patients years after the study, the question of recurrence was not possible in the current study.

* Line 161-163. "We focused on the patients with carcinoma and pooled those patients that received chemotherapy with those that received chemotherapy and radiation to improve our statistical power." Radiation influences the gut microbiota by itself (see for instance Cui et al EMBO Mol Med 2017 and Gerassy-Vainberg et al, Gut 2017), so I don't think that the pooling is scientifically justified.

We have repeated the analysis without pooling in the section, "Difficult to identify effects of specific treatments on the change in the microbiota". This did not change our initial findings of no difference between groups.

**Additional points**

* Fig. 1A shows the diversity pre-post for the 3 lesion types. It might help to have the same data from a control cohort (that provided samples at the same time intervals) to evaluate which portion of the change is actually do to the temporal variation in the gut microbiota and which portion relates to treatment. Here, all we can say is that carcinoma treatment and associated events induces greater changes than adenoma and advanced adenoma ones.

We agree with the reviewer that this additional data would be useful to differentiate between temporal variability and those due to treatment specifically. Unfortunately, we do not have the suggested data for a control population sampled along the same time period. Such data would also need to be sequenced alongside the samples and extracted at the same time under the same conditions to minimize both kit biases and sequence batch effects. We have added a section to the discussion to make sure that it is clear Figure 1 can only specify that greater changes occurred between those with CRC versus those with adenoma or advanced adenoma and that it is unclear if all groups changes were greater than what would be expected due to normal temporal variation (L175-177). However, only the adenoma group had a median that aligns closely with within individual temporal variation reported elsewhere (PMID 25061514).

* Lines 226-229. "The most exciting future direction from the current study is the possibility that markers within the microbiota could be used to evaluate the effect of treatment and predict recurrence for those diagnosed with carcinoma. If such an approach is effective, it might be possible to target the microbiota as part of adjuvant therapy." A biomarker does not always play a key role in the disease.

We agree with the reviewers comment about biomarkers which was originally why we choose to use the word "might" in the second stated sentence. In an effort to more clearly emphasize this point we have modified the sentence to include "if the biomarkers identified play a key role in the disease process."

* The study cohort (67 individuals, before and post treatment) is clearly presented. However, the presentation of the training cohort does not seem clear to me. From Table 2, we don't know if all samples are from post or pre-treatment. This cohort is not presented in the mat and meth section. From line 124, I guess they are all post-treatment? When were they collected?

We have added two sentences to help clear this confusion up to the methods section under "Model Building" (L296-298).

* No information is provided on the type of treatment (which chemo for instance).

To help provide more information on this area we have added three new rows to table 1 (Surgery Only, Surgery & Chemotherapy, Surgery, Chemotherapy, & Radiation). We have also added a section titled "Treatment" to the methods to explain in more detail the types of treatment individuals underwent for the removal of lesions.

* Sampling described in Baxter et al Genome Med 2016 seems to suggest that samples were stored on ice for at least 24h before freezing. We cannot exclude that this sampling condition may have impact the gut microbiota composition. I would suggest mentioning at least the sampling condition in the main text of the manuscript.

We have taken the reviewers suggestion and added their suggestion to the methods section (L245-247).

* It might help to better describe Figure 2. What does the light grey line stand for?

It is not immediately clear which grey lines the reviewer is referring to. We were using them as grid lines to aid in lining up the correct OTU and log10MDA value. In light of the confusion we have removed the gray lines and added a brief description of the graph to the figure legend.

* From the results section, it appears that no definitive statement can be made regarding the effect of a specific treatment on the gut microbiota due to the low number of samples, which I think is correct; however such statement are made at the end of the introduction (lines 84-85) and in the abstract (lines 29-33).

We have softened the language at these points in the manuscript (see L123-125 and L231-235).

* Lines 69-71: "If the microbial community drives tumorigenesis then one would hypothesize that treatment to remove a lesion would affect the microbiota and risk of recurrence." I think the sentence should be rephrased, as it is known that treatment to remove lesion affects risk of recurrence.

This sentence has been edited as suggested.

* Line 83 : typo

This sentence has been edited as suggested.

* Line 142 : "showing are marked"

This sentence has been edited as suggested.

## Reviewer #2

* The statement in the abstract "There were large changes to the bacterial communities associated with treatment across the three groups" seems inconsistent with the statement "The change in the abundance of individual OTUs to treatment was not consistent within diagnosis groups (P-value > 0.05)." If I understand the authors claims correctly, the changes within treatment group could be detected with machine learning approaches (random forest) but not inferential statistics. Moreover, Figs. 1B-1C seem to suggest that the PCOA of pre-and post- treatment are superimposable. Given this, it seems a better summary sentence would be that there were larger changes associated with colorectal cancer but much smaller, and more difficult to detect, changes associated with adenomas and advanced adenomas.

This sentence has been edited as suggested.

* "Only patients who had carcinomas experienced a significant decrease in positive probability of having a lesion after treatment (P-value < 0.05), indicating that the microbial milieu of the colon more closely resembled that of a normal colon." This sentence in the abstract is confusing. Do the authors mean that only patients with carcinomas had a significant decrease in the similarity of the microbial community to other patients in a previous cohort that also had colorectal cancer?

We have edited this sentence to clarify that the microbiota of patients who initially had carcinomas changed to more closely resemble those of patients with normal colons (L19-27).

* I find it a little odd that for the adenomas and advanced adenomas groups, there was no difference between pre-treatment and treatment by change in the beta diversity metric (lines 94-96), PERMANOVA (line 102), number of observed OTUs, Shannon evenness, and Shannon diversity (line 104), and OTUs that were significantly different in the pre and post-treatment groups (line 106). It is of concern that differences in adenomas and advanced adenomas were only detectable by machine learning (Random Forest) techniques. Could the authors use permutation tests (randomly permuting the pre- and post- treatment labels) and rerun their Random Forest pipeline over each permutation to ensure that the results for adenomas and advanced adenomas are not due to over-fitting? In particular, I am concerned that the results could reflect the optimization of the mtry parameter described in lines 277-280. Any permutation scheme would need to randomly assign treatment and pre-treatment parameters and then run the entire optimization scheme over the 80/20 splits and through the rest of the pipeline. It has been demonstrated in the literature that optimization over cross-training can still produce substantial overfitting ( e.g. https://doi.org/10.1093/bioinformatics/btr591; "Optimized application of penalized regression methods to diverse genomic data") and given the lack of any other observable differences between adenomas and post-adenomas, it seems likely that that kind of overfitting has occurred here.

We shared the reviewer's concern that the model might be over-fit given the small n associated with each of the groups. To be clear, the mtry hyperparameter was not selected from the test data but instead was chosen from the most common optimized mtry from the 100 different splits of the data and then applied to all the data to generate the final model. The reported AUC values are the average of the AUCs for the held out data. It should also be noted that the methods focused on in the paper cited by the reviewer use linear regression approaches and did not evaluate the non-linear Random Forest algorithm. These methods are considerably more prone to over-fitting than the Random Forest algorithm. Regardless, to help guard against over-fitting we employed both cross-validation and reduction of the model to only use 10 OTUs. This by no means guaranteed that over-fitting did not occur we think that assigning random labels to the Random Forest model would not answer this question. First, Random Forest is not strictly based on a permutation scheme. Second, when randomizing labels to create the model, obtaining a good AUC would only show that Random Forest is good at identifying patterns in the data set. The reason that this would not be a completely accurate assessment of whether our specific model is over-fit with respect to the treatment models is because it still would not allow us to assess whether our group of data is truly representative or not. Permuting the data with random labels after model creation would not necessarily help either since if the model was over-fit then it should be an outlier value across the many permutations. Even if the AUC value falls within the range of random labels that have been generated it still would not mean the model is not over-fit. Since in small n it could be expected that Random Forest would be able to find some set of OTUs or variables by which it could successfully split the data by. However, in light of this possibility of models being over-fit though we have toned down some of the language used to describe the treatment models. Finally, it is not overly surprising that we would only detect differences at the OTU level. Alpha and beta diversity metrics look at the entire community whereas the OTU-based approach described here focuses on the subset of the communities that are indicative of the different groups. Given the large inter-individual variation in the diversity and structure of microbial communities, it is reasonable to expect that inter-individual variation would be larger than the effect size. What the Random Forest algorithm seeks to accomplish is to focus on the conserved parts of the communities to identify features that distinguish the groups.

* I would also think that language in the abstract and conclusion could better emphasize how small the effect was for adenomas and advanced adenomas. Given that "treatment" for adenomas is presumably just removing the adenomas during colonoscopy, it is reasonable to expect a difference in the microbial community? Can the authors more explicitly state what treatment was for the adenomas and pre-adenomas groups?

To improve the clarity, we have added a section in the methods called "Treatment". In this section we explicitly state what type of treatment was performed for the removal of lesions. We also amended table 1 to include a breakdown of different treatments by diagnosis group. We also amended the first sentence in the conclusion of the abstract (L30-32) and the discussion section (L175-177) to make it clear that the effect size for the adenoma groups was small.

* Likewise, in lines 130-133, could the authors perform a permutation test to give p-values for the result of reproducibility with previous cohorts. And in lines 149-152, could a permutation test be used to provide a p-value for statement "The positive probability for the pre and post-treatment samples from patients diagnosed with carcinoma significantly decreased with treatment, suggesting a shift toward a normal microbiota for most individuals".

We think that there may be some confusion on what the range of AUCs reported in the text represent. The ranges reported in the text are the values of AUCs obtained after 100 different 80/20 splits of 10-fold cross validation. With the solid line in Figure S1 representing the AUC for the full set after 10-fold cross-validation. It is possible to do a permutation test but it would simply test whether our full set falls within the range of 100 different 80/20 splits of the same data. With that being said our AUCs for the CRC data set are similar to what has been independently reported previously in the literature (PMID 25432777).

The P-values for the paired Wilcoxon tests have been added to the results section (L152).