The Fecal Microbiome Before and After Treatment for Colorectal Adenoma or Carcinoma

Running Title: Human Microbiome before and after Colorectal Cancer

Marc A Sze[1], Nielson T Baxter[2], Mack T Ruffin IV[3], Mary AM Rogers[2], and 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

2 Department of Internal Medicine, University of Michigan, Ann Arbor, MI

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

Abstract

Background: Colorectal cancer (CRC) is a worldwide health problem and research suggests a correlation between the fecal bacterial microbiome and CRC. Despite this, very little is known about what happens to the microbiome after treatment for an adenoma or carcinoma. This study tested the hypothesis that treatment for adenoma or carcinoma results in changes towards a normal bacterial community. Specifically, we tried to identify components within the community that were different before and after treatment of adenoma, advanced adenoma, and carcinoma.

Results: There was a larger change in the bacterial community in response to treatment for carcinoma versus adenoma (P-value < 0.05) but not carcinoma versus advanced adenoma (P-value > 0.05). There was a trend for increasingly less community similarity, between samples pre- and post-treatment from adenoma to advanced adenoma to carcinoma. Despite this, no difference was found in the relative abundance of any specific OTU before and after treatment for adenoma, advanced adenoma, or carcinoma groups (P-value > 0.05). Using Random Forest models to assess whether changes in post-treatment samples were towards a normal community, only those with carcinoma had a significant decrease in positive probability (P-value < 0.05); providing further evidence that treatment has the greatest effect in those with carcinoma. The adenoma model used a total of 62 OTUs, the SRN model used a total of 61 OTUs, and the carcinoma model used a total of 59 OTUs. A total of 26 OTUs were common to all three models with many classifying to commensal bacteria (e.g. *Lachnospiraceae*, *Bacteroides*, *Anaerostipes*, *Blautia*, and *Dorea*). Both chemotherapy and radiation did not provide any additional changes to the bacterial community in those treated for carcinoma (P-value > 0.05).

Conclusions: Our data partially supports the hypothesis that the microbiome changes after treatment towards a normal community. Individuals with carcinoma had more drastic

2

differences to the overall community than those with adenoma. Common OTUs to all models were overwhelmingly from commensal bacteria, suggesting that these bacteria may be important in initial polyp formation, development of advanced adenoma, and transition to carcinoma.

Background

Colorectal cancer (CRC) is currently the third most common cause of cancer deaths [1,2]. The rate of disease mortality has seen a significant decrease, thanks mainly to improvements in screening [1]. However, despite this improvement there are still approximately 50,000 deaths from the disease per year [2]. Some of these deaths are due to disease recurrence, with current estimates that 20-30% of those undergoing resection having a CRC recurrence [3]. This is important since approximately 30-50% of those who have a recurrence will die of CRC [4]. To reduce the burden on the healthcare system, finding ways to better stratify those at highest risk of recurrence is of great importance.

Recent studies in humans and mouse models have shown that altered membership and structure of the gut microbiome correlate with CRC pathogenesis [5–7]. Further, bacterial communities on the mucosa are altered between normal and tumor tissue [8]. Collectively, these studies provide a tantalizing link between our gut bacteria and CRC and suggest the possibility of using biomarkers to diagnose disease. Indeed, models created with microbiome data and clinical tests, such as Fecal Immunoglobulin Test (FIT), result in good predictions of CRC [9,10]. While these studies show how changed microbial communities or invasion by inflammatory species can impact CRC progression [11], they provide very little information as to whether these communities change and rebound towards normal after successful treatment of adenoma, advanced adenoma, or carcinoma.

Providing an answer to this question is important because it has far reaching implications on both how the bacterial community causes the formation of more polyps [5,6] and the ability to use the microbiome as a predictive screening tool [9,10]. Understanding polyp formation and transition to advanced adenoma and then carcinoma is crucial to being able to understand how to prevent CRC occurrence. Response of the community to treatment is also equally important to predictive models designed for screening purposes since an

5

unresponsive community would provide little additional information for important events, such as recurrence [12].

Using pre- (initial) and post- (follow up) treatment samples we tested the hypothesis that treatment causes detectable changes towards a normal microbiome in those with adenoma, advanced adenoma, and carincoma. First, we assessed differences between samples pre- and post-treatment in adenoma, advanced adenoma, or carcinoma using alpha or beta diversity metrics. Second, we explored whether models built to classify adenoma, advanced adenoma, or carcinoma versus normal were able to identify specific community members that differed between pre- and post-treatment samples. We also used these models to assess whether changes in the community were toward a more normal microbiome. Finally, we assessed both whether surgery for adenomas and advanced adenomas provided larger community changes and whether chemotherapy or radiation provided additional changes to the microbiome over surgical resection. This study aims to provide evidence as to whether the altered CRC microbiome persists or shifts back towards normal after such interventions.

Results

The Bacterial Community: Within our 67-person cohort we tested whether the microbiome in patients with adenoma (n = 22), advanced adenoma (n = 19), or carcinoma (n = 26) had any broad differences between pre- and post-treatment samples. We found that carcinoma patients had a more dissimilar bacterial community between these samples than those with adenoma (P-value < 0.001) [Figure 1A]. Although no significant differences were observed between advanced adenoma and carcinoma there was an increase in the dissimilarity of pre- and post-treatment samples from adenoma (0.55 ± 0.21 (mean ± SD)) to advanced adenoma (0.65 ± 0.25) to carcinoma (0.78 ± 0.15) [Figure 1A]. The bacterial community structure pre- post-treatment was visualized using NMDS for adenoma [Figure 1B] (PERMANOVA > 0.05), advanced adenoma [Figure 1C] (PERMANOVA > 0.05), and carcinoma [Figure 1D] (PERMANOVA < 0.05). Interestingly, when pre- and post-treatment samples were compared, regardless of whether they were adenoma or carcinoma, there was no significant overall difference in beta diversity (PERMANOVA > 0.05). There was no difference between pre- and post-treatment samples when investigating alpha diversity metrics for adenoma, advanced adenoma, or carcinoma [Table S1]. Additionally, there was also no difference in the relative abundance of any specific OTU between these samples for adenoma, advanced adenoma, or carcinoma only [Figure S1].

Adenoma Model: Using normal and adenoma individuals from a separate cohort we created an adenoma model to help assess whether those with adenoma in this cohort changed towards normal. The range of model AUC's from 100 runs of 20 repeated 10-fold cross-validation was 0.62 - 0.72 with the AUC of the model used for classification having an AUC of 0.65 [Figure S2A]. There was a total of 62 OTUs in this model with the vast majority classifying to bacteria typically thought of as commensal [Figure S3A]. There was a significant difference between the actual and predicted disease classification (P-value < 0.05). There was also no significant decrease in the positive probability of adenoma

7

between pre- and post-treatment samples (P-value > 0.05) [Figure 2A].

Advanced Adenoma Model: Using normal and advanced adenoma individuals from a separate cohort we created an advanced adenoma model to help assess whether those with advanced adenoma in this cohort changed towards normal. The range of model AUC's from 100 runs of 20 repeated 10-fold cross-validation was 0.68 - 0.77 with the AUC of the model used for classification having an AUC of 0.73 [Figure S2B]. There was a total of 61 OTUs in the advanced adenoma model [Figure S3B]. Similar to the adenoma model the vast majority of OTUs classified to bacteria typically thought of as commensal. Also similar to the adenoma model there was a significant difference between the actual and predicted disease classification (P-value < 0.05) and no significant decrease in the positive probability of advanced adenoma between pre- and post-treatment samples (P-value > 0.05) [Figure 2B].

Carcinoma Model: Using normal and carcinoma individuals from a separate cohort we created a carcinoma model to help assess whether those with carcinoma in this cohort changed towards normal. The range of model AUC's from 100 runs of 20 repeated 10-fold cross-validation was 0.84 - 0.9 with the AUC of the model used for classification being 0.88 [Figure S2C]. Interestingly, the AUCs improved from adenoma to advanced adenoma to carcinoma [Figure 2]. There was a total of 59 OTUs in the carcinoma model [Figure S3C]. Similar to the adenoma and advanced adenoma models the vast majority of OTUs classified to bacteria typically thought of as commensal but OTUs that also classified to *Fusobacterium*, *Porphyromonas*, and *Parvimonas* were also important for carcinoma classification [Figure S3C]. Also, like the adenoma and advanced adenoma models there was a significant difference between the actual and predicted disease classification (P-value < 0.05). There was a significant decrease in the positive probability of carcinoma between pre- and post-treatment samples (P-value < 0.05) [Figure 2C]; suggesting that the carcinoma samples changed towards normal after treatment, unlike

8

either adenoma or advanced adenoma. The one individual still positive for carcinoma after treatment had an increase in carcinoma positive proability between their pre- and post-treatment sample [Figure 2C].

Common OTUs: Next, to identify which OTUs were important at all stages of disease, we identified common predictive OTUs within the adenoma, advanced adenoma, and carcinoma models. When we compared the three different models with each other there were a total of 26 common OTUs. Some of the most common taxonomic identifications belonged to *Bacteroides*, *Blautia*, *Anaerostipes*, *Lachnospiraceae*, and *Dorea*. The vast majority of the OTUs that were common between these models had classifications to bacteria typically thought of as commensal [Table S2]. The rank of these OTUs, based on median mean decrease in accuracy (MDA), varied higly depending on whether they were in the adenoma, advanced adenoma, or carcinoma model [Figure 3]. As an example, *Bacteroides* (OTU85) was ranked 14th in the carcinoma model but in both the adenoma and advanced adenoma model ranked 35th and 41st respectively [Figure 3].

Treatment Affects on Community: After observing these changes from treatment we assessed the possible confounding impact of chemotherapy or radiation, in the carcinoma group, and surgical resection, in the adenoma group, on the observed results. In the carcinoma group neither chemotherapy nor radiation provided any additional change from initial sample over surgery alone (P-value > 0.05) [Table S3]. For the adenoma group there was a single difference in observed OTUs (sobs) between those that received surgerical resection and those that had regular polyp removal (P-value < 0.05) [Table S4]. For the surgical resection comparison, adenoma and advanced adenoma were combined due to the low number of resection occurances in these two groups. Using a fisher exact test there was no difference in the proportion of those receiving surgerical resection between the adenoma and advanced adenoma groups (P-value > 0.05). This data suggests that the microbiome changes observed in the carcinoma group were mostly a result from surgerical

150 resection and not from chemotherapy or radiation.

Discussion

This study builds upon previous work from numerous labs that have considered both how the bacterial community between those with and without CRC differs and how it might be used as an early screening tool [9,10,13–15]. Here we show that the microbiome changes towards normal after treatment for carcinoma and that chemotherapy and radiation did not provide any additional change. Although some of the important OTUs classified to genera from bacteria considered the "usual suspects" (e.g. *Fusobacterium*, *Porphyromonas*, and *Parvimonas*) many did not. The majority of important OTUs had taxonomic classifications for resident gut microbes and were common for the adenoma, advanced adenoma, and carcinoma models. This suggests that members within the commensal community may be the first that change during CRC pathogenesis. These subtle changes, in turn, could be the first step in allowing more inflammatory bacteria to gain a foothold within the colon [11].

Unlike previous studies on the microbiome and CRC, ours focuses on comparing the microbiome during recovery from treatment in adenoma, advanced adenoma, and carcinoma groups. Although there were differences for genera associated with specific bacterium linked with CRC [Figure S3], most of the important OTUs taxonomically classified to commensal bacteria [Figure S3]. Although these changes may be subtle, due to the lack of significant differences in the bacterial community pre- and post-treatment in adenoma and advanced adenoma [Figure 1]; they support the hypothesis that the first members of the community to change and potentially stay changed even after treatment are those that are commensal bacteria.

Many of the common OTUs that we identified taxonomically classified to potential butyrate producers (e.g. *Clostridiales*, *Roseburia*, and *Anaerostipes*) [Table S2]. Other OTUs classified to bacteria that are inhibited by polyphenols (e.g. *Bacteroides*). Both butyrate and polyphenols are thought to be protective against cancer, in part by reducing

11

inflammation [16]. These protective compounds are derived from the breakdown of fiber, fruits, and vegetables by resident gut microbes. One example of this potential diet-microbiome-inflammation-polyp axis is that *Bacteroides*, which was highly prevalent in our models, are known to be increased in those with high non-meat based protein consumption [17]. High protein consumption in general has been linked with an increased CRC risk [18]. Conversely, *Bacteroides* are inhibited by polyphenols which are derived from fruits and vegetables [19]. Our data fits with the hypothesis that the microbial metabolites from breakdown products within our own diet could not only help to shape the existing community but also have an effect on CRC risk and disease progression. Within this context the commensal community may be an important modifiable risk factor for monitoring and preventing CRC recurrence after treatment.

A limitation in our study was that there was a significant difference in the time elapsed in the collection of the post-treatment sample between adenoma or advanced adenoma versus carcinoma (P-value < 0.05), with time passed being less for adenoma (255 ± 42 days) and advanced adenoma (250 ± 41) than carcinoma (351 ± 102). These results would indicate that some of the differences observed between the carcinoma and adenoma groups could be due to differences in collection time. Specifically, it could confound the observation that carcinomas changed more than adenomas [Figure 1]. However, there are two reasons that this may not be the case. First, the advanced adenoma group had a higher dissimilarity than adenoma but lower dissimilarity then carcinoma and the collection time to their follow up samples was less than the adenoma group. Second, this confounding would not affect the observations where models were used since they were built using a different cohort [Figures 2 & S2-S3].

Another limitation of this study was that it drew heavily from those with Caucasian ancestry making it possible that the observations may not be representative of those with either Asian or African ancestry. Although our training and test set are relatively large we still run

the risk of over-fitting or having a model that may not be representative of other populations. We have done our best to safeguard against this by not only running 10-fold cross validation but also having over 100 different 80/20 splits to try and mimic the type of variation that might be expected to occur.

Building off of our results, an area for future research, is that within our study we do not know whether individuals who were still classified as positive by the carcinoma model eventually had a subsequent CRC recurrence. This information would help to strengthen the case for this model keeping numerous individuals above the cutoff threshold even though at follow up they were diagnosed as no longer having carcinoma. It would also provide additional evidence that the microbiome could be used as a risk stratification tool in monitoring recurrence risk and whether different interventions could potentially change this community and lower the probability of future reccurence.

Despite the stated shortcomings our findings add to the existing scientific knowledge on CRC and the microbiome: That there is a measurable difference in the bacterial community after adenoma, advanced adenoma, or carcinoma treatment. Further, the ability for machine learning algorithms to take OTU data and successfully lower positive probability of carcinoma after treatment provides evidence that there are specific signatures, attributable to both inflammatory and resident commensal organisms, associated with treatment. Our data provides evidence that commensal bacteria may be important in the development of polyps, potential transition of advanced adenoma to carcinoma, and recovery of the microbiome in CRC following treatment.

Methods

Study Design and Patient Sampling: Sampling and design have been previously reported in Baxter, et al [9]. Briefly, study exclusion involved those who had already undergone surgery, radiation, or chemotherapy, had colorectal cancer before a baseline fecal sample could be obtained, had IBD, a known hereditary non-polyposis colorectal cancer, or familial adenomatous polyposis. Samples used to build the models for prediction were collected either prior to a colonoscopy or between 1 - 2 weeks after. The bacterial community has been shown to normalize back to a pre-colonoscopy community within this time period [20]. Our training cohort consisted of a total of 423 individuals [Table 1]. Our study cohort consisted of 67 individuals with an initial sample as described and a follow up sample obtained between 188 - 546 days after treatment of lesion [Table 2]. This study was approved by the University of Michigan Institutional Review Board. All study participants provided informed consent and the study itself conformed to the guidelines set out by the Helsinki Declaration.

16S rRNA Gene Sequencing: Sequencing was completed as described by Kozich, et al. [21]. DNA extraction used the 96-well Soil DNA isolation kit (MO BIO Laboratories) and an epMotion 5075 automated pipetting system (Eppendorf). The V4 variable region was amplified and the resulting product was split between three sequencing runs with normal, adenoma, and carcinoma evenly represented on each run. Each group was randomly assigned to avoid biases based on sample collection location. The pre- and post-treatment samples were sequenced on the same run.

Sequence Processing: The mothur software package (v1.37.5) was used to process the 16S rRNA gene sequences and has been previously described [21]. The general workflow using mothur was: Paired-end reads were first merged into contigs, quality filtered, aligned to the SILVA database, screened for chimeras, classified with a naive Bayesian classifier

14

using the Ribosomal Database Project (RDP), and clustered into Operational Taxonomic Units (OTUs) using a 97% similarity cutoff with an average neighbor clustering algorithm. The number of sequences for each sample was rarefied to 10523 to minimize uneven sampling.

Model Building: The Random Forest [22] algorithm was used to create the model used to create the three models used. The adenoma model classified normal versus adenoma, advanced adenoma was normal versus advanced adenoma, and carcinoma was normal versus carcinoma. The toal number of individuals in this data set was 423 individuals. There were a total of 239 individuals in the adenoma model, 262 individuals in the advanced adenoma model, and 266 indivdiuals in the carcinoma model [Table 1]. Each model was then applied to our 67-person cohort testing prediction of adenoma intial (adenoma n = 22) versus adenoma follow up (adenoma n = 0), advanced adenoma initial (advanced adenoma n = 19) versus advanced adenoma follow up (advanced adenoma n = 0), carcinoma initial (carinoma n = 26) versus carcinoma follow up (carcinoma n = 1).

The model included only OTU data obtained from 16S rRNA sequencing. Non-binary data was checked for near zero variance and OTUs that had near zero variance were removed. This pre-processing was performed with the R package caret (v6.0.73). Optimization of the mtry hyper-parameter involved making 100 different 80/20 (train/test) splits of the data where normal and adenoma, normal and advanced adenoma, or normal and carcinoma were represented in the same proportion within both the whole data set and the 80/20 split. For each of the different splits, 20 repeated 10-fold cross validation was performed on the 80% component to optimize the mtry hyper-parameter by maximizing the AUC (Area Under the Curve of the Receiver Operator Characteristic). The resulting model was then tested on the hold out data obtained from the 20% component. All three models had an optimized mtry of 2.

Assessment of the most important OTUs to the model involved counting the number of

15

times an OTU was present in the top 10% of mean decrease in accuracy (MDA) for each of the 100 different splits run. This was then followed with filtering of this list to variables that were only present in more than 50% of these 100 runs. The final collated list of variables was then run through the mtry optimization again. Once the ideal mtry was found the entire sample set specific to normal versus adenoma, normal versus advanced adenoma, or normal versus carcinoma was used to create the final Random Forest model on which classifications on the 67-person cohort was completed. For all three models the final optimized mtry was 2

The default cutoff of 0.5 was used as the threshold to classify individuals as positive or negative for lesion. The hyper-parameter, mtry, defines the number of variables to investigate at each split before a new division of the data was created with the Random Forest model.

Statistical Analysis: The R software package (v3.3.2) was used for all statistical analysis. Comparisons between bacterial community structure utilized PERMANOVA [23] in the vegan package (v2.4.1). Comparisons between probabilities as well as overall OTU differences between pre- and post-treatment samples utilized a paired Wilcoxson ranked sum test. Where multiple comparison testing was appropriate, a Benjamini-Hochberg (BH) correction was applied [24] and a corrected P-value of less than 0.05 was considered significant. Unless otherwise stated the P-values reported are those that were BH corrected. Model rank importance was determined by obtining the median MDA from the 100 20 repeated 10-fold cross validation and then ranking from largest to smallest MDA.

Analysis Overview: We first tested whether there were any differences between pre- and post-treatment samples in alpha and beta diversity based on adenoma, SRN, or carcinoma. We then tested all OTUs for differences between pre- and post-treatment samples. We next used our specific models for adenoma, SRN, and carcinoma to test classification accuracy, response towards a normal microbiome, and common OTUs used across models. Finally,

16

for the adenoma group differences between those that received surgery or not was tested while for the carcinoma group differences between those receiving chemotherapy and radiation was tested.

Reproducible Methods: A detailed and reproducible description of how the data were processed and analyzed can be found at https://github.com/SchlossLab/Sze_followUps_ 2017. Raw sequences have been deposited into the NCBI Sequence Read Archive (SRP062005 and SRP096978) and the necessary metadata can be found at https://www. ncbi.nlm.nih.gov/Traces/study/ and searching the respective SRA study accession.

Figure 1: General Differences between Adenoma, Advanced Adenoma, and Carcinoma Groups After Treatment. A) A significant difference was found between the adenoma and carcinoma group for thetayc (P-value = NULL). B) NMDS of the pre- and post-treatment samples for the adenoma group. C) NMDS of the pre- and post-treatment samples for the advanced adenoma group. D) NMDS of the pre- and post-treatment samples for the carcinoma group.

Figure 2: Treatment Response Based on Models Built for Adenoma, SRN, or Carcinoma. A) Positive probability change from initial to follow up sample in those with adenoma. B) Positive probability change from initial to follow up sample in those with SRN. C) Positive probability change from initial to follow up sample in those with carcinoma..

Figure 3: Common OTUs Model Rank Based on Median MDA. For each common OTU the lowest taxonomic identification and importance rank for each model run is shown.

Table 1: Demographic Data of Training Cohort

Table 2: Demographic Data of Pre and Post Treatment Cohort

Figure S1: Distribution of P-values from Paired Wilcoxson Analysis of All OTUs Before and After Treatment

Figure S2: ROC Curves of the Adenoma, Advanced Adenoma, and Carcinoma Models. A) Adenoma ROC curve: The light green shaded areas represent the range of values of a 100 different 80/20 splits of the test set data and the dark green line represents the model using 100% of the data set and what was used for subsequent classification. B) Advanced Adenoma ROC curve: The light yellow shaded areas represent the range of values of a 100 different 80/20 splits of the test set data and the dark yellow line represents the model using 100% of the data set and what was used for subsequent classification. C) Carcinoma ROC curve: The light red shaded areas represent the range of values of a 100 different 80/20 splits of the test set data and the dark red line represents the model using 100% of the data set and what was used for subsequent classification.

Figure S3: Summary of Important Variables for the Adenoma, Advanced Adenoma, and Carcinoma Models. A) MDA of the most important variables in the adenoma model. The dark green point represents the mean and the lighter green points are the value of each of the 100 different runs. B) Summary of Important Variables in the advanced adenoma model. MDA of the most important variables in the SRN model. The dark yellow point represents the mean and the lighter yellow points are the value of each of the 100 different runs. C) MDA of the most important variables in the carcinoma model. The dark red point represents the mean and the lighter red points are the value of each of the 100 different runs.

Declarations

Ethics approval and consent to participate

The University of Michigan Institutional Review Board approved this study, and all subjects provided informed consent. This study conformed to the guidelines of the Helsinki Declaration.

Consent for publication

Not applicable.

Availability of data and material

A detailed and reproducible description of how the data were processed and analyzed can be found at https://github.com/SchlossLab/Sze_followUps_2017. Raw sequences have been deposited into the NCBI Sequence Read Archive (SRP062005 and SRP096978) and the necessary metadata can be found at https://www.ncbi.nlm.nih.gov/Traces/study/ and searching the respective SRA study accession.

Competing Interests

All authors declare that they do not have any relevant competing interests to report.

Authors' contributions

All authors were involved in the conception and design of the study. MAS analyzed the data. NTB processed samples and analyzed the data. All authors interpreted the data. MAS and PDS wrote the manuscript. All authors reviewed and revised the manuscript. All authors read and approved the final manuscript.

22

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA: a cancer journal for clinicians. 2010;60:277–300.

2. Haggar FA, Boushey RP. Colorectal cancer epidemiology: Incidence, mortality, survival, and risk factors. Clinics in Colon and Rectal Surgery. 2009;22:191–7.

3. Hellinger MD, Santiago CA. Reoperation for recurrent colorectal cancer. Clinics in Colon and Rectal Surgery. 2006;19:228–36.

4. Ryuk JP, Choi G-S, Park JS, Kim HJ, Park SY, Yoon GS, et al. Predictive factors and the prognosis of recurrence of colorectal cancer within 2 years after curative resection. Annals of Surgical Treatment and Research. 2014;86:143–51.

5. Zackular JP, Baxter NT, Chen GY, Schloss PD. Manipulation of the Gut Microbiota Reveals Role in Colon Tumorigenesis. mSphere. 2016;1.

6. Arthur JC, Gharaibeh RZ, Mühlbauer M, Perez-Chanona E, Uronis JM, McCafferty J, et al. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. Nature Communications. 2014;5:4724.

7. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, et al. The gut microbiome modulates colon tumorigenesis. mBio. 2013;4:e00692–00613.

8. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. Proceedings of the National Academy of Sciences of the United States of America. 2014;111:18321–6.

9. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Medicine.

395 2016;8:37.

396 10. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of
397 fecal microbiota for early-stage detection of colorectal cancer. Molecular Systems Biology.
398 2014;10:766.

399 11. Flynn KJ, Baxter NT, Schloss PD. Metabolic and Community Synergy of Oral Bacteria
400 in Colorectal Cancer. mSphere. 2016;1.

401 12. Hassan C, Repici A, Sharma P, Correale L, Zullo A, Bretthauer M, et al. Efficacy
402 and safety of endoscopic resection of large colorectal polyps: A systematic review and
403 meta-analysis. Gut. 2016;65:806–20.

404 13. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of
405 faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer.
406 Gut. 2017;66:70–8.

407 14. Zackular JP, Rogers MAM, Ruffin MT, Schloss PD. The human gut microbiome as
408 a screening tool for colorectal cancer. Cancer Prevention Research (Philadelphia, Pa.).
409 2014;7:1112–21.

410 15. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, et al.
411 Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome. 2013;1:16.

412 16. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal
413 cancer. Nature Reviews Microbiology [Internet]. 2014 [cited 2017 Feb 14];12:661–72.
414 Available from: http://www.nature.com/doifinder/10.1038/nrmicro3344

415 17. Zhu Y, Lin X, Li H, Li Y, Shi X, Zhao F, et al. Intake of Meat Proteins Substantially
416 Increased the Relative Abundance of Genus Lactobacillus in Rat Feces. PloS One.

2016;11:e0152678.

18. Mu C, Yang Y, Luo Z, Guan L, Zhu W. The Colonic Microbiome and Epithelial Transcriptome Are Altered in Rats Fed a High-Protein Diet Compared with a Normal-Protein Diet. The Journal of Nutrition. 2016;146:474–83.

19. Ozdal T, Sela DA, Xiao J, Boyacioglu D, Chen F, Capanoglu E. The Reciprocal Interactions between Polyphenols and Gut Microbiota and Effects on Bioaccessibility. Nutrients [Internet]. 2016 [cited 2017 Feb 14];8:78. Available from: http://www.mdpi.com/2072-6643/8/2/78

20. O'Brien CL, Allison GE, Grimpen F, Pavli P. Impact of colonoscopy bowel preparation on intestinal microbiota. PloS One. 2013;8:e62815.

21. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology. 2013;79:5112–20.

22. Breiman L. Random Forests. Machine Learning [Internet]. 2001 [cited 2013 Feb 7];45:5–32. Available from: http://link.springer.com/article/10.1023/A%3A1010933404324 http://link.springer.com/article/10.1023%2FA%3A1010933404324?LI=true

23. Anderson MJ, Walsh DCI. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecological Monographs [Internet]. 2013 [cited 2017 Jan 5];83:557–74. Available from: http://doi.wiley.com/10.1890/12-2010.1

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