

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

Running Title: Human Microbiome before and after Colorectal Cancer

Marc A Sze¹, Nielson T Baxter², Mack T Ruffin IV³, Mary AM Rogers², 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, M.

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

Abstract

Background. Colorectal cancer (CRC) is a worldwide health problem. Despite growing evidence that members of the gut microbiota can drive tumorigenesis, 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 alters the abundance of bacterial populations associated with with disease to those associated with a normal colon. We tested this hypothesis by sequencing the 16S rRNA genes in the feces of 67 individuals before and after treatment for adenoma (N=22), advanced adenoma (N=19), and carcinoma (N=26).

Results. There were large changes the the bacterial communities associated with treatment across the three groups. The communities from patients with carcinomas changed significantly more than those with adenoma following treatment (P-value=5.4e-05); however, there was not a significant difference between those with advanced adenoma and those with adenoma or carcinoma (P-value>0.05). Although treatment brought about large intrapersonal changes, the change in the abundance of individual OTUs to treatment was not consistent within diagnosis groups (P-value > 0.05). Because the distribution of OTUs across patients and diagnosis groups was patchy, we used the Random Forest machine learning algorithm to identify groups of OTUs that allowed us to successfully distinguish between pre- and post-treatment samples for each of the diagnosis groups. However, across the three models, there was little overlap between the OTUs that were indicative of treatment. Next, we used a larger cohort that contained individuals with normal colons and those with adenomas, advanced adenomas, and carcinomas to determine whether individuals who underwent a treatment were more likely to have OTUs associated with normal colons. We again built Random Forest models and measured the change in the positive probability of having one of the three diagnoses. Although we could clearly differentiate pre- and post-treatment communities from the three diagnosis groups, only

27 those patients that initially had carcinomas experienced a significant decrease in positive
28 probability of having a carcinoma ($P\text{-value} < 0.05$). Finally, tested whether the type of
29 treatment impacted the microbiota of those diagnosed with carcinomas and were unable
30 to detect any significant differences in characteristics of these communities between
31 individuals treated with chemotherapy and those treated with chemotherapy and radiation
32 ($P\text{-value} > 0.05$).

33 **Conclusions.** Further exploration of the relationship between diagnosis, treatment, and
34 the impact on the microbiome will yield improvements in disease management.

³⁵ **Keywords**

³⁶ bacterial microbiome; colorectal cancer; polyps; FIT; post-surgery; risk factor.

Background

Colorectal cancer (CRC) is the third most common cause of cancer deaths in the United States [1,2]. Disease mortality has significantly decreased, thanks to improvements in screening [1]. Despite this improvement, there are still approximately 50,000 deaths from the disease per year [2]. Current estimates indicate that 20-30% of those who undergo treatment will experience recurrence and 30-50% of those patients will die [3,4]. Identification of methods to assess patients' risk of recurrence is of great importance to reduce mortality and healthcare costs.

There is growing evidence that the gut microbiota is involved in the progression of CRC. Mouse-based studies have identified populations of *Bacteroides fragilis*, *Escherichia coli*, and *Fusobacterium nucleatum* that alter disease progression [5–9]. Furthermore, studies that alter the structure of the microbiota through the use of antibiotics or inoculation of germ free mice with human feces has shown that varying community compositions can result in varied tumor burden [10–12]. Collectively, these studies hypothesize that the microbiota is altering the amount of inflammation in the colon and with it the rate of tumorigenesis [13].

Building on this evidence, several human-based studies have identified unique signatures of colonic lesions [14–19]. One line of research has identified community-level differences between those bacteria that are found on and adjacent to colonic lesions and have supported a role for *Bacteroides fragilis*, *Escherichia coli*, and *Fusobacterium nucleatum* in tumorigenesis [20–22]. Others have proposed feces-based biomarkers that could be used to diagnose the presence of colonic adenomas and carcinomas [23–25]. These studies have associated *Fusobacterium nucleatum* and other oral pathogens with lesions. They have also noted that the loss of bacteria generally thought to produce short chain fatty acids, which can suppress inflammation, is associated with colonic lesions. Gut bacteria have a role in tumorigenesis and there is promise that these populations may be useful

biomarkers for aiding in the early detection of disease.

Despite advances in understanding the role between the gut microbiota and colonic tumorigenesis, we still do not understand how treatments including resection, chemotherapy, and/or radiation affect the composition of the gut microbiota. If the community drives tumorigenesis then one would hypothesize that treatment to remove a lesion would affect the microbiota and risk of recurrence. To test this hypothesis, we addressed two related questions: Does treatment affect the colonic microbiota in a predictable manner? If so, does the treatment alter the community to more closely resemble that of individuals with normal colons?

We answered these questions by sequencing the V4 region of 16S rRNA genes amplified from fecal samples of individuals with adenoma, advanced adenoma, and carcinomas pre- and post-treatment. We used both classical community analysis to compare the alpha and beta-diversity of communities pre- and post treatment. Next, we generated Random Forest models to identify bacterial populations that were indicative of treatment for each diagnosis group. Finally, we measured the predictive probability of diagnosis models to determine whether treatment would alter the probability of being diagnosed as having a normal colon or the colon of the original diagnosis group. Ultimately, we found that treatment alters the composition of the gut microbiota; however, we only observed a shift from diseased to normal colon for those initially diagnosed with carcinomas. Understanding how the community responds to these treatments could be a valuable tool for identifying biomarkers to quantify the risk of recurrence.

Results

Treatment alters the bacterial community structure of patients diagnosed with

colonic lesions. Within our 67-person cohort we tested whether the microbiome of

patients with adenoma (N=22), advanced adenoma (N=19), or carcinoma (N=26) had

any broad differences between pre- and post-treatment samples (Table 1). The structure

of the microbial communities of the pre- and post-treatment samples differed [Figure

1A]. We found that the communities obtained pre- and post-treatment among the patients

with carcinomas changed significantly more than those patients with adenoma (P-value

< 0.001). There were no significant differences in the amount of change observed

between the patients with adenoma and advanced adenoma or between the patients

with advanced adenoma and carcinoma ($P>0.05$). Next, we tested whether there was

a consistent direction in the change in the community structure between the pre and

post-treatment samples for each of the diagnosis groups [Figure 1BCD]. We only observed

a consistent shift in community structure for the patients with carcinoma when using

a PERMANOVA test (adenoma P-value 0.999, advanced adenoma P-value 0.945, and

carcinoma P-value 0.005). Finally, we measured the number of observed OTUs, Shannon

evenness, and Shannon diversity in the pre and post-treatment samples and did not

observe a significant change for any of the diagnosis groups (P-value >0.05) [Table S1].

The effects of treatment are not consistent across treatment groups. We used two

approaches to identify those bacterial populations that change between the two samples

for each diagnosis group. First, we sought to identify individual OTUs that could account for

the change in overall community structure. However, using a paired Wilcoxon test we were

unable to identify any OTUs that were significantly different in the pre and post-treatment

groups [Figure S1]. It is likely that high inter-individual variation and the patchy distribution

of OTUs across individuals limited the statistical power of the test. To overcome these

problems we developed Random Forest models to identify collections of OTUs that would

allow us to differentiate between pre and post-treatment samples from each of the diagnosis groups. Because of the relatively small number of subject in each group, we restricted our models to only incorporate 10 OTUs. Despite this restriction, the models performed well (adenoma AUC=0.69 - 0.92, advanced adenoma AUC=0.8 - 1, carcinoma AUC=0.82 - 0.98). Interestingly, the 10 OTUs that were used for each model had little overlap with each other [Figure 2]. These results support the earlier community-wide analysis where we observed that the treatment had an impact on the overall community structure; however, the effect of treatment was not consistent across patients and diagnosis groups.

Post-treatment samples from patients with carcinoma share Next, we determined whether treatment changed the microbiota in a way that the post-treatment communities resembled that of patients with normal colons. To test this, we used an expanded cohort of 423 individuals that were diagnosed under the same protocol as having normal colons or colons with adenoma, advanced adenoma, or carcinoma (Table 2). We then constructed Random Forest models to classify the pre and post-treatment samples as having their original diagnosis or having a normal colon. The models performed well (adenoma AUC=0.62 - 0.72, advanced adenoma AUC=0.68 - 0.77, carcinoma AUC=0.84 - 0.9; Figure S2). The OTUs that were most important in the models for differentiating between patients with normal colons and those with adenomas or advanced adenomas were dominated by populations commonly considered as commensals (e.g. *Faecalibacterium*, *Lachnospiraceae*, *Bacteroides*, *Dorea*, *Anaerostipes*, and *Roseburia*) [Figure S3]. Although many of these OTUs were also included in the model differentiating between patients with normal colons and those with carcinoma, this model also included OTUs affiliated with populations that have previously been associated with carcinoma (*Fusobacterium*, *Porphyromonas*, *Parvimonas*) [Figure S3]. Finally, we applied these three models to the pre and post-treatment samples for each diagnosis group and quantified the change in the positive probability of the model. A decrease in the positive probability would indicate that the colon more closely resembled that of a patient with a normal colon. There was

not a significant change in the positive probability for the adenoma or advanced adenoma groups [Figure 3]. The positive probability for the pre and post-treatment samples from patients diagnosed with carcinoma significantly decreased. In fact, only 6 of the 26 patients (23.08%) that were initially diagnosed with a carcinoma had a higher positive probability after treatment. One of those was re-diagnosed with carcinoma on the follow up visit. These results indicate that although there are changes in the microbiota associated with treatment, those experienced by the patients in the carcinoma group were the only ones where the change was directed towards what would be found in a normal colon

Difficult to identify effects of specific treatments on the change in the microbiota.

The type of treatment that the patients received varied across diagnosis groups. Those with adenomas and advanced adenomas received surgical resection (adenoma, N=4; advanced adenoma, N=4) or polyp removal during colonoscopy (adenoma, N=18; advanced adenoma, N=15) and those with carcinomas received surgical resection (N=12), surgical resection with chemotherapy (N=9), and surgical resection with chemotherapy and radiation (N=5). 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. We did not observe a significant difference in the effect of these treatments on the number of observed OTUs, Shannon diversity, or Shannon evenness (P-value > 0.05). Furthermore, there was not a significant difference in the effect of the treatments on the amount of change in the community structure (P= 1). Finally, the change in the positive probability was not significantly different between the two treatment groups (P=1). Due to the relatively small number of samples in each treatment group, it is difficult to make a definitive statement regarding the specific type of treatment on the amount of change in the structure of the microbiota

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 [23–27]. 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 [13].

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

inflammation [28]. 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 [29]. High protein consumption in general has been linked with an increased CRC risk [30]. Conversely, *Bacteroides* are inhibited by polyphenols which are derived from fruits and vegetables [31]. 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 than 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 recurrence.

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 [23]. 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 [32]. 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. [33]. 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 [33]. 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

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 [34] 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 total 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 individuals in the carcinoma model [Table 1]. Each model was then applied to our 67-person cohort testing prediction of adenoma initial (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 (carcinoma 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

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 [35] 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 Wilcoxon ranked sum test. Where multiple comparison testing was appropriate, a Benjamini-Hochberg (BH) correction was applied [36] 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 obtaining 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,

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/Size_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.

317 ****Figure 1: General Differences between Adenoma, Advanced Adenoma, and Carcinoma**
318 **Groups After Treatment.*** A) A significant difference was found between the adenoma and
319 carcinoma group for thetacy (P-value = NULL). Solid black points represent the median
320 value for each diagnosis group. B) NMDS of the pre- and post-treatment samples for
321 the adenoma group. C) NMDS of the pre- and post-treatment samples for the advanced
322 adenoma group. D) NMDS of the pre- and post-treatment samples for the carcinoma
323 group.

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

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

330 **Table 1: Demographic Data of Training Cohort**

331 **Table 2: Demographic Data of Pre and Post Treatment Cohort**

Figure S1: Distribution of P-values from Paired Wilcoxon 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/Size_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.

Funding

This study was supported by funding from the National Institutes of Health to P. Schloss (R01GM099514, P30DK034933) and to the Early Detection Research Network (U01CA86400).

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.

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 also like to thank Amanda Elmore for reviewing and correcting code error and providing feedback on manuscript drafts. We would also like to thank Nicholas Lesniak for providing feedback on manuscript drafts.

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA: a cancer journal for clinicians*. 2010;60:277–300.
2. Haggard 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. Goodwin AC, Destefano Shields CE, Wu S, Huso DL, Wu X, Murray-Stewart TR, et al. Polyamine catabolism contributes to enterotoxigenic *Bacteroides fragilis*-induced colon tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108:15354–9.
6. Abed J, Emgård JEM, Zamir G, Faroja M, Almogy G, Grenov A, et al. Fap2 Mediates *Fusobacterium nucleatum* Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host & Microbe*. 2016;20:215–25.
7. 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.
8. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the

tumor-immune microenvironment. *Cell Host & Microbe*. 2013;14:207–15.

9. Wu S, Rhee K-J, Albesiano E, Rabizadeh S, Wu X, Yen H-R, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nature Medicine*. 2009;15:1016–22.

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

11. 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.

12. Baxter NT, Zackular JP, Chen GY, Schloss PD. Structure of the gut microbiome following colonization with human feces determines colonic tumor burden. *Microbiome*. 2014;2:20.

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

14. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *The ISME journal*. 2012;6:320–9.

15. Chen H-M, Yu Y-N, Wang J-L, Lin Y-W, Kong X, Yang C-Q, et al. Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. *The American Journal of Clinical Nutrition*. 2013;97:1044–52.

16. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PloS One*. 2012;7:e39743.

17. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, et al. Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas.

427 Gut Microbes. 2010;1:138–47.

428 18. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic
429 analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome*
430 *Research*. 2012;22:292–8.

431 19. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome
432 development along the colorectal adenoma-carcinoma sequence. *Nature Communications*.
433 2015;6:6528.

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

437 21. Mima K, Sukawa Y, Nishihara R, Qian ZR, Yamauchi M, Inamura K, et al.
438 *Fusobacterium nucleatum* and T Cells in Colorectal Carcinoma. *JAMA oncology*.
439 2015;1:653–61.

440 22. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, et al.
441 Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science (New*
442 *York, N.Y.)*. 2012;338:120–3.

443 23. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD. Microbiota-based model improves the
444 sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine*.
445 2016;8:37.

446 24. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of
447 fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*.
448 2014;10:766.

449 25. Zackular JP, Rogers MAM, Ruffin MT, Schloss PD. The human gut microbiome as

a screening tool for colorectal cancer. *Cancer Prevention Research (Philadelphia, Pa.)*.
2014;7:1112–21.

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

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

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

29. Zhu Y, Lin X, Li H, Li Y, Shi X, Zhao F, et al. Intake of Meat Proteins Substantially
Increased the Relative Abundance of Genus *Lactobacillus* in Rat Feces. *PloS One*.
2016;11:e0152678.

30. 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.

31. 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>

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

33. 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.

34. 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>

35. 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>

36. 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.