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) continues to be a worldwide health problem with previous research suggesting that a link may exist between the fecal bacterial microbiome and CRC. The overall objective of our study was to test the hypothesis that changes in the bacterial microbiome occur after lesion (i.e. adenoma or carcinoma) removal. Specifically, we wanted to identify components within the community that were different before and after removal of said lesion.

Results: The bacterial microbiome changed more in response to lesion removal in carcinoma cases compared to adenoma cases (P-value < 0.05). There was no difference for either the adenoma or carcinoma group in the relative abundance of any OTU between pre and post lesion removal (P-value > 0.05). A model built to classify lesion had an AUC range of 0.811 - 0.866 and follow up samples had a decrease in the positive probability of lesion (P-value < 0.05) suggesting a movement towards a more normal bacterial community. A second model built to classify initial samples had an AUC range of 0.641 - 0.805 and had a decrease in positive probability for the follow up samples to be an initial sample (P-value < 0.05). The lesion model used a total of 53 variables while the initial sample model used a total of 70 variables. A total of 23 OTUs were common to both models with the majority of these classifying to commensal bacteria (e.g. *Bacteroides*, *Clostridiales*, *Blautia*, and *Ruminococcaceae*).

Conclusions: Our data supports the hypothesis that there are changes in the bacterial microbiome following colorectal cancer lesion removal. Individuals with carcinoma have more drastic differences to the overall community then those with adenoma. Changes to commensal bacteria were some of the most important variables for model classification, suggesting that these bacteria may be central to initial polyp formation and transition to carcinoma.

Keywords

bacterial microbiome; colorectal cancer; polyps; FIT; post surgery; risk factors

Background

Colorectal cancer (CRC) continues to be a leading cause of cancer related deaths and is currently the third most common cause of cancer deaths [1,2]. Over the last few years 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].

Over the last few years studies have shown how the bacterial microbiome [3] or specific members within it [4] could be directly involved with the pathogenesis of CRC. There has also been research into how the bacterial microbiome could be altered directly on tumor tissue itself [5]. These studies have helped to provide a tantalizing link between the bacterial microbiome and CRC. Building off of these findings, there has been promising work on the bacterial microbiome and its ability to complement existing screening methods such as Fecal Immunoglobulin Test (FIT) or act alone as a screening tool [6,7]. Although these studies suggest that the bacterial microbiome might change after treatment there remains limited information on the bacterial microbiome response to lesion (adenoma or carcinoma) removal.

In this study we tested the hypothesis that the bacterial microbiome changes between pre (initial) and post (follow up) removal of lesion (adenoma or carcinoma). We analyzed changes in alpha and beta diversity as well as the relative abundance of specific Operational Taxonomic Units (OTUs). We utilized Random Forest to build two models. The first model was built to classify lesion and normal (non-lesion) samples while the second was built to classify initial and follow up samples. Subsequent observations on how these models, as well as specific OTUs within them, performed pre and post lesion removal helped inform us as to whether initial and follow up samples were changing and whether it was towards a more normal bacterial microbiome or not. We also investigated the two models for similar important OTUs to identify the crucial OTUs for not only classifying lesion or normal samples but also initial and follow up.

Results

Bacterial Community and FIT: Within our 67 person cohort we first wanted to test whether there were any broad differences between initial and follow up samples based on lesion being either adenoma (n = 41) or carcinoma (n = 26). We found that the bacterial community in those with carcinoma were more dissimilar to their initial sample than those with adenoma (P-value < 0.001) [Figure 1A]. We also found that there were larger changes in fecal blood (measured by FIT) for those with carcinoma versus adenoma (P-value < 0.0001) [Figure 1B]. The bacterial community structure before and after surgery was visualized using NMDS for both adenoma [Figure 1C] (PERMANOVA > 0.05) and carcinoma [Figure 1D] (PERMANOVA < 0.05). Interestingly, when initial and follow up samples were compared, regardless of whether the lesions were adenoma or carcinoma, there was no significant overall difference in beta diversity (PERMANOVA > 0.05). When investigating alpha diversity metrics there was no difference found between initial and follow up samples for lesion, adenoma only, or carcinoma only for any metric tested [Table S1]. We also observed that there was no difference in the relative abundance of any OTU between initial and follow up samples for lesion, adenoma only, or carcinoma only [Figure S1].

Carcinoma Associated Bacteria: Previous literature has suggested that a number of oral microbes may be important in CRC pathogenesis [8]. So we next examined whether there were changes in previously well described carcinoma associated OTUs, such as *Porphyromonas asaccharolytica* (Otu000202), *Fusobacterium nucleatum* (Otu000442), *Parvimonas micra* (Otu001273), and *Peptostreptococcus stomatis* (Otu001682). We first observed that only a small percentage of those with adenoma or carcinoma were positive or had a relative abundance above 0.1% for any of these respective OTUs [Figure 2]. Despite this, those with carcinoma had a decrease in relative abundance from initial to follow up for *Parvimonas micra* (P-value < 0.05) and *Porphyromonas asaccharolytica*

(P-value < 0.05) [Figure 2A]. In contrast, there was no difference in relative abundance in any of these OTUs when considering only those with adenoma [Figure 2B].

The Lesion Model: We next wanted to identify if there were any common bacterial microbiome changes in individuals with adenoma and carcinoma versus normal controls. We investigated this by creating a model to classify samples as lesion versus normal based on the bacterial community and FIT measurements. This model had an AUC range of 0.811 - 0.866 after 100 iterations of 20 repeated 10-fold cross validations. The ROC curve for the final lesion model used was within the observed range of the 100 different test set AUC iterations [Figure 3A]. There were a total of 53 variables that were used in this model [Figure 3B]. The FIT measurement for fecal blood resulted in the largest decrease in MDA while the OTU with the largest MDA was *Lachnospiraceae* (Otu000015) [Figure 3B].

If there were common OTUs that could separate adenoma and carcinoma from normal controls, we would expect to find a decrease in the positive probability of the follow up sample to be a lesion. This is what we observed for the lesion model (P-value < 0.001). When we separated individuals based on whether they had an adenoma or carcinoma there was only a decrease in positive probability for the carcinoma group [Figure 3C] (P-value < 0.001) and not for the adenoma group [Figure 3D] (P-value > 0.05). We also observed that there were no significant differences between the predicted and actual calls (P-value > 0.05). The lesion model was also able to correctly classify the one individual who still had a carcinoma on follow up [Figure 3C].

The Initial Sample Model: After building a model to classify based on lesion we built a separate model specifically to be able to classify whether samples were initial (before lesion was removed) samples based on the bacterial community and FIT measurements. The initial sample model had an AUC range of 0.641 to 0.805 after 100 iterations of 20 repeated 10-fold cross validations. The test set AUC range for this model performed better then the training set AUCs. There was a marked decrease in the ROC curve for the final model used when compared to the 100 test set AUC iterations [Figure 4A]. There were a total of 70 variables that were used for this model [Figure 4B]. The variable that resulted in the largest MDA was *Pseudomonas* (Otu000438) while FIT measurement for fecal blood resulted in the sixth largest decrease in MDA [Figure 4B].

If there were common OTUs that could separate initial from follow up sample regardless of whether the lesion was adenoma or carcinoma we would expect to find a decrease in the positive probability of the follow up sample to be an initial sample. This is what we observed for the initial sample model (P-value < 0.001). When we separated individuals based on whether they had an adenoma or carcinoma there was a decrease in positive probability for both the carcinoma group [Figure 4C] (P-value < 0.001) and for the adenoma group [Figure 4D] (P-value < 0.001). For this model there was no difference between the predicted and actual classifications (P-value > 0.05).

Common OTUs to both Models: We next wanted to compare the similarity between the OTU variables used in either model. The main purpose was to identify which OTUs were important not only for the classification of lesion but also for the classification of initial or follow up sample. Potentially, these specific OTUs are the most important with respect to the bacterial microbiome response to removal of lesion. When we compared the two different models with each other there were a total of 23 common OTUs. Some of the most common taxonomic identifications belonged to *Bacteroides*, *Clostridiales*, *Blautia*, and *Ruminococcaceae*. The vast majority of these OTUs had classifications to bacteria typically thought of as commensal [Table S2].

Treatment Differences: After observing these changes in the bacterial community and positive probability we wanted to assess whether additional treatments, such as chemotherapy and radiation, could have an impact on the results that we observed. There was only a significant difference for change in positive probability for those treated with chemotherapy for the initial sample model (P-value < 0.05). This suggests that follow up samples for those treated for chemotherapy my have had a larger change from the initial sample then those without such treatment. All other variables that were tested showed no difference based on whether chemotherapy or radiation was used [Table S3].

Discussion

This study builds upon previous work from numerous labs that have looked into the bacterial microbiome as a potential screening tool [6,7,9–11] by exploring what happens to the bacterial community after lesion removal. Based on previous work by Arthur, et al. [12] it may not be surprising to have E.coli as one of the most important OTUs and one that was common to both models. Interestingly, many of the most important OTUs had taxonomic identification for resident gut microbes. This could suggest that the bacterial community is one of the first components that could change during the pathogenesis of disease. These bacterial microbiome changes could be the first step in allowing more inflammatory bacterium to gain a foothold within the colon [8].

From our results there were large observed differences in the bacterial microbiome in samples before and after lesion removal based on whether the individual had an adenoma or carcinoma. In individuals with carcinoma compared to adenoma there were much larger differences between initial and follow up samples based on the community beta diversity and in fecal blood as measured by FIT [Figure 1]. However, there were no differences between initial and follow up samples for any alpha diversity metric measured regardless of whether the individual had an adenoma or carcinoma [Table S1]. There was also no differences in relative abundance of any specific OTU for lesion, adenoma only, or carcinoma only [Figure S1].

Although there were no differences when investigating all OTUs, when looking specifically at four OTUs that taxonomically classified to previously suggested carcinoma associated microbes we found that only 2/4 had a decrease in relative abundance between initial and follow up for those with carcinoma and 0/4 had differences for those with adenoma. This data would suggest that these specific OTUs may be important in the transition of an adenoma to a carcinoma but less so in the initiation of an adenoma from benign tissue.

We next created a model that incorporated FIT and the bacterial microbiome to be able to classify lesions (adenoma and carcinoma). Based on this model we found that the follow up samples were closer to normal then the initial samples due to a decrease in positive probability for lesion and that the commonly associated CRC bacteria were not highly represented within this model with the exception of *Porphyromonas asaccharolytica* [Figure 3]. Although there was a detectable change towards what would be expected for normal controls, it should be noted that the follow up samples may not be a completely normal bacterial microbiome.

After creating the lesion mode we then created a second model to classify initial versus follow up samples. We found that this model was able to accurately classify initial versus follow up samples suggesting that regardless of adenoma or carcinoma there are distinct common changes within the bacterial microbiome that occurs after lesion removal [Figure 4]. Both models had OTUs that overwhelmingly belonged to commensal bacteria. Providing additional information on the importance of commensal bacteria was that there were a total of 23 OTUs in common to both models and the vast majority belonged to regular residents of our gut community [Table S2].

Within our study there was a significant difference for the time elapsed in the collection of the follow up sample between adenoma and carcinoma (uncorrected P-value < 0.05), with time passed being less for adenoma (253 +/- 41.3 days) than carcinoma (351 +/- 102 days). These results would indicate that the findings described were specific to the surgical intervention and that some of the differences observed between carcinoma and adenoma samples could be due to differences in collection time between samples for the two different groups. Specifically, it could confound the observation that carcinomas changed more than adenomas [Figure 1A & 1D]. This confounding though would not affect the observations where these individuals were grouped together [Figure 3 & 4].

Curiously, we observed that the typical CRC associated bacteria were not predictive within our models. There are a number of reasons why this may have occurred. First, is that they were not present in enough individuals to be able to classify those with and without disease with a high degree of accuracy. Second, is that our Random Forest models were able to gather the same information from measures such as FIT or other OTUs. It is also possible that both of these explanations could have played a role. Regardless, our observations would suggest that an individual's resident bacteria have a large role to play in disease initiation and could change in a way that allows predictive models to lower the positive probability of a lesion after surgery [Figure 3C & 3D]. It should be noted that our study does not argue against the importance of these CRC associated bacteria in the pathogenesis of disease but rather that they are not the main bacteria changing after removal of lesion. In fact, it is possible that these CRC associated bacteria are important in the transition from adenoma to carcinoma and would be one explanation as to why in our data we not only see high initial relative abundances in carcinoma and not adenoma individuals but also large decreases in relative abundance in some of those with carcinoma but not in those with adenoma after surgery [Figure 2].

Many of the common OTUs between the two models had OTUs that taxonomically classified to potential butyrate producers [Table S2]. Another batch of OTUs classified to bacteria that can either degrade polyphenols or are inhibited by them. Both butyrate and polyphenols are thought to be protective against cancer in part by reducing inflammation [13]. 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 [14]. High protein consumption in general has been linked with an increased CRC risk [15]. Conversely, *Bacteroides* are inhibited by polyphenols which are derived from fruits and vegetables [16]. 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.

One limitation of our study is that we do not know whether individuals who were still classified as positive by the lesion model eventually had a subsequent CRC diagnosis. This information would help to strengthen the case for our lesion model keeping a number of individuals above the cutoff threshold even though at follow up they were diagnosed as no longer having a lesion. Another limitation is that we do not know if adding modern tests such as the stool DNA test [17] could help improve our overall AUC. This study also 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've 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.

Interestingly, within the initial sample model the test data performed better than the training data. This may have occurred because the training AUC determined from 20 repeated 10 fold cross validation removed samples at random and did not take into account that they were matched samples. Another potential reason is that the model itself may be over-fit since the total number of samples was not that large. However, the lesion model did not suffer from these discrepancies. Further independent studies need to be carried out to verify our findings since we are dealing with correlations that may not be truly representative of the pathogenesis of disease.

Despite these limitations our findings add to the existing scientific knowledge on CRC and the bacterial microbiome: That there is a measurable difference in the bacterial community after adenoma or carcinoma removal. Further, the ability for machine learning algorithms to take bacterial microbiome data and successfully lower positive probability after either adenoma or carcinoma removal provides evidence that there are specific signatures, mostly attributable to commensal organisms, associated with these lesions. Our data provides evidence that commensal bacteria may be important in the development of polyps and also potentially the transition from adenoma to carcinoma.

Methods

Study Design and Patient Sampling: The sampling and design were similar to that reported in Baxter, et al [6]. In brief, 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 microbiome has been shown to normalize back to a pre-colonscopy community within this time period [18]. Our follow up data set had a total of 67 individuals that not only had a sample as described but also a follow up sample between 188 - 546 days after lesion removal and treatment had been completed. 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.

FIT and 16S rRNA Gene Sequencing: FIT was analyzed as previously published using both OC FIT-CHEK and OC-Auto Micro 80 automated system (Polymedco Inc.) [19]. 16S rRNA gene sequencing was completed as previously described by Kozich, et al. [20]. 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.

Sequence Processing: The mothur software package (v1.37.5) was used to process the 16S rRNA gene sequences. This process has been previously described [20]. The general processing workflow using mothur was as follows: 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 in an attempt to minimize uneven sampling.

Lesion Model Creation: The Random Forest [21] algorithm was used to create the model used for prediction of lesion (adenoma or carcinoma) with the main training and testing of the model completed on an independent data set of 423 individuals. This model was then applied to our follow up data set of 67 individuals. It should be noted that all individuals with an adenoma or carcinoma were grouped together to form the lesion group and the model was not created to find differences between normal, adenoma, and carcinoma but rather differences between both adenoma and carcinoma versus normal.

In brief, the model included data on FIT and the bacterial microbiome. 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 taking the samples and making 100 different 80/20 (train/test) splits of the data where normal and lesion were represented in the same proportion within both the whole data set and the 80/20 split. Each of these splits were then run through 20 repeated 10-fold cross validations to optimize the mtry hyper-parameter by maximizing the AUC (Area Under the Curve of the Receiver Operator Characteristic). This resulting model was then tested on the 20% of the data that was originally held out from this overall process. Next, in order to assess which variables were most important to the model we counted the number of times a variable was present in the top 10% of mean decrease in accuracy (MDA) for each of the 100 different 80/20 split models and then filtered this list to variables that were only present more than 50% of the time. This final collated list of variables was what was considered the most important for the model. This reduced data set was then run through the mtry optimization again. Once the ideal mtry was found the entire 423 sample set was used to create the final Random Forest model on which classifications on the 67-person cohort was completed.

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.

Initial Sample Model Creation: We also investigated whether a model could be created that could identify pre (initial) and post (follow up) lesion removal samples from each other. The main difference was that only the 67-person cohort was used at all stages of model building and classification. Other than this difference the creation of this model and optimization of the mtry hyper-parameter was completed using the same procedure as was used for the lesion model. Instead of classifying samples as positive or negative of lesion this model classified samples as positive or negative for being an initial sample prior to lesion removal.

Statistical Analysis: The R software package (v3.3.2) was used for all statistical analysis. Comparisons between bacterial community structure utilized PERMANOVA [22] in the vegan package (v2.4.1). Comparisons between probabilities as well as overall OTU differences between initial and follow up samples utilized a paired Wilcoxson ranked sum test. Where multiple comparison testing was appropriate, a Benjamini-Hochberg (BH) correction was applied [23] 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.

Analysis Overview: We first wanted to test if there were any differences based on whether the individual had an adenoma or carcinoma. This was done by testing initial and follow up samples for differences in alpha and beta diversity, testing differences in FIT between initial and follow ups, testing all OTUs, and investigating the relative abundance of specific previously associated CRC bacteria (*Fusobacterium nucleatum*, *Parvimonas micra*, *Peptostreptococcus assacharolytica*, and *Porphyromonas stomatis*) based on adenoma and carcinoma. From here the lesion model was then tested for accuracy in prediction and whether it reduced the positive probability of lesion after surgery. We then used the initial sample model to assess whether it could classify samples better then the lesion model and whether it could reduce the positive probability of an initial sample in the follow up samples. Finally, a list of common OTUs were found for the two different models used.

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 the Adenoma and Carcinoma Group. A) A significant difference was found between the adenoma and carcinoma group for thetayc (P-value = 0.000472). Advanced adenomas are denoted as Screen Relevant Neoplasia (SRN). B) A significant difference was found between the adenoma and carcinoma group for change in FIT measurement (P-value = 2.15e-05). Advanced adenomas are denoted as Screen Relevant Neoplasia (SRN). C) NMDS of the initial and follow up samples for the adenoma group. D) NMDS of the initial and follow up samples for the carcinoma group.

Figure 2: Previously Associated CRC Bacteria in Initial and Follow Up Samples. A) Carcinoma initial and follow up samples had an observed significant difference in initial and follow up sample for the OTUs classified as *Parvimonas micra* (P-value = 0.0116) and *Porphyromonas asaccharolytica* (P-value = 0.00842). B) Adenoma initial and follow up samples. There were no significant differences between initial and follow up (P-value = 0.37.

Figure 3: The Lesion Model. A) ROC curve: The shaded areas represents the range of values of a 100 different 80/20 splits of the test set data and the blue line represents the model using 100% of the data set and what was used for subsequent classification. B) Summary of Important Variables. MDA of the most important variables in the lesion model. The black point represents the mean and the different colors are the values of each different run up to 100. C) Positive probability change from initial to follow up sample in those with carcinoma. D) Positive probability change from initial to follow up sample if those with adenoma or advanced adenoma (Screen Relevant Neoplasia (SRN)).

Figure 4: The Initial Sample Model. A) ROC curve: The shaded areas represents the range of values of a 100 different 80/20 splits of the test set data and the blue line represents the model using 100% of the data set and what was used for subsequent classification. B) Summary of Important Variables. MDA of the most important variables in the initial sample model. The black point represents the mean and the different colors are

the values of each different run up to 100. C) Positive probability change from initial to follow up sample in those with carcinoma. D) Positive probability change from initial to follow up sample of those with adenoma or advanced adenoma (Screen Relevant Neoplasia (SRN)).

Figure S1: Distribution of P-values from Paired Wilcoxson Analysis of All OTUs for Initial versus Follow Up

Figure S2: Thetayc Versus Time of Follow up Sample from Initial

Declarations

Ethics approval and consent to participate

Consent for publication

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.

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

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

4. 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.

5. 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.

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

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

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

9. 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.

10. 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.

11. 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.

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

13. 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

14. 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.

15. 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.

16. 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

17. Cotter TG, Burger KN, Devens ME, Simonson JA, Lowrie KL, Heigh RI, et al. Long-Term Follow-up of Patients Having False Positive Multi-target Stool DNA Tests after Negative Screening Colonoscopy: The LONG-HAUL Cohort Study. Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology. 2016;

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

19. Daly JM, Bay CP, Levy BT. Evaluation of fecal immunochemical tests for colorectal cancer screening. Journal of Primary Care & Community Health. 2013;4:245–50.

20. 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.

21. 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

22. 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

23. 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.