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

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. This study tested the hypothesis that treatment for adenoma or carcinoma results in changes to the bacterial community. Specifically, we tried to identify components within the community that were different before and after treatment of adenoma, advanced adenoma (Screen Relevant Neoplasia (SRN)), and carcinoma.

Results: There was a larger change in the bacterial community in response to treatment for carcinoma versus adenoma or SRN cases *(P-value)*. But there was a trend for increasingly less community similarity between initial and follow up samples from adenoma to SRN to carcinoma. Despite this, no difference was found in the relative abundance of any OTU before and after treatment for adenoma, SRN, or carcinoma groups *(P-value)*. Using Random Forest models to assess whether changes in follow up samples were towards a normal community showed that only those with carcinoma had a significant decrease in positive probability *(P-value)*; providing further evidence that treatment has the greatest effect in those with carcinoma. The adenoma model used a total of *model n* OTUs, the SRN model used a total of *model n* OTUs, and the carcinoma model used a total of *model n* OTUs. A total of *common n* OTUs were common to both 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 *(Pvalue)*

Conclusions: Our data partially supports the hypothesis that the bacterial community changes after treatment. Individuals with carcinoma had more drastic differences to the overall community then those with adenoma. Common OTUs to all models were overwhelmingly from commensal bacteria, suggesting that these bacteria may be important

26   to 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].

Recent studies in humans have shown that both the microbiome and specific members within it correlate with CRC pathogenesis [3,4]. Further, bacterial communities have been observed to be altered between normal and tumor tissue [5]. Mouse models of CRC have further demonstrated the importance of the microbiome, both on a community [3,6] and species level [4], for tumorgenesis. Collectively, these studies provide a tantalizing link between our gut bacteria and CRC and suggest that biomarkers using our microbes could be developed. Indeed, builidng models using 16S rRNA gene sequencing along with clinical tests such as Fecal Immunoglobulin Test (FIT) result in good predictions of CRC [7,8]. Although these studies show how our gut bacteria can impact CRC progression via a changed community or invasion by more inflammatory bacteria [9]. They provide very little information as to whether these communities change and rebound towards normal after successful treatment of adenoma, advanced adenoma (Screen Relevent Neoplasia (SRN)), 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 [3,4] and the ability to be able to use the microbiome as a predictive screening tool [7,8]. Understanding polyp formation and transition to SRN and then carinoma is crucial to being able to understand how to prevent CRC occurence. Response of the community to treatment is also eqaully important to predictive models designed for screening purposes since an unresponsive community would provide little additional information for important events,

54 such as recurrence [10].

55 Using pre- (initial) and post- (follow up) treatment samples we tested the hypothesis that
56 treatment causes detectable changes to the microbiome in those with adenoma, SRN,
57 and carincoma. First, we assessed differences between initial and follow up samples in
58 adenoma, SRN, or carinoma using alpha or beta diversity metrics. Second, we explored
59 whether models built to classify adenoma, SRN, or carcinoma versus normal were able
60 to identify specific community members that differed between initial and follow up. We
61 also used these models to assess whether changes in the community were toward a
62 more normal micorbiome. Finally, we assessed both whether surgery for adenomas and
63 SRN provided larger community changes or whether chemotherapy or radiation provided
64 additive changes to the microbiome over surgerical resection. This study helps to provide
65 evidence as to whether treatment can influence the community and if the CRC microbiome,
66 identified in previous studies, persists after such interventions.

Results

68 ***The Bacterial Community:*** Within our 67-person cohort we tested whether those with
69 adenoma *(n =)*, SRN *(n =)*, or carcinoma *(n =)* had any broad differences between
70 their initial and follow up samples. We found that those with carcinoma had a more
71 dissimilar bacterial community between their initial and follow up sample than those with
72 adenoma *(P-value 0.001)* [Figure 1A]. Although no significant differences were observed
73 between SRN and carcinoma there was an increase in the dissimilarity of the initial and
74 follow up samples from adenoma *(mean and SD)* to SRN *(mean and SD)* to carcinoma
75 *(mean and SD)* [Figure 1A]. The bacterial community structure before and after surgery
76 was visualized using NMDS for adenoma [Figure 1B] *(PERMANOVA 0.05)*, SRN [FIgure
77 1C] *(PERMANOVA 0.05)*, and carcinoma [Figure 1C] *(PERMANOVA 0.05)*. Interestingly,
78 when initial and follow up samples were compared, regardless of whether the lesions
79 were adenoma or carcinoma, there was no significant overall difference in beta diversity
80 *(PERMANOVA 0.05)*. There was no difference between initial and follow up samples when
81 investigating alpha diversity metrics for adenoma, SRN, or carcinoma for any metric tested
82 [Table S1]. Additionally, there was also no difference in the relative abundance of any OTU
83 between initial and follow up samples for adenoma, SRN, or carcinoma only [Figure S1].

84 ***Adenoma Model*** The range of model AUC's from 100 runs of 20 repeated 10 fold
85 cross-validation was *from to this* with the AUC of the model used being *AUC value* [Figure
86 S2A]. There was a total of *blank* OTUs in this model with the vast majority classifying to
87 bacteria typically thought of as commensal [Figure S3A]. There was a significant difference
88 between the actual and model predicted group stratification calls *(P-value)*. There was no
89 significant decrease in the positive probability of adenoma between initial and follow up
90 samples *(P_value)* [Figure 2A].

91 ***SRN Model*** The range of model AUC's from 100 runs of 20 repeated 10 fold

7

cross-validation was *from to this* with the AUC of the model used being *AUC value* [Figure S2B]. There was a total of *blank* OTUs in the SRN 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 model predicted group stratification calls *(P-value)*. There was no significant decrease in the positive probability of SRN between initial and follow up samples *(P_value)* [Figure 2B].

Carcinoma Model The range of model AUC's from 100 runs of 20 repeated 10 fold cross-validation was *from to this* with the AUC of the model used being *AUC value* [Figure S2C]. Interestingly, the AUCs improved from adenoma to SRN to carcinoma [Figure 2]. There was a total of *blank* OTUs in the carinoma model [Figure S3C]. Similar to the adenoma and SRN models the vast majority of OTUs classified to bacteria typically thought of as commensal but OTUs that also classified to *Fusobacterium*, *Porphyromonas*, and *Parvimonas* appear to be important for carcinoma classification [Figure S3C]. Also similar to the adenoma and SRN models there was a significant difference between the actual and model predicted group stratification calls *(P-value)*. There was a significant decrease in the positive probability of carcinoma between initial and follow up samples *(P_value)* [Figure 2C]; suggesting that the carcinoma samples changed towards normal after treatment, unlike either adenoma or SRN. The one indivdiual still positive for carcinoma after treatment had an increase in carcinoma positive proability on follow up [Figure 2C].

Adenoma, SRN, and Carcinoma Common OTUs We next wanted to know what predictors within the adenoma, SRN, and carcinoma models were similar to each other. The main purpose was to identify which OTUs could be important at all three stages of disease. When we compared the three different models with each other there were a total of *number* common OTUs. Some of the most common taxonomic identifications belonged to *one*, *two*, *three*, *four*, *five*, *six*, and *seven*. These along with the vast majority of the

8

OTUs that were common between these models had classifications to bacteria typically thought of as commensal [Table S2].

Treatment Affects on Community After observing these these changes from treatment we assessed whether chemotherapy or radiation, in the carcinoma group, and surgery, in the adenoma group, impacted the observed results. In the carcinoma group neither chemotherapy or radiation provided any additive change from initial sample over surgery alone [Table S3]. For the adenoma group there was no difference between those that received surgery and those that did not [Table S4]. For surgery adenoma and SRN were combined due to the low number of surgery occurances. There was no difference in the proportion of those receiving surgery between the adenoma and SRN groups *(P-value)*. This data suggests that microbiome changes observed in the carcinoma group were mostly a result from surgery and not from chemotherapy or radiation.

Discussion

131 This study builds upon previous work from numerous labs that have considered both

132 how the bacterial community between those with and without CRC differ and how it

133 might be used as an early screening tool [7,8,11–13]. Here we show that the bacterial

134 community changes towards normal after treatment for carcinoma and that chemotherapy

135 and radiation did not provide an additive change. Although some of the important OTUs

136 classified to genera from bacteria considered the usual suspects (e.g. *Fusobacterium*,

137 *Porphyromonas*, and *Parvimonas*) many did not. The majority of important OTUs had

138 taxonomic classifications for resident gut microbes and were common for the adenoma,

139 SRN, and carcinoma models. This suggests that members within the commensal

140 community may be the first that change during CRC pathogenesis. These subtle changes,

141 in turn, could be the first step in allowing more inflammatory bacteria to gain a foothold

142 within the colon [9].

143 Unlike previous studies on the microbiome and CRC, ours focuses both on identifying

144 commonalities and differences within adenoma, SRN, and carcinoma groups before and

145 after treatment. Although there were differences for genera associated with specific

146 bacterium linked with CRC [Figure S3]. The majority of important OTUs taxonmically

147 classified to commensal bacteria [Figure S3]. Although these changes may be subtle, due

148 to the lack of signifcant difference in the bacterial community before and after treatment in

149 adenoma and SRN [Figure 1], they support the hypothesis that the first members of the

150 community to change and potentially stay changed even after treatment are those that are

151 commensal bacteria.

152 Many of the common OTUs that we identified taxnomically classified to potential butyrate

153 producers (e.g. *Clostridiales*, *Roseburia*, and *Anaerostipes*)[Table S2]. Other OTUs

154 classified to bacteria that are inhibited by polyphenols (e.g. *Bacteroides*). Both butyrate

155 and polyphenols are thought to be protective against cancer, in part by reducing

156 inflammation [14]. These protective compounds are derived from the breakdown of

157 fiber, fruits, and vegetables by resident gut microbes. One example of this potential

158 diet-microbiome-inflammation-polyp axis is that *Bacteroides*, which was highly prevalent

159 in our models, are known to be increased in those with high non-meat based protein

160 consumption [15]. High protein consumption in general has been linked with an increased

161 CRC risk [16]. Conversely, *Bacteroides* are inhibited by polyphenols which are derived

162 from fruits and vegetables [17]. Our data fits with the hypothesis that the microbial

163 metabolites from breakdown products within our own diet could not only help to shape the

164 existing community but also have an effect on CRC risk and disease progression.

165 A limitation, in our study, was that there was a significant difference in the time elapsed

166 in the collection of the follow up sample between adenoma or SRN versus carcinoma

167 *(uncorrected P-value 0.05)*, with time passed being less for adenoma *(days)* and SRN *(*

168 *days)* than carcinoma *(days).* These results would indicate that some of the differences

169 observed between the carcinoma and adenoma groups could be due to differences in

170 collection time. Specifically, it could confound the observation that carcinomas changed

171 more than adenomas [Figure 1A & 1C]. However, there are two reasons that this may not

172 be the case. First, the SRN group did not have a significant difference and the collection

173 time of their follow up sample was less than the adenoma group. Second, this confounding

174 would not affect the observations where models were used since they were built using a

175 different cohort [Figures 2 & S2-S3].

176 Another limitation was that we do not know whether individuals who were still classified

177 as positive by the carcinoma model eventually had a subsequent CRC recurrence. This

178 information would help to strengthen the case for this model keeping numerous individuals

179 above the cutoff threshold even though at follow up they were diagnosed as no longer

180 having carcinoma. 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.

Despite these 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, SRN 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 inflammtory and resident commensal organisms, associated with treatment. Our data provides evidence that commensal bacteria may be important in the development of polyps and potentially the transition from adenoma to carcinoma.

Methods

Study Design and Patient Sampling: Sampling and design have been previously reported in Baxter, et al [7]. 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-colonscopy community within this time period [18]. 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. [19]. 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 initial and follow up 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 [19]. 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

13

₂₂₀ using the Ribosomal Database Project (RDP), and clustered into Operational Taxonomic

₂₂₁ Units (OTUs) using a 97% similarity cutoff with an average neighbor clustering algorithm.

₂₂₂ The number of sequences for each sample was rarefied to 10523 to minimize uneven

₂₂₃ sampling.

₂₂₄ ***Lesion Model Creation:*** The Random Forest [20] algorithm was used to create the

₂₂₅ model used for prediction of lesion (adenoma or carcinoma) with the main training and

₂₂₆ testing of the model completed on an independent data set of 423 individuals. This model

₂₂₇ was then applied to our 67-person cohort. It should be noted that all individuals with an

₂₂₈ adenoma or carcinoma were grouped together to form the lesion group and the model

₂₂₉ was not created to find differences between normal, adenoma, and carcinoma but rather

₂₃₀ differences between both adenoma and carcinoma versus normal.

₂₃₁ The model included only OTU data obtained from 16S rRNA sequencing. Non-binary data

₂₃₂ was checked for near zero variance and OTUs that had near zero variance were removed.

₂₃₃ This pre-processing was performed with the R package caret (v6.0.73). Optimization of

₂₃₄ the mtry hyper-parameter involved making 100 different 80/20 (train/test) splits of the data

₂₃₅ where normal and lesion were represented in the same proportion within both the whole

₂₃₆ data set and the 80/20 split. For each a 20 repeated 10-fold cross validation was performed

₂₃₇ on 80% component to optimize the mtry hyper-parameter by maximizing the AUC (Area

₂₃₈ Under the Curve of the Receiver Operator Characteristic). The resulting model was then

₂₃₉ tested on the hold out data obtained from the 20% component. Assessment of the most

₂₄₀ important OTUs to the model involved counting the number of times an OTU was present

₂₄₁ in the top 10% of mean decrease in accuracy (MDA) for each of the 100 different splits

₂₄₂ run. This was then followed with filtering of this list to variables that were only present in

₂₄₃ more than 50% of these 100 runs. The final collated list of variables was then run through

₂₄₄ the mtry optimization again. Once the ideal mtry was found the entire 423 sample set was

₂₄₅ 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.

Treatment Model Creation: We also investigated whether a model could be created that could identify pre- (initial) and post- (follow up) treatment samples. 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 treatment.

Statistical Analysis: The R software package (v3.3.2) was used for all statistical analysis. Comparisons between bacterial community structure utilized PERMANOVA [21] 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 [22] 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 tested for 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 all OTUs, and investigating the relative abundance of genera from previously associated CRC bacteria (*Fusobacterium*, *Parvimonas*, *Peptostreptococcus*, and *Porphyromonas*). Next, the lesion model was

15

tested for accuracy in prediction and whether it reduced the positive probability of lesion in follow up samples. We then used the treatment model to assess whether it could classify samples better than the lesion model and whether it could reduce the positive probability of an initial sample in the follow up samples. Common OTUs were found for the two different models used to assess which were important for both models. Finally, differences between those receiving chemotherapy and radiation versus those who received neither were 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, SRN, and Carcinoma Groups After Treatment. A) A significant difference was found between the adenoma and carcinoma group for thetayc (P-value = NULL). Advanced adenomas are denoted as Screen Relevant Neoplasia (SRN). B) NMDS of the initial and follow up samples for the adenoma group. C) NMDS of the initial and follow up 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: The Treatment Model. B) 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)).

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, SRN, and Carcinoma Models. A) Adenoma ROC curve: The light greenshaded 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) SRN 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, SRN, 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 SRN 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 redpoints 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.

Funding

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

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. 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.

7. 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.

8. 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.

9. Flynn KJ, Baxter NT, Schloss PD. Metabolic and Community Synergy of Oral Bacteria in

Colorectal Cancer. mSphere. 2016;1.

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

11. 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.

12. 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.

13. 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.

14. 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

15. 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.

16. 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.

17. 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

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

19. 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.

20. 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

21. 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

22. 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.