IRB Protocol Template

Title: COLON-MD: COlon cancer LONgitudinal study of the Microbial metabolites and Dietary factors that influence response to treatment

Investigators:
PI: Leigh Greathouse (BU), Co-PIs: Lucas Wong (BSWH), Sub I: Rahila Essani, Hitesh Singh

1. Introduction/Background

More than half of all colon cancer is diagnosed at a stage of regional or distant metastasis. The standard treatment for these late stage malignancies includes surgery and adjuvant chemotherapy with a combination of drugs including platinum-based therapy (e.g FOLFOX). These treatments, while demonstrating improvement in survival, often have severe side effects that can be both transient and long-lasting. Although the microbiome field of research is still relatively nascent, mounting evidence indicates that the gut microbiome influences the development and treatment of colon cancer through several factors including modulation of the immune system. Specifically, *F. nucleatum* colonization in murine models of colon cancer ($Apc^{Min/+}$) increases tumor multiplicity by various mechanisms including favoring the infiltration of tumor-promoting myeloid cells to create a pro-inflammatory environment. However, the microbiome field still lacks critical *prospective* studies to determine causation, which is the focus of this proposal.

Recently, multiple murine studies demonstrated that treatment efficacy depends in part on the microbiome-immune system interactions. Specifically, destruction of the microbiome through use of antibiotics resulted in a compromised efficacy of immune stimulating and platinum-based chemotherapeutics. In addition, commensal bacteria can enhance treatment through increased secretion of TNF-α in tumor-associated myeloid cells. Supporting this evidence in patients with melanoma, investigators were able to use the initial microbial community structure (shot-gun DNA sequencing) to identify bacterial pathways (polyamine synthesis, B vitamin metabolism) that were able correctly predict colitis development after treatment with the immunotherapeutic agent ipilimumbab (anti-CTLA-4) (Dubin, 2016). Several fundamental questions therefore remain unanswered 1) how do prophylactic antibiotics prior to treatment change the microbiome and its function in relationship to with treatment outcome? 2) does the microbial community structure prior to treatment influence or predict treatment outcome, survival or recurrence in colorectal cancer (CRC) patients? 3) how does the microbiome-immune system relationship change during anti-cancer therapy? We propose to address these last two questions in a prospective incident cohort of colorectal cancer using patients recruited from the Baylor Scott and White Healthcare System. This study will collect multiple biospecimens (stool, blood, tissue) before, during and after treatment. These data will enable us to measure the microbiome-immune system relationship at multiple time points to determine the effect on side-effect type and severity, and on colon cancer recurrence (stage II/III) and survival. The results from this study will 1) assist patients and clinicians in making informed treatment decisions, and 2) potentially lead to new therapeutics to increase survival, reduce treatment side effects and prevent recurrence of colon cancer.

2. Significance

Why is the microbiome important in colon cancer treatment and outcome? Overall, colorectal cancer (CRC) accounts for approximately 142,000 new cancer cases and 50,000 cancer deaths annually, making it the second most lethal cancer in the U.S. (SEER). Alarmingly, a recent report indicates that CRC is on the rise among those younger than 55. Unfortunately, more than half of all colon cancer is diagnosed at a stage of regional or distant metastasis. The standard treatment for these late stage malignancies includes surgery and adjuvant chemotherapy with a combination of drugs including antibiotics and platinum-based therapy. One of the methods that has shown promise for identifying early stage colon cancer, response to treatment is the microbiome of the gastrointestinal tract (GI)[1-4] . The quantity of microorganisms present in the human gastrointestinal (GI) tract are approximately equal to the number of human cells, yet their collective genomes, the microbiome, exceeds our own genome by over100-fold [5] [6]. Therefore, the

potential of the microbiome to affect cancer etiology, inflammation, treatment and survival is substantial. Several lines of evidence already support this idea, demonstrating an intimate relationship between the microbiome, its structure, and metabolic products and the immune system. However, these studies have only been conducted in animal models of cancer. While case-control studies in humans have demonstrated an association between the microbiome and immune system in individuals with cancer, we lack an understanding of the causal effects; how the microbiome affects the immune system response and colon cancer treatment, recurrence or survival. This gap in understanding of causation can thus only be addressed in a prospective study of colon cancer, which is a specific directive from the Whitehouse Microbiome Initiative.

By prospectively characterizing the colon cancer microbiome and immune system responses, we can address multiple outstanding questions. First, does an individual's core or initial microbiome/community structure predict treatment response, recurrence or survival? Second, how much of an effect does the microbiome have on the immune response during and after cancer treatment, and how long does it take the microbiome to recover (if it does)? Does, does gut barrier integrity or biofilm formation predict treatment outcome?

Answers to these important question, can be used to not only enrich our understanding of the mechanisms important to the microbiome-immune system relationship, but will directly inform future decision in colon cancer treatment. Most important, these results may uncover potential therapeutics for cancer treatment and side-affect mitigation, such as chemotherapy-induced diarrhea, which affects over 50% of patients undergoing anti-cancer therapy for CRC. For example, it may be prudent to first restructure the microbiome to establish a more conducive environment prior to treatment to enable the most effective treatment response. Further, it may be possible to leverage the immune system by engineering probiotics that deliver anti-inflammatory cytokines that ameliorate inflammation after treatment. **Conducting this study is a significant part of the key steps in elucidating causal mechanisms underlying the impact of the microbiome-immune system relationship on efficacy of anti-cancer therapy, and determining the impact of diet and metabolite fluctuations in predicting recurrence and survival.**

3. Objectives & Specific Aims

Objective: This is a formative study to lay the groundwork for establishing a large cohort of incident colon cancer to examine the microbiome-immune relationship in colon cancer outcome. Our hypothesis is that colon cancer treatment, recurrence and survival are compromised by treatment-induced dysbiosis and dependent upon pre-treatment microbial structure. We plan to address this hypothesis with the following aims:

Aim 1: Determine if the gut microbiome and microbial metabolites are altered in response to cancer treatment and correlate with CID incidence and severity in stage II/III colon cancer

Sub-Aim1a: Quantify the gut microbiota prior to surgery and treatment, and at 3 time points during FOLFOX, and correlate with CID incidence and severity

Sub-Aim1b: Quantify microbial pathway abundance and fecal metabolite levels prior to, and during treatment, and determine their associations with incidence and severity of CID

Sub-Aim1c: Quantify circulating and tumor tissue cytokine levels prior to, and during treatment, and determine their associations with the gut microbiota, and incidence and severity of CID

Aim 2: Determine if the dietary history or dietar intake during treatment is associated with CID incidence and severity in stage II/III colon cancer

Primary outcomes: CID incidence and severity and gut microbiome and metabolome changes. The patients with colon cancer, serving as their own controls, will provide stool samples that enable clinicians to predict toxicity of anti-cancer therapy in stage II/III colon cancer patients.
Secondary outcomes: Circulating cytokines, dietary history and changes during treatment, stool index changes

4. Methodology

4.1 Study Design.

While incident colon cancer studies have been conducted, none have integrated the analysis of the microbiome with other clinically relevant factors. With the recent evidence from animal models demonstrating the importance of the microbiome-immune relationship in shaping the efficacy of anti-cancer treatment, a prospective human study is critical to understand how and if these results translate to human colon cancer. In general, however, there is a paucity of prospective human microbiome studies addressing the microbiome-host relationship. Thus, this study design will address several outstanding issues in the fields of cancer and microbial ecology.

In order to understand the causative factors in response to colon cancer treatment and identify predictive biomarkers, and prospective cohort study is necessary. As previously discussed, the outstanding questions that will be answered through this type of study design are 1) whether the initial microbiome/community structure predicts treatment toxicity, and 2) the impact of the diet and microbiome on the immune response during and after cancer treatment, and its resiliency after cessation of treatment. By collecting multiple types of specimens before, during and after treatment with extensive follow up, we will be able to answer the first questions regarding predictive biomarkers that individually or in combination predict response to treatment. We will also be able to use these data to identify mechanisms in the causal pathway of response to treatment that can be used in future studies to improve survival or reduce recurrence.

This will be a large prospective cohort of incident colon cancer. For this protocol, there is no group assignment, randomization or placebo group. Study Day 0 for this protocol will be considered when the patient is diagnosed with colon cancer after consent, and separately Treatment Day 1 is the day on which the patient starts their first chemotherapy infusion (lasting 12 weeks). All stool samples collected will be considered beyond standard of care (SOC), however, all blood samples will be taken as part of SOC, with the exception that an 2 additional vials of blood will be taken for immune marker analysis. Briefly, samples for this project will include:

A. After pre-study consent, the day prior to colonoscopy for colon cancer screening/detection, a stool (patient-collected) sample will be collected prior to bowel prep for colonoscopy. If, for some reason, this is not possible to do by the patient, then the physician will collect a rectal swab during colonoscopy (**Visit 1**). One stool sample (self-collected) the day prior to the first outpatient appointment before start of treatment (between Days 21-28) – Pre-Treatment (**Visit 2**) 3. Once IV chemotherapy begins, this will be designated at treatment day 1 (TD1). After the IV chemo rounds 2 (**Visit 3, treatment day 15**), 6 (**Visit 4, treatment day 71**) and 10 (**Visit 5, treatment day 128**), stool will be collected by the patient at home the day before starting chemo and on the 2nd day (or within 72 hours to account for constipation/diarrhea) after the last treatment day of that round of IV chemo for that session. Specifically, one stool will be collected on TD 15 and 19 (or the day before the next IV chemo round starts) after the 2nd round of IV chemo; on TD, 71 and 75 (or the day before the next IV chemo round starts) after the 6th round of IV chemo; and on TD 128 and 131 (or the day before the next round of IV chemo) after 10th round of IV chemo treatment. Blood draw is SOC, however, stool samples will be requested in addition to SOC.

B. A blood sample taken at the same time as blood is draw for standard clinical treatment when the patient goes in for the clinical visit or treatment by the REA during usual blood draw times (Surgery Day and on the outpatient visit prior to treatment; Treatment Days 15, 71, 128).

C. Tissue samples from the patient's resected tumor and non-tumor adjacent tissue after resection surgery will be collected for 1) fresh/frozen and 2) Carnoy's fixitive. Specifically, the physician will take a larger piece of both tumor and non-tumor adjacent tissue so that a small piece of each can be stored for analysis. After specimen collection and ID coding, microbial DNA will be isolated from the microbiome specimens and fecal and circulating metabolites identified. The DNA will be used to conduct microbial DNA sequencing (metagenomics) in accordance with existing protocols through the Integrated Human Microbiome Project (iHMP) study of IBD. The fecal metabolites will be used

to produce spectral profiles for quantification using protocols from the iHMP; performed at Baylor University in the laboratory of Dr. Ramon Lavado. RNA will be isolated from tissue samples and separate section preserved in Carnoy's to determine biofilm formation and specific bacterial species present. RNA will be used to conduct gene expression assays (cytokines), and blood samples will be used to conduct serum cytokine assays and blood metabolite analysis.

Sample Population:
Inclusion/exclusion criteria – the study population will be derived from the two participating BSW hospital systems (Waco and Temple, TX) and will be selected to participate in the study on the basis of the following criteria (inclusion): Age >18 and ≤79 years, written informed consent, ability to perform informed consent, FOLFOX treatment and Stage II/III colon cancer diagnosis. To ensure fitness for recurrence detection and anti-cancer therapy, the age was restricted to >18 so as not to include minors, and ≤79 years based on life expectancy with colon cancer diagnosis. Individuals will be excluded from participating if any of the following criteria are met: not diagnosed with stage II or III colon cancer, pregnancy, Lynch syndrome or FAP diagnosis, inability or refusal to perform informed consent, inability to comply with the follow-up program, history of previous colon cancer diagnosis, previous treatment with antibiotics in the last month, previous bowel resection.

Sample size - Based on current data from BSW, the recurrence rate for stage II/III colon cancer is 20%, with a 10% 5-year survival rate for stage IV colon cancer. Given our study design, we conducted power calculations for each group. To meet the number required to detect chemotherapy induced diarrhea and mucositis associated with microbial and metabolite abundance, we need to collect at least 560 stool samples from 112 people (Observational Study). To meet the power requirements for detection of differentially abundant microbes and metabolites associated with chemotherapy-induced CID we estimate needing at least 100 stage II/III patients. The specific power calculation for these groups are provided below.

Taxanomic differential abundance analysis
Power calculations for differential abundance analysis in human microbiome studies can be challenging, as there is limited data with highly consistent sample preparation available and uncertainty in the overall distribution of bacterial taxa throughout the human population[7,8]. Recently, a CRC stool-based microbiome study of a large cohort was performed by Baxter and colleagues[3]. This publicly available data serves as a potential resource for power-analyses by allowing for empirical calculation of the expected variability of taxa in both healthy and CRC-associated gut microbiomes.

We obtained the raw 16S rRNA sequence data from the Baxter study including patients with CRC (n=120) and healthy controls (n=172). After preprocessing and subsampling high-quality reads to an equivalent depth of 10,000 per sample, we identified 92 genus-level groups that each represented (on average) at least 0.25% of sequences from the CRC or healthy patient groups. We then computed the corresponding variance in relative abundance for each of these 92 taxa, and used the empirical values to perform a comprehensive power-analysis.

As expected, increased statistical power was found with higher abundance genera, so here we focus on power results for organisms reflecting an average of 0.25-1.0% of 16S sequences in the gut microbiome because detecting differences in these rare taxa may be critical to our study. Among the genera found at this low level, if we assume equal variance in two normally-distributed populations and require a significance level of α=0.05, we expect that with a comparison of 40 vs 100 individuals, we will be able to detect a relative increase of 200% (e.g. 0.5% mean abundance increased to 1.5%), with a statistical power of 92.3% (*pwr.t2n.test* in R).

Considering the genera with average frequency >1%, we estimate that with a 40 vs 100 sample comparison, we will be able to detect a relative increase of 200% (e.g. 1% mean abundance increased to 3%, or 2% mean abundance increased to 6%), with an expected statistical power of 99.6%.

Table summarizing results for power in detecting differential abundance:

			Group Sizes			
		Taxa Abundance	20vs100	40vs100	50vs100	75vs100
Relative difference	50%	0.25-1.0%	20.5%	30.6%	34.2%	40.6%
		>1.0%	40.0%	54.9%	59.3%	66.4%
	100%	0.25-1.0%	52.7%	67.2%	71.0%	76.5%
		>1.0%	77.7%	88.8%	91.2%	94.5%
	200%	0.25-1.0%	84.6%	92.3%	93.9%	96.0%
		>1.0%	97.9%	99.6%	99.8%	99.9%
	300%	0.25-1.0%	95.0%	98.2%	98.7%	99.3%
		>1.0%	99.9%	100.0%	100.0%	100.0%

4.3 Procedures/Methods.

Subject recruitment and enrollment:
Subjects for this study will be between the ages of 18 and 79 with a colon cancer diagnosis, and will be recruited from one of two participating sites (Waco, Temple). We plan to include at least 20% of the subjects from minorities. This will be accomplished by monthly checks on enrollment demographics by REA and discussed at each clinical study meeting to ensure this goal is met.

The REA will pre-screen patients that are scheduled to undergo colonoscopy to identify patients as a potential study participant. Patients scheduled to undergo colonscopy for suspected colon cancer will be approached by the REA or Physician to ask if they would like to participate in the study and enrolled if they give consent. These informational sources will provide a short description of the study purpose and requirements for participation. Written informed consent will be obtained by the REA from patients agreeing to participate in the study and before commencement of any study activities. Once identified, the REA will screen the patients and obtain consent for study entry. After consent, participating patients will be provided with a study identification number. This number will be utilized on the data collection form, sample labeling and dietary assessments to promote confidentiality throughout the study procedures. At the consent visit, patients will complete a data collection form utilized to confirm study eligibility and record interval history at the sampling time points. Next, they will be given a stool collection kit to take home along with their bowl prep kit, to collect stool prior to bowel prep. If a positive CRC diagnosis is made through the pathology department and confirmed by GI physicians at BSWH facilities, they would continue as a participant in study. This method will allow the collection of stool specimen prior to colonoscopy. **If they are not diagnosed with stage II/III CRC, subjects will be asked if the stool sample can be used in the study as a control. Subjects will be given the option to agree or decline on the prescreening consent form.** Dr. Greathouse will direct the longitudinal specimen collections, assessments and retention. This consent information will be kept at BSWH by either the REA in a password protected file only accessible by the REA or Dr. Wong. We plan to recruit 112 stage II/III patients, (110% of what is needed according to power calculations). Recruitment and sample collection will be supported by the clinical practice sites in Temple and Waco, REAs, and facilitated by Principle Investigator Dr. Leigh Greathouse PhD, MPH, RD and Physician Drs. Lucas Wong, MD.

Study procedures:

This study will collect two sets of data from subjects at baseline (pre-treatment) and on-treatment, as outlined in the Study Design section above.

After specimen collection and coding, microbial DNA will be isolated from the microbiome specimens and fecal metabolites identified. The DNA will be used to conduct DNA sequencing (metagenomics) in accordance with existing protocols through the Integrated Human Microbiome Project (iHMP) study of IBD. The fecal metabolites will be used to produce spectral profiles for quantification using protocols from the iHMP. RNA will be isolated from tissue samples and separate section preserved in Carnoy's to determine biofilm formation. RNA will be used to conduct gene expression assays to confirm taxa identified from sequencing, and blood samples will be used to conduct cytokine assay.

In compliance with BSW protocols, subjects will be screened after consent to verify eligibility. No interventions, device or drugs will be used as part of this observational study. Each participant will serve as his or her own control for the observational (and large cohort) study since this is a prospective cohort design, thus no randomization procedures will be employed. In addition, the questions being asked (e.g. whether the stool microbiome is predictive of treatment toxicity) are dependent upon the participant's baseline data.

This study will consist of potentially five study visits as described below.

The first study visit (2 hours) will occur at the BSWH center at which they were initially given the diagnosis of CRC. This visit will occur after consent (after colonoscopy and stool collection at baseline) and screening has occurred and prior to surgery or anti-cancer treatment is initiated. For stool samples, subjects will be given a bag with collection devices, a **Stool Collection History Form,** and instructions (**Comprehensive Stool Kit FED EX Ready)** for collecting stool and contacting the REA for stool drop off or pick up (**instruction for Retruning Your Specimen Samples)**. The patients will also be provided with a mailer box if they would rather mail in their specimen, which is stable on ice packs (provided with sstool kit) for 24 hours. Prior to Visit 2, patients will complete 2 online dietary assessment tools, the food frequency questionnaire called the **DHQIII (an online survey that takes 30-60 minutes**) and **24-hour recall called the ASA24 (an online survey that takes 20-30 minutes**), at home prior to their outpatient visit (after surgery) and prior to the beginning of treatment. If the participant needs assistance with either of these surveys, the REA will guide the patient through the online assessment from their computer. This dietary assessment tool is well validated and has been used in numerous large population studies for collection dietary pattern information. Participants will also complete a medical history questionnaire (**COLON-MD Questionnaire 2020-2023**) in an interview with one of the BSW REAs at the time of their first visit.

On the day of surgery, blood will be collected as part of SOC. During surgery, tissue will also be collected from resection, tumor and non-tumor adjacent (2 of each; one for frozen and one for fixation).

The second study visit (1 hour) will take place approximately three to four weeks (weeks 3 or 4) after surgery in the outpatient clinic. During this study visit, an additional 2 vials of blood will be taken at the same time as part of the routine clinical visit, and the patient will be given another stool sample kit to take home to collect stool and instructions. The patient will again be asked to take the 24-hour recall online **(ASA24)** complete it at home online within 72 hours of producing the stool sample, and complete the **Stool Collection Form**.

The third through fifth study visits (1 hour each) will take place on the 2nd(TD 15), 6th (TD 71) and 10th (TD 128) rounds of IV chemo of treatment (i.e. chemotherapy) at the BSWH center at which they are having their treatment. During this study visit, 2 blood vial samples will be taken as part of the routine clinical visit and the patient will be given two additional stool sample kits to take home to collect stool on the day before chemo starts and before chemo begins and the 2nd day and after their chemotherapy round is complete (i.e. if the infusion starts on Monday, the subject collect a stool sample on that Friday). After

each study visit, the patient will produce 2 stool samples (day before chemo starts and on the 2nd day after their chemo round ends) and drop it off at their treatment hospital or mail (FedEx) it into to the sequencing center at Baylor University. Again, they will be asked to complete the online 24 hour dietary recall **(ASA24)** at home within 72 hours of producting the stool samples. This is the same procedure that will be followed in the fourth and fifth study visits. During each of these visits, the REA or GI physician will ask about incidence and severity of chemotherapy induced diarrhea (**CTCAE – Gastrointestinal/Diarrhea, Grade 1-5)**.

The REA will either call the subject or send a text message (if subject agrees) through Twillo after each visit to remind the subject of the days that stool collection and 24 hour (ASA24) recalls should occur.

The REA will obtain interval information from the participant according to the protocol and will provide subjects with stipends for each stool collection provided, which will include $20 for each specimen collected (stool) up to $160. If the patient was unable to provide a stool sample at the designated time point due to treatment side-effects they will still be given the compensation for completing the study. Thus, if patients complete all sample collections and all follow up visits, they may receive up to $160.

Only Dr. Greathouse and the REA will have access to the data-collection sheets and registration forms that will be kept in a locked cabinet at the clinical site.

Primary data endpoint: At the end of the observational study we will assess incidence, duration and severity of chemotherapy-induced diarrhea.

Secondary data endpoints: The secondary outcomes will include other side effects, including oral mucositis and constipation, circulating cytokine levels, fecal/tissue/blood metabolites, nutrient intake, and changes in microbiome from baseline.

Data Collection

Stool microbiome:
Stool will be obtained at home by the patient prior to colonscopy and bowel prep, after surgery but prior to FOLFOX treatment, and on the 2nd day after TD15, TD71 and TD128. The patient will obtain stool samples according to the directions provided in the kit (**Comprehensive Stool Kit FED EX Ready**). Subjects will collect stool in the stool hat after first voiding urine, and will use the collection vial with scoop to obtain a stool sample. The patient will then put the labeled stool collection vial into the biohazard bag and put the bag into the shipping box along with three frozen ice packs. The patient will either drop the stool samples in the box with ice packs to their treatment hospital or call for FedEx pick up at home. The samples will be recived at the Greathouse lab at Baylor University (BSL2) and will be be shipped to Dr. Joseph Petrosino at the Alkek Metagenomics and Microbiome Research Center in Houston, TX at Baylor College of Medicine. This will occur for each time point in the study for stool collection.

DNA extraction: For fresh extraction, samples will arrive in PowerBead tubes provided. For banked approach samples, initial thawing on ice with subsequent sample addition to PowerBead tubes. Nucleic acid is freed with extensive vortexing, and subsequent transfer of supernatant to collection tubes for extraction using filter-based centrifugation in a modified MoBio protocol (HMP protocol).

Sequencing overview: The predominant metagenomics sequencing strategy will focus on 16S rDNA MiSeq sequencing using the V4 region primers. This takes advantage of existing and fully functional technology and pipelines with proven reproducibility and quality reads in a self-sampled population studies. In addition to bacteria, archaea will be surveyed by the addition of different primer sets to amplify archaeal 16S rDNA sequences, and fungi will be identified by 18S rRNA gene/ITS-based sequencing. Sequencing of the bacterial 16S rRNA gene targets will be performed in collaboration with

the existing HMP Genome Sequencing Centers (Baylor College of Medicine). Proper positive and negative controls will be in place to control for contamination.

Bioinformatics overview: Read data will be filtered and analyzed using our validated 16S rRNA analysis pipeline in the Greathouse lab, which is housed on the Kodiak server at Baylor University. Sequencing analysis will be conducted using the QIIME bioinformatics pipeline and operational taxonomic unit (OTU) classification will be obtained. Specifically, 16S rRNA gene sequence data will be conducted by first demulitplexing fastq files generated by the MiSeq platform first screened for length and quality using QIIME. Reads containing more than five consecutive low quality base calls (Phred < Q20) will be truncated at the beginning of the low-quality region. Passing sequences will be required to have high quality base calls (≥ Phred Q20) along with a minimum of 75% of the read length to be included. After primer removal, final sequences containing ambiguous bases (Ns) or lengths less than 150bp will be removed. High quality sequences will be screened for spurious PhiX contaminant using BLASTN with a word size of 16. Reads will be assessed for chimeras using USEARCH61 (de novo mode, 97% identity threshold for clustering). Non-chimeric sequences will be screened for contaminant chloroplast and mitochondria using the RDP naïve Bayesian classifier, as well as non-specific human genome contaminant using Bowtie2 against the UCSC hg19 reference sequence. Finally, sequences will be evaluated for residual contaminants using BLASTN searches of the GreenGenes database (v13.5). High-quality passing sequences will subsequently be clustered into operational taxonomic units using the open-reference operational taxonomic unit (OTU) picking methodology implemented within QIIME using default parameters and the GreenGenes database (99% OTUs) supplemented by reference sequences from the SILVA database (v111). Prior to downstream diversity analyses, taxa with <5% abundance in all samples will be dropped from downstream analysis. Alpha diversity estimators and beta-diversity metrics will be computed in QIIME with differential abundance analyses performed in R. Mann-Whitney tests corrected for multiple testing (Benjamini–Hochberg (FDR)) will be used to conduct initial comparisons between time points.

Metabolite Analysis:
The stool samples collected and stored in 95% ethanol will undergo metabolite analysis (using the iHMP protocol; https://ibdmdb.org/cb/browser/Clinical%20Protocol) in the laboratory of Dr. Lavado (Collaborator, BU) using high-performance liquid chromatography for SCFAs and indoles according to previous methods, as well as, putative metabolites identified from the PIPHILLIN algorithm. We will use a chromatographic approach based on Tuomola et al., to quantify these indoles and their other analogs. Stool samples will also be analyzed for SCFA (acetic acid, propionic acid, butyric acid, isobutyric acid, valeric acid, and isovaleric acid) with liquid chromatography associated to mass spectrometry.

Circulating Cytokines:
Levels of circulating cytokines at baseline (prior to surgery), prior to FOLFOX (treatment naive), and during treatment will be captured at the same time as regular blood draws to minimize patient risk and discomfort. Clinical staff will be provided with labels for each patient attached to their chart to apply to the tubes at the time of blood draw. All clinical specimens will be labeled with the participant code, date of sample acquisition and type of sample (e.g. blood 01), then placed in the clinical laboratory refrigerator within 2h after collection to maintain storage at 4°C until processing serum. Serum samples will be centrifuged and immediately stored at -80°C until shipped to the Greathouse lab in batch for processing. Cytokine measures will be conducted using the Luminex technology.

Tissue Biofilm Quantification:
Invasive biofilms (>200 micrometers) will be assessed after CRC and paired normal tissue is collected from enrolled patients undergoing surgery at Baylor Scott and White. The pathologist will use aseptic techniques to process the tissues immediately after tissue collection. All tissue not needed for pathologic diagnosis will be rapidly preserved in formalin, Carnoy's solution and RNAlater for analysis. For patients in this study, two bowel preparations will be used and recorded (mechanical bowel preparation [Miralax™], or Fleet Phospo-soda™ enema). Pre-operative intravenous antibiotics are administered in all

cases (cefotetan or clindamycin/gentamycin). Oligonucleotide probes will be synthesized and conjugated at the 5' end to the fluorophore Cy3, Alexa 405, Alexa 488, Oregon Green 514, Alexa 514, Alexa 555, Rhodamine Red X, Texas Red X, Alexa 633 or Alexa 647 (Invitrogen Life Technologies). Twelve probes will be applied to 5 μιη thick Carnoy 's solution-fixed paraffin-embedded tissue sections. All samples collected will also be stained with the Eub338 universal bacterial probe to detect all bacteria present on the sample. Successive sections will be stained with Periodic acid Schiff (PAS) to confirm mucus presence and preservation. Samples that are determined to have a bacterial presence by universal probe will be analyzed using group and species specific probes. Slides will be de-waxed following standard procedures. Probes will be applied to slides at a concentration of 2 pmol/ul in prewarmed hybridization buffer (900 mM NaCl, 20 mM Tris pH 7.5, 0.01% SDS, 20% formamide). Slides will be incubated at 46°C in a humid chamber for 2 hours, and washed at 48°C for 15 minutes in wash buffer (215 mM NaCl, 20 mM Tris pH 7.5, 5 mM EDTA). Slides will be dipped in water, then in 100% ethanol, air-dried, and coverslips mounted using ProLong Gold antifade reagent (Life Technologies). Biofilm bacterial density and depth will be measured using slides hybridized with the universal bacterial probe, Eub338, and imaged at lOOOx magnification with a Nikon E800 microscope and Nikon NIS elements viewing software.

Dietary Analysis:
Subject will complete dietary assessments within a 24 hour window of time of stool collection for each stool collection timepoint listed in figure 1. Through the National Cancer Institute (NCI) ASA24-2016 web-portal, the project PI's will register the study allowing utilization of the free ASA24-2016 Dietary Recall instrument. Study participants will be instructed to complete the ASA24-2016 within 24 hours of stool collection at each time point. Participants will also be provided instructions for login to the ASA24-2014 using their confidential study ID. Similarly, the PI will register the study for use of the FFQ to allow the online use of the Diet History Questionnaire. The study participants will be instructed to also complete the online DHQIII prior to the first FOLFOX treatment. The ASA24-2016 asks about food consumed within the last 24 hours based on the on USDA's Automated Multiple-Pass Method (AMPM). Participants have approximately five screens to detail their food intake with prompts to remember food in between meals, source of preparation for the food and a question that asks if food intake was reflective of the participants' typical diet. The recall is respondent-driven in that the detailed questions asked are based on a series of linked question pathways.[50] ASA24 also does not collect any identifying data directly from respondents. However, IP address information is accessed for the purpose of routing information between the server and the respondent's computer -- often the IP address is that of the user's Internet Service Provider (ISP). IP addresses are not stored or tracked by ASA24. However, logs of connections are kept for audit trail purposes. This information is not mined in any way but would be available if there were a legal obligation to release it. Similarly, the Food Frequency Questionnaire (FFQ) or DHQIII also is usually self-administered and has recently become available as a web-based tool, specifically for the Diet History Questionnaire (DHQ). The DHQ was developed and validated by the NCI in two separate studies.[51,52] Recently, DHQ III Ihas a food list that has been updated from DHQ I based on more recent dietary data and consists of 134 food items and 8 dietary supplement questions. This web-based version takes <1 hour to complete and can be analyzed with the Diet*Calc software developed by the NCI. This version will be completed by study participants at the time point prior to FOLFOX treatment.

Medical and Symptom History:
Specific subject information will be collected via review of the medical record(s). Specifically, data will be collected apart from that captured on the medical record, which will include minimal information on socioeconomic status (education, income), tobacco and alcohol history and medication history (**COLON-MD Questionnaire 2020-2023)**. At each clinical visit after surgery and during the treatment time points, the REA will assess mucositis and diarrhea grade using the **CTCAE – Gastrointestinal/Diarrhea, Grade 1-5**.

Risks & Benefits:

The benefits of the study are improved survival for patients, less toxicities for patient on chemotherapy, and improved quality of life. Also, this study will help contribute to the growing body of knowledge regarding colon cancer and could help future patients diagnosed with colon cancer. Since we are collecting stool specimens which does not expose patient to any adverse procedures or drug or potential side effects there is very minimal risk, and substantial benefit for the patient and for society in general.

Data handling and storage:

The research data will be stored on a secure server at Baylor University. Once samples are collected, identifiers will be removed the Baylor Scott and White clinical research coordinator/staff and only the Study ID will be retained. Subject's data and all sequencing data will be maintained in a coded fashion on secure servers at Baylor University. Subject confidentiality will be strictly held in trust by the participating investigators, their staff, and the sponsor(s) and their agents. This confidentiality will be extended to cover testing of biological specimens in addition to the clinical information relating to participating subjects. The study documentation, data and all other information generated will be held in strict confidence. No information concerning the study or the data will be released to any unauthorized third party without prior written approval of the sponsor. The study monitor or other authorized representatives of the sponsor may inspect all documents and records required to be maintained by the Investigator, including but not limited to, medical records (office, clinic or hospital) and pharmacy records for the subjects in this study. PHI provision will be observed with all patients information will be coded on the study except on consents that only KSP will have access to. Subjects on the study will be given an ID that only the PI will have deoding key to. The nature of the information collected from subjects may involve risk to their privacy.

4.4 Data Analysis.

Classification of taxa associated with side effects or severity: Recently, in an animal model of colon tumorigenesis (AOM/DSS), investigators used the initial microbial community structure (16s rRNA sequencing) to predict tumor burden after exposure to various antibiotics during colon carcinogenesis. In order to conduct a similar analysis and reduce data dimensionality, we will first perform cluster modeling using Dirichlet multinomial mixtures (Holmes et al. 2012) to identify microbial clusters (community state types). To determine which taxa or taxonomic cluster are the best predictors of treatment side effects (CID, mucositis) and severity (stage 3-4), we will conduct 10-fold cross-validation using the RandomForest package (Version 3.6.4) in R (R-studio, Version 0.98.1091; Boston, MA). Specifically, 5,000 trees will be used and the parameter mtry will be set at 100. The taxa with the highest importance scores will be used for further analysis. This procedure will be tested by splitting the data into test and training sets, then running Random Forest using the same procedure described above on the training dataset and comparing the predictive accuracy of the top taxa (based on importance scores) in the test set. Based on this 10-fold cross-validation, we will be able to calculate sensitivity and specificity. These top taxa will then be used to generate receiver operator curves (ROC) and calculate area under the curve (AUC). Each microbial cluster or taxa will undergo independent logistic regression testing to calculate area under the ROC curve using Stata (v13). The command roccomp will be used to compare statistical difference between AUCs and determine the best combination of taxa for classification of CID or mucositis and for severity.

Microbiome-Diet Analysis: To evaluate the impact of diet on the stool and tissue microbiome, we will use the dietary analysis generated from the ASA24 and DHQIII, which includes the Healthy Eating Index (HEI). Prior to analysis, we will conduct normality testing for each variable using Shapiro-Wilk test. For any variables that do not pass, a logarithm transformation will be performed. In order to identify any potentially important diet or nutrient-microbiome associations, we will conduct correlation analysis between dietary macronutrients (carbohydrates, fats, proteins) or micronutrients (vitamins and minerals)

and microbial clusters (community state types) at baseline using Kruskall-Wallis, and for individual taxa using Pairwise Spearman correlations. We will use PERMANOVA for comparing beta-diversity (weighted UniFrac distances) at baseline by levels of macro- and micronutrients. DESeq2, Kruskall-Wallis and Mann-Whitney tests will be conducted to identify significant differences in tissue and stool taxa or microbial clusters according HEI scores at baseline. The Benjamini and Hochberg false discovery rate (FDR) will be used to correct for multiple testing. Taxa or microbial clusters found to be significantly differentially abundant between treatment time points and baseline will be used to conduct 'random forest' modeling, as previously described, to identify dietary factors that best predict changes in these taxa or clusters. Additionally, we will investigate the tissue and stool microbiome-diet relationship using principal components analysis to identify significant trends among the dietary factors using coinertia analysis. To investigate the microbiome-diet relations during treatment, we will use the microbial clusters and calculate the $\log_2$ fold change in the clusters from baseline. We will then determine how much the clusters change from baseline according to dichotomous values of macronutrients.

Microbiome-treatment analysis: To evaluate differential abundance comparing treatment time points to baseline (prior to surgery and post-surgery/pre-treatment), DESeq2 will be used to control operational taxonomic unit (OTU) over-dispersion between replicates; OTUs will be considered significantly differentially abundant between time points if the false discovery rate (FDR) corrected p-value is below 0.1. A change in relative abundance will be defined as a fold change between mean log relative abundance during treatment phase (week 4, 14, 24) and baseline (prior to surgery). Taxa found to be significantly differentially abundant between treatment time points and baseline (FDR corrected p-value <0.05) will be put into an adjusted regression model to control for age, education, income level, marital status, and activity level, diet quality (HEI), tumor stage, medical history with p-values corrected for multiple hypothesis testing using FDR adjustment.

Community state type modeling of treatment effect: As describe earlier, in order to reduce data dimensionality and improve power (due to small sample size), we will perform cluster modeling using Dirichlet multinomial mixtures, as well as, use the partitioning around the medioid (pam in R) algorithm. To calculate the dissimilarity matrix for this analysis, we will first calculate weighted UniFrac distance, then this matrix will be dimensionally reduced by choosing the most significant eigenvectors from PCoA analysis. Community state types (CST), which are defined as 'clusters of community samples with similar compositions of microbial taxa', will be obtained from this analysis. Statistical difference in CST abundance between treatment and baseline will be conducted using median values. Plots of CSTs for each subject over time will be used to identify associations between CSTs and incidence of CID, mucositis and severity of side effects. In order to model the dynamics of the CST transitions, we will use Markov chain to estimate the transition probabilities, as demonstrated for vaginal CST associated with pre-term birth [9]. Transition rates between CSTs for each set of paired samples will be calculated using maximum-likelihood estimates. Additionally, differential relative abundance analysis of CSTs between baseline (pre-surgery and pre-treatment) and treatment time points will be conducted using linear mixed models to account for the repeated measurements at the 4 time points and missing observations. To evaluate differences in beta diversity between treatment and baseline the PERMANOVA test will be used. The inferred function of the microbiota will be obtained using the bioinformatics tool PICRUSt, which uses 16S rRNA marker genes to predict presence of gene families and construct a composite metagenome. This method has been used successfully in several similar studies to predict key metabolic functions associated with taxa important in distinguishing disease from healthy controls.

Metabolite-CID analysis: In a similar manner to the Microbiome-CID analysis, we will conduct logistic regression analysis between CID (present/absent) and metabolites at baseline, and at individual time points controlling for multiple hypothesis testing (FDR <0.05). Further, principal component analysis and partial least squares-discriminant analysis will be conducted to determine the metabolic alteration between CID and nonCID subjects after mean centering and unit variance scaling. The variable importance scores of all of the peaks for the tenfold cross-validated PLS-DA model will be considered for

peak selection and biomarker classification. Additionally, metabolites found to be significantly different between treatment time points and baseline will be put into an adjusted linear mixed model regression.

Microbiome-Cytokine Analysis:
To identify key microbes associated with inflammation in colon cancer we will conduct serum cytokine profiling using the Luminex array platform. Plasma samples will be taken from patients prior to surgery and stored at -80° C. Plasma cytokine concentrations (32 cytokines) will be measured in the laboratory of Dr. Greathouse and Dr. Brid Ryan, who has expertise in conducting cytokine assays and analysis. We will use the Luminex Plate Array. All samples and standards will be run in duplicate and averaged, and calibration curves will be run on each plate using the standards provided. Prior to analysis, we will conduct normality testing for each variable using Shapiro-Wilk test. For any variables that do not pass, a logarithm transformation will be performed. Analysis of circulating cytokine levels will be analyzed using the Kruskall-Wallis test across CSTs by time point, as well as, by linear mixed models to account for the repeated measurements and missing observations using quartiles or dichotomous values of cytokine levels. We will also perform correlation analysis using Spearman correlation on measures of alpha diversity and cytokine levels at baseline, and PERMANOVA for measurement of beta-diversity stratified by level of cytokine

5. Human Subjects

Protection of Human Subjects

A. Human Subjects Involvement and Characteristics

Human subjects involvement: Individuals with a positive diagnosis upon histological examination for colon cancer will be recruited from the Baylor Scott and White McClinton Cancer Center in Waco, TX and The BSWH Temple, TX locations. The Baylor Scott & White McClinton Cancer Center (BSWMCC), in Waco, Texas, and BSWH Clinic in Temple, TX provides ambulatory medical, radiation, and supportive hematology/oncology services to adults five days per week. In addition, the BSWMCC/BSWH Temple Clinics provides on-site laboratory and pharmacy services in conjunction with medical oncology and hematology care and has infusion services. The CRC Group at BSWMCC and BSWH Temple is responsible for the orchestration of day-to-day aspects of clinical research such as patient eligibility, appointments, informed consent, data collection and transcription of information from medical records onto special data collection forms.

- Patients enrolled in this study represent a vulnerable population of terminally or critically ill individuals. For this reason, those staff, Physicians and REAs, will explain with care and clarity that as a patient they will not experience any personal medical benefit from their participation in this study, but that findings may benefit future patients undergoing similar treatment. It will also be made clear that this is an observational study and will in no way impact their treatment or standard of care, and if they choose not to participate will also not affect their treatment in any way.

- Characteristics of subject population: We will recruit 112 stage II/III patients who fit our inclusion criteria (see inclusion/exclusion criteria). Baseline characteristics of current patients undergoing treatment at BSWMCC reflect the demographics of the greater Temple and Waco area: 44.7% White, 30.4% Hispanic and 21.5% Black, 1.7% Asian. During 2016, 109 total CRC cancer cases were treated (Table 1).

-

Table 1. 2016 Colorectal Adenocarcinomas (BSWH)*

Characteristic	N (%)
Stage	
I	3(3)
II	25(23)

III	56(51)
IV	25(23)
Sex	
M	43(39)
F	66(61)
Race/Ethnicity	
Non-Hispanic white	74(68)
African-American	21(19)
Hispanic	11(10)
American Indian	3(3)

*Combined Temple and Waco locations

B. Potential Risks

- Potential risk: potential risks associated with this study are no higher than the risk of completing these procedures for the anti-cancer treatment they will undergoing. The alternatives to this procedure introduces greater risk and requires discomfort of a rectal exam and additional clinical visit. Subjects may choose to not participate in the voluntary study.
- Potential benefits: patients may benefit from this study by learning about their dietary history and regular dietary habits, and the potential effects on their response to therapy and type of and severity of side-effects. They may be able to use this information to reduce side-effects that may be a result of diet-treatment interactions.

References

1 Zackular, J. P., Baxter, N. T., Chen, G. Y. & Schloss, P. D. Manipulation of the Gut Microbiota Reveals Role in Colon Tumorigenesis. *mSphere* **1**, doi:10.1128/mSphere.00001-15 (2016).

2 Nakatsu, G. *et al.* Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* **6**, 8727, doi:10.1038/ncomms9727 (2015).

3 Baxter, N. T., Ruffin, M. T. t., Rogers, M. A. & Schloss, P. D. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med* **8**, 37, doi:10.1186/s13073-016-0290-3 (2016).

4 Kurita, A. *et al.* Streptomycin alleviates irinotecan-induced delayed-onset diarrhea in rats by a mechanism other than inhibition of beta-glucuronidase activity in intestinal lumen. *Cancer Chemother Pharmacol* **67**, 201-213, doi:10.1007/s00280-010-1310-4 (2011).

5 Sender, R., Fuchs, S. & Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* **164**, 337-340, doi:10.1016/j.cell.2016.01.013 (2016).

6 Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* **32**, 834-841, doi:10.1038/nbt.2942 (2014).

7 Gevers, D. *et al.* The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS biology* **10**, e1001377, doi:10.1371/journal.pbio.1001377 (2012).

8 Knight, R. *et al.* Unlocking the potential of metagenomics through replicated experimental design. *Nature biotechnology* **30**, 513-520, doi:10.1038/nbt.2235 (2012).

9 DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11060-11065, doi:10.1073/pnas.1502875112 (2015).