AWARD NUMBER: W81XWH-18-1-0112

TITLE: Preventing Adverse Patient Responses to Cancer Chemotherapeutics

PRINCIPAL INVESTIGATOR: Libusha Kelly

CONTRACTING ORGANIZATION: Albert Einstein College of Medicine , Bronx, NY

REPORT DATE: June 2021

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
June 2021	Annual Report	01Jun2020-31May2021

4. TITLE AND SUBTITLE

Preventing Adverse Patient Responses to Cancer Chemotherapeutics

5a. CONTRACT NUMBER
W81XWH-18-1-0112

5b. GRANT NUMBER
CA171019

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Libusha Kelly

E-Mail: libusha.kelly@einstein.yu.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Albert Einstein College of
Medicine
1300 Morris Park Ave,
Bronx, NY, 10461

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this project is to prevent adverse patient responses to the cancer drug irinotecan by analyzing the gut microbiomes of patients. The scope of this project is to study irinotecan metabolism and the microbiome over time using fecal samples from healthy individuals and metastatic colorectal cancer patients. We are happy to report a number of significant results for this year. We have sequenced the metagenomes of 39 fecal samples collected over time from six metastatic colorectal cancer patients. This unique dataset will be of substantial interest to the community because there are no available temporal metagenomic datasets for colorectal cancer patients and no datasets that include data on adverse events and treatment regimens. We note that our microbiome enzyme of interest, beta-glucuronidase, which can interact with metabolites of the colorectal cancer drug irinotecan, changes its distribution in one patient during transient diarrhea. Among our publications for this year, our work describing a novel computational approach, topological data analysis, to identify metastable states and state transitions in microbiome data that are linked to patient outcomes was published (Chang, VanInsberghe, and Kelly, *npj Biofilms and Microbiomes*, 2020) and will be used to analyze the temporal colorectal cancer patient samples sequenced this year. We published an additional three original research papers this year which acknowledged DoD funding, bringing the total number of peer-reviewed publications this grant has supported to 11.

15. SUBJECT TERMS

Colorectal cancer, drug metabolism, microbiome, carbohydrate active enzymes, phase II drug metabolism, metabolomics, metagenomics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	23	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

INTRODUCTION:

The microbiome shapes the metabolic and immunological landscape of individuals in health and disease. Its plasticity can be leveraged for therapeutic interventions and to improve therapeutic outcomes. Recent studies have implicated gut microbiome metabolism at the gene and species level in driving the variability in patient drug response and toxicity. One of few therapeutic drugs for which we have a mechanistic understanding of how the gut microbiome influences drug metabolism is the colorectal cancer chemotherapeutic and prodrug irinotecan (CPT-11). CPT-11, in combination with fluorouracil and leucovorin, is one of three first- line treatments for metastatic colorectal cancer. Reactivation of the drug by beta-glucuronidases (BGs) in the gut can lead to severe diarrhea in patients. We hypothesize that individuals with high gut-driven turnover of SN-38G are at heightened risk for ADRs and can be identified via microbiome-based pretherapy analysis. Our overall objective is to identify patients at high risk for adverse events by non-invasive fecal sampling. The results will provide a clinical forecast for therapy in high-risk patients.

1. KEYWORDS:

Colorectal cancer, drug metabolism, microbiome, carbohydrate active enzymes, phase II drug metabolism, metabolomics, metagenomics

2. ACCOMPLISHMENTS:

What were the major goals of the project?

For all major tasks I have included only subtasks that were designed to be completed within the first 24 months of the project, I have not included tasks that were already completed in months 1-12 of the project.

Major Task 1: Quantify CPT-11 metabolites in healthy and metastatic colorectal cancer Patients

Subtask 4: Collect fecal samples from 20 healthy individuals and quantify metabolite production over time. We target 5 samples per individual. Samples will be used for both metabolite analysis and for metagenomic sequencing. For both Subtask 4 and Subtask 5, concentrations of SN-38G, SN-38 and the ISTD in the fecal extracts will be determined used, the Agilent G6490 Triple Quadrupole Mass Spectrometer. We will examine our mass spectrometry data for any additional, closely structurally related, metabolites of CPT-11 that have not been previously described (**~25% COMPLETE, we have collected 22 individual fecal samples from healthy individuals, due to COVID-19 we have not collected additional samples this year**)

Subtask 5: Collect fecal samples from 20 metastatic colorectal cancer patients, targeting 5 samples per patient, and quantify metabolite production over time per Subtask 1 **(~40% COMPLETE, we have collected 39 samples total from six metastatic colorectal cancer patients and have developed an in-house pipeline to quantify metabolite production which has been tested and validated on patient samples.)**

Subtask 6: Correlate adverse responses in metastatic colorectal cancer patients with metabolite production. Parametric and nonparametric tests will be used to identify significant differences between adverse events quantified as continuous outcomes (number of instances of diarrhea, severity of diarrhea), and chi-square tests will used to compare categorical outcomes (diarrhea/no diarrhea). All statistical tests will be two-tailed, and p values of less than 0.05 will be considered statistically significant. Analyses will be performed with the R statistical software package. (~**50% COMPLETE, we have developed and tested R scripts for the statistical analysis and we have developed an in-house pipeline to quantify metabolite production which has been tested and validated on patient samples.)**

Major Task 2: Quantify beta-glucuronidase abundance and taxonomy in colorectal cancer patients over time

Subtask 1: Sequence fecal metagenomes of 20 metastatic colorectal cancer patients (from Major Task 1) using Illumina NextSeq sequencing, with a target of 3.5 M paired end reads and 1 Gb sequence per sample (**40% COMPLETE, we have sequenced 39 samples taken over time from six metastatic colorectal cancer patients**)

Subtask 2: Correlate beta-glucuronidase abundance with adverse responses to CPT-11 by comparing all reads with our in-house database of beta-glucuronidases and comparing the relative abundance of specific beta-glucuronidases in patients who suffer adverse responses (diarrhea >= grade 3) with those patients who do not suffer adverse responses (**25% COMPLETE, we have collected adverse response data from patients for which we have 27 samples, we have extracted DNA from the 27 samples and prepared it for sequencing, and we have our computational analysis pipeline ready for analysis**)

Major Task 3: Activity-based protein profiling of functionally active human gut microbiome β-glucuronidases (**0% COMPLETE, our efforts to overcome the problems with this task were hampered severely by COVID-19. We propose moving to a proteomics-based approach to overcome these hurdles, this potential alteration is discussed below in CHANGES/PROBLEMS**)

Subtask 1: Optimize synthesis of a custom fluorescently labeled SN-38G probe in collaboration with the Einstein Chemical Synthesis core. We currently have low, impure yields that are not yet sufficient for our experiments

Subtask 2: Validate uptake of SN38-G labeled probe using positive and negative controls. Positive control: E. coli strain ATCC 25922 which can convert SN-38G to S38. Negative

control: E. coli strain BW18812 (ΔuidA), which lacks the BG gene and thus should not convert SN-38G to S38G. Successfully sort labeled cells via flow cytometry

Subtask 3: Optimize flow sorting of fecal samples; specifically identify optimal sample concentrations and buffer conditions to reliably sort these very heterogeneous samples. Optimize sorting of cells that uptake the labeled probe by defining appropriate parameterizations for sorting and by quantifying the populations of cells that have taken up the labeled probe

Major Task 4: Quantify microbiome gene expression during SN-38G exposure. (0% COMPLETE, the problems with this task are discussed below. COVID-19 hampered our ability to overcome the problems associated with this task. We are back in the lab and beginning to attempt new extraction methods to deal with this roadblock.)

Subtask 1: Amend fresh fecal samples from the same 6 healthy volunteers referenced in Major Task 3, 3 high and 3 low metabolizers, with SN-38G, extract RNA at timepoints corresponding to known metabolism of SN- SN-38G. Sequence total RNA with a target of 12.5 M PE reads and 3.8 Gb of sequence per sample.

Subtask 2: Analyze RNASeq data to identify genes significantly associated with SN-38G -> SN38 conversion using the SAMSA pipeline, an open source tool which breaks down metatranscriptome data by organism and transcript function.

What was accomplished under these goals?
For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

Major activities

- We have sequenced the metagenomes of 39 fecal samples collected over time from six metastatic colorectal cancer patients. This is a unique dataset that will be of substantial interest to the community because, to our knowledge, there are no available temporal metagenomic datasets for colorectal cancer patients and no available temporal datasets that include data on adverse events and treatment regimens.

- We have begun to analyze these 39 metagenomic datasets with an eye towards BG composition, temporal stability, and associations with patient metadata such as adverse events and changes in treatment regimen. We detail our preliminary studies below.

- We have begun to quantify an observation made previously that oral bacteria found in the gut are a hallmark of colorectal cancer. Our large meta-analysis of microbiome datasets statistically associates oral bacteria in the gut with disease and our proteomics data,

described below, suggests that oral bacteria are not mere travelers to the gut in colorectal cancer but are also metabolically active.

- We have collected the first proteomics datasets from one colorectal cancer patient in our study. Again, as with our metagenomes, to our knowledge there are no temporal proteomics datasets for colorectal cancer patients and we are beginning to analyze the microbial and functional properties using protein information of our patient samples. We note that we observe putative oral bacterial proteins in our dataset, strongly supporting the hypothesis that these bacteria are functional and influencing the microbial ecosystem of colorectal cancer patient guts.

- During this funding cycle we have published three original research papers and one invited commentary. All research papers acknowledge this funding from the Department of Defense. This brings the number of manuscripts that this grant has been acknowledged on since the commencement of funding to twelve total publications.

Specific objectives

We conducted longitudinal sampling on six colorectal cancer patients who are on regimens that include irinotecan. Samples were prepared for sequencing and metabolomic analysis and an in-house method was tested to assess metabolite production in patient fecal samples (**Major Task 1**, **Major Task 2**). DNA was extracted from patient samples, prepared for sequencing, and sequenced. Computational analysis of these datasets has begun and has already identified differences in patient microbiomes over time that are detailed below. Proteomics datasets were collected from three patient samples to further assay function. We apply computational methods for meta-analysis of microbiome data to this data and identify oral microbes as hallmarks of colorectal cancer (**Major Task 2**). We continue to struggle with RNA extraction from incubated fecal samples but have reached out to experts at Rutgers and are testing a new protocol (**Major Task 4**). The COVID-19 pandemic has delayed our efforts to improve our approach to study microbial uptake of glucuronidated substrates (**Major Task 3**). Both of these issues are discussed in detail in "5 CHANGES/PROBLEMS".

Significant results

We are happy to report a number of significant results for this year. We have successfully sequenced longitudinal samples for six colorectal cancer patients. Three of the six patients suffer from sporadic diarrhea during treatment. Three patients did not report diarrhea. We have begun to analyze this data to distinguish microbiome features associated with diarrheal events from those in patients who do not suffer diarrhea.

The Bronx, our lab's home, was an early epicenter of the COVID-19 pandemic in the USA. We conducted temporal genomic surveillance of SARS-CoV-2 genomes across the Bronx from March-October 2020. Mapping the trajectories of variants, we found that while some have become 'endemic' to the Bronx, other, novel variants rose in prevalence in the late summer/early fall. Geographically resolved genomes enabled us to distinguish between a case of reinfection and a case of persistent infection. We propose that limited, targeted, temporal genomic surveillance has clinical and epidemiological utility in managing the ongoing COVID pandemic.

This work was submitted to medrXiv and is currently undergoing peer review (Fels, et al, *medrXiv*, 2021). I want to note that as the final corresponding author, I was responsible for leading the large, multidisciplinary, mixed clinical and basic research team that conducted the work as well as guiding the results to publication.

Despite the pandemic, we have continued our history of strong publications on microbiome metabolism and computational microbiome analysis this year. We were invited to write a commentary on a paper describing novel microbial chemical transformations (Wolfson and Kelly, *eLife*, 2020). Our work describing a novel computational approach, topological data analysis, to identify metastable states and state transitions in microbiome data that are linked to patient outcomes was published (Chang, VanInsberghe, and Kelly, *npj Biofilms and Microbiomes*, 2020) and will be used to analyze the temporal colorectal cancer patient samples.

I contributed genome sequence analysis to a beautiful paper from my mentor for this grant, Dr. Sridhar Mani, that demonstrated for the first time that swarming behavior in bacteria is enriched during intestinal stress and can ameliorate intestinal damage in mice (De, et al, *Gastroenterology* 2021). I

Finally, I am the corresponding author on a publication demonstrating that the degradation of host translational machinery drives tRNA acquisition in viruses of bacteria (Yang and Fang, et. al, *Cell Systems*, 2021) As viruses of bacteria remain an understudied component of the gut microbiome, we will apply our approaches to identify functional virus/host interactions in our newly sequenced temporal colorectal cancer patient samples to ask to what extent viral DNA associates with colorectal cancer and with irinotecan treatment. Each of these publications acknowledges funding from the DoD.

Collection, sequencing, and analysis of samples and associated metadata from colorectal cancer patients.

We have now successfully collected and sequenced 39 longitudinal fecal samples for six colorectal cancer patients. Our whole-community shotgun metagenomic sequencing approach means that we can classify reads both by microbial taxonomy and by function. Three of six patients suffer from sporadic diarrhea during treatment (**Figure 1**).

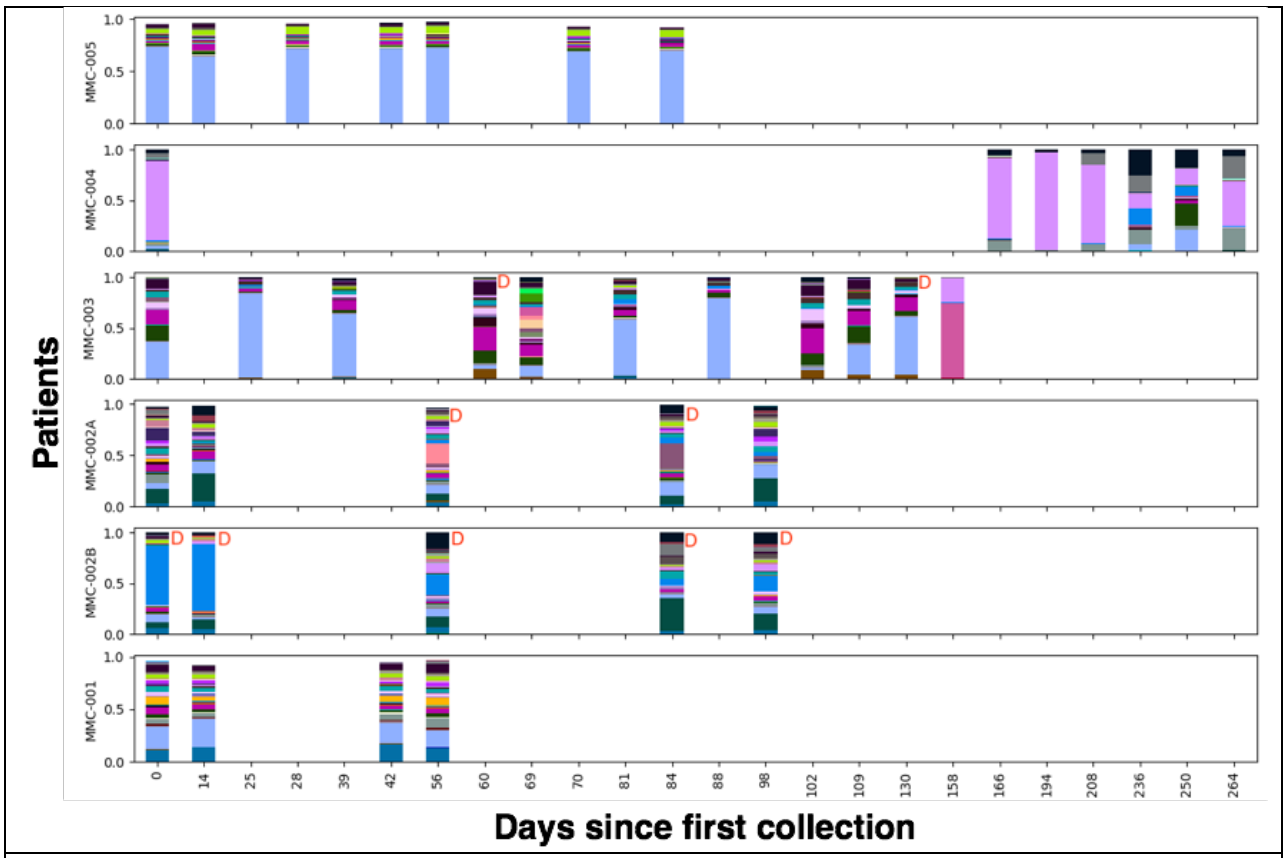

Figure 1. Longitudinal sampling of colorectal cancer patients on regimens containing irinotecan. Patient samples were collected over several months. Whole community metagenomic sequencing was performed and the relative abundance of genus level taxonomic classifications for each patient's sequences are shown. Color indicates bacterial genus. Diarrhea is indicated with a red "D".

Importantly, the longitudinal sampling means that patients can act as their own controls, therefore, even if there are not overlapping microbes between two patients that are associated with the propensity for diarrhea, we can ask, for each patient, which microbes change when the patient suffers from diarrhea and which microbes are stable over time?

In **Figure 1**, it is apparent even by eye that some patients have a more 'stable', i.e, unchanging microbiome over the course of treatment (patient MMC-005, the top row; and patient MMC-001, the bottom row). Neither of these patients reported diarrhea. In contrast, patients MMC-003 and MMC-002A all suffered sporadic diarrhea and had major shifts in their genus level microbial abundances. MMC-002B reported diarrhea for all four of their collected samples. We will interrogate these microbiome changes to ask what microbes and functions are shifting as patients report diarrhea. We will conduct these analyses at the levels of **1)** microbial taxonomy (genus classification of sequencing reads is shown in **Figure 1**, we will also assess species level and strain level differences); **2)** functional gene and pathway profiling by mapping all reads to large functional databases including eggNOG (http://eggnog45.embl.de/) and KEGG (https://www.genome.jp/kegg/); **3)** BG sequence profiling by mapping all reads to human microbiome BG enzymes as we have done previously.

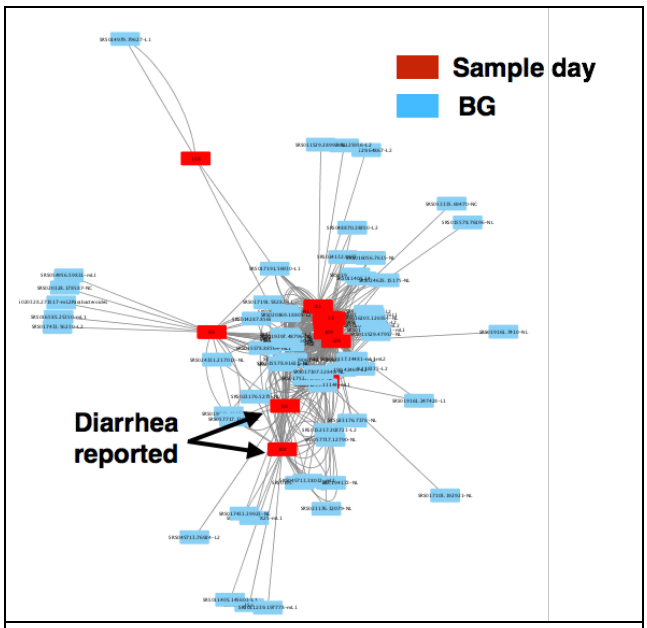

Figure 2. BG sequences in a single colorectal cancer patient sampled 11 times over the course of treatment. Nodes represent sample day (red) or a specific BG gene from the Redinbo human microbiome BG database (279 sequences total). An edge is drawn between a day and a gene if that gene was present on that day of sampling. This patient experienced two bouts of diarrhea, those two days appear close together in space due to shared BG sequences.

With this unique dataset we can also ask: how does treatment with irinotecan influence the taxonomic and functional composition of patient microbiomes over time? There are no other available studies that address this question. Here, per the overall goal of this Career Development Award, we can assess whether there are any similarities in our patients over the course of treatment with irinotecan, despite their clearly diverse microbiomes (**Figure 1**). As a preliminary analysis I worked with our new medical student volunteer, Amanda Zong, to quantify changes in the microbiomes. We identified genuses with 10% relative abundance differences, a substantial change, between samples in our six patients. For five of six patients Bacteroides fluctuated in relative abundance over the course of treatment. For two of six patients Prevotella fluctuated in relative abundance over the course of treatment. These fluctuations could indicate a conserve microbiome response to irinotecan therapy, as they appear to happen several times as patients are treated. We will address these fluctuations on a functional level, that is, what genes and pathways are altered in patients over the course of treatment and do these correlate with the genus level fluctuations we have already identified?

Identifying beta-glucuronidase genes that are associated with sporadic diarrhea.

For patient MMC-003 we have begun testing our BG annotation and analysis pipeline. Briefly, We next utilized Dr. Matt Redinbo's human microbiome BG database (Biernat, et. al., *Scientific Reports*, 2019) to conduct a protein BLAST of our metagenomic proteins against the BG database with our previously published protocol (Guthrie, et. al, 2017) to identify, for each of 11 samples from this patient, what BGs were present. We then constructed a network in which nodes represent either a sample date (**Figure 2**, red nodes) or a BG gene (**Figure 2**, blue nodes) and we draw a line between a BG and a sample date if that particular BG appeared on that particular sample date. The overall network thus represents the collection of BGs represented in this patient over all 11 samples.

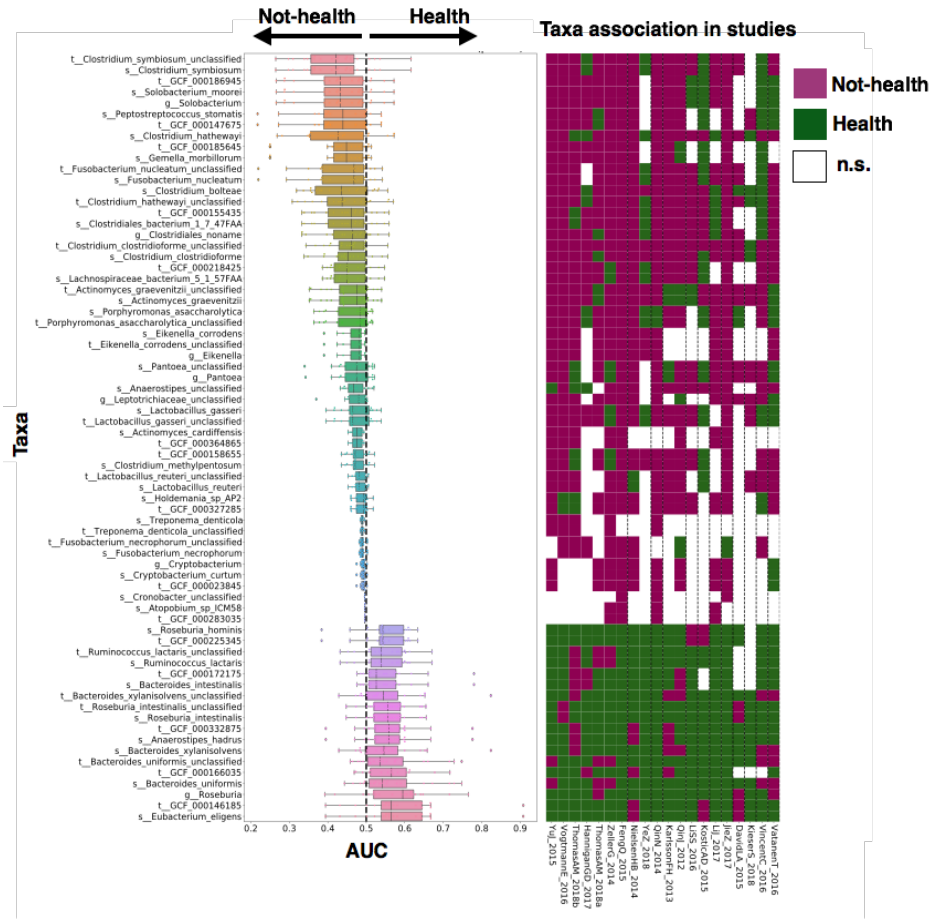

Figure 3 Taxa significantly associated with health or not-health. For each study in our dataset (20 total), we performed Mann–Whitney U (**MWU**) tests for association between study outcome and control group. Each of these studies was a case–control study of the stool microbiome in a disease setting; p-values were adjusted using false-discovery rate (FDR) correction for each study. The **MWU** test gives a p-value and an AUC score, which represents the probability that a health sample will be ranked higher than a not-health sample and, therefore, measures the effect size and direction of effect. We used a two-sided t-test to measure the null hypothesis that the mean AUC for a given taxon is 0.5 (meaning no effect). We used a cutoff of AUC > 0.5 to define health-associated taxa and AUC < 0.5 to define not-health-associated taxa. Significant taxa (left) and significant associations of the taxa in each study (right); taxa associated with health are colored green, taxa associated with disease, red.

Several aspects of this preliminary analysis are of interest. **1)** the dates on which the patient reported diarrhea are close together in space because they share many BGs, in addition many of these shared BGs do not appear on other days of sampling where no diarrhea was reported; **2)** There appears to be a 'core' BG set that is shared with almost all samples, which may represent a BG fingerprint for this individual. We will ask, as we begin to analyze our other 5 sets of patient samples, **1)** whether individuals have a unique BG fingerprint; and **2)** whether some BGs are consistently present during diarrhea and absent during non-diarrhea.

Machine learning approaches to identify signatures of health and disease.

In work published last year we developed multiclass classifiers to predict specific diseases using microbiome data(Khan and Kelly, *Pacific Symposium on Biocomputing*, 2020). Specifically, we utilized 5,643 aggregated, annotated, whole-community metagenomes to implement the first multiclass microbiome disease classifier of this scale, which was able to discriminate between 18 different disease states and a healthy state. We implemented three types of classifiers: a feed-forward deep neural network (**DNN**), a graph convolutional network (**GCN**), and a random

forest (**RF**). In the multiclass problem, instead of giving the classifier the standard problem of distinguishing a particular disease state (i.e., **CRC**) from a healthy state, we instead gave the classifier the following, more difficult problem: given a dataset containing 18 different diseases and healthy controls, identify which disease this sample originated from or whether the sample is healthy. The classifier has only a 1/N (where N = phenotype number) chance of randomly selecting the correct answer, compared with a 1 in 2 chance when selecting between a single disease and healthy controls, assuming that the samples represent a balanced phenotypic distribution, which we set. Overall, our classifiers were able to discriminate between 18 different disease states, including colorectal cancer, and a healthy state with greater than 70% accuracy.

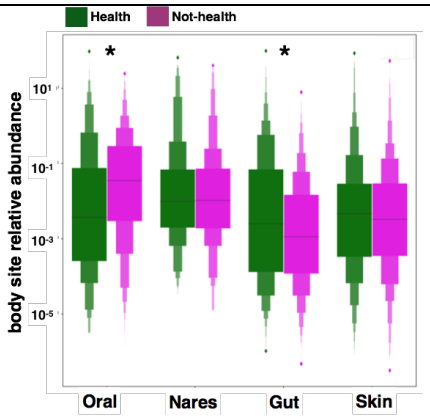

Figure 4. Oral bacteria, when found in the gut, associate with not-health. We calculated the relative abundance of health/not-health marker taxa microbes found in four body sites: oral, nares, gut, and skin. Gut marker taxa that are prevalent in the oral microbiome are associated with disease, whereas gut bacteria marker taxa, when found in the gut, are associated with health (Student's t-test, * indicates $p<0.05$).

We wished to extend this work to identify microbes that were general predictors of health and disease. Using 20 stool microbiome studies from the curatedMetagenomicData dataset, we computed associations between each taxon and disease/control and identified approximately 200 taxa that were significantly associated with either health or disease from this dataset at ($p<0.05$, Mann-Whitney U test). We show a $p<0.01$ cutoff in **Figure 3** to visualize the significant marker taxa (left) and the significant associations of the taxa in each study (right) with health and not-health states.

A startling observation was that among our marker taxa, those that are oral bacteria, when found in the gut, are predictive of not-health (**Figure 4**). A greater prevalence of oral bacteria was noted in the guts of colorectal cancer patients but to date there have been no studies addressing whether these bacteria are simply being translocated to the gut from the mouth and do not engraft due to inflammation or, instead, if they are functionally active in the gut.

Oral microbes as active members of colorectal cancer patient microbiomes.

Due to ongoing difficulties with **Major Task 3** Dr. Kelly contacted a former MIT colleague, Jacob Waldbauer, Neubauer Family Associate Professor in the department of Geophysical Sciences at the University of Chicago, to ask if he'd be willing to collaborate on using proteomics to identify active BG enzymes in colorectal cancer patient guts. Dr. Waldbauer, who is a world-recognized expert in proteomics and had previously done proteomics on fossilized giant sloth coprolites, asked us to send a few protein samples to assess the feasibility of this collaboration.

Dr. Waldbauer was able to identify 700 high-quality proteins in one patient sample, suggesting that this approach was feasible. While we did not find BG enzymes in this preliminary work, we did find proteins that are likely from oral bacteria, including *Porphyromonadaceae*, in this dataset. In combination with the computational results demonstrating the prevalence of oral

bacteria in the gut in disease states, we argue that utilizing proteomics to study our colorectal cancer patient samples is a productive, novel, and potentially transformative approach to understand microbiome function in colorectal cancer. I will arrange to speak with the PO for this award to ask if we can potentially utilize this proteomics approach, in collaboration with Dr. Waldbauer, in lieu of our original **Major Task 3.**

An in-house spectrophotometric approach to quantify microbiome metabolism of SN-38G. Ruth Hauptman, an MSTP student supported on this grant, implemented a protocol from the Redinbo group (Jariwala et. al., *ACS Chemical Biology*, 2020) that allows us to record the concentration of SN-38G over time using a spectrophotometer. Because SN-38G is fluorogenic we can measure the relative fluorescent units (RFU) that it emits and compare it to the RFU of standards with a known concentration. At excitation 260 and emission 450 the RFU of SN-38G is ~200 fold greater than SN-38; this allows us to distinguish between the two metabolites using a spectrophotometer. When SN-38G is converted into SN-38 the RFU is decreased. We have assessed this protocol with two different concentrations of SN-38G incubated with protein extracted from fecal samples and get clear separation between the experiment (SN-38G+active fecal protein) and negative controls (SN-38G+denatured fecal protein, SN-38G+BSA, active protein alone). Based on these results we are confident that we can get reliable SN-38G concentrations using this method. It will both improve our throughput, as we can do it in house, and it is substantially cheaper than our former LC-MS/MS methods.

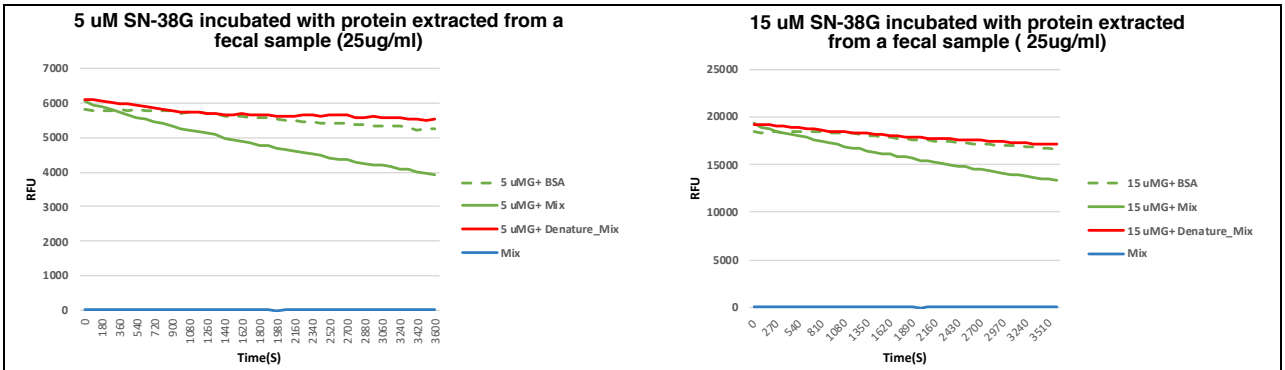

Figure 5. LCMS strategy for assessing SN-38G metabolism by fecal samples. In the above graphs there are two concentrations of SN-38G (left, 5uM; right 15 uM) incubated with the same density of protein (25ug/ml) extracted from a fecal sample. The green solid line is SN-38G incubated with an intact protein mix extracted from a fecal sample (this is the experiment). The red line is the SN-38G incubated with a denatured protein extracted from a sample (negative control 1). The green dashed line is SN-38G incubated with BSA instead of intact protein (negative control 2). The blue line is intact protein mix without SN-38G (negative control 3).

4) other achievements

- Dr. Kelly was awarded a 2021 Radcliffe Fellowship.
- Dr. Kelly was named a 2021 Google Cloud Project Research Innovator.

- Dr. Leah Guthrie, the first graduate student supported under this grant, is currently a postdoc with Justin Sonnenburg at Stanford and was named a 2021 HHMI Hannah Gray Fellow. This is a huge honor for Leah and will greatly enhance her career opportunities!

What opportunities for training and professional development has the project provided?

Subtasks 1-5 were met in the first year of the grant.

Subtask 6: Meet monthly with Dr. Mani. Dr. Kelly continues to meet monthly with Dr. Mani and is an author on one of Dr. Mani's papers assessing the role of bacterial swarming in inflammation, this paper was recently published in *Gastroenterology*.

Additional professional development activities:

Dr. Kelly proposed, developed, and ran a selected symposium in June 2021 at the international World Microbe Forum meeting, a joint, remote meeting from the American Society for Microbiology and the Federation of European Microbiological Societies. Dr. Kelly's symposium: "What makes a microbiome healthy?" was attended (based on chat box comments) by microbiologists from all over the world and had a lively question and answer session.

How were the results disseminated to communities of interest?

Results were disseminated *via* publications and formal scientific talks as described. In addition to these, Dr. Kelly will speak about colorectal cancer and the microbiome at the Einstein - Montefiore Virtual Summer High School Program, which will connect 17 high school students from the Bronx with Einstein researchers over the summer in support of STEM education. Dr. Kelly also wrote an opinion piece titled "Hunting a Changing Virus" detailing the dire need for a broad and nimble sequencing program to track, anticipate, and hopefully crush the SARS-Cov-2 virus and the inevitable future pathogens to come. This piece is to be published in *The Scientist*,

a professional magazine for life scientists, later in the summer and will hopefully reach a both broad swath of the life sciences community and interested members of the public.

What do you plan to do during the next reporting period to accomplish the goals?

- Recruit remaining colorectal cancer patients (**Major Task 1**). Current issues with the COVID-19 pandemic may slow recruitment but we are still aiming to meet our goal of 20 patients.
- Obtain metabolomic data for all samples. As described in the previous annual report, we have a protocol in place for determination of SN-38 and SN-38G concentrations in fecal samples (**Major Task 2).**
- Sequence and analyze remaining colorectal cancer patient metagenomes (**Major Task 2**).
- If it is acceptable to the Department of Defense, we will conduct proteomics on our colorectal cancer patient samples to identify functional BG enzymes and more deeply explore the functional role of oral microbes in the guts of cancer patients (**Major Task 3**). (See "5 CHANGES/PROBLEMS")
- Solve the problem of isolating sufficient RNA from fecal cultures (**Major Task 4**). (See "5 CHANGES/PROBLEMS")

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

Our findings from this year will make an impact on the base of knowledge in the field of microbiome influences on patient outcomes as follows:

Our research demonstrating that machine learning/artificial intelligence approaches can reliably and robustly classify disease in a difficult, multi-class, classification setting lays the foundation for routine clinical use of the microbiome as a diagnostic. We will use these classifiers to predict adverse responses to irinotecan in our patient population using their microbiome data.

We further demonstrated this year that our approaches can identify robust signatures of health and not-health and revealed a potential role of oral microbes in the gut in colorectal cancer that we followed up with proteomics conducted by Dr. Jacob Waldbauer at the University of

Chicago. Our proteomics analysis suggests that oral microbes in the gut are functionally active and producing proteins, a novel observation that requires further validation and study.

Our novel, topological data analysis-based approach to analyzing microbiome time-series data identifies 'states', that is, microbiome compositions, and state transitions in microbiomes that can be linked to clinical outcomes. As noted above, our approach overcomes several challenges in microbial time series analysis: **1)** it makes minimal assumptions regarding biological mechanisms, and is applicable to systems as diverse as the human gut and the ocean; **2)** it uses all available information regarding the similarity between samples of microbial communities; and **3)** it inherently scales with the amount and dimensionality of data. We therefore anticipate that many other microbiome studies will benefit from this approach. All code for this project is publicly available. With our newly sequenced temporal metagenomic datasets in hand we will use this approach to understand whether there are particular microbiome compositions that are associated with adverse responses to irinotecan treatment.

Finally, our temporal, metadata annotated, whole community metagenomic sequence data from colorectal cancer patients receiving treatment regimens that include irinotecan is a unique resource for both the colorectal cancer and the microbiome scientific communities. We anticipate that this dataset will be widely used and distributed and we hope that it will contribute substantially to our understanding of microbiome drug metabolism in colorectal cancer.

What was the impact on other disciplines?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Our machine learning and our topological data analysis approaches for classifying and visualizing microbiome data are relevant and applicable to any microbiome dataset. We note that in the topological data analysis paper (Chang, Van Insberghe, and Kelly, *npj Biofilms and Microbiomes*, 2021) we include analysis of two longitudinal marine datasets. We therefore think this approach will be useful in any discipline that considers microbial communities, from the human body, to the soil, to the oceans.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

Nothing to report.

Actual or anticipated problems or delays and actions or plans to resolve them

1. We have been unable to synthesize stable fluorescent probes for SN38-G (**Major Task 3**). We previously suggested that as an alternative we intend to utilize a purchasable probe to study microbial uptake of glucuronidated substrates with similarity to SN-38G. However, in considering this approach we are concerned that the probes will not accurately reflect drug metabolism. We have collected preliminary proteomics data for one patient, as described above, suggesting that proteomics approaches may provide a valuable lens through which to understand microbiome function in colorectal cancer. We propose to collaborate with Dr. Waldbauer to generate proteomics data for the remaining 38 patient samples and all samples to be collected in the final year of this project. If this is acceptable to the DoD, we will analyze these samples to identify **1)** functional BG enzymes; **2)** the role of oral microbes in the guts of colorectal cancer patients.

2. We have had difficulty extracting RNA from fecal samples (**Major Task 4**). We have attempted to work with Genewiz (http://www.genewiz/com) for mRNA extraction and purification from fecal samples; however while Genewiz has successfully extracted mRNA from our samples it was not of sufficient quality and quantity for mRNA sequencing. We next attempted a new protocol that has been used to extract mRNA from soil and fecal samples that was shared with us by colleagues at Rutgers. This approach has successfully extracted sufficient mRNA from test fecal samples, we are currently

working on cleaning up the mRNA such that it will be appropriate for sequencing. The COVID-19 pandemic slowed down our progress on this task but we are confident that we can overcome this hurdle for fecal samples from this project.

3. Our colorectal cancer patient recruitment was substantially delayed by the COVID-19 pandemic.

Changes that had a significant impact on expenditures

Nothing to report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

Nothing to report.

Significant changes in use or care of vertebrate animals

Nothing to report.

Significant changes in use of biohazards and/or select agents

Nothing to report.

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Authors who are members of the Kelly lab are **bolded**. Corresponding authorship is indicated with *.

a. Degradation of host translational machinery drives tRNA acquisition in viruses.

   Yang JY, Fang W, Miranda-Sanchez F, **Brown JM**, Kauffman KM, Acevero CM, Bartel DP, Polz MF*, **Kelly L*.** *Cell Syst*. 2021 Jun 15:S2405-4712(21)00205-2. doi: 10.1016/j.cels.2021.05.019. PMID: 34143976. Federal support acknowledged: YES

b. Endotoxin acts synergistically with C. difficile toxin B to increase IL-1 β production: A potential role for the intestinal biome in modifying the severity of C. difficile colitis. Htwe P, Aung HH, Kywe B, Niang PT, Oo TS, Monhandas S, **Kelly L**, Goldman DL. *J Infect Dis*. 2021 Mar 29:jiab165. doi: 10.1093/infdis/jiab165. PMID: 33780547. Federal support acknowledged: NO

c. Bacterial Swarmers Enriched During Intestinal Stress Ameliorate Damage. De A, Chen W, Li H, Wright JR, Lamendella R, Lukin DJ, Szymczak WA, Sun K, **Kelly L**, Ghosh S, Kearns DB, He Z, Jobin C, Luo X, Byju A, Chatterjee S, San Yeoh B, Vijay-Kumar M, Tang JX, Prajapati M, Bartnikas TB, Mani S. *Gastroenterology*. 2021 Jul;161(1):211-224. doi: 10.1053/j.gastro.2021.03.017. PMID: 33741315. Federal support acknowledged: YES

d. Genomic surveillance of SARS-CoV-2 in the Bronx enables clinical and epidemiological inference. Fels JM, **Khan S**, **Forster R**, Skalina KA, Sirichand S, Fox AS, Bergman A, Mitchell WB, Wolgast LR, Szymczak WA, Bortz RH, Dieterle ME, Florez C, Haslwanter D, Jangra RK, Laudermilch E, Wirchnianski AS, Barnhill J, Goldman DL*, Khine H*, Goldstein DY*, Daily JP*, Chandran K*, **Kelly L*.** *medRxiv*. 2021 Feb 10:2021.02.08.21250641. doi: 10.1101/2021.02.08.21250641. Preprint. PMID: 33594384. Federal support acknowledged: YES

e. Finding phenazine. **Kelly L**, **Wolfson SJ**. *Elife*. 2020 Oct 27;9:e62983. doi: 10.7554/eLife.62983. PMID: 33108273. Federal support acknowledged: NO

Books or other non-periodical, one-time publications.

Nothing to report

Other publications, conference papers and presentations.

Conference presentations

1. **Invited speaker**, Microbiome Movement: Drug Development, October 19-21, 2020 (Zoom)
2. **Invited speaker**, Pharmacology 2020, British Pharmacological Society, December 14-18, 2020 (Zoom)
3. **Symposium leader**, "What makes a microbiome 'healthy'?" World Microbe Forum, American Society for Microbiology and the Federation of European Microbiological Societies, June 20-24, 2021 (Zoom)

- **Website(s) or other Internet site(s)**
List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

Kelly lab Github:
https://github.com/kellylab
This site is the repository for code and data used in all published analyses.

- **Technologies or techniques**

Nothing to report.

- **Inventions, patent applications, and/or licenses**

Nothing to report.

- **Other Products**

Nothing to report.

7 PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: **Libusha Kelly**, no change.

Name: **Ruth Hauptman**, no change.

Funding Support: **No other support.**

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to report.

What other organizations were involved as partners?

Nothing to report.

8 SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS:

QUAD CHARTS:

9 APPENDICES: