AWARD NUMBER: W81XWH-18-1-0749

TITLE: Mammary Tissue Host-Microbiome Interactions in Breast Cancer Development

PRINCIPAL INVESTIGATOR: Leah T. Stiemsma

CONTRACTING ORGANIZATION: Pepperdine University; Malibu, CA

REPORT DATE: October 2020

TYPE OF REPORT: Annual Report

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

DISTRIBUTION STATEMENT: Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE October 2020	**2. REPORT TYPE** Annual	**3. DATES COVERED** 30Sep2019-29Sep2020

4. TITLE AND SUBTITLE Mammary Tissue Host-Microbiome Interactions in Breast Cancer Development	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER W81XWH-18-1-0749
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Leah T Stiemsma E-Mail: leah.stiemsma@pepperdine.edu	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Pepperdine University 24255 Pacific Coast Hwy Malibu CA, 90263	**8. PERFORMING ORGANIZATION REPORT**
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

Through this project we will determine the role of the mammary tissue microbiome in breast cancer development using 16S ribosomal RNA sequencing and dual-transcriptomic sequencing. In the first two years of this project we have selected and received 165 samples from the Susan G. Komen and Indiana University Simon Cancer Center Tissue Banks. We completed 16S rRNA sequencing on DNA isolated from all samples in this cohort and note a distinct microbial compositional signature that is associated with breast cancer development. We anticipate submitting this work for publication by February 2021. Due to COVID-19, the RNA isolations from these samples have been delayed until we can return to campus. We anticipate completing RNA isolations in Summer 2021 and will begin our analysis of the RNA sequencing data in Fall 2021. Regardless of these delays, the project is well underway. Results from this work will be key in characterizing host-microbiome cross-talk in the pathogenesis of breast tumor development.

15. SUBJECT TERMS

Mammary tissue, microbiome, bacteria, breast cancer

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	Unclassified	28	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

1. INTRODUCTION

This study aims to characterize the microbiome of mammary tissue donated from three groups of women: healthy women (healthy tissue, n = 50), women who were healthy at the time of tissue donation but later developed breast cancer (pre-diagnostic tissue, n = 15), and women with breast cancer. From the women with breast cancer we will study two tissue types: tissue adjacent to the malignant tumor (adjacent normal, n = 50) and tumor tissue (n = 50). We will employ 16S ribosomal RNA gene sequencing to analyze the composition of the mammary tissue microbiome among these four tissue types. In addition to studying the composition of the mammary microbiome, we also aim to study its function and interaction with its human host. We will perform dual-transcriptomic sequencing, whereby we extract and sequence mammary tissue microbial community RNA and human RNA from the same tissue. Using a multivariate linear modeling strategy, we aim to determine which bacterial taxa and genes are associated with the human transcriptome. Further, we will analyze whether these associations are modified by cancer status. Collectively, these analyses will determine the mechanistic role of the mammary tissue microbiome in breast cancer development.

2. KEYWORDS

Mammary tissue, microbiome, bacteria, breast cancer, pre-diagnostic, adjacent normal, tumor, 16S ribosomal RNA, meta-transcriptome, next generation sequencing

3. ACCOMPLISHMENTS

3a. What were the major goals of the project?

Training:
Major Task 1: Training and educational development in breast cancer research

- Subtask 1: Present research at Pepperdine Natural Sciences Divisional Seminars, attend UCLA microbiome club meetings (quarterly), attend seminars at the Jonsson Comprehensive Cancer Center. **- Ongoing**
- Subtask 2: Attend one national scientific meeting in a relevant scientific field in Y2 and in Y3. **- Ongoing**
- Subtask 3: Maintain regular mentorship meetings with Dr. Brewster, Dr. Binder, and Dr. Marino. - **Ongoing**
- Subtask 4: Audit relevant courses in cancer epidemiology and computational biology in Y2 and Y3. **– Ongoing**
- Subtask 5: Attend an RNA-sequencing workshop in Summer of Y1 **– Completed: month 11**

Research:

Specific Aim 1: There is a distinct mammary tissue microbiome compositional signature associated with breast cancer development. – Complete. Publication planned for submission by February 2021**.**

Major Task 1: Selection and shipping of samples - Complete

- Subtask 1: Selecting and requesting breast tissue samples from the Komen Tissue Bank (KTB), ordering supplies and equipment. – **Complete.**
- Ship tissue specimens to Dr. Stiemsma and prepare for DNA/RNA isolation (from IU to Pepperdine). – **Complete.**

Major Task 2: 16S rRNA sequencing of breast tissue samples – Complete.

- Subtask 1: DNA will be isolated from breast tissue samples at Pepperdine. – **Complete.**
- Subtask 2: 16S rRNA sequencing of DNA samples at UCLA microbiome core – **Complete.**
- Subtask 2: Preprocessing and analysis of 16S rRNA gene sequencing data – **Complete.**
- Subtask 4: Manuscript preparation and submission – **Ongoing.** Anticipated submission by February 2021**.**

Specific Aim 2: Distinct microbial functional variations are associated with healthy, pre-cancerous, and cancerous tissue. – To Do

Major Task 1: RNA extraction – Ongoing – delayed due to COVID-19 university shutdown. Anticipated completion: May 2021**.** Further details discussed in '*Actual or anticipated problems or delays and actions or plans to resolve them*'.

Major Task 2: Dual-RNA sequencing – To Do – delayed due to COVID-19 university shut down.

- Subtask 1: Dual-RNA sequencing performed at USC Molecular Genomics Core – **To Do**. Anticipated completion: July 2021.
- Subtask 2: Sequence preprocessing and filtering to identify the microbial transcriptomic profile – **To Do**. Anticipated completion: August 2021.
- Subtask 3: Analysis of associations of microbial functional groups and cancer status – **To Do**. Anticipated completion: October 2021.
- Subtask 4: Integration and analysis of microbial transcriptome with microbiome 16S data using MaAsLin (multivariate analysis with linear modeling) and analysis of effect modification by cancer status - **To Do**. Anticipated completion: October 2021
- Subtask 5: Manuscript preparation and submission – **To Do**. Anticipated manuscript submission in December 2021.

Specific Aim 3: The microbiome interacts with the host transcriptome to drive cancer development - To Do – delayed due to COVID-19 university shut down.

Major Task 1: Selection of genes associated with immune, anti-apoptotic, and

pro-proliferative pathways and association with cancer status - To Do. Anticipated completion in October 2021 (alongside transcriptomic analysis in aim II)

Major Task 2: Integration and analysis of microbial and host transcriptomic profiles using MaAsLin and analysis of effect modification by cancer status - **To Do**. Anticipated completion in October 2021 (alongside transcriptomic analysis in aim II).

- Subtask 1: Manuscript preparation and submission – **To Do.** Anticipated manuscript submission in December 2021.

3b. What was accomplished under these goals?

All major accomplishments are bolded.

AIM I: There is a distinct mammary tissue microbiome compositional signature associated with breast cancer development.

Sample Selection: **We selected and have received 165 tissue samples from the Susan G. Komen (KTB) and Indiana University Simon Cancer Center (IUSCC) tissue banks.** We used a cutoff of 100mg of fresh, frozen tissue as our primary requirement before applying additional selection criteria.

The following criteria were used to identify samples from 50 healthy controls and 15 pre-diagnostic subjects from the KTB: no prior breast surgery within three months of donating, no lactation or pregnancy at the time of sample collection, no personal history of cancer or benign breast disease, and no antibiotic use at the time of sample collection. Similar criteria were used to identify 50 adjacent normal and 50 tumor samples from the IUSCC (no lactation or pregnancy at the time of sample collection, no prior breast disease (where possible), and no antibiotic use at the time of sample collection) with the exception of lack of breast surgery and no personal history of cancer. Of these 100 samples, 46 are from the same women (matched adjacent normal and tumor samples). We preferentially selected samples from women with both adjacent normal and tumor samples available. However, these matched pairs were limited given our selection criteria and samples with enough tissue available for our planned analyses.

Dr. Marino (Indiana University) was key in identifying these samples and was responsible for shipping the samples to Dr. Stiemsma at Pepperdine University in Malibu, California. Dr. Stiemsma received all the samples by mid-June 2019. In addition to the tissue samples identified from these women, the following metadata was requested (received in August 2019) from the tissue banks: age, prior breastfeeding, prior breast surgery, BMI category, parity, menopausal status, cancer therapy, medications, BRCA1/2 positive, age at menarche, age at first birth, alcohol consumption, postmenopausal hormone therapy, diet, and physical activity. **Characteristics of this cohort are outlined in Table I** (**Appendix I**).

DNA Isolation: **Metagenomic DNA was isolated from 50mg of all 165 of these tissue samples** using the AllPrep PowerFecal DNA/RNA kit from Qiagen. We modified the lysis procedure to include 50uL of proteinase K, an hour-long heating step, and bead beating homogenization to ensure isolation of both human and microbial DNA.

16S rRNA Sequencing: **metagenomic DNA from 165 samples, 1 positive control, 8 negative controls, and 13 duplicates was submitted to the Host Microbe Systems Core at the University of California, Davis for targeted amplicon sequencing of the 16S ribosomal RNA (rRNA) gene in February 2020**. DNA was subjected to a nested-pcr for 16S enrichment. In this way universal 16S primers flanking the entire 16S gene enrich for 16S. Amplified 16S DNA is subjected to a second round of PCR to amplify the V3 – V4 region of the 16S gene. Amplicons were sequenced via Illumina Miseq paired-end sequencing.

Preprocessing of sequences: **Sequence reads were preprocessed and quality filtered by size prior to construct an amplicon sequence variant (ASVs) table using the DADA2 pipeline (1).** These ASVs were aligned to the SILVA database to assign taxonomy. The R package, *decontam*, was used to identify contaminant sequences using a prevalence-based approach, wherein reads from samples were compared to reads derived from eight blank control samples (2). Reads reaching a probability threshold of 0.1 in the blank controls were removed from the sample reads in the dataset. The total read count and the read counts between the four groups (relative abundance and alpha diversity) was assessed before and after removal of contaminants **(Figure 1).** The removal of contaminant taxa had no significant effect on the alpha diversity of the samples or relative abundance of taxa classified in the top 100 ASVs **(Figure 1).**

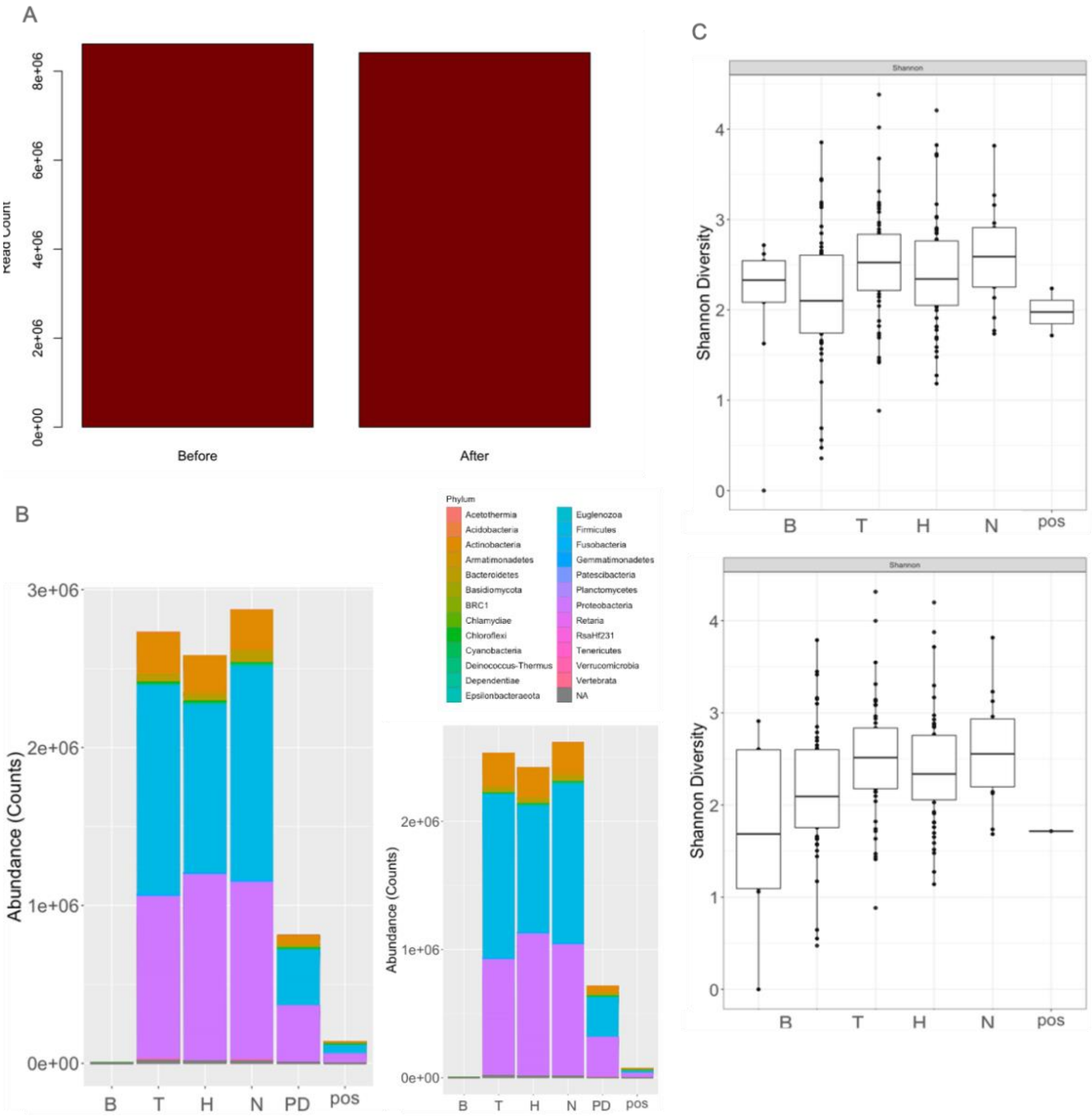


Figure 1. Preliminary Analyses of Human Mammary Tissue and Removal of Contaminant Taxa. Total 16S rRNA metagenomic sequencing data was compared to blank controls. Reads reaching a probability threshold of 0.1 in the blank controls were removed from the sample reads in the dataset. A) Total read count prior to and following application of *Decontam* analysis code. B) Total read counts of samples classified into four cancer status groups (n = 50 healthy tissue (H), n = 50 from normal adjacent tissue (N), n = 15 from pre-diagnostic tissue (PD), n = 50 from tumor tissue (T)) obtained through analysis of 16S rRNA sequencing along with a blank control (B) and positive control (pos) prior to and following *Decontam*. C) Alpha diversity of four cancer status groups (n = 50 healthy tissue (H), n = 50 from normal adjacent tissue (N), n = 15 from pre-diagnostic tissue (PD), n = 50 from tumor tissue (T)), B, and pos obtained through analysis of 16S rRNA sequencing prior to and following *Decontam*.

Analysis of 16S sequencing data: **We analyzed alpha and beta diversity and relative abundance between the four groups (healthy, pre-diagnostic, normal adjacent, and tumor) using the *Phyloseq* package in R** (3)**. Differentially abundant taxa were identified in the top 100 ASVs using *Deseq2*** (4)**.** Below we outline some of the preliminary findings from this analysis.

Figure 2 highlights differences in the mammary tissue microbiome among these four subsets. We did not identify strong global differences in the composition of the microbiome (**Figure 2C and D**). However, analysis of phylum and family relative abundance suggests there are distinct microbial populations associated with each of these cancer groups. At the phylum level, we note a decrease in Bacteroidetes in the prediagnostic and tumor groups relative to the healthy controls and adjacent normal tissue. At the family level, there is an apparent decrease in the abundance of Xanthomonodaceae between the healthy controls, prediagnostics, and cancer tissue subsets (tumor and adjacent normal), with the highest abundance of this taxa being present in the healthy control mammary tissue. The abundance of Bacillaceae also appears to increase in abundance, with the healthy control tissue displaying very low abundance (if any) of this taxon, the prediagnostics displaying 2 and 5% of this taxon, and the cancer subsets (adjacent normal and tumor) displaying between 10 and 15% of this taxon. Comparing the prediagnostic subset to our healthy control subset specifically, there also appears to be an outgrowth of Staphylococcaceae and Brevibacteriaceae.

We also analyzed the breast tissue microbiome among these four groups for differentially abundant taxa using *Deseq2* (4). *Deseq2* is traditionally applied to RNA-sequencing data, but has been adapted for microbiome analysis (4). Using *Deseq2*, we identified 46 taxa as differentially abundant relative to the healthy controls (**Figure 3**). Altogether, **this data suggests a distinct mammary tissue microbiome signature that is associated with the development of breast cancer.** Future functional analyses (RNA sequencing and metagenome prediction) will provide valuable insight into the host-microbiome interactions that may drive breast cancer development.

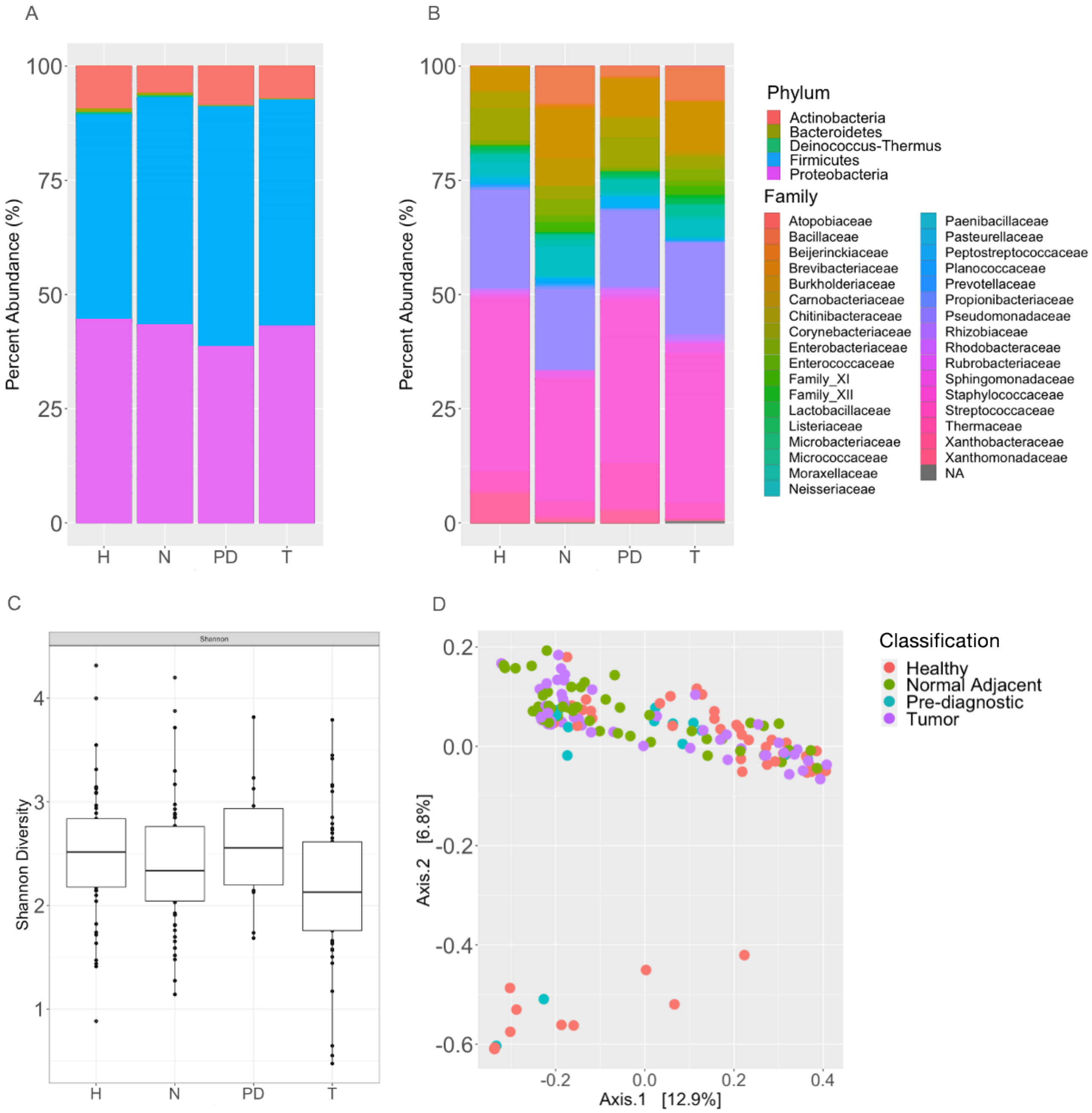


Figure 2. Relative Abundance and Diversity Analyses of Human Mammary Tissue.
A) Phylum B) and family relative abundances for the top 100 mammary tissue ASVs for subjects classified into four cancer status groups (n = 50 healthy tissue (H), n = 50 from normal adjacent tissue (N), n = 15 from pre-diagnostic tissue (PD), n = 50 from tumor tissue (T)). C) Alpha diversity and D) beta diversity of subjects classified into four cancer status groups (n = 50 healthy tissue (H), n = 50 from normal adjacent tissue (N), n = 15 from pre-diagnostic tissue (PD), n = 50 from tumor tissue (T)).

Figure 3. Differentially Abundant Taxa In Human Mammary Tissue. Significantly differentially abundant taxa for subjects classified into four cancer status groups (n = 50 from healthy tissue, n = 50 from normal adjacent tissue, n = 15 from pre-diagnostic tissue, n = 50 from tumor tissue). This data is presented as log2 fold changes relative to the healthy control group.

Another novel facet of our study is our ability to characterize the microbiome of truly healthy breast tissue. Current studies have compared breast tissue adjacent to malignant tumors to tissue adjacent to benign tumors or tissue from breast reduction or enhancement surgeries. Neither of these tissue types are ideal for accurate comparisons of healthy mammary tissue to cancerous tissue. Breast tissue from breast alteration surgeries has significant histological abnormalities when compared to tissue voluntarily donated from healthy women (5). **Figure 4** outlines a preliminary assessment of the healthy mammary tissue microbiome composition. As expected, we note strong variability between subjects. However, common taxa identified in all the healthy tissue samples include Corynebacteriaceae, Staphylococcaceae, and Pseudomonaceae (**Figure 4B**).

The prediagnostic subset also displays strong variability between subjects as expected. However, these is less Corynebacteriaceae among the prediagnostic subset. Further, we note that Staphylococcaceae remains a common taxon among this group (**Figure 5 A and C**). There is also no clear trend in bacterial abundance observed when taking into account years between biopsy and diagnosis of breast cancer (**Figure 5 B and D**). We are excited to review this data further as we **prepare a publication for submission by February 2021**.

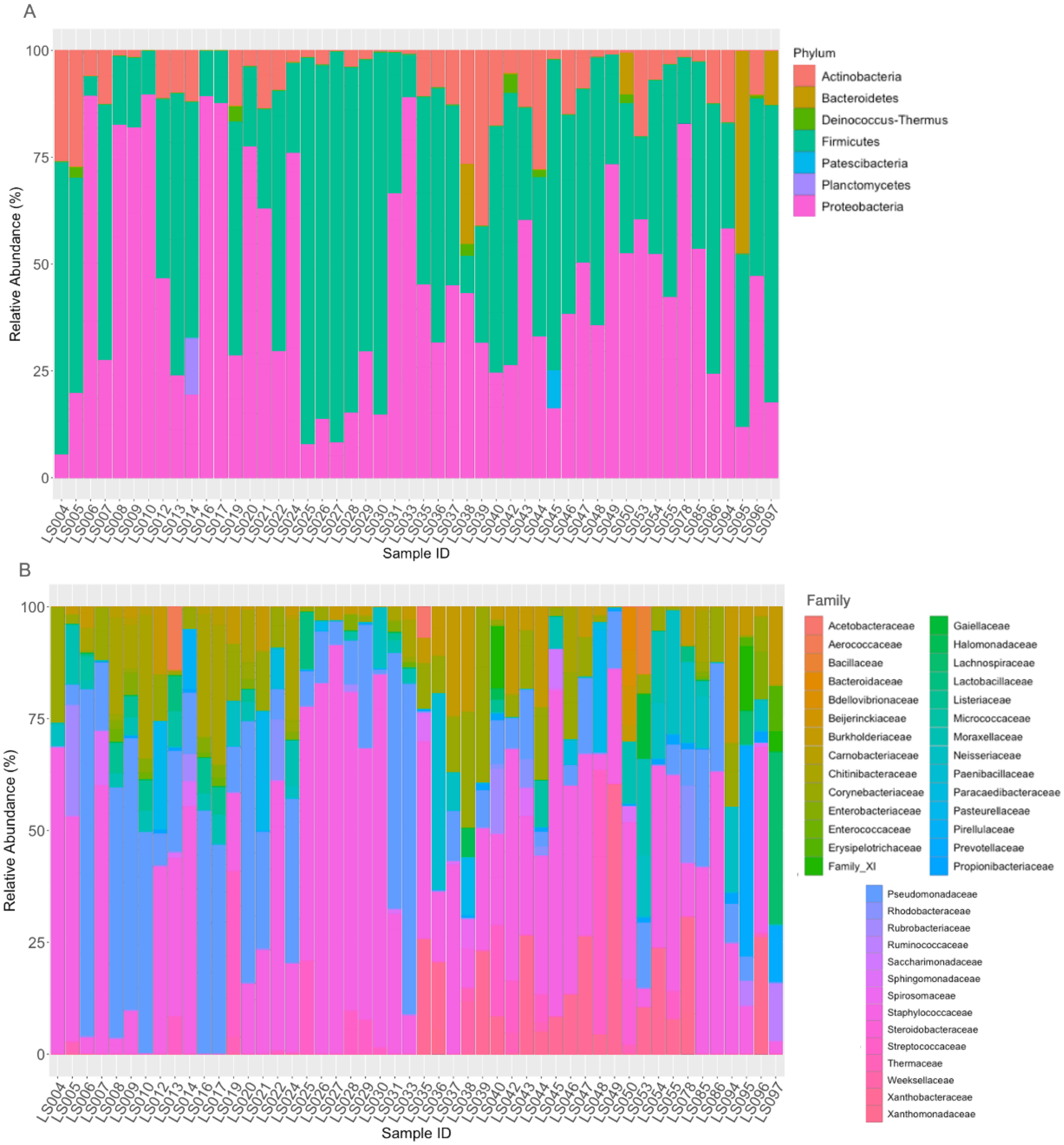


Figure 4. Analyses of Microbiome Relative Abundance in Healthy Human Mammary Tissue. A) Relative phylum abundance of healthy human mammary tissue (n = 50). B) Relative family abundance of healthy human mammary tissue (n = 50).

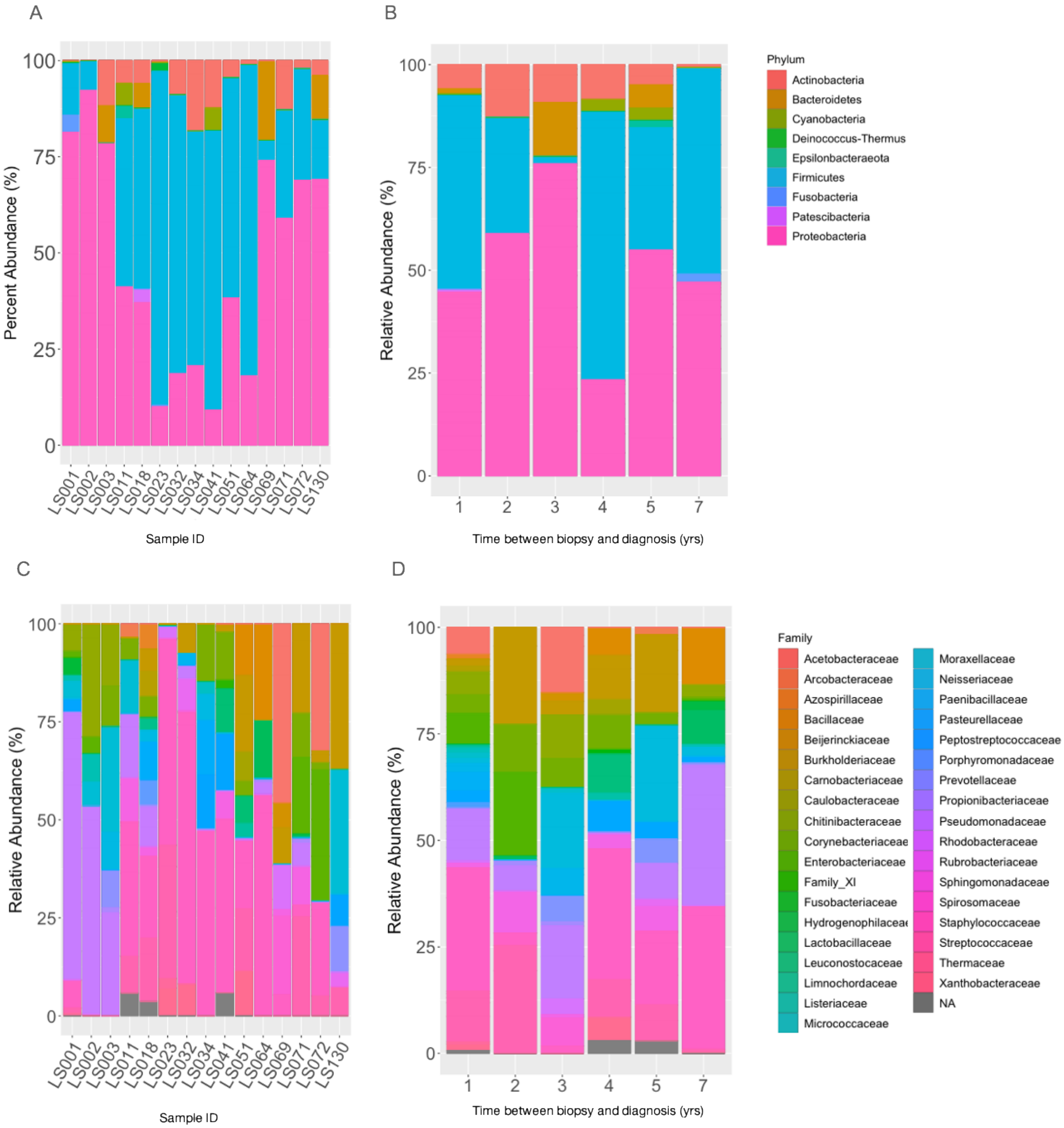


Figure 5. Analyses of Microbiome Relative Abundance in Prediagnostic Human Mammary Tissue. A) Phylum relative abundance between prediagnostic individuals. B) Phylum relative abundance based on the time between biopsy and diagnosis of breast cancer. C) Family relative abundance between prediagnostic individuals. D) Family relative abundances based on the time between biopsy and diagnosis of breast cancer.

AIM II: Distinct microbial functional variations are associated with healthy, pre-cancerous, and cancerous tissue.

RNA Isolation and meta-transcriptomic sequencing: Pilot analysis of RNA sequencing from five breast tissue samples showed significant variability in the amount of ribosomal RNA between the samples. This inconsistency suggests RNA quality that is too low to conduct dual-RNA sequencing (**Appendix I, Table 2**). We expected this and intended to begin isolating RNA from the remaining tissue for these samples in Summer 2020. However, due to COVID-19, we have not been able to access the laboratory since March 2020. We anticipate returning to campus in February 2021, and will begin RNA isolations at this time. In this way, we will be able to submit total RNA for sequencing in July 2021. And begin our analysis of this data and preparation of a manuscript highlighting results from aims II and III in October 2021. *We anticipate requesting a no-cost extension to ensure completion of this analysis.*

qPCR analyses: Using the Qiagen AllPrep PowerFecal DNA/RNA kit, we did isolate DNA and RNA in parallel from the same tissue sample. RNA isolated in this way is not of high enough quality to conduct dual-transcriptomic sequencing; however, it is of high enough quality for use in qPCR to analyze specific host and microbial genes. Annie Kump, a student in my lab generated cDNA for 60 of these samples (15 per group; healthy, prediagnostic, normal adjacent, tumor). Using this cDNA, we conducted qPCR analyses to assess expression of G-protein coupled receptors in breast tissue (*GPR41*, *GPR43*, and *GPR109A,* **Figure 6**). These are receptors for short chain fatty acids, bacterial-derived metabolites. This serves as a functional assessment of the mammary tissue microbiome. We did not identify any statistically significant differences among the four subsets of participants (healthy, prediagnostic, adjacent normal, and tumor tissue), suggesting an alternate pathway of host-microbiome interaction in mammary tissue. **This represents one avenue of microbiome function in mammary tissue that we will analyze further after completing RNA-sequencing on these samples.**

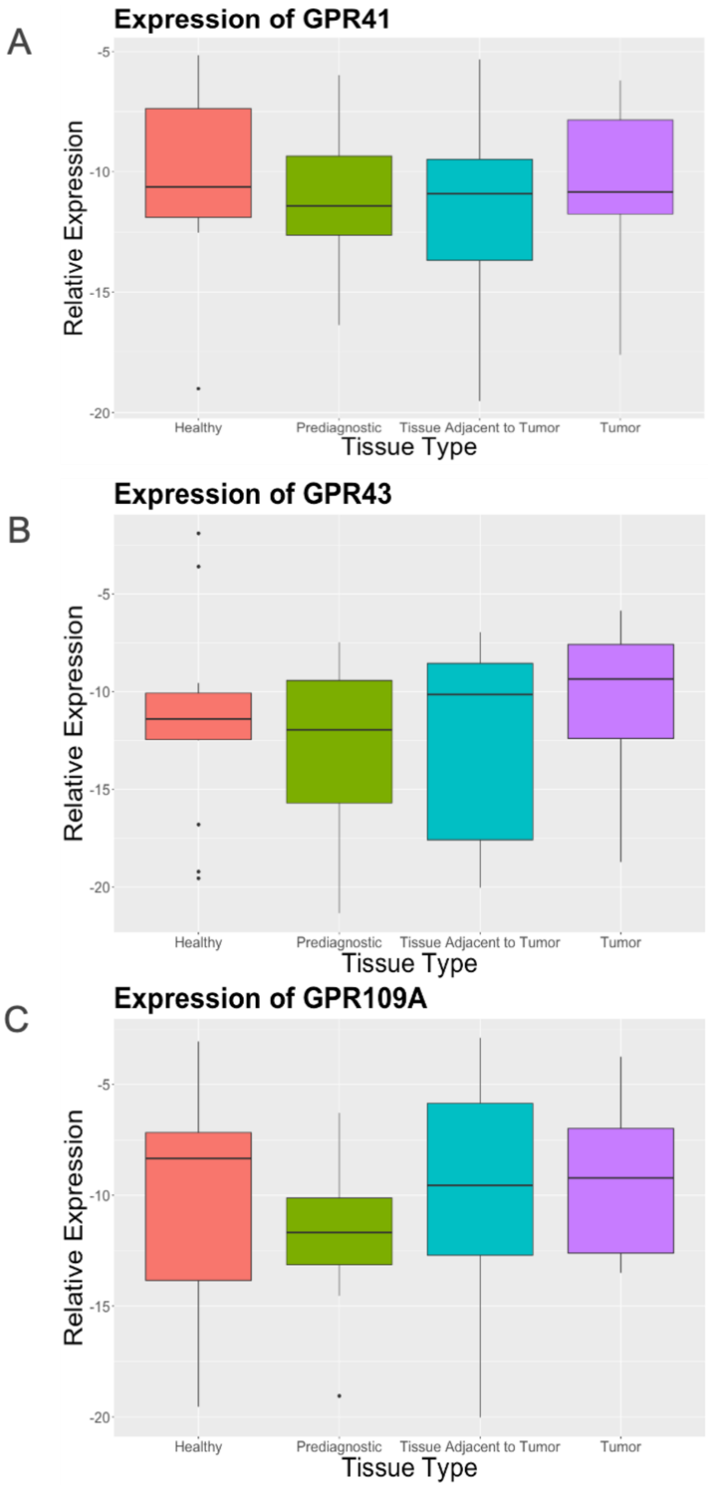


Figure 6. Expression of *GPR41*, *GPR43*, and *GPR109A* in Breast Tissue. A) *GPR41* expression levels between healthy tissue (n=11), normal tissue adjacent to a tumor (n=10), pre-diagnostic tissue (n=12), and tumor tissue (n=15). P value > 0.05. B) *GPR43* expression levels between healthy tissue (n=15), normal tissue adjacent to a tumor (n=15), pre-diagnostic tissue (n=15), and tumor tissue (n=15). P value > 0.05. C) *GPR109A* expression levels between healthy tissue (n=15), normal tissue adjacent to a tumor (n=15), pre-diagnostic tissue (n=15), and tumor tissue (n=15). P value > 0.05.

AIM III: The microbiome interacts with the host transcriptome to drive cancer development.

All planned analyses for this aim require the 16S and meta-transcriptomic sequencing data. We are excited to discuss details of this analysis in future reports.

3c. What opportunities for training and professional development has the project provided?

Conference Attendance in 2020: I attended the American Association for Cancer Research – The Microbiome, Viruses, and Cancer in February 2020 with three of my research students. All three students presented posters related to this breast cancer microbiome study.

1. Kump A, **Stiemsma LT.** Expression of *GPR41*, *GPR43*, and *GPR109* in breast tissue: an investigation into the expression of human receptors in releation to breast cancer development. Abstract & Poster, *American Association for Cancer Research – The Microbiome, Viruses, & Cancer.* Orlando, FL. February 2020.

2. Gabel J, Hoskinson C, Kump A, Michels KB, Marino N, **Stiemsma LT.** The mammary tissue microbiome in breast cáncer development. Abstract & Poster, *American Association for Cancer Research – The Microbiome, Viruses, & Cancer.* Orlando, FL. February 2020.

Student Mentoring: I have been able to engage a number of undergraduate students in this research over the past two years. Specifically, in Summer 2019 I participated in Pepperdine's Summer Undergraduate Research in Biology Program (SURB). Through this program, I mentored and supervised three undergraduate students, all of whom contributed to this research project through DNA/RNA isolation, establishment of protocols, or quantitative PCR of host and microbial genes. Two of these summer students continued to work on this project and one new student is joined the lab during the academic year 2019 - 2020. In Summer 2020, I mentored Courtney Hoskinson through the 16S rRNA sequencing for this project. A publication outlining these analyses (Appendix A) is aimed for submission by February 2021.

Student mentoring is a critical part of my career as an independent investigator at Pepperdine University. I recently transitioned to a tenure-track assistant professorship at Pepperdine. My participation in programs such as SURB and my mentorship of students thus far in my career is key to further establishing my research program in my new role at the university.

Workshops: In August 2019, I attended the UC Davis Bioinformatics Core RNA Sequencing Workshop. This was an exceptional workshop, which provided me with invaluable information regarding RNA isolation and preparation for sequencing, sequence pre-processing, and analysis of RNA sequencing data. I was also able to discuss the goals of my project one-on-one with instructors at this workshop, which gave me valuable insight into the best approaches for sequencing these samples and the tools necessary to do so. Additionally, I learned the computational pipeline to preprocess meta-transcriptomic data and analyze the data for differential gene expression.

3d. How were the results disseminated to communities of interest?

Dissemination of Research Through Conference Attendance in 2020: I attended the American Association for Cancer Research – The Microbiome, Viruses, and Cancer in February 2020 with three of my research students. All three students presented posters related to this breast cancer microbiome study.

> Kump A, **Stiemsma LT.** Expression of *GPR41*, *GPR43*, and *GPR109* in breast tissue: an investigation into the expression of human receptors in releation to breast cancer development. Abstract & Poster, *American Association for Cancer Research – The Microbiome, Viruses, & Cancer.* Orlando, FL. February 2020.
>
> Gabel J, Hoskinson C, Kump A, Michels KB, Marino N, **Stiemsma LT.** The mammary tissue microbiome in breast cáncer development. Abstract & Poster, *American Association for Cancer Research – The Microbiome, Viruses, & Cancer.* Orlando, FL. February 2020.

Dissemination of Research As An Invited Speaker, The University of Hawaii:

Stiemsma LT. An Introduction to Microbiome Analysis. Analysis of -Omics Data Working Group Meeting. University of Hawaii.

Dissemination of Research Through the Summer Undergraduate Research in Biology Program at Pepperdine (Summer 2019): In Summer 2019, I participated as a mentor in the Summer Undergraduate Research in Biology Program (SURB) at Pepperdine. Through this program, undergraduates gain experience working in a research lab; developing hypotheses, writing proposals, giving presentations, and conducting lab work. I was a mentor to three undergraduate students (Annie Kump, Daniel Herrera, and Courtney Hoskinson) through this program.

Throughout the summer, these students engaged with 10 other SURB students, Pepperdine faculty, and with non-Pepperdine scientists (guest speakers), discussing their results and receiving feedback/troubleshooting support. In addition, at the end of the Summer (July 2019), my students gave oral presentations open to the Pepperdine community, and presented posters at the SURB Poster Session and Summer Undergraduate Research Banquet. Through this banquet, our research was disseminated to Pepperdine faculty, research students, family members, local ecological staff/scientists, Pepperdine staff, and Pepperdine administrators.

3e. What do you plan to do during the next reporting period to accomplish the goals?

AIM I: There is a distinct mammary tissue microbiome compositional signature associated with breast cancer development.

> Microbiome Compositional Analysis Manuscript Preparation/Submission: Anticipated date for manuscript submission by February 2021.

AIM II: Distinct microbial functional variations are associated with healthy, pre-cancerous, and cancerous tissue.

Functional Metagenome Prediction: We are expanding on our 16S compositional analysis by using Piphillin to predict the functional metagenome (6). In this way, we can assess the function of the bacteriome based on our 16S compositional sequencing. The functional metagenome can be aligned to the KEGG database and organized in to biological pathways for further analysis of microbiome function (7). This data will be validated and expanded upon in our transcriptomic analyses (details for these analyses are described below).

RNA Isolation and meta-transcriptomic sequencing: Upon our return to campus in early Spring 2020 (anticipated for February 2020) my students and I will isolate RNA from the remaining tissue available for the cohort (~50mg per person). We will ensure that the RNA is high quality via gel analysis prior to submitting a small subset of samples for another pilot sequencing run.

Based on our 16S rRNA sequencing results, the majority of the sequence reads will be human. This is why a nested PCR method was used for the compositional 16S analysis. Originally, we planned to do a double ribosomal RNA depletion to ensure that we could capture both human and microbial mRNA. However, some of the kits to do this depletion are discontinued. Hence, we will need to conduct another pilot analysis in which we finalize the enrichment method for bacterial and host RNA before applying this technique to the rest of our cohort.

Analysis and Manuscript preparation/submission: We anticipate completing the RNA isolations and this pilot assessment by July 2021. We will then begin our analysis of this data and preparation of a manuscript highlighting results from aims II and III in October 2021. *We anticipate requesting a no-cost extension to ensure completion of this analysis.*

AIM III: The microbiome interacts with the host transcriptome to drive cancer development.

All planned analyses for this aim require the 16S and meta-transcriptomic sequencing data. We are excited to discuss details of this analysis in future reports.

3. IMPACT

4a. What was the impact on the development of the principal discipline(s) of the project?

Using these highly innovative techniques, this work will inform future studies focusing on breast cancer development, specifically studies of the breast tissue microbiome. W have identified specific breast tissue bacteria that may be key in breast cancer development. Thus, it may be possible to utilize the mammary microbiome as a therapeutic target or biomarker for breast cancer development. Our future analyses of the transcriptome will inform how the microbiome interacts with the human host and whether these interactions may contribute to or protect against breast cancer risk. Ultimately, future clinical applications of our work include targeting breast tissue bacteria or genes prior to or after breast cancer onset. As this has not been previously investigated, our work represents a crucial step toward ending breast cancer.

4b. What was the impact on other disciplines?

<u>Nested-PCR for 16S enrichment:</u> We applied a nested-PCR strategy to enrich for the 16S gene in metagenomic DNA isolated from the breast tissue samples. Our pilot analysis of this method compared to the traditional one-step PCR method for 16S enrichment suggested enhanced ability to study the bacterial microbiome in low-biomass samples. This pilot analysis was presented at the American Association for Cancer Research National Meeting – The Microbiome, Viruses, and Cancer and was well received by others in this field. We anticipate that publication of this work in peer-reviewed journals will inform the methodology used for 16S enrichment in other microbiome studies of low-biomass samples.

<u>Impact on clinicians:</u> In the last decade, the role of the microbiome has been highlighted in many human diseases. Through research of the human microbiome, new potential avenues for treatment and protection from diseases have been developed. However, the role of the mammary microbiome in breast cancer and general breast health has not been fully elucidated. Our study will play a key role in determining how these microbes are involved in this disease, which will impact future treatment options and preventive strategies used by clinicians in this field.

<u>Impact on fields of computational biology and bioinformatics:</u> Through this project, we will be employing a novel sequencing technique (dual-transcriptomic sequencing) by which we sequence the host and microbial transcriptome. In this way, we sequence all of the host and microbial genes and differentiate them *in silico.* This technique has yet to be performed in the context of the breast tissue microbiome, and is primarily used to study host-pathogen interactions rather than host-microbiome interactions. Thus, our execution of this technique will inform the fields of computational biology and bioinformatics of the technical/computational requirements necessary to perform this type of analysis.

4c. What was the impact on technology transfer?

Nothing to report.

4d. What was the impact on society beyond science and technology?

Undergraduate Student Involvement: This work is contributing to our mechanistic understanding of breast cancer, while providing undergraduate students at Pepperdine with clinically relevant research opportunities. These experiences are pivotal in their career development. Annie Kump and Jaelyn Gabel (two Pepperdine undergraduate who have also contributed to this project over the past year) are on a medical school track, with a research emphasis. This project that involves human data and samples has been instrumental in informing them of human-focused research that combines basic science with clinical medicine. Additionally, through her work on this project, Courtney Hoskinson has learned skills in computational biology and molecular biology that she will hone further in her graduate program at the University of British Columbia.

Beyond the current impact of this project, we anticipate that this project will inform treatment and prevention strategies for breast cancer, which will have a significant impact on patients and families affected with breast cancer.

4. CHANGES/PROBLEMS:

5a. Changes in approach and reasons for change

Nothing to report.

5b. Actual or anticipated problems or delays and actions or plans to resolve them

There are no changes to the scope of this project and the project is well underway. However, we are experiencing significant delays due to COVID-19 shutdowns. In accordance with LA county guidelines, the university has restricted campus access since March 2020. This has impacted our ability to conduct RNA isolations (necessary for the analyses outlined in aims II and III), which we intended to do in Summer 2020. As such, we have delayed this portion of the project until we are able to access campus (potentially in January 2021). This delay has not resulted in any additional expenses. However, we do anticipate requesting a no-cost extension to complete this work.

DNA from the 165 samples was submitted for 16S rRNA sequencing to the UC Davis Host Microbe Systems Core lab in February 2020. The COVID-19 shut downs across the nation resulted in some delays in receiving our sequencing results from this submission (expected in April 2020, delayed until July 2020). However, with the inability to work on campus, we transitioned our focus this past summer and fall solely to the analysis of this microbiome sequencing data (outlined in this report, appendix I). A publication outlining this work will be submitted by February 2020.

5c. Changes that had a significant impact on expenditures

We do not anticipate completing the RNA sequencing until end of summer 2021. Thus, the funds set aside for these portions of the project in year 1, will not be spent until year 3.

I have also transitioned from a restricted-faculty position (Visiting Assistant Professor) to a tenure-track position (Assistant Professor of Biology). As such, we have requested some budget reallocations to account for my increase in salary. These reallocations will have no impact on the scope of the project.

5d. 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.

Pepperdine Institutional Review Board Approval Dates: August 2018 – end of study, assuming no significant changes.

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.

5. PRODUCTS:

Publications, conference papers, and presentations

Journal publications

Nothing to report.

Books or other non-periodical, one-time publications

Nothing to report.

Other publications, conference papers and presentations.

1. Stiemsma LT, Nakamura RE, Nguyen JG, Michels KB. Does Consumption of Fermented Foods Modify the Human Gut Microbiota? The Journal of Nutrition. 2020 Jul 1;150(7):1680–92.

2. Stiemsma LT. <u>An Introduction to Microbiome Analysis.</u> Analysis of -Omics Data Working Group Meeting. University of Hawaii. August 2020.

3. Kump A, Stiemsma LT**.** Expression of *GPR41*, *GPR43*, and *GPR109* in breast tissue: an investigation into the expression of human receptors in releation to breast cancer development. Abstract & Poster, *American Association for Cancer Research – The Microbiome, Viruses, & Cancer.* Orlando, FL. February 2020.

4. Gabel J, Hoskinson C, Kump A, Michels KB, Marino N, Stiemsma LT**.** The mammary tissue microbiome in breast cáncer development. Abstract & Poster, *American Association for Cancer Research – The Microbiome, Viruses, & Cancer.* Orlando, FL. February 2020.

5. <u>Herrera D</u>, <u>Kump A</u>, Stiemsma LT. <u>Breastfeeding and its effects on the microbiome of human breast tissue.</u> Poster, *SURB Poster Session*, *Pepperdine University*. Malibu, CA; 2019

Website(s) or other Internet site(s)
Nothing to report.

Technologies or techniques

Nothing to report.

Inventions, patent applications, and/or licenses

Nothing to report.

Other Products

Nothing to report.

6. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

7a. What individuals have worked on the project?

Name:	Leah Stiemsma
Project Role	PI
Nearest Person Month Worked:	4.38 calendar months & 2.74 summer months
Contribution to the Project:	Dr. Stiemsma directs the overall project. Dr. Stiemsma works alongside the undergraduate students identified below to prepare all samples for microbiome and transcriptome analysis (DNA and RNA extraction). She will also analyze the microbiome sequencing data and, with guidance from Dr. Marino (Indiana University) and Dr. Binder (UCLA), will analyze associations between the breast tissue transcriptome and microbiome.
Funding Support:	Breakthrough Fellowship Award – Department of Defense

Name:	Jay Brewster
Project Role	Mentor
Nearest Person Month Worked:	
Contribution to the Project:	Mentor
Funding Support:	

Name:	Alexandra M. Binder
Project Role	Mentor
Nearest Person Month Worked:	0.24 calendar months
Contribution to the Project:	Mentor, statistical support
Funding Support:	K07 from NIH/NCI

Name:	Natascia Marino
Project Role	Mentor
Nearest Person Month Worked:	0.24 calendar months
Contribution to the Project:	Mentor, bioinformatic support to analyze the RNA sequencing data, and assisted with sample selection.
Funding Support:	

Name:	Annie Kump

Project Role	Undergraduate Student
Nearest Person Month Worked:	12.00 Calendar Months
Contribution to the Project:	Annie contributes by isolating DNA and RNA from the tissue samples.
Funding Support:	Undergraduate research fellowship

Name:	Courtney Hoskinson
Project Role	Undergraduate Student
Nearest Person Month Worked:	12.00 calendar Months
Contribution to the Project:	Courtney contributes by isolating DNA and RNA from the tissue samples. She also spent Summer 2020 analyzing the 16S rRNa sequencing data.
Funding Support:	Undergraduate research fellowship

Name:	Daniel Herrera
Project Role	Undergraduate Student
Nearest Person Month Worked:	3.00 calendar Months
Contribution to the Project:	Daniel contributed by isolating DNA and RNA from the tissue samples.
Funding Support:	Undergraduate research fellowship

Name:	Jaelyn Gabel
Project Role	Undergraduate Student
Nearest Person Month Worked:	12.00 calendar Months
Contribution to the Project:	Jaelyn contributes by isolating DNA and RNA from the tissue samples. She will also assist Dr. Stiemsma with the 16S rRNA sequencing analysis.
Funding Support:	Undergraduate research fellowship

7b. Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Jay Brewster, Divisional Dean of Natural Science, is now serving as a mentor on this project.

7c. What other organizations were involved as partners?

Organization Name: Indiana University
Location: Indiana, United States
Partner’s Contribution: Collaboration/Sub-contract

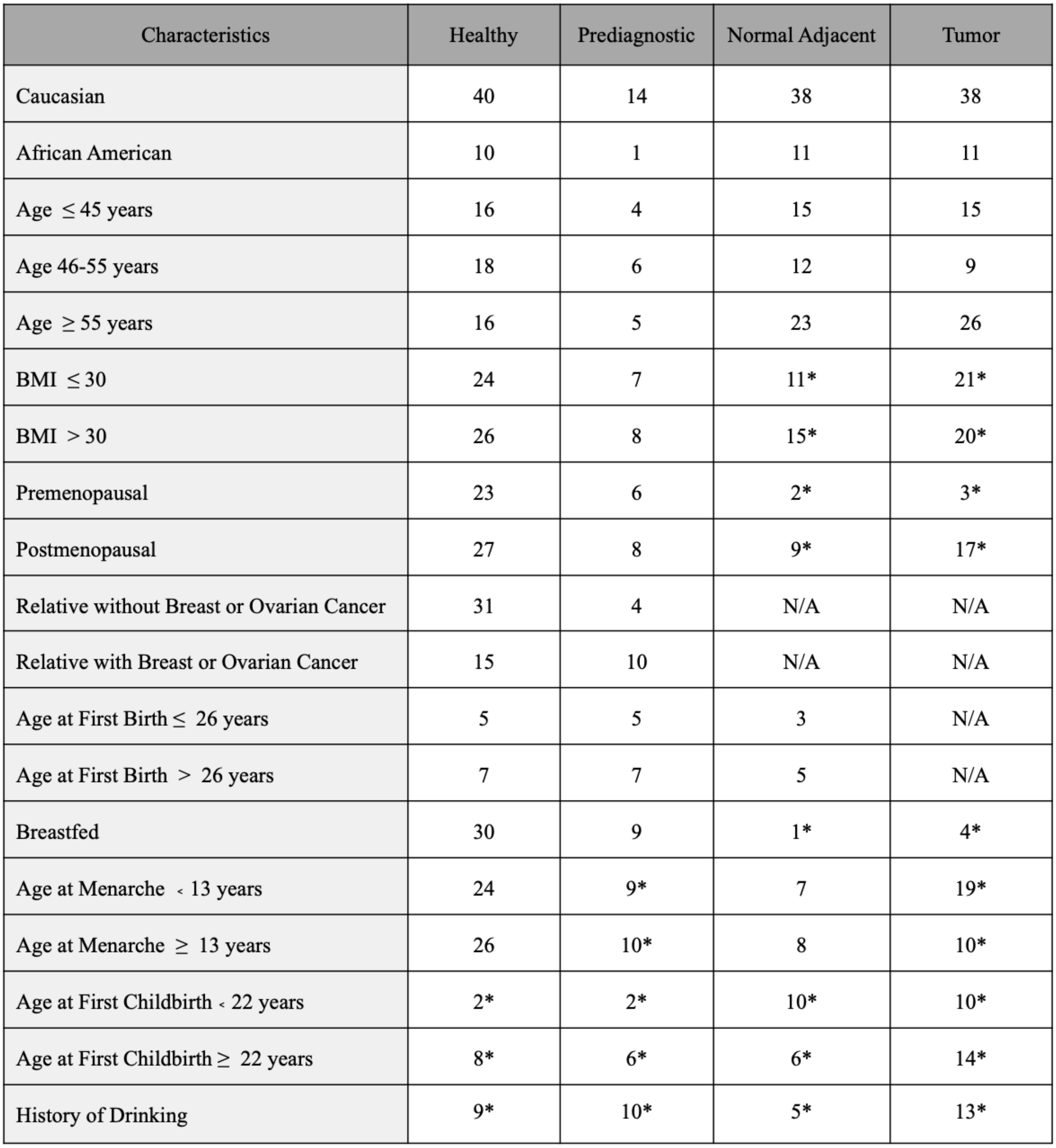
Organization Name: University of Hawaii
Location: Hawaii, United States
Partner’s Contribution: Collaboration/Mentorship

7. SPECIAL REPORTING REQUIREMENTS

Not applicable.

Appendix I:

Table 1. Metadata of Human Mammary Tissue Donors.



Characteristics	Healthy	Prediagnostic	Normal Adjacent	Tumor
Caucasian	40	14	38	38
African American	10	1	11	11
Age ≤ 45 years	16	4	15	15
Age 46-55 years	18	6	12	9
Age ≥ 55 years	16	5	23	26
BMI ≤ 30	24	7	11*	21*
BMI > 30	26	8	15*	20*
Premenopausal	23	6	2*	3*
Postmenopausal	27	8	9*	17*
Relative without Breast or Ovarian Cancer	31	4	N/A	N/A
Relative with Breast or Ovarian Cancer	15	10	N/A	N/A
Age at First Birth ≤ 26 years	5	5	3	N/A
Age at First Birth > 26 years	7	7	5	N/A
Breastfed	30	9	1*	4*
Age at Menarche < 13 years	24	9*	7	19*
Age at Menarche ≥ 13 years	26	10*	8	10*
Age at First Childbirth < 22 years	2*	2*	10*	10*
Age at First Childbirth ≥ 22 years	8*	6*	6*	14*
History of Drinking	9*	10*	5*	13*

Metadata of Human Mammary Tissue Donors. Metadata of patients from whom samples classified into four cancer status groups were collected (n = 50 healthy tissue (H), n = 50 from normal adjacent tissue (N), n = 15 from pre-diagnostic tissue (PD), n = 50 from tumor tissue (T)). *Limited metadata available from donors.

Table 2: Percent of Ribosomal RNA in Samples from Pilot RNA-Seq Analysis

Sample	Percent rRNA
A	43%
B	81%
C	69%
D	61%
E	39%

Citations:

1. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul;13(7):581–3.

2. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018 Dec;6(1):226.

3. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. Watson M, editor. PLoS ONE. 2013 Apr 22;8(4):e61217.

4. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec;15(12):550.

5. Degnim AC, Visscher DW, Hoskin TL, Frost MH, Vierkant RA, Vachon CM, et al. Histologic findings in normal breast tissues: comparison to reduction mammaplasty and benign breast disease tissues. Breast Cancer Res Treat. 2012 May;133(1):169–77.

6. Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, Dabbagh K, et al. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. He Z, editor. PLoS ONE. 2016 Nov 7;11(11):e0166104.

7. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D457–62.