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 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
October 2021	Annual	30Sep2020-29Sep2021

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 and functional signature that is associated with breast cancer development. We are hopeful our manuscript outlining these findings will be accepted in late Fall 2021. We are currently revising this manuscript for resubmission to mSystems. We have opted to transition to conduct whole metagenome sequencing to assess genetic predisposition to mammary microbiome colonization, based on findings from our 16S microbiome analysis. Due to COVID-19, analyses outlined in aims II and III were delayed due to Pepperdine campus being under restricted access from March 2020 to May 2021. Regardless of these delays, the project is well underway and we are confident we will complete these analyses by Summer 2022. 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	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	33	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

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 aim to perform metatranscriptomic sequencing and/or shotgun metagenomic sequencing, to assess host-microbiome interactions in breast tissue. Using a multivariate linear modeling strategy, we aim to determine which bacterial taxa and genes are associated with the human transcriptome and genome. 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. Manuscript attached - in revision for resubmission to the journal, *mSystems*.

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. Manuscript in revision.**

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

Major Task 1: RNA/DNA extraction – DNA and RNA are extracted. **Complete.**

Major Task 2: Metatranscriptomics/Whole Genome Metagenomics – To Do – delayed due to COVID-19 university shut down.

- Subtask 1: Metatranscriptomic/whole genome metagenomic sequencing - ongoing. Anticipated completion: July 2022.
- Subtask 2: Sequence preprocessing and filtering – **To Do**. Anticipated completion: Spring 2022.
- Subtask 3: Analysis of associations of microbial functional groups and cancer status/cancer gene variants – **To Do**. Anticipated completion: early Summer 2022.
- Subtask 4: Integration and analysis of microbial metagenome with microbiome 16S data using MaAsLin (multivariate analysis with linear modeling) and analysis of effect modification by cancer status - **To Do**. Anticipated completion: Early Summer 2022.
- Subtask 5: Manuscript preparation and submission – **To Do**. Anticipated manuscript submission in Late Summer/early Fall 2022.

Specific Aim 3: The microbiome interacts with the host transcriptome/genome 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 Late Summer 2022 (alongside analyses in aim II).

Major Task 2: Integration and analysis of microbial and host transcriptomic/metagenomic profiles using MaAsLin and analysis of effect modification by cancer status - **To Do**. Anticipated completion in Late Summer 2022 (alongside analyses in aim II).

- Subtask 1: Manuscript preparation and submission – **To Do.** Anticipated manuscript submission in Late Summer/Early Fall 2022.

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.

The abstract from our manuscript with an overview of our findings is copied below. Please see appendix I for figures outlining our analyses. Figures 1 and 2 outline analyses pertaining to this aim. Table 1 provides an overview of the cohort characteristics.

Hoskinson C, Zheng K, Gabel J, Kump A, German R, Podicheti R, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems.* October 2021.

Abstract: Microbiota studies have reported changes in the microbial composition of the breast upon cancer development. However, results are inconsistent and limited to the late phase of cancer development. Here, we analyzed and compared the resident bacterial taxa of histologically normal breast tissue (healthy, H, n = 50) and compared this tissue with tissue donated prior to (pre-diagnostic, PD, n = 15) and after (adjacent normal, AN, n = 51 and tumor, T, n = 49) breast cancer diagnosis. For microbiota compositional analysis, DNA isolated from 165 tissue samples was submitted for Illumina Miseq paired-end sequencing of the V3 - V4 region of the 16S gene. To infer bacterial function in breast cancer, we predicted the functional bacteriome from the 16S sequencing data using Piphillin. **Bacterial compositional analysis revealed an intermediary taxonomic signature in the PD tissue relative to that of H tissue, represented by shifts in *Bacillaceae*, *Burkholderiaceae*, *Corynebacteriaceae*, *Enterobactereaceae*, *Xanthobacteriaceae*, and *Staphylococcaceae*. This compositional signature is enhanced in AN and T tissues relative to H tissue.** Analysis of differentially abundant predicted metabolic pathways highlights significant metabolic reprogramming of the microbiota of PD, AN, and T tissue, relative to H tissue. Further, correlation analysis between host transcriptome profiling and microbial taxa and genes in PD tissue highlights

enhanced interaction between the human host and mammary microbiota prior to breast cancer diagnosis. These findings suggest that compositional shifts in bacterial abundance and metabolic reprogramming of the breast tissue microbiota are early events in breast cancer development, and potentially linked with cancer susceptibility.

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

We used Piphillin, a metagenome prediction software, to predict the metabolic capacity of the mammary microbiota in healthy, prediagnostic, and cancerous tissues. Results from this analysis are outlined in our manuscript. We also conducted correlation analyses between the host transcriptome and our microbiome data.

The abstract from our manuscript with an overview of our findings is copied below. Please see appendix I for figures outlining our analyses (figures 3 and 4 outline analyses pertaining to this aim).

Hoskinson C, Zheng K, Gabel J, Kump A, German R, Podicheti R, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems.* October 2021.

Abstract: Microbiota studies have reported changes in the microbial composition of the breast upon cancer development. However, results are inconsistent and limited to the late phase of cancer development. Here, we analyzed and compared the resident bacterial taxa of histologically normal breast tissue (healthy, H, n = 50) and compared this tissue with tissue donated prior to (pre-diagnostic, PD, n = 15) and after (adjacent normal, AN, n = 51 and tumor, T, n = 49) breast cancer diagnosis. For microbiota compositional analysis, DNA isolated from 165 tissue samples was submitted for Illumina Miseq paired-end sequencing of the V3 - V4 region of the 16S gene. To infer bacterial function in breast cancer, we predicted the functional bacteriome from the 16S sequencing data using Piphillin. Bacterial compositional analysis revealed an intermediary taxonomic signature in the PD tissue relative to that of H tissue, represented by shifts in *Bacillaceae*, *Burkholderiaceae*, *Corynebacteriaceae*, *Enterobactereaceae*, *Xanthobacteriaceae*, and *Staphylococcaceae*. This compositional signature is enhanced in AN and T tissues relative to H tissue. **Analysis of differentially abundant predicted metabolic pathways highlights significant metabolic reprogramming of the microbiota of PD, AN, and T tissue, relative to H tissue. Further, correlation analysis between host transcriptome profiling and microbial taxa and genes in PD tissue highlights enhanced interaction between the human host and mammary microbiota prior to breast cancer diagnosis.** These findings suggest that compositional shifts in bacterial abundance and metabolic reprogramming of the breast tissue microbiota are early events in breast cancer development, and potentially linked with cancer susceptibility.

RNA Isolation and meta-transcriptomic sequencing: Due to COVID-19, we were not able to access the laboratory to continue this work until June 2021. This past summer, my research student, Lindsey Marian, and I, worked to assess the quality of RNA samples that we previously isolated. The RNA isolated from these samples is of low quality (RIN score 4 or less). Likely this is due to how the samples were originally stored at the tissue banks. As we were able to conduct correlation analysis with RNA isolated from a small subset of these tissues by our collaborator, we are focusing on whole metagenomic sequencing of these tissue samples for the remainder of this work.

DNA Isolation and meta-genomic sequencing: We modified our scope of work to include a whole metagenome analysis of the mammary tissue. Our data suggests that the microbiota may be responding to the changing microenvironment of the host tissue rather than driving breast cancer forward. To address this further, we conducted a pilot whole metagenome analysis on 5 breast tissue samples and one negative extraction control to ensure proper sequencing depth of these samples. This pilot analysis identified low amounts of microbial sequence reads in all of our samples and we are working with the UC Davis sequencing core to analyze the microbiome data in correlation with the human genome sequences.

qPCR analyses: Using the Qiagen AllPrep PowerFecal DNA/RNA kit, we did isolate DNA and RNA in parallel from the same tissue sample. 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 1**). 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 whole metagenome sequencing on these samples.**

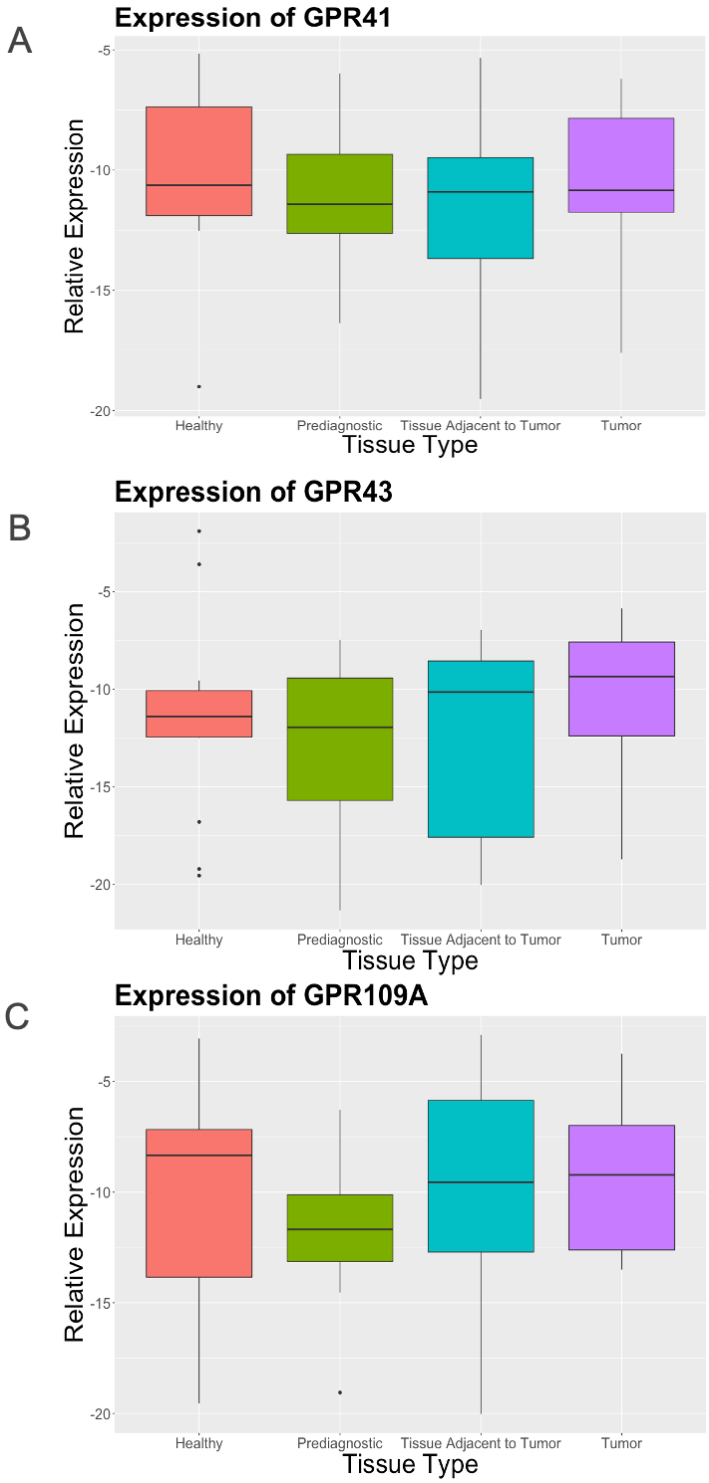


Figure 1. Expression of *GPR41*, *GPR43*, and *GPR109A* in Breast Tissue. A) *GPR41* expression levels between healthy tissue (n=11), normal tissue adjacent to a tumor (n=10), pre-diagnostic tissue (n=12), and tumor tissue (n=15). P value > 0.05. B) *GPR43* expression levels between healthy tissue (n=15), 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.

Host-microbe correlation analyses:
We have done preliminary analyses on this question, which are outlined in our manuscript. **Please see appendix I for figures outlining our analyses.**
Figure 4 outlines the correlations.

Hoskinson C, Zheng K, Gabel J, Kump A, German R, Podicheti R, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems.* October 2021.

All additional planned analyses for this aim require the 16S and whole genome metagenomic 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 has joined the lab during the academic year 2019 - 2020. In Summer 2020, I mentored Courtney Hoskinson through the 16S rRNA sequencing for this

project (Hoskinson C*, Gabel J*, Kump A*, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems.* October 2021. In Summer 2021, I mentored two additional students through Pepperdine's SURB program - Kelly Zheng contributed to key analyses outlined under AIM III and Lindsey Marian has been key to preparing RNA and DNA for submission for transcriptomic and metagenomic sequencing (AIM II).

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 and 2022: 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.

> *Hoskinson C, Zheng K, Gabel J, Kump A, German R, Podicheti R, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. Abstract submitted to the Keystone Symposium: The human microbiome, ecology and evolution.

*In February 2022, I plan to attend and present our manuscript at the Keystone Symposium, The Human Microbiome: Ecology and Evolution in BAMF, Alberta.

> Zheng K, Stiemsma LT. Interactions between the breast microbiota and host gene expression in breast cancer development. Oral Presentation. Pepperdine University Waves Classic Golf Tournament. October 2021.
>
> Zheng K, Stiemsma LT. Interactions between the breast microbiota and host gene expression in breast cancer development. Oral Presentation. SURB Symposium, July 2021.

Kump A, Stiemsma LT**.** Expression of *GPR41*, *GPR43*, and *GPR109* in breast tissue: an investigation into the expression of human receptors in relation 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 and 2021): 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 - 11 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 and 2021), my students gave oral presentations open to the Pepperdine community, and in July 2019, 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.

Hoskinson C*, Gabel J*, Kump A*, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems.* October 2021.

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

Whole metagenome sequencing: We conducted a pilot whole metagenome analysis on 5 breast tissue samples and one negative extraction control to ensure proper sequencing depth of these samples. This pilot analysis identified low amounts of microbial sequence reads in all of our samples and we are working with the UC Davis sequencing core to analyze the microbiome data in correlation with the human genome sequences. We aim to submit the full cohort of samples for sequencing in January 2022.

Analysis and Manuscript preparation/submission: We anticipate completing these analyses in Summer 2022 and will prepare a manuscript for publication at that time.

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

All additional 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. We 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?

Nested-PCR for 16S enrichment: 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.

Impact on clinicians: 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.

Impact on fields of computational biology and bioinformatics: Through this project, we will be employing a novel sequencing technique (dual-metatranscriptomic 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 undergraduates 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

Based on our 16S compositional and predicted bacteriome analysis (Hoskinson C*, Gabel J*, Kump A*, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems*. October 2021.), we added a whole-genome metagenomic approach to assess whether the breast tissue microbiota is associated with genetic variations in the host tissue. Our data suggests significant dysregulation of the bacteriome likely driven by tumor development (contrary to our original hypothesis, which suggested microbial contributions to tumor development). Our expansion in our technical approach will allow us to better understand the relationship between the mammary microbiota and the human host as it pertains to breast cancer development.

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 had restricted campus access from March 2020 - May 2021. This has impacted our ability to conduct wet-lab procedures necessary for aims II and III. This delay has not resulted in any additional expenses. However, we did request a no-cost-extension to complete this work (project extended to Fall 2022).

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 in 2020 solely to the analysis of this microbiome sequencing data (outlined in Hoskinson C, Zheng K, Gabel J, Kump A, German R, Podicheti R, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems*. October 2021). We focused summer 2021 on aims II and III and expect sequencing results to address these aims in Summer 2022.

5c. Changes that had a significant impact on expenditures

We do not anticipate completing the whole metagenome sequencing until late Spring 2022. Thus, the funds set aside for these portions of the project in year 1, will not be spent until year 4.

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

Hoskinson C*, Zheng K, Gabel J*, Kump A*, German R, Podicheti R, Marino N, Stiemsma LT. The composition and functional potential of the human mammary microbiota prior to and following breast tumor development. In revision for resubmission to *mSystems*. October 2021

Books or other non-periodical, one-time publications

Nothing to report.

Other publications, conference papers and presentations.

1. Zheng K, Hoskinson C, Marino NM, Stiemsma LT. Interactions between the breast microbiota and host gene expression in breast cancer development. SURB 2021 Presentation. Malibu, CA.

2. Patrick DM, Sbihi H, Dai DLY, Al Mamun A, Rasali D, Rose C, Marra F, Boutin RC, Petersen C, Stiemsma LT, et al. Decreasing antibiotic use, the gut microbiota, and asthma incidence in children: evidence from population-based and prospective cohort studies. Lancet Respir Med. 2020 Nov;8(11):1094–105.

3. Stiemsma LT, Davis SD, Brewster JL. Analysis of Microbial water contamination, soil microbial community structure, and soil respiration in a collaborative first-year Students as Scholars Program (SAS). Front Microbiol. 2020 Dec 17;11:590035.

4. 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.

5. Stiemsma LT. An Introduction to Microbiome Analysis. Analysis of -Omics Data Working Group Meeting. University of Hawaii. August 2020.

6. 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.

7. 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.

8. Cait A, Cardenas E, Dimitriu PA, Amenyogbe N, Dai D, Cait J, Sbihi H, **Stiemsma LT,** et al. Reduced genetic potential for butyrate fermentation in the gut microbiome of infants who develop allergic sensitization. J Allergy Clin Immunol. 2019 Dec;144(6):1638-1647.e3.

9. Herrera D, Kump A, Stiemsma LT. Breastfeeding and its effects on the microbiome of human breast tissue. 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.

10. 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:	
Contribution to the Project:	Mentor, statistical support
Funding Support:	K07 from NIH/NCI

Name:	Natascia Marino

Project Role	Mentor
Nearest Person Month Worked:	
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

Name:	Lindsey Marian
Project Role	Undergraduate Student

Nearest Person Month Worked:	12.00 calendar Months
Contribution to the Project:	Jaelyn contributes by isolating RNA from the tissue samples and assessing RNA quality for transcriptome sequencing.
Funding Support:	Undergraduate research fellowship

Name:	Kelly Zheng
Project Role	Undergraduate Student
Nearest Person Month Worked:	12.00 calendar Months
Contribution to the Project:	Kelly contributed computational analysis to analyze links between host and microbial gene expression and microbial composition.
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?
Nothing to report.

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

11. **SPECIAL REPORTING REQUIREMENTS**

Not applicable.

Appendix I:

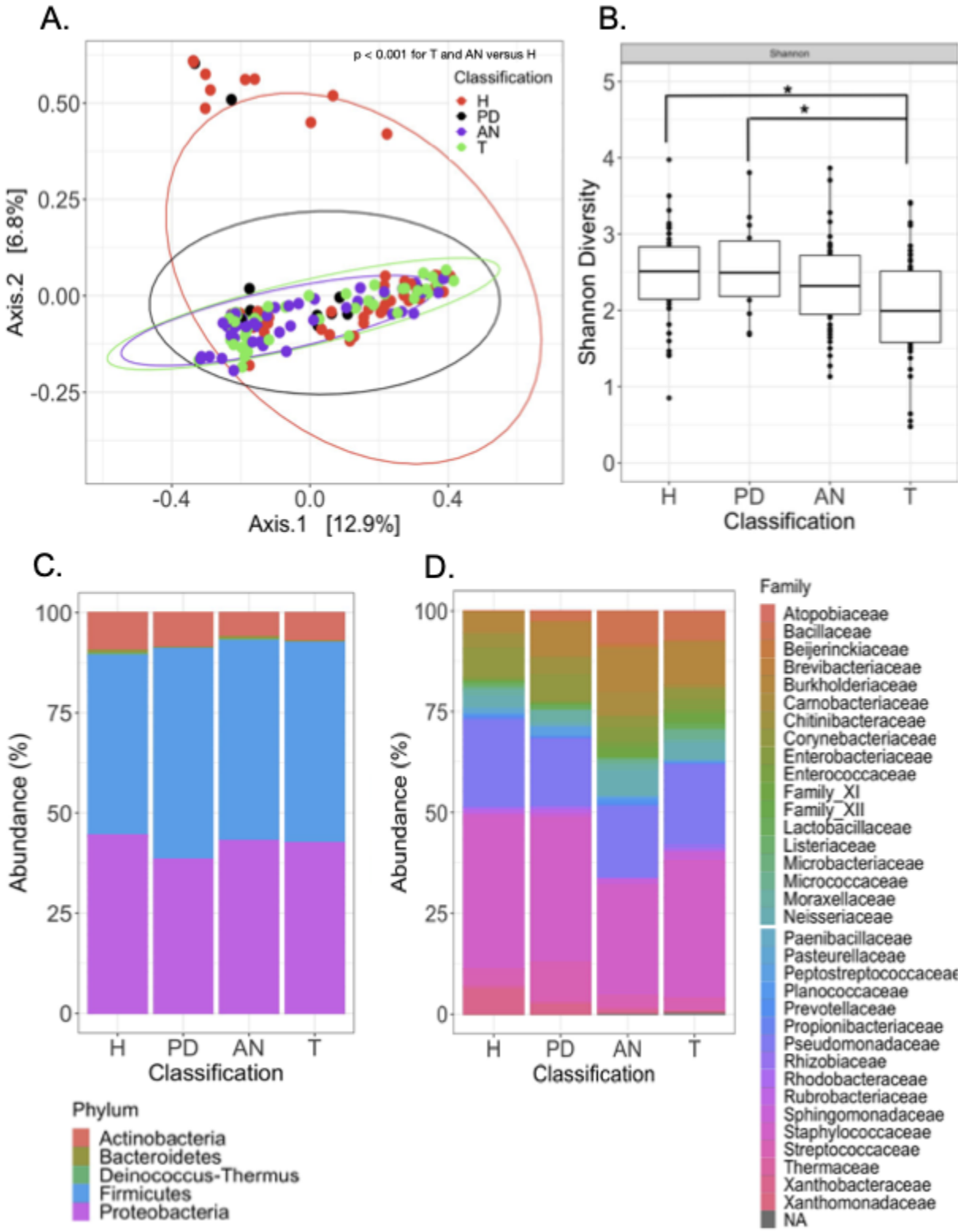
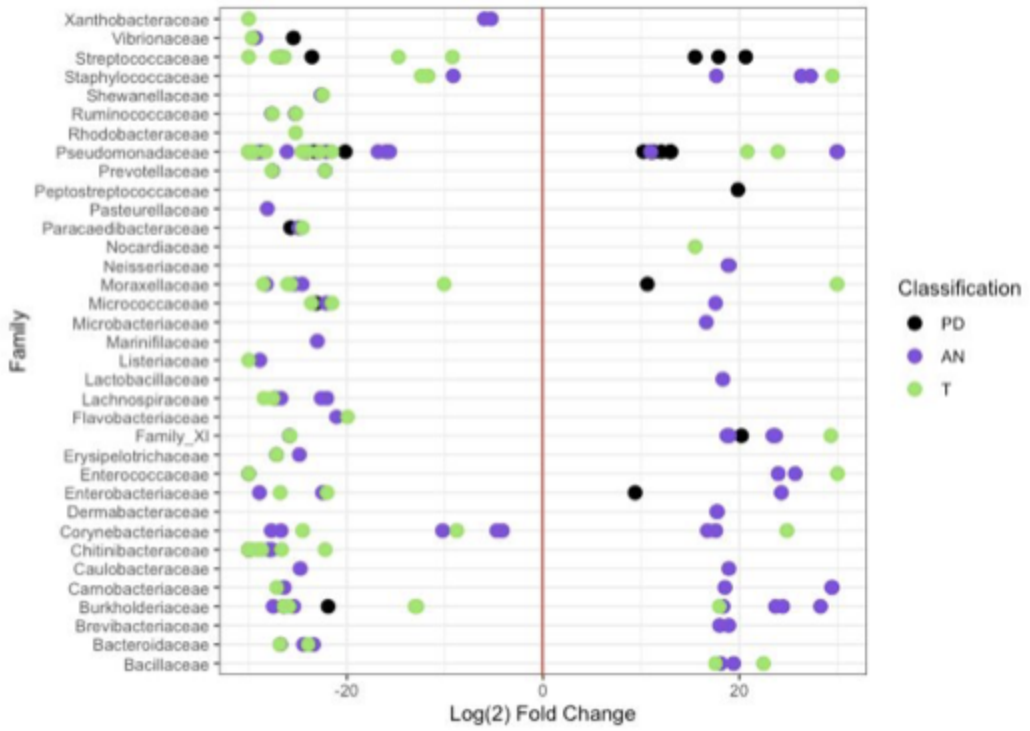

A.
p < 0.001 for T and AN versus H
Classification
H
PD
AN
T
Axis.2 [6.8%]
Axis.1 [12.9%]
B.
Shannon
Shannon Diversity
Classification
C.
Abundance (%)
Classification
Phylum
Actinobacteria
Bacteroidetes
Deinococcus-Thermus
Firmicutes
Proteobacteria
D.
Abundance (%)
Classification
Family
Atopobiaceae
Bacillaceae
Beijerinckiaceae
Brevibacteriaceae
Burkholderiaceae
Carnobacteriaceae
Chitinibacteraceae
Corynebacteriaceae
Enterobacteriaceae
Enterococcaceae
Family_XI
Family_XII
Lactobacillaceae
Listeriaceae
Microbacteriaceae
Micrococcaceae
Moraxellaceae
Neisseriaceae
Paenibacillaceae
Pasteurellaceae
Peptostreptococcaceae
Planococcaceae
Prevotellaceae
Propionibacteriaceae
Pseudomonadaceae
Rhizobiaceae
Rhodobacteraceae
Rubrobacteriaceae
Sphingomonadaceae
Staphylococcaceae
Streptococcaceae
Thermaceae
Xanthobacteraceae
Xanthomonadaceae
NA

Figure 1. Variations in microbial diversity between healthy patients, pre-diagnostic patients, and patients with cancer. A) Principal Coordinates Analysis (PCoA) of the mammary microbiota across the four tissue types ($p < 0.001$ for T and AN versus H). **B)** Alpha diversity (Shannon diversity index) of the mammary microbiota across the four tissue types. **C)** Phylum and **D)** family relative abundances for the top 100 mammary tissue ASVs across the four tissue types. n healthy = 50 (H), n pre-diagnostic = 15 (PD), n adjacent normal = 51 (AN), n tumor = 49 (T). * = $p<0.05$, ** = $p<0.01$.

A.


Xanthobacteraceae
Vibrionaceae
Streptococcaceae
Staphylococcaceae
Shewanellaceae
Ruminococcaceae
Rhodobacteraceae
Pseudomonadaceae
Prevotellaceae
Peptostreptococcaceae
Pasteurellaceae
Paracaedibacteraceae
Nocardiaceae
Neisseriaceae
Moraxellaceae
Micrococcaceae
Microbacteriaceae
Marinifilaceae
Listeriaceae
Lactobacillaceae
Lachnospiraceae
Flavobacteriaceae
Family_XI
Erysipelotrichaceae
Enterococcaceae
Enterobacteriaceae
Dermabacteraceae
Corynebacteriaceae
Chitinibacteraceae
Caulobacteraceae
Camobacteriaceae
Burkholderiaceae
Brevibacteriaceae
Bacteroidaceae
Bacillaceae
Family
-20
0
20
Log(2) Fold Change
Classification
PD
AN
T

B.

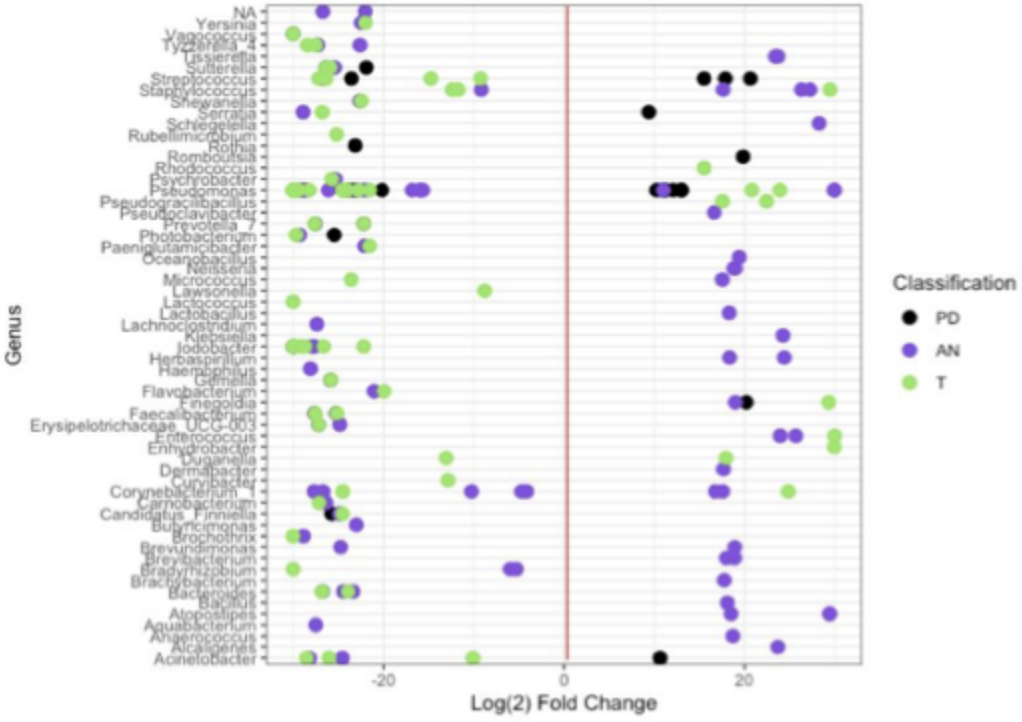

NA
Yersinia
Vagococcus
Tyzzerella_4
Ussiereta
Sutterella
Streptococcus
Staphylococcus
Shewanella
Serratia
Schlegelella
Rubellimicrobium
Rothia
Rombousia
Rhodococcus
Psychrobacter
Pseudomonas
Pseudogracilibacillus
Pseudoclavibacter
Prevotella_7
Photobacterium
Paenoglutamicibacter
Oceanobacillus
Neisseria
Micrococcus
Lawsonella
Lactococcus
Lactobacillus
Lachnoclostridium
Klebsiella
Jodobacter
Herbaspirillum
Haemophilus
Gemella
Flavobacterium
Finegoldia
Faecalibacterium
Erysipelotrichaceae_UCG-003
Enterococcus
Enhydrobacter
Duganella
Dermabacter
Curvibacter
Corynebacterium_1
Carnobacterium
Candidatus_Finniella
Burkholderia
Brochothrix
Brevundimonas
Brevibacterium
Bradyrhizobium
Bracteribacterium
Bacteroides
Bacillus
Atopostipes
Aquabacterium
Ahargococcus
Alcaligenes
Acinetobacter
Genus
-20
0
20
Log(2) Fold Change
Classification
PD
AN
T

Figure 2. Differentially abundant taxa in human mammary tissue. A) Family and **B)** genus abundance in PD tissue (black), AN tissue (purple), and T tissue (green) as compared to H tissue (red reference line) identified through the *Deseq2* R package with a 0.01 adjusted p value (Benjamin-Hochberg adjustment) cut-off. Each dot corresponds to a specific ASV classified to each taxon. Taxa abundances are expressed as transformed log2 fold changes.

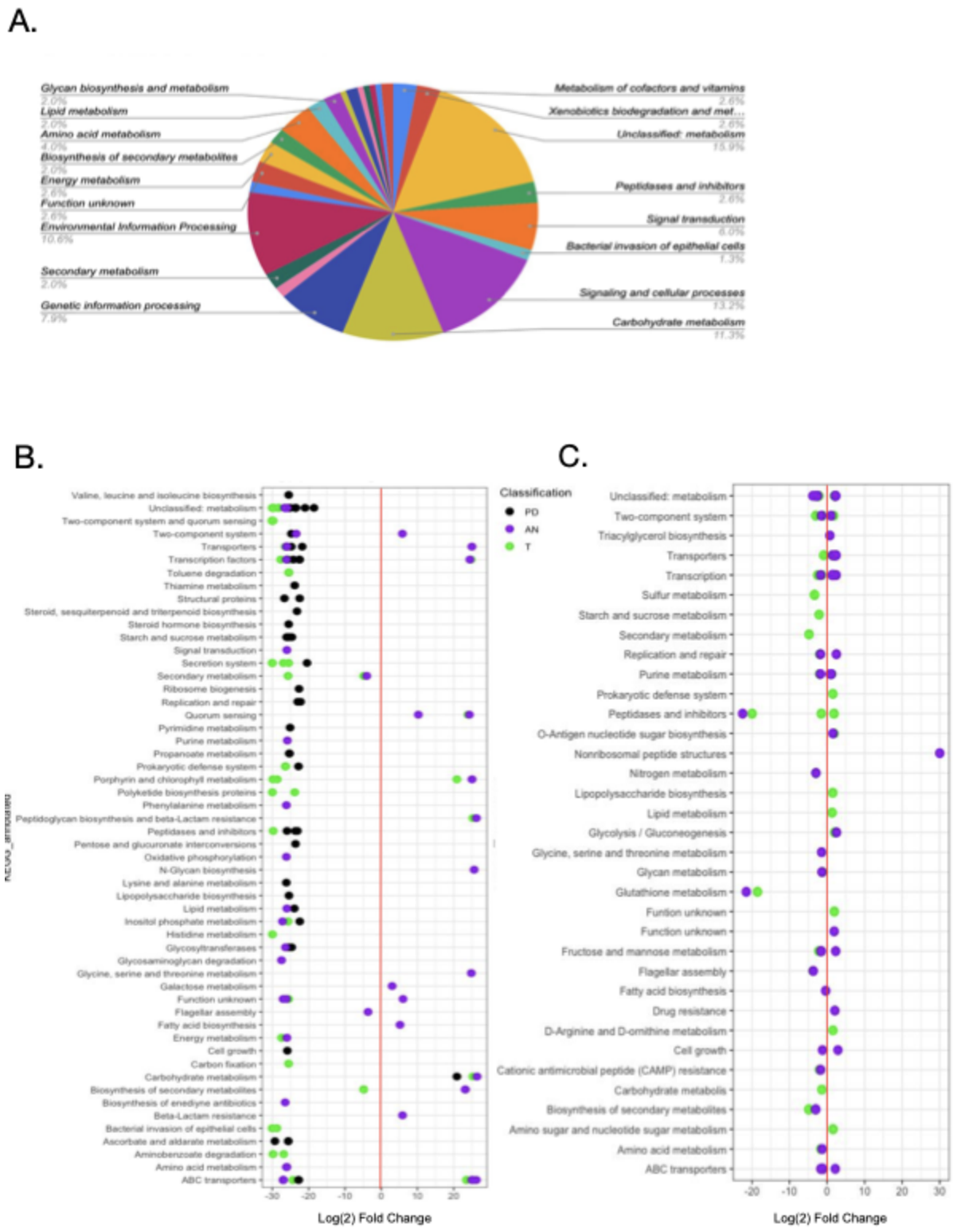


Figure 3. Differentially abundant genes in the mammary tissue functional bacteriome. A) Bacterial KEGG pathways in mammary tissue related to breast cancer development. KOs were

predicted from the 16S amplicon sequencing data using Piphillin and identified as differentially abundant through the *Deseq2* R package with a 0.01 alpha cut-off (Benjamin-Hochberg adjustment). Percentages of significant differentially abundant KOs organized into pathways are represented here. **B)** Statistically significant KOs (0.01 alpha cutoff) organized into pathways and identified as differentially abundant in PD (black), AN (purple), and T (green) tissues as compared to H tissue (red reference line). KO abundances were transformed to log2 fold changes. **C)** Statistically significant KOs (0.01 alpha cutoff) when adjusted for age and race organized into pathways identified as differentially abundant in AN (purple) and T (green) tissues as compared to H tissue.

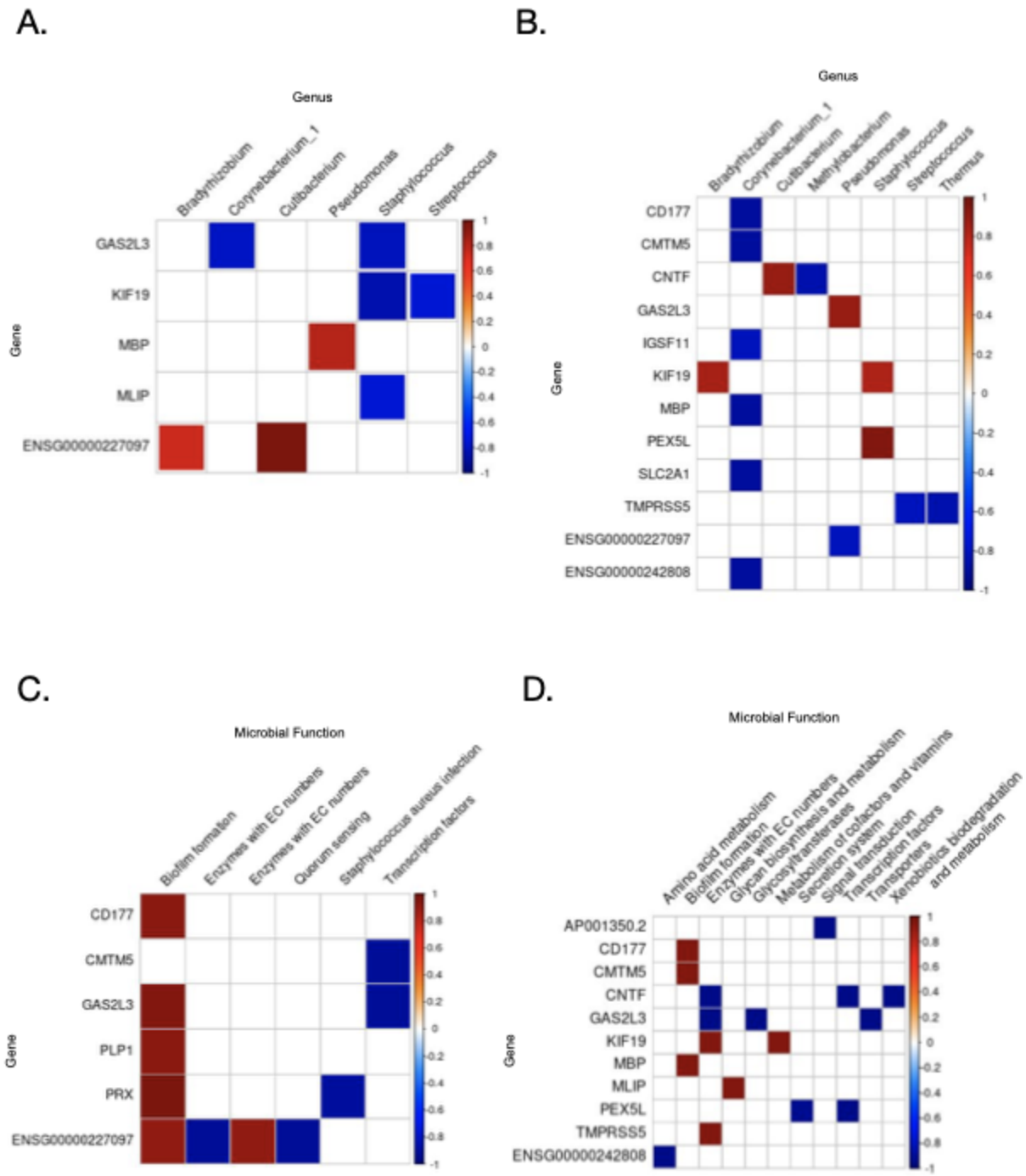


Figure 4. Heatmaps representing correlations between human genes and mammary tissue microbiome. A) Significant gene and microbe correlations in the H group and **B)** the PD group identified based on Spearman coefficients ($p<0.05$, unadjusted). The magnitude of significant

correlations is represented with a colored square while a white square represents a correlation with no significance. **C)** Significant correlations between genes and bacterial KOs identified in the H group and **D)** the PD group based on Spearman coefficients ($p<0.01$, unadjusted). In some cases, multiple KOs represent the same biological pathway. All KO - gene correlations are represented in the supplementary tables (S1 - S4). $n_H = 7$, $n_{PD} = 6$.

Tables:

Table I. Cohort characteristics and regression analysis of metadata and cancer status.

Variable Category	n, %	*P* for Healthy versus Cancer Groups	*P* for Healthy versus Pre-diagnostic Groups
Age Range			
27-45	41, 29.078%	0.6884	0.274
46-56	46, 32.624%	0.4209	0.352
57-82	54, 38.298%	†	†
Race			
African American	28, 19.858%	0.5949	0.993
White	112, 79.433%	†	†
Menopausal Status			
Uterine ablation	1, 0.706%	NA	0.997
Post-menopausal	55, 39.007%	†	†
Premenopausal	36, 25.532%	0.0379 *	0.462
BMI Category			
Normal weight	24, 17.021%	0.1749	0.806
Obese	61, 43.262%	†	†
Overweight	29, 20.567%	0.8258	0.381
Underweight	2, 1.418%	0.9915	0.752
History of Cancer			
Yes	1, 0.709%	NA	0.996
No	64, 45.39%	NA	†
Status at time of donation			
Healthy	50 (35.46%)	NA	NA
Pre-diagnostic (Tissue collected prior to cancer diagnosis)	15 (10.64%)	NA	NA

Diagnosed (Has developed breast cancer)	76 (53.9%)	NA	NA

† Reference group.

*Significant, $p < 0.05$.

Citations for the technical report:

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.