Final Report: The Relationship between Microbiome Composition and Development of Breast Cancer

Charlotte Lopp

December 2, 2017

# Introduction

Breast cancer affects one in eight women in their lifetime and according to the American Cancer Society, it is the second leading cause of cancer death in women with only lung cancer killing more women each year. Over the past twenty years $5.5 billion has been dedicated to breast cancer research but the origins of a majority of breast cancer cases remain unknown (Madigan *et al.*, 1995). Part of what is known is that migration studies have shown an increased incidence of breast cancer among migrants and their descendants after they move from a region of low breast cancer risk to a region of high risk (Le *et al.*, 2002, Shimizu *et al.* (1991)). Environmental factors as well as genetics play a role in disease development but which environmental factors and the extent of their influence is still being explored (Urbaniak *et al.*, 2016); diet, age and genetic disposition are established risk factors (Xuan *et al.*, 2014). In fact, there are many epidemiological evidences and research studies that have suggested that diet plays an important role in breast cancer prevention and progression (Aragón *et al.*, 2014). Specifically, there are recent studies that suggest that eating fermented foods such as probiotic yogurt play a role in providing protection (Lakritz *et al.*, 2014). The recent appreciation of the influential role that microbiota has on human health and disease has led to the investigation of if microbes play a role in breast cancers and if they do, to what extent (Xuan *et al.*, 2014).

The human microbiota refers to the collection of microbes inhabiting the human body; microbes inhabiting the human body outnumber human cells 10:1 (Xuan *et al.*, 2014). This microbiota plays an integral role in human development and changes in an individual's composition have or an imbalance has been implicated in various human diseases. Individuals with periodonitis (Darveau, 2010, Ximénez-Fyvie *et al.* (2000)), inflammatory bowel disease (Frank *et al.*, 2007), psoriasis (Gao *et al.*, 2008), asthma (Hilty *et al.*, 2010), diabetes (Larsen *et al.*, 2010), bacterial vaginosis (Hummelen *et al.*, 2010), and colorectal cancer (Mira-Pascual *et al.*, 2015) have different bacterial communities than individuals. Although it is unclear whether the these microbial differences are a consequence or a cause of the disease, there is evidence in favor of a causal relationship, as healthy animals transplanted with feces from those with obesity, colitis, or colorectal cancer then go on to develop disease (Garrett *et al.*, 2007, Turnbaugh *et al.* (2009), Zackular *et al.* (2013)). In obese individuals, the ratio of Firmicutes to Bacteroidetes in the colon is significantly higher than in lean individuals (Turnbaugh *et al.*, 2006, Ley *et al.* (2006)). Placing these obese individuls on low-fat diets resulted in a decrease in this ratio, although not to the levels seen in lean individuals (Ley *et al.*, 2006). In colon cancer, the overabundance of a single bacterial species *Fusobacterium nucleatum* correlates with disease and increased likelihood of lymph node metastasis (Castellarin *et al.*, 2012). In contrast to the detrimental affects of *Fusobacterium nucleatum*, the bacterium *Bacteroidetes fragilis* exerts a protective effect against colitis by modulating inflammatory immune responses in the gut (Mazmanian *et al.*, 2008). These recent studies have made it increasingly apparent that microbiota composition as well as specific bacterial species can have either preventive or progressive effects in terms of maintaining health or encouraging disease development, respectively (Xuan *et al.*, 2014). Thus, we are led to the question of what bacterial species, if any, have effect on breast cancer and how are these effects are manifested.

The first step in answering this question is determining that there is in fact a microbiome within the mammary tissue - a question that was addressed by Urbanick et al. (2014). In the female mammary gland, milk has been shown to contain bacterial species which ostensibly reach the ducts from the skin (Urbaniak *et al.*, 2014). The rationale of Urbaniak et al. (2014) was that given the nutrient-rich fatty composition of the female breast, the widespread vasculature and lympathics, and the diffuse location of the lobules and ducts leading from the nipple, bacteria would be widespread within the mammary glands, irrespective of lactation. Using 16S rRNA sequencing and culture, they analyzed breast tissue from 81 women with and without cancer in Canada and Ireland. A diverse bacteria population was detected within the tissues collected from sites all around the breast in women raging from 18 to 90 years old, not all of whom had a history of lactation. They stated that they successfully confirmed a breast microbiome in this study and added the study by Xuan et al. (Xuan *et al.*, 2014) that established the same conclusion. They found that *Proteobacteria* was the most abundant phylum in breast tissue where members of this phylum make up only a small portion of the overall bacterial community - a case unlike those seen in the vagina, oral cavity, bladder, skin, and gastrointestinal tract. This finding was consistent with other pieces of literature and suggested that breast tissue may in fact have a unique microbiota, distinct from that found at other body sites. The higher abundance of *Proteobacteria* and *Firmicutes* (specifically the class *Bacilli*) was perhaps due to host microbial adaptation to the fatty acid environment in the tissue. Another thing of interest was that *Proteobacteria* is also the principle phylum in human milk (Ward *et al.*, 2013) along with many of the same bacteria that they had detected in the tissue which raises the possibility that the tissue microbiota could be a source of bacterial inocula for babies. Despite having some difficulties with culturing different bacterial strains, all the bacteria that were cultured were also detected by 16s rRNA sequencing, supporting the argument that the DNA amplimers were from viable bcateria and not solely from remnant bcaterial DNA (Urbaniak *et al.*, 2014).

The use of 16S rRNA gene sequences for bacterial identification has been by far the most common housekeeeping genetic marker used for a multitude of reasons (Janda and Abbott, 2007). The highest ranking of these include: (1) its presence in almost all bacteria, often existing as a multigene family, or operons; (2) the function of the 16S rRNA gene over time has not changed, suggesting that random sequence changes are a more accurate measure of time (evolution); and (3) the 16S rRNA gene (1,500 bp) is large enough for informatics purposes (Patel, 2001). 16S rRNA gene sequence analysis can better identify poorly described, rarely isolated, or phenotypically aberrant strains, can be routinely used for identification of mycobacteria, and can lead to the recognition of novel pathogens and noncultured bacteria (Clarridge, 2004). It is easy to use and performs well as well in comparison to the more cumbersome manipulations of involving the "gold standard" for proposed new species DNA-DNA hybridization investigations (Janda and Abbott, 2007). One of the most attractive aspects of 16S rRNA gene sequence informatics is to provide genus and species identification for isolates that do not fit any recognized biochemical profiles, for strains generating only a "low likelihood" or "acceptable" identification according to commercial systems, or for taxa that are rarely associated with human infectious diseases (Janda and Abbott, 2007). The Illumina MiSeq platform (San Diego, CA, USA) provides researchers with a scalable, high-throughput and streamlined sequencing platform to survey community composition from clinical and environmental samples (Fadrosh *et al.*, 2014). Thus, Urbaniak et al. (2016) used this method to identify taxa that may affect breast cancer in the relatively newly established breast microbiome.

As they had already established that breast tissue microbiome exists in Canadian and Irish women (Urbaniak *et al.*, 2014), Urbaniak et al. then went on to determine whether this local microbiome plays a role in modulating the risk of breast cancer development (Urbaniak *et al.*, 2016). From a new patient study group, they examined the breast microbiota of 70 women who varied across having breast cancer (normal adjacent tissue collected) or benign tumors (normal adjacent tissue collected) or were disease free. They isolated bacteria from cancer patients and then characterized and examined them for their abilities to induce DNA damage. The endeavour of Urbaniak et al. was to differentiate between baterial profiles in breast tissue of healthy women and those with breast cancer and see which bacteria influenced the progression or prevention of breast cancer. We expected to see a trend of increased abundance for certain bacteria in cancer patients and a trend of increased abundance for certain bacteria in healthy patients. If this trend is observed, we were curious about the overall relationship that this bacteria had between tissue types - will one with higher expression in patients with cancerous tumors be seen in minimal abundance in healhty patients, etc.

# Methods

## Study design

Urbaniak et al. collected fresh breast tissue from 71 women ranging from 19 to 90 years old who were undergoing breast surgery at St. Joseph's Hospital in London, Ontario, Canada. Ethicla approval was obtained from the Western Research Ethics Board and Lawson Health Research Institute, London, Ontario, Canada. Subject provided written consent for sample collection and subsequent analysis. 58 of these women underwent lumpectomies or mastectomes for either benign (*n* = 13) or cancerous (*n* = 45) tumors, and 23 were free of disease and underwent either breast reductions or enhancements. For those women with tumors, the tissue obtained for analysis was collected outside the marginal zone, approximately 5 cm away from the tumor. None of the subjects had been on antibiotics for at least 3 months prior to collection.

## Sample origin and sequencing

After excision, fresh tissue was immediately placed in a sterile vial on ice and homogenized within 30 minutes of collection. After tissue homogenates in sealed containers were thawed on ice, procedures for DNA isolation ensued. For the 16S rRNA gene sequencing, PCR amplification, the genomic DNA isolated from the clinical samples was amplified using barcoded primers that amplified the V6 hypervariable region of the 16S rRNA gene (70 bp long): V6-forward, 5′ACACTCTTTCCCTACACGACGCTCTTCCGATCTnnnn(8)CWACGCGARGAACCTTACC3′; and V6-reverse, 5′CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTnnnn(8)ACRACACGAGCTGACGAC3′. The pooled PCR purified sample was then paired-end sequenced on the Illumina Mi-Seq platform using a 150 cycle kit with a paired-end 80-bp run at the London Regional Genomics Center, London, Ontario, Canada, following standard operating procedures. Custom Perl (Lindner and Hoffmann, 2004) and Bash (Badal and Sempau, 2006) scripts were used to to demultiplex the reads and assign barcoded reads to individual samples followed by multiple layers of filtering.

## Computational

Computation with this data began with creating a bashscript with code which downloaded the list of 68 files (from the NCBI Sequence Read Archive study number SRP076038) in the run table to the project's data directory. Before we could do anything else, we have to load the following libraries: mctoolsr (Leff, 2015), seqinr (Charif and Lobry, 2016), citr (Wickham and Chang, 2015a), dplyr (Wickham and Francois, 2015), tidyr (Wickham, 2014), knitr (Xie, 2013), dada2 (Callahan *et al.*, 2016), ggplot2 (Wickham and Chang, 2015b) and phyloseq (McMurdie and Holmes, 2013). After looking at the quality profiles of all 68 samples, we adjusted the filter and trim parameters to a maximum length of 100 and a minimum length of 55. We then created a table of all the samples of their original lengths and then their output lengths after trimming to ensure that everything looked correct. The next steps was plotting the error models we built to ensure that the error models match the data. Next, we got rid of any duplicate sequences and then ran DADA2. Then, we created a species matrix ("sequence\_table"), made a histogram in order to look at the distribution of the trimmed and denoised sequences, and removed any chimeras. We proceeded to make a table of the outputs of each step ("track"), assigned taxonomy ("taxa"), and then used this taxonomy to create a phylogenetic tree with Geneious (Drummond *et al.*, 2011). Finally, we created a phyloseq object from DADA2 outputs and saved everything to a file ("output/phyloseq\_obj.RData") which was then loaded into the Rmd file for the final report using the 'load()' function.

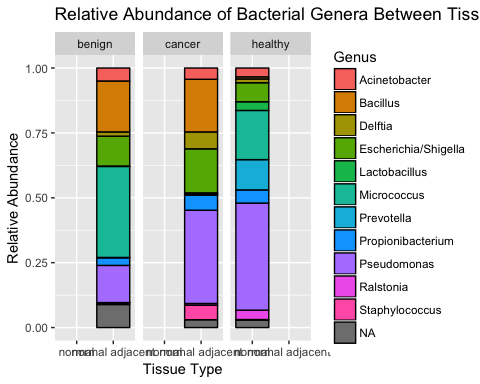
# Results

## Warning: package 'dplyr' was built under R version 3.4.2

## Warning: package 'tidyr' was built under R version 3.4.2

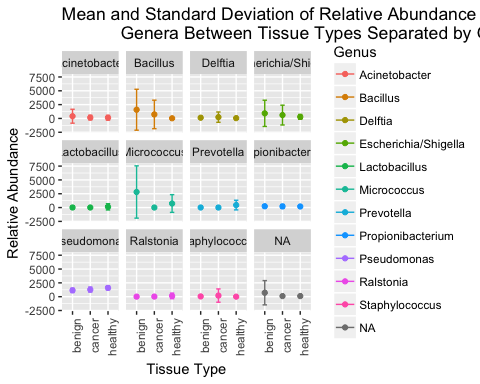
## Warning: package 'Rcpp' was built under R version 3.4.2

In order to look at bacterial relative abundance in a meaningful way, we had to first narrow down the number of genera we were looking at by separating those with the highest abundance. Figure 1 shows the 11 most abundant genera seen in normal adjacent tissue in patients with benign tumors, normal adjacent tissue in patients with cancerous tumors, and normal tissue in healthy patients. The relative abundance seen in patients wih benign tumors, from greatest to least, is *Micrococcus*, *Bacillus*, *Pseudomonas*, *Escherichia/Shigella*, *Acinetobacter*, *Propionibacterium*, and *Delftia*. Looking at patients with cancerous tumours, we see the following relative abundance: *Pseudomonas*, *Bacillus*, *Escherichia/Shigella*, *Delftia*, *Propionibacterium*, *Staphylococcus*, and *Acinetobacter*. The relative abundance in healthy patients is as follows: *Pseudomonas*, *Micrococcus*, *Prevotella*, *Escherichia/Shigella*, *Propionibacterium*, *Acinetobacter*, *Ralstonia*, *Lactobacillus*, *Delftia* and *Bacillus*. Every bacterial genera is represented in healthy patients except for *Staphylococcus*.



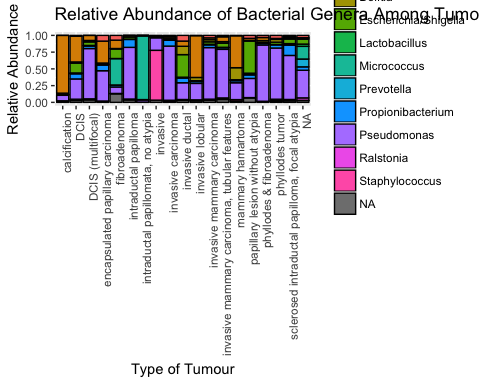
**Figure 1**: This figure shows the relative abundance of bacterial genera among tissue types.

While Figure 1 allows for visualization of the bacterial abundance relative to one another within a tissue type, the relative abundance of a bacterial genus across all three tissue is not as easily discernable. Thus, our next step was to calculate the mean and standard deviation of each bacterial abundance across all three tissue types for comparison amongst one another (Figure 2). A higher abundance is seen in cancer patients than healthy patients for *Acinetobacter*, *Bacillus*, *Delftia*, *Escherichia/Shigella*, *Micrococcus*, *Staphylococcus*, and *Propionibacterium* (although this difference is barely discernable). A higher abundance is seen in healthy patients than cancer patients for *Lactobacillus*, *Prevotella*, and *Ralstonia*. There is no noticeable difference in terms of abundance between cancer or healthy patients for *Pseudomonas*. The largest margins of difference in relative abundance between cancer patients and healthy patients are seen in *Bacillus*, *Escherichia/Shigella*, and *Micrococcus*. However, for *Micrococcus* benign tumor patients have the greatest relative abundance, then healthy patients have the next highest, followed by cancer tumor patients who have a relative abundance of zero.



**Figure 2**: This figure shows the relative abundance via mean and standard deviation of bacterial genera between tissue types, futher separated by genus.

We were then curious about the relationship between the relative abundance of bacterial genera and the type of tumor that the cancer patients had (Figure 3). *Pseudomonas* is seen in a majority abundance for most of the tumour types; for "calcification," "fibroadenoma," and "intraductal papillomata, no atypia", it is seen in its lowest levels. *Bacillus* is the dominant bacterial genera in terms of abudance for "calcification," "invasive lobular" and "mammary hamartoma." However, it is present in every tumor type except for "intraductal papillomata, no atypia," and "invasive."" *Staphylococcus* is the large majority of bacterial genera abundance for tumor type "invasive." It is also seen in most tumor type except for "DCIS (multifocal)," "encapsulated papillary carcinoma," "fibroadenoma," "intraductal papillomata, no atypia," "papillary lesion without atypia," "phyllodes & fibroadenoma," and "N/A" (which represent the healthy patients). For "intraductal papillomata, no atypia," almost all of the bacterial genera abundance is *Micrococcus*. The only other tumor where it is a majority of the abundance is for tumor type "fibroadenoma." *Lactobacillus* is only seen in "sclerosed intraductal pailloma, focal atypia" and "papillary lesion without atypia" in addition to healthy patients. *Propionibacterium* is seen in every single tumor type except for "indtraductal papilloma." *Escherichia/Shigella* is seen in every single tumor type except for "intraductal papillomata, no atypia" and its seen in highest abundance for "invasive ductal" and "papillary lesion without atypia." *Actinobacter* is seen in relatively low abundance for almost all tumor types except "calcification," "DCIS (multifocal)," "intraductal papilloma" and "intraductal papillomata, no atypia." *Ralstonia* is seen in most of the tumor types except for "calcification," "encapsulated papillary carcinoma," "fibroadenoma," "intraductal papillomata, no atypia," "invasive," "mammary hamartoma," "papillary lesion without atypia" and "phyllodes & fibroadenoma." *Delftia* is seen in every tumor type except "intraductal papillomata, no atypia" and "invasive carcinoma." It is seen in its highest abundance for "invasive ductal" and "mammary hamartoma." Finally, *Prevotella* is seen only in "papillary lesion without atypia," "mammary hamartoma" and "sclerosed intraductal papilloma, focal atypia" in addition to healthy patients.

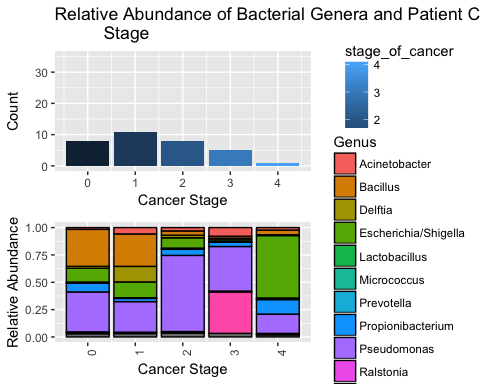


**Figure 3**: This figure shows the relative abundance of bacterial genera among the different types of tumors seen in the cancer patients.

After seeing the relationship between relative abundance of bacterial genera and tumor type, we were interested to see what the relationship was between relative abundance of bacterial genera and cancer stage (Figure 4). Here we see that for stage cancer 4, the relative abundance from highest to lowest is *Escherichia/Shigella*, *Pseudomonas*, *Propionibacterium*, *Bacillus* and *Acinetobacter*. Stage 4 has the lowest number of patients, the number being one, and has the highest relative abundance of *Escherichia/Shigella* of all of the cancer stages. Stage 3 cancer, which had the second lowest number of patients, is made up of primarily *Staphylococcus* and *Pseudomonas* followed by smaller amounts of *Acinetobacter*, *Propionibacterium*, *Escherichia/Shigella* and *Bacillus*. This stage is the only stage to exhibit the presence of *Staphylococcus*. Stage 2 cancer, which had the same number of patients as stage 0 cancer, has *Pseudomonas* as the most abundant bacteria. The rest of the abundance from largest amount to smallest is *Escherichia/Shigella*, *Propionibacterium*, *Bacillus*, *Acinetobacter*, and *Delftia*. For cancer stage 1, which had the most patients, the abundance is *Pseudomonas*, *Bacillus*, *Escherichia/Shigella*, *Delftia*, *Acinetobacter*, and *Propionibacterium*. Finally, for stage 0 cancer, the abundance is *Pseudomonas*, *Bacillus*, *Escherichia/Shigella*, *Propionibacterium*, *Delftia*, and *Acinetobacter*.

## Warning: Removed 1 rows containing missing values (geom\_col).

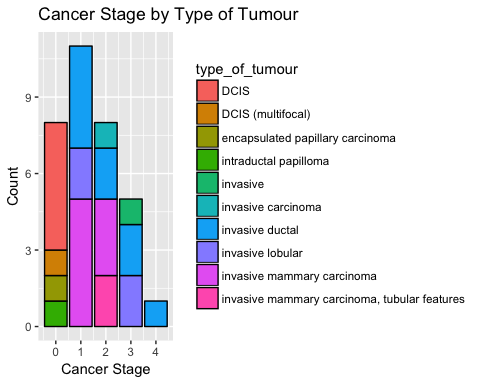
## Warning: Removed 12 rows containing missing values (position\_stack).



**Figure 4**: This figure shows the relative abundance of bacterial genera for each cancer stage.

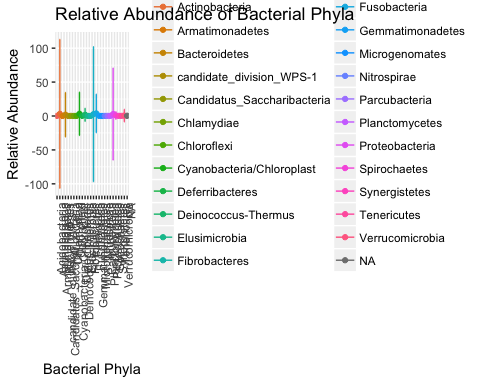
Once we saw the relationship between relative abundance of bacterial genera and cancer stage, we wondered how tumor types were distributed across the cancer stages (Figure 5). Since we had already established what the relative abundance of bacterial genera was for each type of tumor, we believed this new figure would shed some more light on the roles that certain bacteria play. In cancer stage 4, the only type of tumor is "invasive ductal." For cancer stage 3, there's "invasive lobular," "invasive ductal," and "invasive." "Invasive mammary carcinoma, tubular features," "invasive mammary carcinoma," "invasive ductal," and "invasive carsinoma" are seen in cancer stage 2. In cancer stage 1, we see the majority being either "invasive mammary carcinoma" and "invasive ductal" with the rest being "invasive lobular." Most of cancer stage 0 is "DCIS" with the rest being "DCIS (multifocal," "encapsulated papillary carcinoma," and "intraductal papilloma." "Invasive ductal" is seen in every cancer stage except for stage 0.

## Warning: Removed 9 rows containing missing values (position\_stack).



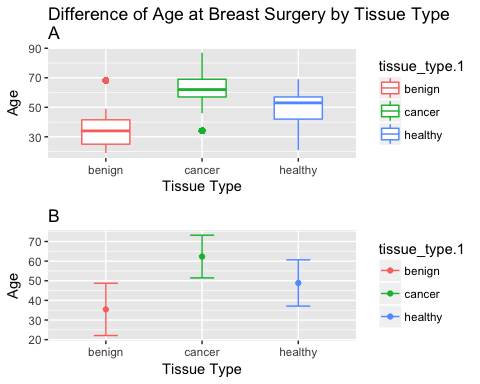
**Figure 5**: This figure shows the relationship between cancer stage and type of tumor.

After seeing different relationships involving relative abudance of bacterial genera, we thought it would be benefical to see the relative abundance among bacterial phyla (Figure 6). Visualizing this relationship would also be helpful in relating the findings in this paper with other literature on the same subject. The largest abundance is in *Actinobacteria* followed by *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Cyanobacteria/Chloroplast* which are about the same, *Fusobacteria* coming in not far behind, *Verrucomicrobia*, and *Spirochaetes*.



**Figure 6**: This shows the relative abundance of bacterial phyla.

Finally, we wanted to look at how the patients were distributed across tissue type in terms of age (Figure 7). In 7A we see that the median for patients with benign tumors is just above 30, while for patients with cancerous tumors is just above 60, and for healthy patients it is just above 50. The thick line within the box represents the median. The box itself signifies the first and third quartile meaning that the box represents where 50% of the samples lie. The whiskers show the miniumum and maximum of the data points. In 7B we see that the mean is about 35 for patients with benign tumors, for patients with cancerous tumors its just above 60 and for healthy patients its just below 50. The point represents the mean while the bars represent the standard deviation.



**Figure 7**: This figure shows the relationship between age and tissue type.

# Discussion

According to the American Cancer Society, there are more than 200,000 cases of breast cancer in the United States per year. The second leading cause of cancer death in women, breast cancer affects on in eight women in their lifetime. Environmental factors as well as genetics play a role in breast cancer development; factors such as diet, age and genetic disposition have been established as risk factors. The recent appreciation of the influential role that microbiota has on human health and disease has led to the investigation of if microbes play a role in breast cancers and if they do, to what extent. In 2014, Urbaniak et al. established that breast tissue microbiome exists in Canadian and Irish women. To follow that study, they decided to look at whether this local microbiome plays a role in modulating the risk of breast cancer development. The endeavour of Urbaniak et al. in their 2016 study was to differentiate between baterial profiles in breast tissue of healthy women and those with breast cancer and see which bacteria influenced the progression or prevention of breast cancer.

*Bacillus* and *Escherichia/Shigella* have higher abundances in cancer patients (which includes both groups of patients with cancerous tumors and patients with benign tumors) than healthy patients (Figure 1). This is suggestive that these bacterial genera are associated with the development of tumors (whether they are benign or cancerous) as their respective presences in healthy patients are greatly decreased. *Prevotella* and *Lactobacillus* are only seen in healthy patients whcih suggests that these two bacterial genera are associated with a lack of tumor growth - whether this is a preventative relationship in terms of increasing the abundance of these bacteria provides a protective measure against the development of breast cancer has yet to be determined. *Micrococcus* is present in patients with benign tumors and healthy patients with a relatively somewhat large abundance while in cancer patients it is not seen. This raises the question of it this bacteria plyas a role in making the tumors benign instead of malignant. Another difference between patients with benign tumors versus cancerous tumors is the presence of *Staphylococcus;* its present in a much higher relative abundance in patients with cancerous tumors in comparison to those with benign tumors (for which it appears barely at all). When looking at the makeup of each tissue type's microbiome profile, it is interesting to note that the profile for patients with benign tumors more closely resembles that of the profile of patients with cancerous tumors than the profile of healthy patients. This obervation raises the question as to why these women with benign tumors do not have cancer. As aforementioned, the presence of *Micrococcus* or the absence of *Staphylococcus* may lead to tumors being benign. However, the solution may lie in more idiosyncratic factors that lead to enhancement or reduction of certain tumor behaviors that lead to benign tumors rather than malignant ones. To take a closer look at the relative abundance, we calculated the mean and standard deviation of each bacterial abundance across all three tissue types for comparison within each bacterial genera (Figure 2).

What we see in Figure 2 futher supports our observations in Figure 1; there is a higher abundance of *Bacillus*, *Escherichia/Shigella*, *Delftia* and *Staphylococcus* in patients with cancerous tumors while a higher abundance of *Prevotella* and *Lactobacillus* is seen in healthy patients. *Acinetobacter* and *Micrococcus* have higher abundances in patients with benign tumors. In order to further investigate the questions raised earlier, we looked to other literature for possible answers. We found that *Bacillus*, in particular a *B. cereus* strain (all *Bacillus* strains cultured from the breast cancer patients, in the study our data is from, were of the species *B. cereus*), metabolizes the hormone progesterone into 5-alpha-pregnane-3,20-dione (5αP) (Ojanotko-Harri *et al.*, 1990). 5αP is higher in breast tumors than in healthy breast tissue (Wiebe *et al.*, 2000) and is believed to promote tumor development by stimulating cell proliferation (Wiebe *et al.*, 2000, wiebe2006progesterone). *Escherichia/Shigella* and *Staphylococcus* from normal adjacent tissue of breast cancer patients displayed the ability to to induce DNA double-stranded breaks by all isolates (Urbaniak *et al.*, 2016). Double-strand breaks are the most detrimental type of DNA damage and are caused by genotoxins, reactive oxygen species, and ionizing radiation (Lees-Miller and Meek, 2003). DNA-strand breaks caused by bacteria, such a certain strains of *E. coli*, have been shown to induce chromosomal instability with prolonged exposure (Cuevas-Ramos *et al.*, 2010, Toller *et al.* (2011)). *Lactobacillus* is a bacteria that exhibits anticarcinogenic properties and may play a role in prevention (Urbaniak *et al.*, 2016). Natural killer (NK) cells are vital in controlling tumor growth, with epidemiological studies showing that low NK cell activity (from peripheral blood mononuclear cells [PBMC]) is associated with an increased incidence of breast cancer (Imai *et al.*, 2000, Strayer *et al.* (1984)). *Lactobacillus* has been shown to activate murine splenic NK cells, enhancing cellular immunity (Kosaka *et al.*, 2012). It may also modulate cellular immunity by maintaining the cytotoxic activity of resident NK cells (Carrega *et al.*, 2014), thus helping to prevent cancer development. *Prevotella* produces the short-chain fatty acid (SCFA) propionate, which, like other SCFA, has many beneficial health effects in the gut, one of them being the ability to regulate colorectal tumor growth (Hosseini *et al.*, 2011). *Micrococcus* DNA has been used to induce interferons, activate natural killer cells and inhibit tumor growth (Yamamoto *et al.*, 1992). In one study, *Actinobacter* bacteria was used in order to derive deferoxamine (DFO), an iron chelator, which is clinically used as an iron-chelating drug in diseases of iron overload and make complex with iron ions (Dashtizadeh and Baharara, 2015). This iron depletion by iron chelators, especialy deferoxamine in some cancerous cells, leads to different expression of some of the molecules involved in the cell cycle and so, due to the iron depletion, apoptosis is induced (Richardson *et al.*, 2009). With this understanding of the roles that these bacteria play, the trends we are seeing in terms of the relationship between their relative abundances and tumor types and tissue types make complete sense. It also provokes the idea of investigating further whether the modulation the abundances of these bacteria would affect which tumor types were seen in patients, whether the tumors were benign or cancerour, or even if it would affect if tumors were even developed.

In follow-up to that query, we then looked at the relationship between relative abundance of bacterial genera and the tumor types seen in patients (Figure 3). It is interesting to note that *Bacillus* and *Escherichia/Shigella* were seen in almost every tumor type except two and one respectively. This makes sense considering the roles they play in tumor growth, cell proliferation, and DNA-strand breaks. It was surprising to see *Lactobacillus* and *Prevotella* appear in any cancer patient tumor types at all although they were present in very few and in very low abundances. "Intraductal papillomata, no atypia" was almost entirely dominated *Micrococcus* while "invasive" was almost entirely dominated by *Staphylococcus*. Observing this trend, we were curious if the high abundance of these bacteria were in any way causal to their respective tumor types or were merely a result of those tumor types. Also, there is a very high abundance of *Bacillus* in "calcification," "invasive lobular," and "mammary hamartoma" (in addition to it being present in almost every tumor type) whereas there is practically no *Bacillus* present in "N/A" which would be the healthy patients without any sort of tumor. Would decreasing relative abundance of *Bacillus* decrease the occurrence of these three types of tumors? Would it have an even greater effect in that there would be a decrease of tumors? Or would there be no affect at all? The same goes for "invasive" and "intraductal papillomata, no atypia" - would a decrease in *Staphylococcus* and *Micrococcus* result in a decreased occurrence of these tumors or even no tumor development at all?

Another relationship we were interested in was relative abundance of bacterial genera and cancer stage (Figure 4). When looking at this relationship, we see that the highest abundance of *Escherichia/Shigella* is in stage 4 patients while *Staphylococcus* dominates stage 3. *Escherichia/Shigella* is the majority after *Pseudomonas* for stage 2 cancer and *Bacillus* is the majority for both cancer stages 0 and 1. Considering that these three bacteria (*Escherichia/Shigella*, *Staphylococcus*, and *Bacillus*) have been the leading established bacteria so far in terms of being indicative or perhaps even promoting tumor growth, it is hardly surprising to see them present as a majority when looking at cancer stages. However, perhaps this trend is indicative of the progression that the microbiome goes through as patients progress through the cancer stages - first they have a majority of *Bacillus* which then decreases for stage 1 while *Escherichia/Shigella* increases, *Bacillus* and *Escherichia/Shigella* both decrease slightly for stage 3, stage 4 is characterized by a majority of *Staphylococcus*, and finally in stage 5 *Escherichia/Shigella* takes the large majority. It is worth noting, however, that although the bacteria are represented by their relative abundance, how many patients that are in each cancer stage still matters; stage 4 only has one patient and so if there were more patients, like over 10 as in stage 1, there may be a different bacteria composition than the one we see here. Another factor is the type of tumors represented in each cancer stage (Figure 5).

If the correlation in terms of microbiota composition lies in tumor type, it would be worth looking into the occurrence of each tumor type in each cancer stage. Here we see that "invasive ductal" tumors are in every cancer stage except for stage 0 and since we saw elevated levels of *Escherichia/Shigella* in Figure 4 for stage 4, that elevation must come from "invasive ductal" tumor type. Looking back at Figure 3, we see that "invasive ductal" does indeed have the second highest abundance of *Escherichia/Shigella* of all of the types of tumors. "Invasive lobular" is seen in stage 1 and 3 and "invasive mammary carcinoma" is seen in stages 1 and 2. Looking back at Figure 3, we see that "invasive lobular" had the second highest abundance of *Bacillus* while "invasive mammary carcinoma" had a majority of *Pseudomonas* with very little of anything else. This further explains the trends that we see in Figure 4 but also suggests that in order to better explore this idea of a correlation between microbiome composition and cancer stage, we would need to have more of an equal spread of patients across the different cancer stages to have more representative trends.

As it had been mentioned in other related literature as well as the previous study by Urbaniak et al. (Urbaniak *et al.*, 2014), we decided to also look at the relative abundance of the bacteria but by phyla (Figure 6). *Proteobacteria* had been expected to be one of the most if not the most abundance phylum since it is the principle phylum in human milk (Ward *et al.*, 2013) and is suspected, along with *Firmicutes*, to be prevalent due to host microbial adaptation to the fatty acid environment in the tissue (Xuan *et al.*, 2014). We see that top three phyla in terms of largest abundance is *Actinobacteria*, *Firmicutes*, and *Proteobacteria*. We were surprised to see that *Actinobacteria* was, by a discernable margin, the highest occurring phyla in the study. While this most likely has to do with the relationship it has with the human body, *Actinobacteria* is also one of the largest taxonomic units among the 18 major lineages currently recognized with the domain Bacteria (Ventura *et al.*, 2007). In terms of the role *Actinobacteria* plays in the human body, actinomycetes, which are a bacteria belonging to the phylum *Actinobacteria*, are producers of a large number of natural products with different biological activities which includes antitumor properties (Olano *et al.*, 2009). The way they exert antitumor activity is by inducing apoptosis (which is much like the role [@(Richardson *et al.*, 2009) we saw in the bacterial genus *Actinobacter* earlier in our investigation) through DNA cleavage mediated by topoisomerase I or II inhibition, mitochondria permeabilization, inhibition of key enzymes involved in signal transduction like proteases, or cellular metabolism and in some cases by inhibiting tumor-induced angiogenesis (Olano *et al.*, 2009). Thus, considering its antitumor activity, its interesting to see its presence in every single cancer stage (Figure 4), almost every single tumor type (Figure 3), and in each of the tissue types (Figure 1). It makes sense that it would be seen in a higher abundance in patients with benign tumors as *Actinobacteria* may also play a role in keeping tumors non-malignant (Figure 2). It is interesting, however, that it is not seen in a higher presence in healthy individuals. Its presence across all tissue types and tumor types and cancer stages could also be attributed to the fact that it is just one of the largest taxonomic units in the Bacteria domain.

Finally, we were interested in whether age played any influential role in the results that we saw in this study so we compared the ages of patients by tissue type (Figure 7). Here we that the mean and median age of patients with cancerous tumors is 62 years, for patients with benign tumors its 38 years and 36 years respectively, and for healthy patients its 49 years and 53 years respectively. So, there is a difference between tissue type groups. However, since there was no difference in microbiome composition between patients with benign tumors and patients with cancerous tumors (Figure 1), as was discussed earlier, we did not conclude that the dfferences observed between the healthy individuals and those with cancerous tumors were due to an age difference.

While this study did open new and intriguing avenues for future study, it is important to note that this study is not wholly conclusive as there were disparities in terms of patient balance in terms of age, cancer stage, and tissue type. This leads to potentially misrepresentation of the data and not entirely accurate conclusions. However, for human studies, 71 patients is a respectable number considering the cost and overall difficulty of conducting such trials. Another thing to note is that their entire study came from women who were having breast surgery at St. Joseph's Hospital in London, Ontario, Canada which means that this study is somewhat limited in terms of scope. Despite these limitations, though, this study does raise interesting questions in terms of whether or not the relationships observed are causal and if they are, is modulation of their abundance and control or manipulation of microbiome composition a potential avenue for breast cancer treatment or prevention.

# Sources Cited

Aragón,F. *et al.* (2014) Modification in the diet can induce beneficial effects against breast cancer. *World journal of clinical oncology*, **5**, 455.

Badal,A. and Sempau,J. (2006) A package of linux scripts for the parallelization of monte carlo simulations. *Computer physics communications*, **175**, 440–450.

Callahan,B.J. *et al.* (2016) DADA2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, **13**, 581–583.

Carrega,P. *et al.* (2014) CD56brightperforinlow noncytotoxic human nk cells are abundant in both healthy and neoplastic solid tissues and recirculate to secondary lymphoid organs via afferent lymph. *The Journal of Immunology*, **192**, 3805–3815.

Castellarin,M. *et al.* (2012) Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome research*, **22**, 299–306.

Charif,D. and Lobry,J. (2016) Seqinr: Biological sequences retrieval and analysis. *New York, US. URL https://cran. r-project. org/web/packages/seqinr/index. html*.

Clarridge,J.E. (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical microbiology reviews*, **17**, 840–862.

Cuevas-Ramos,G. *et al.* (2010) Escherichia coli induces dna damage in vivo and triggers genomic instability in mammalian cells. *Proceedings of the National Academy of Sciences*, **107**, 11537–11542.

Darveau,R.P. (2010) Periodontitis: A polymicrobial disruption of host homeostasis. *Nature Reviews Microbiology*, **8**, 481–490.

Dashtizadeh,A. and Baharara,J. (2015) Investigation of interaction between deferoxamine and low frequency electromagnetic field on angiogenesis in chick embryo. *Zahedan Journal of Research in Medical Sciences*, **17**.

Drummond,A. *et al.* (2011) Geneious, version 5.4. *Geneious, Auckland, New Zealand*.

Fadrosh,D.W. *et al.* (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the illumina miseq platform. *Microbiome*, **2**, 6.

Frank,D.N. *et al.* (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences*, **104**, 13780–13785.

Gao,Z. *et al.* (2008) Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. *PloS one*, **3**, e2719.

Garrett,W.S. *et al.* (2007) Communicable ulcerative colitis induced by t-bet deficiency in the innate immune system. *Cell*, **131**, 33–45.

Hilty,M. *et al.* (2010) Disordered microbial communities in asthmatic airways. *PloS one*, **5**, e8578.

Hosseini,E. *et al.* (2011) Propionate as a health-promoting microbial metabolite in the human gut. *Nutrition Reviews*, **69**, 245–258.

Hummelen,R. *et al.* (2010) Deep sequencing of the vaginal microbiota of women with hiv. *PloS one*, **5**, e12078.

Imai,K. *et al.* (2000) Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: An 11-year follow-up study of a general population. *The Lancet*, **356**, 1795–1799.

Janda,J.M. and Abbott,S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of clinical microbiology*, **45**, 2761–2764.

Kosaka,A. *et al.* (2012) Lactococcus lactis subsp. cremoris fc triggers ifn- production from nk and t cells via il-12 and il-18. *International immunopharmacology*, **14**, 729–733.

Lakritz,J.R. *et al.* (2014) Beneficial bacteria stimulate host immune cells to counteract dietary and genetic predisposition to mammary cancer in mice. *International journal of cancer*, **135**, 529–540.

Larsen,N. *et al.* (2010) Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PloS one*, **5**, e9085.

Le,G.M. *et al.* (2002) Cancer incidence patterns among vietnamese in the united states and ha noi, vietnam. *International journal of cancer*, **102**, 412–417.

Lees-Miller,S. and Meek,K. (2003) Repair of dna double strand breaks by non-homologous end joining. *Biochimie*, **85**, 1161–1173.

Leff,J. (2015) Mctoolsr: Microbial community data analysis tools.

Ley,R.E. *et al.* (2006) Microbial ecology: Human gut microbes associated with obesity. *Nature*, **444**, 1022–1023.

Lindner,T.H. and Hoffmann,K. (2004) EasyLINKAGE: A perl script for easy and automated two-/multi-point linkage analyses. *Bioinformatics*, **21**, 405–407.

Madigan,M.P. *et al.* (1995) Proportion of breast cancer cases in the united states explained by well-established risk factors. *JNCI: Journal of the National Cancer Institute*, **87**, 1681–1685.

Mazmanian,S.K. *et al.* (2008) A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature*, **453**, 620–625.

McMurdie,P.J. and Holmes,S. (2013) Phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, **8**, e61217.

Mira-Pascual,L. *et al.* (2015) Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. *Journal of gastroenterology*, **50**, 167–179.

Ojanotko-Harri,A. *et al.* (1990) Metabolism of progesterone and testosterone by bacillus cereus strain socransky 67 and streptococcus mutans strain ingbritt. *Molecular Oral Microbiology*, **5**, 237–239.

Olano,C. *et al.* (2009) Antitumor compounds from marine actinomycetes. *Marine drugs*, **7**, 210–248.

Patel,J.B. (2001) 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis*, **6**, 313–321.

Richardson,D. *et al.* (2009) Cancer cell iron metabolism and the development of potent iron chelators as anti-tumour agents. *Biochimica et Biophysica Acta (BBA)-General Subjects*, **1790**, 702–717.

Shimizu,H. *et al.* (1991) Cancers of the prostate and breast among japanese and white immigrants in los angeles county. *British journal of cancer*, **63**, 963–966.

Strayer,D.R. *et al.* (1984) Low natural cytotoxicity of peripheral blood mononuclear cells in individuals with high familial incidences of cancer. *Cancer research*, **44**, 370–374.

Toller,I.M. *et al.* (2011) Carcinogenic bacterial pathogen helicobacter pylori triggers dna double-strand breaks and a dna damage response in its host cells. *Proceedings of the National Academy of Sciences*, **108**, 14944–14949.

Turnbaugh,P.J. *et al.* (2009) A core gut microbiome in obese and lean twins. *nature*, **457**, 480–484.

Turnbaugh,P.J. *et al.* (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *nature*, **444**, 1027–131.

Urbaniak,C. *et al.* (2014) Microbiota of human breast tissue. *Applied and environmental microbiology*, **80**, 3007–3014.

Urbaniak,C. *et al.* (2016) The microbiota of breast tissue and its association with breast cancer. *Applied and environmental microbiology*, **82**, 5039–5048.

Ventura,M. *et al.* (2007) Genomics of actinobacteria: Tracing the evolutionary history of an ancient phylum. *Microbiology and molecular biology reviews*, **71**, 495–548.

Ward,T.L. *et al.* (2013) Human milk metagenome: A functional capacity analysis. *BMC microbiology*, **13**, 116.

Wickham,H. (2014) Tidyr: Easily tidy data with spread () and gather () functions. *R package version 0.2. 0*.

Wickham,H. and Chang,W. (2015a) Devtools: Tools to make developing r code easier. *R package version*, **1**.

Wickham,H. and Chang,W. (2015b) Ggplot2: An implementation of the grammar of graphics. 2015. *URL http://CRAN. R-project. org/package= ggplot2. R package version*, **1**.

Wickham,H. and Francois,R. (2015) Dplyr: A grammar of data manipulation. *R package version 0.4*, **1**, 20.

Wiebe,J.P. *et al.* (2000) The 4-pregnene and 5-pregnane progesterone metabolites formed in nontumorous and tumorous breast tissue have opposite effects on breast cell proliferation and adhesion. *Cancer research*, **60**, 936–943.

Xie,Y. (2013) Knitr: A general-purpose package for dynamic report generation in r. *R package version*, **1**, 1.

Ximénez-Fyvie,L.A. *et al.* (2000) Comparison of the microbiota of supra-and subgingival plaque in health and periodontitis. *Journal of clinical periodontology*, **27**, 648–657.

Xuan,C. *et al.* (2014) Microbial dysbiosis is associated with human breast cancer. *PloS one*, **9**, e83744.

Yamamoto,S. *et al.* (1992) DNA from bacteria, but not from vertebrates, induces interferons, activates natural killer cells and inhibits tumor growth. *Microbiology and immunology*, **36**, 983–997.

Zackular,J.P. *et al.* (2013) The gut microbiome modulates colon tumorigenesis. *MBio*, **4**, e00692–13.