May 2016

Characterization of microbial communities in the mammary glands and the influence of this microbiota on health and disease

Camilla Urbaniak
The University of Western Ontario

Supervisor
Dr. Gregor Reid
The University of Western Ontario

Graduate Program in Microbiology and Immunology

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Camilla Urbaniak 2016

Follow this and additional works at: http://ir.lib.uwo.ca/etd

Part of the Medical Sciences Commons

Recommended Citation
http://ir.lib.uwo.ca/etd/3738
Abstract

The collection of microbes and their genomic activity within us, referred to as the human microbiome, is crucial in maintaining health, and disruptions in composition and/or activity have been associated with various diseases. Considering the importance of breast milk in neonatal development and the high rate of breast cancer in women, the objective of my thesis was to obtain a comprehensive analysis of the microbiota of the mammary glands and to examine the influence of these organisms on health and disease.

Human milk is an important source of bacteria for the developing infant. While a variety of studies over the last 20 years have provided a greater appreciation for how diverse and variable the milk microbiota is, little is known about these bacterial communities. Many studies have emphasized the adverse effects of antibiotics on the infant microbiota, however the impact of drugs on the lactating mother has not been studied. In Chapter 2 we report the changes caused by chemotherapy (over a 4 month period) on the milk microbiota and metabolome of a woman undergoing treatment for Hodgkin’s lymphoma. In Chapter 3, we show from a study of 39 Canadian women that neither gestation, mode of delivery nor gender of the child explains the diversity of the milk microbiota.

As milk is not sterile and given the nutrient-rich fatty composition of the breast and its exposure to the external environment via the nipple, we hypothesized that breast tissue has its own indigenous microbiota. In Chapter 4, we confirmed this in a study of breast tissue from 81 Canadian and Irish women, with the discovery of diverse bacterial communities dominated by the phylum Proteobacteria. To determine whether this tissue microbiota could play a role in breast cancer development, we compared, in Chapter 5,
bacterial profiles in tissue between women with breast cancer and those who were
disease free. We show that differences do exist and that women with cancer have higher
numbers of bacteria with the ability to induce host DNA damage.

This work forms the platform for further studies that examine how breast milk and tissue
are first colonized and how these microbiotas can be manipulated to promote optimal
maternal and infant health.

**Keywords**

Human microbiome, breast milk microbiome, breast tissue microbiome, 16S rRNA gene
sequencing, breast cancer, DNA damage
Co-Authorship Statement

Chapter 2: Effect of chemotherapy on the microbiota and metabolome of human milk, a case report
I co-designed the study, recruited and collected milk samples from lactating mothers, extracted and amplified the milk DNA for sequencing on the Ion Torrent platform, analyzed 16S rRNA gene sequencing data and wrote the manuscript. Amy McMillan ran the milk samples on GC-MS and performed the metabolome analysis.

Chapter 3: Human milk microbiota profiles in relation to birthing method, gestation and infant gender
I co-designed the study, recruited and collected milk samples from lactating mothers, extracted and amplified the milk DNA for sequencing on the Illumina MiSeq platform, analyzed 16S rRNA gene sequencing data and wrote the manuscript.

Chapter 4: Microbiota of human breast tissue
I recruited the Canadian subjects, extracted DNA from the tissue samples and plated tissue homogenates for culture analysis. Joanne Cummins recruited and collected breast tissue from Irish women and extracted DNA from these Irish samples.

I amplified the extracted DNA from both the Canadian and Irish samples and pooled the samples in equimolar amounts in preparation for sequencing. I also analyzed the 16S rRNA gene sequencing data and wrote the manuscript.

Chapter 5: The microbiota of breast tissue and its association with breast cancer
I co-designed the study, recruited subjects, extracted the DNA from the samples, amplified the DNA for sequencing on the Illumina MiSeq platform, analyzed 16S rRNA gene sequencing data, performed the γH2AX DNA damage assay and wrote the manuscript.
Acknowledgments

It is fascinating how small, seemingly unassuming events, can have a substantial impact on an individual’s life. For me, those are memories of late evenings in high school, sitting in the Family room with my Dad while everyone else was asleep, working on math and physics problems together. It is during these late night sessions that I began to appreciate the inquiry process of science and the excitement that comes with questioning the unknown. I would never have been able to do this without you Dad. Your versatile talents never cease to amaze me- you are a computer engineer with a PhD in Physics, who can remodel a full kitchen and basement like an expert and whom can fix any problem or answer any question that I have. I hope that one day I can be at least half as talented as you.

To mom: Thank-you for dedicating your time and energy into allowing me to experience so many facets of life. Because of you I am creative, compassionate and adventurous. No matter what roadblocks come my way, I am well equipped to deal with them because of you.

To Jahan: I cannot express how lucky I am to have met you in undergrad. Your wit and humour has had me doubled over in laughter more times then I can count. I could not imagine my life without you. Thank-you for everything!

To Jane: Thank-you for being my rock and my shoulder to complain on. Thank-you for being there through all the tough choices and decisions I have had to make. Your zest for life is infectious and I hope that we continue to have many more adventures together.
To Ania: You are one of the smartest people I know (along with Jane). You have been a constant source of knowledge and feedback throughout the years. Thank-you for being receptive of my “thought” questions and a friend I can trust with all the moments in my life, good or bad. I hope we continue to be friends for many more years to come.

To Jorum: Who would have thought that when I walked through the front door of the lab in 2005 we would become such great friends. You have been my best supporter, champion and advocate and have believed in me during times when I didn’t believe in myself. There are not enough words to express my gratitude for all your kindness. I hope that one day we can again work together.

There have been so many people that have come into my life throughout the years. Whether it has just been in passing or for a longer period of time, you have all made a difference in my life and have made me who I am today.

Finally, I would like to thank my committee members, Muriel Brackstone, Greg Gloor and Jim Koropatnick for their mentorship and guidance throughout my PhD. I would especially like to thank my supervisor, Gregor Reid, for his support. His dedication to his students is truly inspirational and without him I would not have learned and have achieved as much as I have. You are one of a kind and I am truly blessed to have been able to do a PhD with you.

~“Many of life’s failures are people who did not realize how close they were to success when they gave up” - Thomas A. Edison~
# Table of Contents

Abstract................................................................................................................................................. ii  
Co-Authorship Statement..................................................................................................................... iv  
Acknowledgments................................................................................................................................. v  
Table of Contents................................................................................................................................... vii  
List of Tables ........................................................................................................................................... xii  
List of Figures ......................................................................................................................................... xiii  
List of Abbreviations ............................................................................................................................. xvi  
Chapter 1.............................................................................................................................................. 1  
  1 General introduction......................................................................................................................... 1  
    1.1 Functional anatomy of the female breast.................................................................................... 1  
      1.1.1 Embryogenesis.................................................................................................................. 1  
      1.1.2 Puberty............................................................................................................................ 2  
      1.1.3 Reproduction.................................................................................................................... 2  
      1.1.4 Structure of ducts and alveoli ......................................................................................... 5  
    1.2 Human microbiome .................................................................................................................... 5  
      1.2.1 Influence of the microbiota on health and disease......................................................... 6  
      1.2.2 Early life factors that shape our microbiota................................................................. 8  
    1.3 Human milk microbiota ............................................................................................................. 9  
      1.3.1 Importance of milk bacteria in neonatal and long term health................................. 10  
      1.3.2 Where do breast milk bacteria come from? ................................................................. 13  
    1.4 Human microbiome and breast cancer.................................................................................... 17  
      1.4.1 Causative role of bacteria in breast cancer development........................................... 17  
      1.4.2 Protective role of bacteria in breast cancer development............................................. 19  

1.4.3 Bacteria in the breast and their role in breast cancer development........... 20
1.5 High-throughput sequencing to analyze the human microbiota ...................... 21
1.5.1 16S rRNA gene targeted sequencing ......................................................... 22
1.6 Scope and objectives ................................................................................ 24
1.7 References ................................................................................................. 27

Chapter 2 ........................................................................................................ 42

2 Effect of chemotherapy on the microbiota and metabolome of human milk, a case report ................................................................. 42
2.1 Introduction ................................................................................................. 42
2.2 Methods ..................................................................................................... 45
2.2.1 Clinical samples and study design ......................................................... 45
2.2.2 Milk collection and processing .............................................................. 45
2.2.3 DNA isolation ......................................................................................... 45
2.2.4 Ion Torrent V6 16S rRNA sequencing .................................................. 46
2.2.5 GC-MS sample preparation ................................................................ 49
2.3 Results ...................................................................................................... 51
2.4 Discussion ................................................................................................. 60
2.5 Conclusion ................................................................................................. 63
2.6 References ................................................................................................. 64

Chapter 3 ........................................................................................................ 70

3 Human milk microbiota profiles in relation to birthing method, gestation and infant gender................................................................. 70
3.1 Background ............................................................................................... 70
3.2 Methods .................................................................................................... 71
3.2.1 Milk collection and processing ............................................................. 71
3.2.2 DNA isolation ...................................................................................... 72
5.3 Results........................................................................................................... 128

5.3.1 Microbiota analysis...................................................................................... 128

5.3.2 Assessment of DNA damage ability of breast tissue isolates .......... 142

5.4 Discussion........................................................................................................ 148

5.5 Conclusion........................................................................................................ 152

5.6 References........................................................................................................ 153

Chapter 6.............................................................................................................. 158

6 Draft genome sequence of a Bacillus cereus strain isolated from non-malignant
breast tissue from a woman with breast cancer................................................. 158

6.1 Genome announcement.................................................................................... 158

6.2 References........................................................................................................ 163

Chapter 7.............................................................................................................. 164

7 Discussion............................................................................................................ 164

7.1 Verification of a breast tissue microbiome...................................................... 164

7.1.1 Why do these bacteria not induce infections or inflammation .......... 165

7.2 Factors contributing to the microbial differences between breast cancer
patients and healthy controls.............................................................................. 169

7.3 The influence of the breast tissue microbiome on the response to
chemotherapy........................................................................................................ 172

7.3.1 The ability of bacteria to bio-transform drugs ........................................ 172

7.3.2 The ability of bacteria to modulate the immune system which
indirectly effects drug therapy ........................................................................... 173

7.4 Future directions ............................................................................................. 175

7.5 Concluding remarks......................................................................................... 176

7.6 References........................................................................................................ 178

8 Appendix............................................................................................................ 182

8.1 Optimization of DNA extraction methods for breast milk and breast tissue..... 182
List of Tables

Table 1-1. Bacteria identified in human milk by both culture dependent and independent techniques. ........................................................................................................................................ 12

Table 2-1. Comparison of relative abundances of different genera detected in milk between the chemotherapy and non-treatment groups.......................................................... 56

Table 2-2. Metabolites significantly altered by chemotherapy........................................ 58

Table 5-1. Summary of p-values generated by MIRKAT ................................................. 134

Table 5-2. Comparison of relative abundances of bacteria in breast tissue between women with benign tumors, cancerous tumors and healthy controls. .................. 138

Table 6-1. JSpecies analysis comparing B. cereus CU-1 to other B. cereus isolates with fully sequenced genomes. ........................................................................................................ 160

Table 6-2. Potential proteins involved in cancer development or protection. ............ 162

Table 7-1. Effects of bacteria on chemotherapeutic drug cytotoxicity in vitro .......... 173

Table 8-1. Primers used in the studies discussed in the appendix ................................. 188

Table 8-2. Recipe for making denaturing polyacrylamide gel solutions .................... 188

Table 8-3. Recipe for AEC stain for immunohistochemistry ........................................ 194
List of Figures

Figure 1-1. Stages of mammary gland development .......................................................... 4

Figure 1-2. Possible mechanism of bacterial translocation from the gut to the breast ..... 16

Figure 2-1. Summary of sample collection ......................................................................... 44

Figure 2-2. Changes in bacterial diversity as a result of chemotherapy ......................... 52

Figure 2-3. 16S rRNA sequencing analysis of bacteria in human milk ......................... 53

Figure 2-4. Comparison of relative proportions of bacterial taxa between groups ....... 55

Figure 2-5. Principal component analysis of metabolites in breast milk at week 0 and 
during chemotherapy ....................................................................................................... 59

Figure 3-1. Principal components analysis biplot comparing milk samples with controls. 
............................................................................................................................................ 74

Figure 3-2. Breast milk microbiota in 39 Canadian women identified by 16S rRNA 
amplicon sequencing ........................................................................................................ 79

Figure 3-3. Five most relatively abundant genera in human milk ............................... 80

Figure 3-4. Percent abundance of bacterial phyla in breast milk identified by 16S rRNA 
amplicon sequencing ........................................................................................................ 81

Figure 3-5. Principal coordinate (PCoA) plots comparing bacterial profiles based on 
gestation, mode of delivery and gender ........................................................................... 83
Figure 3-6. Principal coordinate analysis based on generalized UniFrac distances. ....... 85

Figure 4-1. Location within the breast of tissue samples collected for analysis........... 102

Figure 4-2. Breast tissue microbiota in 43 Canadian women identified by 16S rRNA sequencing.................................................................................................................. 104

Figure 4-3. Breast tissue microbiota in 38 Irish women identified by 16S rRNA sequencing.................................................................................................................. 106

Figure 4-4. Bacterial diversity of the breast tissue microbiota .................................. 108

Figure 4-5. Percent abundance of different bacterial phyla in breast tissue identified by 16S rRNA sequencing.................................................................................................................. 110

Figure 4-6. UPGMA hierarchical clustering comparing Canadian tissue samples with their respective environmental controls .............................................................. 112

Figure 5-1. Assessment of DNA contamination towards the breast tissue microbiota .. 130

Figure 5-2. Examining the relatedness of bacterial profiles in tissue samples to that of the environmental controls.................................................................................................................. 131

Figure 5-3. Breast tissue microbiota identified by 16S rRNA gene amplicon sequencing. ............................................................................................................................................... 132

Figure 5-4. Comparison of bacterial profiles between breast cancer patients and healthy controls........................................................................................................................................... 135
Figure 5-5. Differences in relative abundances of taxa exist between healthy and cancer patients .......................................................... 137

Figure 5-6. Comparison of bacterial profiles between tumour tissue and normal adjacent tissue .......................................................... 140

Figure 5-7. DNA damage ability of E.coli breast tissue isolates from cancer patients. 144

Figure 5-8. DNA damage ability of bacteria isolated from breast tissue of cancer patients .......................................................... 147

Figure 6-1. Schematic of protein functional groups .......................................................... 159

Figure 8-1. Schematic of DNA extraction protocol for bacterial detection in breast milk .......................................................... 184

Figure 8-2. Schematic of DNA extraction for bacterial detection in breast tissue ....... 185

Figure 8-3. Denaturing gradient gel electrophoresis analysis on breast tissue .......... 189

Figure 8-4. Methylene blue stain of breast tissue cryosections ................................. 192

Figure 8-5. Immuno-staining of breast tissue cryosections ........................................... 194

Figure 8-6. qPCR results showing the relative fold change of the genes involved in the breakdown of lactosylceramide to ceramide ......................................................... 202
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDEEx</td>
<td>ANOVA-like differential expression</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>BLAST</td>
<td>basic local alignment tool</td>
</tr>
<tr>
<td>BLN</td>
<td>bronchial lymph nodes</td>
</tr>
<tr>
<td>BRCA</td>
<td>breast cancer susceptibility gene</td>
</tr>
<tr>
<td>°C</td>
<td>degree Celsius</td>
</tr>
<tr>
<td>C section</td>
<td>cesarean section</td>
</tr>
<tr>
<td>CBA</td>
<td>columbia blood agar</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CFU</td>
<td>colony forming units</td>
</tr>
<tr>
<td>clr</td>
<td>centered log-ratio</td>
</tr>
<tr>
<td>CRC</td>
<td>colorectal cancer</td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>double deionized water</td>
</tr>
<tr>
<td>DGGE</td>
<td>denaturing gradient gel electrophoresis</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleoside triphosphate</td>
</tr>
<tr>
<td>DSB</td>
<td>double stranded breaks</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>GC-MS</td>
<td>gas chromatography-mass spectrometry</td>
</tr>
<tr>
<td>GIT</td>
<td>gastrointestinal tract</td>
</tr>
<tr>
<td>HMO</td>
<td>human milk oligosaccharides</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>HMP</td>
<td>human microbiome project</td>
</tr>
<tr>
<td>HTS</td>
<td>high throughput sequencing</td>
</tr>
<tr>
<td>IBD</td>
<td>inflammatory bowel disease</td>
</tr>
<tr>
<td>Ig</td>
<td>immunoglobulin</td>
</tr>
<tr>
<td>IL</td>
<td>interleukin</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>Mb</td>
<td>megabases</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>ml</td>
<td>milliliter</td>
</tr>
<tr>
<td>MLN</td>
<td>mesenteric lymph nodes</td>
</tr>
<tr>
<td>MOI</td>
<td>multiplicity of infection</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>MRS</td>
<td>de Man, Rogosa and Sharpe</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NK</td>
<td>natural killer</td>
</tr>
<tr>
<td>OTU</td>
<td>operational taxonomic unit</td>
</tr>
<tr>
<td>PBMC</td>
<td>peripheral blood mononuclear cell</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCA</td>
<td>principal components analysis</td>
</tr>
<tr>
<td>PCoA</td>
<td>principal coordinates analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>RAST</td>
<td>rapid annotation using subsystem technology</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>RDP</td>
<td>ribosomal database project</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>rotations per minute</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>s.c.</td>
<td>sub-cutaneous</td>
</tr>
<tr>
<td>SCFA</td>
<td>short chain fatty acid</td>
</tr>
<tr>
<td>TEB</td>
<td>terminal end bud</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>UPGMA</td>
<td>unweighted pair group method with arithmetic mean</td>
</tr>
<tr>
<td>UV</td>
<td>ultra violet</td>
</tr>
<tr>
<td>VCAM-1</td>
<td>vascular cell adhesion protein 1</td>
</tr>
<tr>
<td>γ</td>
<td>gamma</td>
</tr>
<tr>
<td>µ</td>
<td>micro</td>
</tr>
</tbody>
</table>
Chapter 1

1 General introduction

1.1 Functional anatomy of the female breast

The mammary gland, as the “mammary” part implies, is an organ unique to mammals, with the main purpose of delivering nutrients and protective immunity to the baby in the form of milk. But before milk can be produced and secreted, a complex and coordinated process of gland morphogenesis occurs during three stages of human development; embryogenesis, puberty and reproduction.

1.1.1 Embryogenesis

Human mammary gland morphogenesis starts with the formation of the milk line (lateral thickening of the ectoderm) which then gives rise to a single pair of placodes (Robinson, 2007). These placodes later transform into bulbs of epithelial cells, which invaginate into the underlying mesenchyme becoming “mammary buds” (Gjorevski and Nelson, 2011). During the later stages of embryogenesis, the mammary buds start to proliferate and elongate, forming sprouts that push down and out of the mesenchyme and into the stromal fat pad, located within the dermis (Hens and Wysolmerski, 2005). These sprouts concomitantly develop a lumen with an opening to the skin via the nipple (Macias and Hinck, 2012). Once the sprouts have reached the fat pad, ductal branching occurs, regulated by the stromal environment (Pavlovich et al., 2010; Tan et al., 2014; Wiesen et al., 1999). This branched network gives rise to what is known as the “rudimentary ductal tree” consisting of a primary duct and 15-20 secondary branches (Hens and Wysolmerski,
This structure is what is present at birth and will remain as such until puberty (Figure 1-1, top).

### 1.1.2 Puberty

The pubertal stage is characterized by ductal elongation accompanied by secondary branching. Unlike the embryonic stage, which is independent of hormone regulation, this stage is initiated by elevated levels of estrogen (Silberstein et al., 1994), working in concert with estrogen receptor (Feng et al., 2007) and growth hormone (Gallego et al., 2001). At the tips of these elongating and branching ducts is a structure termed the terminal end bud (TEB), which is a highly proliferative cap that “drives” the elongating duct into the fat pad (Hinck and Silberstein, 2005). After successive rounds of elongation and branching, when the limits of the fat pad have been reached, growth ceases and TEBs disappear (Humphreys et al., 1996) (Figure 1-1, bottom)

### 1.1.3 Reproduction

While the hallmark of the pubertal stage is elongation and secondary branching, the hallmark of the reproductive stage (pregnancy/lactation) is formation of extensive tertiary branches (i.e. side branches) and the differentiation of TEBs into alveoli which produce milk during lactogenesis (Robinson et al., 1995) (Figure 1-1, bottom). This process is dependent on progesterone, progesterone receptor and prolactin (Atwood et al., 2000; Gallego et al., 2001; Lydon et al., 1995). During lactogenesis, the breast is composed of mostly glandular tissue with very few adipocytes (Ramsay et al., 2005). While the mammary gland has the ability to synthesize milk throughout pregnancy, this does not occur due to the high levels of progesterone secreted by the placenta, which acts as a negative regulator of milk production (Djiane and Durand, 1977; López-Fontana et al., 2005).
2012). Once progesterone levels have decreased at parturition, copious amounts of milk are produced at approximately 2 days post partum.

Cessation of lactation is followed by a phase known as involution, characterized by apoptosis of milk producing epithelial cells, clearance of these apoptotic cells and milk components and repopulation of the stroma with adipocytes (Alexander et al., 2001; Clarkson et al., 2004; Lund et al., 1996; O’Brien et al., 2012; Walker et al., 1989). It is estimated that about 80% of milk producing cells will undergo regulated apoptosis in only a few days (Watson and Khaled, 2008). To prevent inflammation associated with cellular debris, involution must be efficient and rapid. While clearance of apoptotic bodies is usually associated with professional phagocytes such as macrophages, it has been shown that mammary epithelial cells, rather than macrophages, are involved in removing cellular debris and milk fat globules during post-lactational involution (Monks et al., 2005, 2008).

Although the mammary gland reverts back to its pre-pregnancy state after involution, the ductal network retains a more extensive framework of side branching than nulliparous glands (Fendrick et al., 1998).
Figure 1-1. Stages of mammary gland development

Top panel, embryonic; Bottom panel, post embryonic
1.1.4 Structure of ducts and alveoli

The ducts and alveoli are composed of a bi-layered epithelium consisting of (i) an inner layer of luminal epithelial cells that surround a hollow lumen, which differentiate into milk producing alveoli during pregnancy/lactation and (ii) an outer layer of myoepithelial cells, that upon oxytocin-mediated contraction triggers the release of milk through the ducts (Gjorevski and Nelson, 2011). In 2006, Shackleton and colleagues isolated mammary stem cells from adult mice and demonstrated that a single stem cell could give rise to these luminal and myoepithelial cells and could generate a complete functional gland when transplanted into gland-free mammary fat pads of C3H mice (Shackleton et al., 2006).

This bi-layered epithelium is surrounded by an extracellular basement membrane embedded within the stroma (a.k.a. fat pad) which consists of adipocytes, fibroblasts, blood vessels and immune cells which all play a crucial role in optimal mammary development (Gjorevski and Nelson, 2011). For example, macrophages, eosinophils and mast cells are recruited from the connective tissue and lymph nodes in the stroma to the TEB in the pubertal mammary gland to promote ductal elongation and branching (Gouon-Evans et al., 2000; Lilla and Werb, 2010). These leukocytes are essential, as in their absence, ductal growth is abrogated and the glands retain a neonatal appearance (Gouon-Evans et al., 2000).

1.2 Human microbiome

The term “human microbiome”, was coined in 2001 by Joshua Lederberg who defined it as the “ecological community of commensal, symbiotic and pathogenic microorganisms
that literally share our body space” (Lederberg, 2001). These microorganisms consist of viruses, fungi, archaea and eukaryotes, with bacteria making up the majority of organisms present (99%) (Qin et al., 2010). While the terms “human microbiome” and “human microbiota” are often used interchangeably, the latter refers to the microbial taxa associated with humans, while the former refers to the collection of microbial taxa and their genes (Ursell et al., 2012).

The human microbiota consists of trillions of microbial cells. The gastrointestinal tract (GIT) alone harbours 100 trillion microbes, 1000 different species, 7000 bacterial strains and 3.3 million non-redundant microbial genes (Eckburg et al., 2005; Gill et al., 2006; Qin et al., 2010). On the other hand, some sites have more simple bacterial communities, like the healthy vagina, which is mainly composed of *Lactobacillus* (Ravel et al., 2011). These differences in community structure are likely a result of evolutionary pressures that select for specific organisms able to adapt to a particular niche. This highlights the different functional capacities of these microbes across various body sites.

### 1.2.1 Influence of the microbiota on health and disease

The bacterial communities that colonize various parts of our body are important in promoting health whether it be by synthesizing vitamins the host cannot make, salvaging energy from indigestible compounds, creating a competitive environment to prevent pathogen colonization, promoting maturation and regulation of the immune system, contributing to vascular development and angiogenesis or enhancing the integrity of the epithelial barrier (Hyun et al., 2015; Mazmanian et al., 2005; Ragupathy et al., 2014; Rakoff-Nahoum et al., 2004; Stappenbeck et al., 2002).
The human microbiota in healthy individuals contains a collection of (i) symbionts, which are bacteria that actively promote health, (ii) commensals, which are neither advantageous nor detrimental to the host and (iii) pathobionts, which can promote pathology when conditions are altered in the host (Round and Mazmanian, 2009). When there is microbial disruption, either by reduction of symbionts and/or an overgrowth of pathobionts, a breakdown in homeostasis will occur, leading to disease. This has been observed in periodontitis (Darveau, 2010; Ximénez-Fyvie et al., 2000), inflammatory bowel disease (Frank et al., 2007), psoriasis (Gao et al., 2008), asthma (Hilty et al., 2010), bacterial vaginosis (Hummelen et al., 2010) and colorectal cancer (Mira-Pascual et al., 2014). These changes are not simply a consequence of the diseased state creating an environment that selects for certain bacteria, as studies have shown that healthy animals transplanted with feces from those with obesity, colitis and colorectal cancer then go on to develop disease (Garrett et al., 2007; Turnbaugh et al., 2009; Zackular et al., 2013). These shifts in bacterial profiles not only have consequences at the site of origin but can have distal site effects as well. For example, alterations in the gut microbiota can have effects on the brain (Bercik et al., 2011; Cryan and Dinan, 2012), liver (Boursier and Diehl, 2015; Yoshimoto et al., 2013) and pancreas (Larsen et al., 2010), while microbial shifts in the oral cavity can be detrimental to cardiovascular health (Ettinger et al., 2014).

The human microbiota undergoes dramatic changes during the first year of life at which time it becomes more stable and reflects what the community structure of the adult will be. Thus, the bacteria acquired during infancy, influences the adult microbiota, which in turn influences disease risk in adulthood (Guaraldi and Salvatori, 2012). For this reason,
early life factors that shape the human microbiota play a critical role in health and disease.

1.2.2 Early life factors that shape our microbiota

1.2.2.1 Mode of delivery

As early as 5 min after birth, newborns become colonized with a plethora of bacteria. Unlike adults, whose bacterial composition is structured by body habitat, the microbiota of the newborn is believed to be somewhat uniform across body sites (Dominguez-Bello et al., 2010) until about 3 months of age (Capone et al., 2011). While traditional dogma is that we are all born germfree, the discovery of bacteria in the healthy placenta (Aagaard et al., 2014), amniotic fluid (DiGiulio et al., 2008), umbilical cord blood (Jiménez et al., 2005) and meconium (Gosalbes et al., 2013) suggests that microbial contact could occur in utero. However, a more complex community is acquired during the birthing process and is largely influenced by mode of delivery. Vaginally delivered babies have microbial profiles similar to that of their mother’s vaginal and fecal microbiota, while those born by cesarean section have profiles similar to that of their mother’s skin (Bäckhed et al., 2015; Dominguez-Bello et al., 2010). Infants born by C-section are more often colonized with *Clostridium*, have lower levels of *Bacteroides* and acquire *Bifidobacterium* and *Lactobacillus* later than vaginally delivered infants (Biasucci et al., 2008; Grönlund et al., 1999; Penders et al., 2006). These microbial differences and the consequential immune responses associated with them can last anywhere from 6 months to 7 years of age (Grönlund et al., 1999; Jakobsson et al., 2013; Salminen et al., 2004).
1.2.2.2 Breast feeding

In addition to bacteria acquired during the birthing process, the ingestion of fresh human milk is a major source of microbes for the newborn (Martín et al., 2003, 2012). A characteristic of breast fed infants is a gut microbiota dominated by *Bifidobacterium*. Formula fed infants on the other hand harbour more diverse taxa dominated by *Enterobacteriaceae, Streptococcus, Bacteroides* and *Clostridium* (Bezirtzoglou et al., 2011; Favier et al., 2002). While the high abundance of *Bifidobacterium* in the gut of breast fed infants is partly due to its ability to utilize human milk oligosaccharides (HMOs) as an energy source (Sela et al., 2008), *Bifidobacterium* originating from milk also contribute to its prevalence, as low levels in milk correlate to low levels in the neonate (Grönlund et al., 2007).

1.3 Human milk microbiota

Human milk is generally the first food given to newborns and while the average amount consumed varies with age, one study has shown that 15 week old infants ingest approximately 788 ml of milk per day (Kent et al., 2006). The composition of human milk consists of various fats, proteins, minerals, vitamins, sugars, bioactive compounds, antibodies, immune cells, chemokines/cytokines, anti-microbials and growth factors which all contribute to optimal nutrition, protection and development of the infant (Ballard and Morrow, 2013). Bacteria also contribute in a significant way to the overall make-up of human milk, with the microbiological content routinely recorded as high as $1 \times 10^5$ colony-forming units (CFU) per milliliter (Mense et al., 2014). The earliest studies describing bacterial diversity in human milk utilized simple culture methods and reported high levels of staphylococci and streptococci, originally believed to be contaminants.
Since then, numerous studies using both culture and non-culture methods have detected a plethora of bacterial species in human milk, which are no longer considered contaminants, but normal constituents in the overall composition.

The first non-culture analysis of human milk started in the mid-2000s with the use of denaturing gradient gel electrophoresis and 16S rRNA gene clone library analysis, followed by the first deep sequencing analysis in 2011. These studies confirmed what was observed by culture but also showed a greater diversity of bacteria in milk (Hunt et al., 2011; Martín et al., 2007). While the milk microbiota varies greatly among women, common constituents appear to be *Staphylococcus* and *Streptococcus* (in high abundance), *Corynebacterium* and *Propionibacterium* and at lower levels, *Lactobacillus* and *Bifidobacterium*. A list of all the bacteria identified in human milk to date, is presented in Table 1-1.

### 1.3.1 Importance of milk bacteria in neonatal and long term health

It is widely accepted that breast milk is far superior than formula for optimal physical, psychological and immunological development of the newborn. The World Health Organization recommends exclusive breastfeeding for 6 months, followed by another 2 years of partial breastfeeding, supplemented with solid foods. Infants fed human milk compared to formula have lower incidences of necrotizing enterocolitis and diarrhea in infancy and allergies, asthma, inflammatory bowel disease, type 1 and type II diabetes, obesity and cardiovascular disease in adulthood (Guaraldi and Salvatori, 2012; Stuebe, 2009) as well as better cognitive function throughout life (Anderson et al., 1999). This
protective effect may be due, in part, to the types of bacteria present in milk, as infants fed formula supplemented with probiotics were better protected against some of the above diseases than those just fed formula (Chouraqui et al., 2008; Guaraldi and Salvatori, 2012; Silvers et al., 2012).
Table 1-1. Bacteria identified in human milk by both culture dependent and independent techniques.

<table>
<thead>
<tr>
<th>Genera</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acinetobacter</td>
<td><em>A. calcoaceticus</em></td>
</tr>
<tr>
<td>Akkermansia</td>
<td><em>A. muciniphilia</em></td>
</tr>
<tr>
<td>Bifidobacterium</td>
<td><em>B. breve, B. adolescentis, B. bifidum, B. longum, B. animalis, B. catenulatum</em></td>
</tr>
<tr>
<td>Enterococcus</td>
<td><em>E. faecalis, E. faecium, E. durans, E. gallinarum, E. mundtii, E. hirae</em></td>
</tr>
<tr>
<td>Escherichia</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>Gemella</td>
<td><em>G. haemolysans</em></td>
</tr>
<tr>
<td>Klebsiella</td>
<td><em>K. oxytoca</em></td>
</tr>
<tr>
<td>Kocuria</td>
<td><em>K. kristinae</em></td>
</tr>
<tr>
<td>Lactobacillus</td>
<td><em>L. animalis, L. brevis, L. crispatus, L. fermentum, L. gasseri, L. oris, L. plantarum, L. rhamnosus</em></td>
</tr>
<tr>
<td>Lactococcus</td>
<td><em>L. lactis</em></td>
</tr>
<tr>
<td>Leuconostoc</td>
<td><em>L. citreum, L. fallax, L. mesenteroides</em></td>
</tr>
<tr>
<td>Pediococcus</td>
<td><em>P. pentosaceus</em></td>
</tr>
<tr>
<td>Propionibacterium</td>
<td><em>P. acnes, P. granulosum</em></td>
</tr>
<tr>
<td>Pseudomonas</td>
<td><em>P. synxantha, P. fluorescens</em></td>
</tr>
<tr>
<td>Rothia</td>
<td><em>R. mucilaginosa</em></td>
</tr>
<tr>
<td>Serratia</td>
<td><em>S. proteomaculans</em></td>
</tr>
<tr>
<td>Staphylococcus</td>
<td><em>S. aureus, S. epidermidis, S. haemolyticus, S. hominis, S. pasteuri, S. warneri</em></td>
</tr>
<tr>
<td>Streptococcus</td>
<td><em>S. salivarius, S. mitis, S. galloyticus, S. australis, S. vestibularis, S. parasanguis, S. pneumoniae</em></td>
</tr>
<tr>
<td>Weisella</td>
<td><em>W. cibaria, W. confusa</em></td>
</tr>
</tbody>
</table>

**Other genera detected but not assigned to species**

*Actinomyces, Bacteroides, Blautia, Bradyrhizobium, Brevundimonas, Burkholderia, Clostridium, Corynebacterium, Flavobacterium, Granulicatella, Prevotella, Rlastonia, Sphingomonas, Stenotrophomonas, Veillonella*

Data taken from: (Albesharat et al., 2011; Cabrera-Rubio et al., 2012; Collado et al., 2009; Grönlund et al., 2007; Heikkila and Saris, 2003; Hunt et al., 2011; Jost et al., 2013; Khodayar-Pardo et al., 2014; Martín et al., 2003, 2007; Solis et al., 2010)
1.3.2 Where do breast milk bacteria come from?

Until recently, it was generally believed that bacteria in human milk originated from the skin surrounding the nipple or from the oral cavity of the infant as a result of retrograde backflow during suckling (Ramsay et al., 2004). However, the presence of obligate anaerobes and species associated with the gut, respiratory tract and vagina would suggest additional mechanisms.

This begs the question then, as to what other mechanisms could be responsible for bacterial homing to the breast? Mucosal surfaces are defined as areas of the body exposed to the external environment with a single layer of epithelial cells protecting the lumen from the rest of the body. The mammary glands, along with the GIT, respiratory tract, urogenital tract and glandular tissue (salivary, lacrimal) are considered mucosal surfaces with their own specialized immune system that maintains homeostasis (McGhee and Fujihashi, 2012). It has long been recognized that there is a common mucosal system that allows for trafficking and homing of immune cells between mucosal sites. This has been established through studies showing that (i) immunization at one mucosal site confers protection at another mucosal site (Weisz-Carrington et al., 1979; Wilson and Obradovic, 2014) and that (ii) adoptive transfer of lymphoblasts from the mesenteric lymph nodes (MLN) or bronchial lymph nodes (BLN) into syngeneic recipients home to other mucosal surfaces such as the breast, bronchus, cervix and gut (McDermott and Bienenstock, 1979). It was also shown in the 1970s that cells isolated from the MLN would preferentially home to the breast during lactation (Roux et al., 1977) which is probably due the increased expression of CCL28 during this time (Meurens et al., 2006; Wilson and Butcher, 2004). CCL28 is a chemokine expressed on mammary epithelial cells,
responsible for the recruitment of IgA plasmoblasts and T cells to the mammary glands from the gut or the airways (Lazarus et al., 2003; Wang et al., 2000). This migration and homing of cells from the GIT and the respiratory tract to the breast has been termed the entero-mammary and bronchial-mammary pathway respectively.

While these entero-mammary and bronchial-mammary pathways are important in delivering antigen specific immune cells to the immunologically immature infant in order to provide passive immunity against gastrointestinal and respiratory infections, it could also be a mode by which bacteria migrate to the mammary glands. Indeed, during pregnancy and lactation, increased numbers of bacteria have been detected in the mammary glands compared to virgin mice with these bacteria co-localizing with dendritic cells (Donnet-Hughes et al., 2010). This transfer of bacteria from the GIT to the breast would explain the commonality of strains between a mother’s faeces and her breast milk (Albesharat et al., 2011; Jost et al., 2014) and the detection of probiotic strains in human milk after oral ingestion (Arroyo et al., 2010; Jiménez et al., 2008) (Figure 1-2).

1.3.2.1 Does trafficking occur in the “resting” breast?

Trafficking of bacteria via dendritic cells to the mammary glands has only been proposed in the context of pregnancy and lactation but the fact that immune cells still migrate to the mammary glands in non-gravid and non-lactating women suggests that bacterial translocation to the breast could be happening continuously throughout a woman’s life.

VCAM-1, expressed on mammary epithelial cells, is responsible for B cell migration to the breast during pregnancy (Low et al., 2010). However it is also expressed at high levels irrespective of pregnancy and lactation and it is the main mechanism by which
mouse mammary tumour virus infected lymphocytes reach the non-lactating/non-gravid gland from the gastrointestinal tract (Finke and Acha-Orbea, 2001). High proportions of intralobular B cells and luminal deposits of IgA have also been observed in breast tissue of non-gravid/non-lactating women (Brandtzaeg, 1983; Drife et al., 1976). Most importantly, dendritic cells, which have been the proposed vehicle for bacterial transfer from the GIT to the mammary gland, are present in high numbers in breast lobules of non-lactating/non-pregnant women (Degnim et al., 2014).
Figure 1-2. Possible mechanism of bacterial translocation from the gut to the breast

The lumen of the gut is inhabited by trillions of bacteria and is separated from the interior by a thin layer of cells called the epithelium. DCs can sample bacteria directly from the lumen by the extension of dendrites or can capture bacteria that have translocated across specialized epithelial cells called ‘M cells.’ DCs laden with bacteria can either prime T cells in the Peyer’s patches or migrate to the MLN to prime T-cell reactions there. Unlike primed T cells, DCs do not normally move past the MLN. However, during pregnancy and lactation, it is believed that bacteria-laden DCs migrate out of the MLN and into the mammary glands. These glands are composed of approximately 15–20 lobes that circle around the nipple, albeit not as definitively visible during surgery as depicted in this diagram. Inside these lobes are lobules and at the end of each lobule are tiny alveoli that produce milk. The milk is carried via the ducts into the nipple. DC: Dendritic cell; MLN: Mesenteric lymph node. Figure taken from (Urbaniak et al., 2012)
1.4 Human microbiome and breast cancer

While the breast (via milk) is so vital to the health of the offspring, ironically, it is the cause of one of the deadliest cancers in women. In Canada alone, it is estimated that by the end of this year, 25,400 women will be diagnosed with breast cancer with 5,000 of them dying from the disease (Canadian Cancer Society, 2016). To put into perspective, this equates to 68 women being diagnosed and 14 women dying each day. Based on current trends, it is predicted that these numbers will increase rather than decrease over the next 20 years.

The etiology of breast cancer is still unknown but thought to be due to a combination of both genetic and environmental factors. Support for environmental factors comes from migration studies showing an increased incidence of breast cancer amongst 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). While many environmental predictors have been proposed, such as smoking and diet, one factor not considered at the time of entrance into my PhD program, was the potential of microbes, specifically in the breast, to influence the risk of breast cancer development. The rationale for this idea is presented below.

1.4.1 Causative role of bacteria in breast cancer development

Almost 40 years ago, it was reported that the spontaneous rate of various tumour formations, over a 10-year period, was higher in conventional rats compared to germ-free ones, with mammary tumours being the second most common tumour recorded (Sacksteder, 1976). The role of bacteria in mammary carcinoma was again delineated a
few years later, when it was shown that conventional rats had an increased risk of mammary tumours compared to germ free ones after sub-cutaneous injection with the carcinogen, 3,2'-dimethyl-4-aminobiphenyl (DMAB) (Reddy and Watanabe, 1978).

One mechanism by which bacteria may contribute to disease onset is through production of carcinogenic compounds from dietary components. Numerous studies using germ-free and conventional animals have shown that risk-related components of the human diet interact with the gut microbiota to induce DNA adducts and colon cancer (Hambly et al., 1997; Reddy et al., 1975; Rumney et al., 1993). One of these risk-related dietary components, PhIP (2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine), has been detected in the milk of lactating rats and human milk (DeBruin and Josephy, 2002; Scott et al., 2007).

Another mechanism by which bacteria may promote breast cancer is through chronic inflammation, which is now widely accepted to be associated with the development of cancer (Mantovani et al., 2008). The effect of pathogen-induced inflammation is however not limited to the site of infection. It has been shown that C57BL/6 ApcMin/+ mice, which are genetically predisposed to develop mammary carcinomas, fail to develop disease when housed under specific pathogen free conditions (Moser et al., 1993) but upon gastric administration of *Helicobacter hepaticus* they develop mammary adenosquamous carcinoma as a result of innate immune induction of inflammation (Rao et al., 2006a, 2006b).
1.4.2 Protective role of bacteria in breast cancer development

Certain bacteria have properties that may reduce the risk of breast cancer development. An epidemiological study conducted in The Netherlands involved 323 women showed that the incidence of breast cancer decreased in women who drank fermented milk products, irrespective of multivariable risk factors such as smoking, obesity, age of first pregnancy, parity and alcohol consumption (van’t Veer et al., 1989). *In vivo* and *in vitro* studies further support the protective role of lactic acid bacteria in breast cancer development. An *in vitro* study showed that milk fermented with separate strains of lactic acid bacteria (*Bifidobacterium infantis*, *B. bifidum*, *B.animalis*, *Lactobacillus acidophilus*, *L.paracasei*) had an anti-proliferative effect on the MCF7 breast cancer cell line (Biffi et al., 1997). *In vivo*, it has been shown that mice fed milk fermented with *Lactobacillus helveticus* R389, display delayed or absent tumour growth compared to those that did not receive the probiotic (de Moreno de LeBlanc et al., 2005). These protective effects were proposed to be due to immune modulation by *Lactobacillus*, specifically a reduction in IL-6 and an increase in IL-10 cytokine levels. *Lactobacillus* and *Bifidobacterium* have also been shown to increase the percentage of activated natural killer (NK) cells in peripheral blood mononuclear cells (PBMC) (Chiang et al., 2000; Takeda and Okumura, 2007). This could be beneficial, as women with breast cancer have impaired NK cell function compared to healthy controls (Piroozmand and Hassan, 2010), which is associated with an increased risk of breast cancer development (Imai et al., 2000; Strayer et al., 1984).
1.4.3  Bacteria in the breast and their role in breast cancer development

The studies described above present a rationale for how distally located bacteria may contribute to disease progression or prevention. Below I discuss the possible role of bacteria within the breast in modulating this risk.

In a large scale analysis of 140,000 women in 47 studies from 30 different countries it was shown that for every 12 months of breast-feeding the risk of breast cancer decreased by 4.3% (Collaborative Group on Hormonal Factors in Breast Cancer, 2002). This study did not account for the different molecular subtypes of breast cancer, but another international study, conducted to assess the role of breast feeding in hereditary cancer, showed that women with BRCA1 mutations who breast fed for 1 year or longer had a 45% reduced risk of breast cancer, though these protective effects were not seen in patients with BRCA2 mutations (Jernström et al., 2004; Tryggvadottir et al., 2003).

Interestingly, an *in vivo* model with Caveolin-3⁻/⁻ mice, which have constitutive milk production, showed that mutants were substantially more protected against mammary tumour development than their wild-type counterparts, after tumour injection (Sotgia et al., 2009).

There are many theories as to why lactation could be protective, such as reduction of estrogen levels (Jernström et al., 2004; Lyons et al., 2009), excretion of carcinogens from the breast and differentiation of mammary epithelial cells (Lipworth et al., 2000). Hormone modulation, however, may not be a significant contributor, as breast-feeding is protective in estrogen receptor negative, progesterone receptor negative (ER-PR-) and triple negative tumours (Gaudet et al., 2011; Ursin et al., 2005).
We raise the possibility that bacteria present in the mammary glands during lactation could act independently or in concert with other factors to afford protection. In a study comparing the immunomodulatory potential of two species of *Lactobacillus* isolated from breast milk (*L. fermentum* CECT5716 and *L. salivarius* CECT5713) with two non-milk derived species (*L. fermentum* NCIMB7017511 and *L. salivarius* NCIMB11795) it was shown that the milk-derived strains induced higher levels of IL-10 and activated NK cells from PBMC (Pérez-Cano et al., 2010), two factors that are protective against breast cancer. Possibly, the strains of bacteria growing in breast milk have been selected to provide an evolutionary advantage to women to help control and/or prevent the progression of breast cancer. This will be a difficult hypothesis to prove but one worthy of consideration.

### 1.5 High-throughput sequencing to analyze the human microbiota

The first description of microbes was reported in 1673 by the Dutch scientist Antonie van Leeuwenhoek, with his work considered to be the start of the field of Microbiology. Another pivotal landmark was the development of solid media by Robert Koch in 1881, which allowed him to isolate disease-causing organisms thereby determining the agents responsible for tuberculosis, cholera and anthrax. While culture analysis is routinely used in labs across the world and has been indispensable to microbiologists over the last 130 years, culture methods have limited our understanding of the composition and function of complex microbial communities, as most organisms cannot be grown in the lab (Amann et al., 1995; Stewart, 2012). It is for this reason, that culture-independent methods, most notably high throughput sequencing (HTS), has revolutionized the field of microbial
ecology, allowing us to study the human microbiota and its function at various body sites, in healthy and diseased patients and at different stages of human development (Coburn et al., 2015; Costello et al., 2013; Gill et al., 2006; Grice et al., 2009; Human Microbiome Project Consortium, 2012; Hummelen et al., 2010; Hunt et al., 2011; Koenig et al., 2011; Macklaim et al., 2013; Turbaugh et al., 2009; Wolfe et al., 2012).

1.5.1 16S rRNA gene targeted sequencing

The 16S rRNA gene is part of the small ribosomal subunit of bacteria, an evolutionary conserved, ubiquitous molecule, that makes it useful for phylogenetic studies (Woese and Fox, 1977). While 16S rRNA gene amplicon sequencing is the most popular method to identify and compare bacteria within samples, some groups have opted to utilize other targets, such as cpn60 (Links et al., 2012).

HTS has allowed for the unprecedented ability to process 100-1000s of samples at a time, generating millions of reads at a reasonable cost, however, it is limited by the length of sequence it can produce and the resolution of the phylogenetic assignment. The current version of the Illumina MiSeq platform generates a maximum read length of 600bp, up from 100bp when it was first developed. Because of this, HTS studies do not sequence the full 16S rRNA gene (1.5kb) but specific regions of it, called hypervariable regions. These hypervariable regions (V1-V9), flanked by conserved regions from which amplification primers are designed, are as effective as the full length version at assigning taxonomy and estimating community richness (Andersson et al., 2008; Gloor et al., 2010; Liu et al., 2007).
Currently there is no consensus in the field as to which hypervariable region should be sequenced, because the “best one” will depend on the sequencing platform, databases used for assigning taxonomy and the type of samples used (Soergel et al., 2012). Stool microbiota studies typically use V4 or V3-V5 (which span the V4 region). V3, V4 and V4/V5 have been shown to be more accurate in determining taxonomy compared to V2, V5 alone and V6, regardless of sequencing platform (454 and Illumina MiSeq) or database used to assign taxonomy (i.e. RDP, Greengenes, NCBI) (Claesson et al., 2010; Huse et al., 2008; Liu et al., 2008; Mizrahi-Man et al., 2013). However, in terms of estimating diversity (in the stool), all regions tested performed equally well (Liu et al., 2008). Since V4 can detect a broad range of bacteria and archaea it has been chosen as the target for the Earth Microbiome Project (www.earthmicrobiome.org). V6 is commonly used in vaginal microbiota studies as it can differentiate Lactobacillus species, which dominate the healthy vagina, as well as picking up the various organisms found in women with bacterial vaginoses (Gloor et al., 2010).

V6 was the region of choice for the milk and tissue studies described herein, because unlike primers for the V4 region, it does not cross-react with eukaryotic mitochondrial DNA (personal observations and also noted in (Huys et al., 2008)). In stool samples, where there is a higher proportion of bacterial vs eukaryotic DNA this is not a problem. But in tissue, where eukaryotic DNA predominates, amplification needs to be as specific as possible so as to not saturate the reads with eukaryotic sequences. In addition, V6 but not V4 directed primers, detect Propionibacterium (Kuczynski et al., 2012) a common skin organism that we expected to be present in our samples.
In order to more readily interpret the massive amount of data generated with HTS, 16S rRNA gene sequences that are similar to each other are clustered together into operational taxonomic units (OTUs) which can then be used to estimate microbial diversity (Chen et al., 2013; Schloss and Westcott, 2011). This is based on the assumption that sequences that are similar to each other represent organisms that are more phylogenetically related (He et al., 2015). Sequences are typically clustered into OTUs with 97% similarity (considered the threshold for separating species) (Schloss and Handelsman, 2005), with the most abundant sequence in that group being the representative one used for taxonomy. A table of OTUs and counts for each sample is then used for downstream analysis such as beta diversity (diversity between samples), alpha diversity (diversity within a sample) and comparing relative abundances between groups.

### 1.6 Scope and objectives

The Human Microbiome Project (HMP), funded by the National Institutes of Health, was developed in order to comprehensively characterize the presence and function of bacteria colonizing different body sites, under different conditions. This initiative, and a similar one in Europe, was undertaken because of the overwhelming evidence showing how important microbes can be in maintaining health and how significant aberrations can lead to detrimental outcomes (Peterson et al., 2009). Phase 1 of the project was published in 2012 and reported the diversity of bacterial communities from 242 healthy American adults (males and females) from 15-18 body sites (Human Microbiome Project Consortium, 2012). Unfortunately and surprisingly, this large-scale study did not include analysis of human milk.
While there were milk microbiota studies prior to the HMP and others conducted after its 2012 publication, most of these utilized culture, denaturing gradient gel electrophoresis and/or PCR specific amplification. At the start of this thesis there had only been 1 human milk microbiota study utilizing HTS to characterize its composition (Hunt et al., 2011) and none to determine what factor(s) could shape this bacterial community.

My first objective was thus to characterize the human milk microbiota in Canadian women and to determine what factors shape its composition. In Chapter 2, I present a case report following a woman over 4 months for Hodgkins lymphoma to determine how drugs given distally could affect the milk microbiota - in this case we looked at the effects of the ABVD chemotherapeutic regime (adriamycin, bleomycin, vinblastine, dacarbazine). In Chapter 3, I look at the effects of common pregnancy related factors and its impact on the milk microbiota - specifically gestation (pre-term vs term birth), mode of delivery (cesarean vs vaginal) and gender of the child (male vs female vs twins).

Considering that milk is not sterile, that there is exposure of the nipple to the external environment and that the nutrient rich fatty environment of the breast could support bacterial growth, I hypothesized that breast tissue, irrespective of lactation and pregnancy, could also have its own indigenous microbiota. My second objective was thus to characterize, using HTS and culture, the bacteria in breast tissue. In Chapter 4, I present our findings from tissue samples analyzed from women from two distant demographics: Canada (n=43) and Ireland (n=39).

After showing that a breast microbiome exists, my third objective was to examine whether this microbiome could have a role in the development of breast cancer. The first
aim within this objective was to determine whether bacterial profiles differ between women with breast cancer and those who are disease free (Chapter 5). The rationale for this comes from studies showing differences in microbial profiles between healthy individuals and those with colorectal cancer, with these differences driving disease development. The second aim within this objective was to explore how the breast microbiota could promote disease progression. I did this by looking at the ability of bacteria isolated from breast cancer patients to induce DNA damage using the γH2AX assay (Chapter 5) and by sequencing the genome of a Bacillus cereus strain isolated from a women with stage III invasive breast cancer (Chapter 6). Bacillus cereus was chosen for this analysis as Bacillus species were higher in cancer patients compared to controls and they were the most abundant organism cultured from breast cancer patients.

In a series of Appendices, I describe my work showing (i) the optimization of DNA extraction protocols for low biomass samples, (ii) histological work to try and identify where in the breast bacteria were located and lastly (iii) the properties of Lactobacillus rhamnosus GR-1 (a probiotic) when grown in bovine milk, with an emphasis on a pathway involving breakdown of lactosylceramide to ceramide, a pro-apoptotic signaling molecule involved in controlling aberrant cell growth.
1.7 References


Chapter 2

2 Effect of chemotherapy on the microbiota and metabolome of human milk, a case report

This chapter has been reproduced with permission (Appendix 8.6) from:


A summary of clinical data (Dataset 2-1) and metabolites detected in human milk (Dataset 2-2) is available for download online.

2.1 Introduction

Colonization of the neonatal gut plays a pivotal role in gastrointestinal, immunological and neurological development with one of the initial sources of bacteria coming from the mother’s milk (Martín et al., 2003, 2012). Breast fed infants have been shown to have lower incidences of asthma, diarrhea, and necrotizing enterocolitis (NEC) compared to formula fed infants (Stuebe, 2009). This protective effect may be due, in part, to the types of bacteria present in milk, as infants fed formula supplemented with probiotics were better protected against these conditions compared to those just fed formula (Chouraqui et al., 2008; Guaraldi and Salvatori, 2012; Silvers et al., 2012). The bacteria acquired during infancy can influence disease risk later in life and play a major role in what the composition of the adult microbiome will be (Guaraldi and Salvatori, 2012). Thus, factors that affect the milk microbiota have important health consequences for the child not only during development but also into adulthood. In addition to the microbiota, the metabolites of human milk, such as fatty acids, carbohydrates, proteins and vitamins also
play an important role in infant development and long term health (Agostoni, 2008; Boudry et al., 2010; German et al., 2002; Kunz et al., 2000).

Post delivery, many women are prescribed pharmaceutical agents for various reasons. While most over the counter drugs and antibiotics are not contraindicated during breast feeding (Donaldson and Goodchild, 2012; U.S. National Library of Medicine), when it comes to chemotherapeutics, the recommendation is that breastfeeding should be avoided until the drug has been cleared from the milk (Gartner et al., 2005). In a case report of a 70mg infusion of cisplatin, no detectable levels were found in milk after 66hr (Hays et al., 2013) and in another case study using doxorubicin (trade name Adriamycin), no detectable levels were seen after 72hr (Pistilli et al., 2013). In our particular study, the subject was advised that breastfeeding could resume 12 days after each chemotherapy session.

Here we present the first report on the effects of chemotherapy on microbial and metabolomic profiles in human milk over a 4 month period in a breast feeding woman undergoing treatment for Hodgkin’s lymphoma (Figure 2-1).
**Clinical Information**

Diagnosed with Hodgkin’s lymphoma during 2nd trimester of pregnancy → Undergoes 3 rounds of chemotherapy at 29wks gestation (3rd trimester) → Two weeks after the last session she delivers → Four weeks after delivery, chemo starts again. Milk sample collection starts for the study

**Sample Collection**

<table>
<thead>
<tr>
<th>Week</th>
<th>Sample</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2-1. Summary of sample collection**

Milk samples were collected from a lactating woman undergoing chemotherapy for Hodgkin’s lymphoma. Milk samples were collected every two weeks over a 4-month period. At each session milk was collected 15-30min before (“A”) and after (“B”) chemotherapy. Duration of chemotherapy treatment lasted 2hr. No milk was collected at week 8 due to scheduling conflicts.
2.2 Methods

2.2.1 Clinical samples and study design

Ethical approval was obtained from Western Research Ethics Board and Lawson Health Research Institute, London, Ontario, Canada. Subjects provided written consent for sample collection and subsequent analyses.

2.2.2 Milk collection and processing

Mature milk was collected from a lactating woman undergoing the ABVD chemotherapy regime (A=Adriamycin (40mg), B=Bleomycin (16 units), V= Vinblastine (9.6mg), D= Dacarbazine (600mg) for Hodgkin’s lymphoma at the London Health Sciences Center, London, Ontario. Mature milk was also collected from 8 healthy women recruited from London, Ontario and the surrounding area. Wearing sterile gloves the women cleaned their nipple and surrounding area with sterile saline swabs to reduce the presence of skin bacteria. Milk was collected using a sterile HygieniKit Milk Collection System (Ameda) attached to an electric breast pump. Between 5-15mL of milk was pumped into a sterile tube and kept on ice until transfer to the laboratory which occurred within 1 hour of collection. Samples were aliquoted and stored at -20°C until DNA extraction.

2.2.3 DNA isolation

After thawing on ice, 2 mL of milk were spun down at 13,000g for 10min and the supernatant discarded. The pellet was then homogenized in 1.4mL of ASL buffer (QIAamp® DNA Stool Kit, Qiagen) and 400mg of 0.1mm diameter zirconium-glass beads (BioSpec Products). Mechanical and chemical lyses were performed by bead beading at 4800rpm for 60s, then 60s on ice (repeated twice) (mini-beadbeater-1,
BioSpec Products) and then incubated at 95°C for 5min. Subsequent procedures were performed using the Qiagen QIAamp® DNA Stool Kit according to the manufacturer’s protocol, with the exception of the last step in which the column was eluted with 120µL of elution buffer. DNA was stored at -20°C until further use.

2.2.4 Ion Torrent V6 16S rRNA sequencing

2.2.4.1 PCR amplification

Genomic DNA isolated from the clinical samples, was amplified using the barcoded primers, V6-LT:

5’CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNWCACGCGARAGAACCCTACC3’

and V6-RT: 5’CCTCTCTATGGGCAGTCGGTGATACRACACGAGCTGACGAC

which amplify the V6 hypervariable region of the 16S rRNA gene. The PCR was carried out in a 40µL reaction containing 5µL of DNA template (or nuclease free water as a negative control), 1.5mM MgCl₂, 0.8µM of each primer, 4µL of 10x PCR Buffer (Invitrogen), 0.2mM dNTPs, 0.05U Taq Polymerase (Invitrogen) and 0.15µg/µL of bovine serum albumin. Thermal cycling was carried out in an Eppendorf Mastercyler under the following conditions: Initial denaturation at 95°C for 2min followed by 25 cycles of 95°C for 1min, 55°C for 1min and 72°C for 1min. After amplification, the DNA concentration was measured with the Qubit® 2.0 Fluorometer (Invitrogen) using the broad range assay. Equimolar amounts of each PCR product were pooled together and purified using the QIAquick PCR purification kit (Qiagen). The PCR purified sample was then sent to the London Regional Genomics Center, London, Ontario, Canada for V6 16S rRNA sequencing using the Ion Torrent platform as per the Center’s standard operating procedure.
2.2.4.2 Sequence processing and taxonomic assignment

Custom Perl and Bash scripts were used to de-multiplex the reads and assign barcoded reads to individual samples. Reads were kept if the sequence included a perfect match to the barcode and the V6 16S rRNA gene primers. Reads were clustered by 97% identity into Operational Taxonomic Units (OTUs) using UClust v. 3.0.617 (Edgar, 2010). OTUs that represented ≥1% of the reads in at least one sample were kept, while those that did not meet the cut-off were discarded. Taxonomic assignments for each OTU were made by the ribosomal database project (RDP) SeqMatch tool (Cole et al., 2009). From the top 20 matches to the RDP named isolates database, the full taxonomy was retained for matches with the highest S_ab score. For multiple top matches with equal scores, the highest common taxonomy was retained (e.g. genus level if multiple species matched equally well). Since the maximum number of matches displayed per sequence is 20, the RDP taxonomic assignments were verified by BLAST against the Greengenes named isolates database with an output of 100 hits (DeSantis et al., 2006). Taxonomy was assigned based on hits with the highest % identities and coverage. If multiple hits fulfilled this criterion, classification was re-assigned to a higher common taxonomy. In instances where the highest % identity/coverage yielded a single match, if this were less than 90% and the S_ab score from RDP was less than 0.7, taxonomy was assigned at the Family level instead of at the Genus level. The raw sequencing reads generated in this study have been deposited to the NCBI Short Read Archive (SRA) database (SRP041626).
2.2.4.3 Data analysis

Weighted UniFrac distances were calculated in QIIME (Caporaso et al., 2010) by using a phylogenetic tree of OTU sequences built with FastTree (Price et al., 2009) and based on an OTU sequence alignment with MUSCLE (Edgar, 2004). The QIIME pipeline was also used to calculate Shannon’s diversity index (logarithms with base 2) and to generate principal coordinate analysis (PCoA) plots. Weighted UniFrac distances compare microbial profiles (presence/absence and abundance) between samples (i.e. beta-diversity) (Lozupone and Knight, 2005) while Shannon’s diversity index evaluates the microbial diversity within a sample (i.e. alpha diversity). The higher the Shannon’s diversity index, the more diverse a sample is and a value of zero indicates the presence of only one species (Hill et al., 2003). PCoA plots allows one to visualize the UniFrac distance matrix and plots the values on a set of orthogonal axes which capture the greatest amount of variation between all samples tested. For beta-diversity analyses, the data set was rarified to the lowest read count/sample, which was 734 reads. A summary of clinical data, including total number of sequence reads per sample is shown in Dataset 2-1. Barplots, boxplots and stripcharts were all generated in R (https://www.r-project.org).

2.2.4.4 Statistical analysis

The ALDEx R package version 2 (Fernandes et al., 2014) was used to compare genera between the non-treatment and chemotherapy treatment groups (as portrayed in the boxplots). Microbiome data represents proportional distributions and are thus not independent of each other. This means that a decrease in one organism will inevitably lead to a concomitant increase in another organism. For example, if a sample has two
organisms A (50%) and B (50%) and A is completely killed by an antibiotic, the proportion of B in that sample will now be 100% even if its actual abundance has not changed. The ALDEx R package estimates the technical variation inherent in high-throughput sequencing by Monte-Carlo sampling from a Dirichlet distribution (Jaynes, 2003). The Monte-Carlo replicates are transformed using the centered log-ratio transformation that takes the logarithm of the Monte-Carlo estimates of organism abundances in each sample divided by the per-sample geometric mean organism abundance (Aitchison and Egozcue, 2005). This transformation has several desirable properties that do not exist in proportional data, notably, subcomposition coherence and linear sample independence. Data transformed in this way permit the use of standard statistical tests to determine significance. Values reported in this manuscript represent the expected values of 128 Dirichlet Monte-Carlo instances. A value of zero indicates that organism abundance is equal to the geometric mean abundance. Thus organisms more abundant than the mean will have positive values and those less abundant than the mean will have negative values. Base 2 was used for the logarithm, thus differences between values represent fold changes. Statistical significance for these comparisons was determined by a Mann Whitney U test with $p<0.05$ and a false discovery rate of $<0.1$ using the q values output by the fdrtool R package. The unpaired student’s t-test was used to compare Shannon’s diversity index ($p<0.05$).

2.2.5 GC-MS sample preparation

To extract metabolites 100µL of milk were mixed with 400µL pure methanol. Samples were vortexed for 15s and centrifuged for 10 min at 9000g. Supernatants (200µL) were transferred to GC-MS vials and 2.5µL of ribitol solution (2mg/ml) was added to each vial.
as an internal standard. Samples were dried to completeness using a SpeedVac. After drying, 100µL of 2% methoxyamine•HCl in pyridine (MOX) was added to each sample for derivitization and samples were incubated at 50°C for 90 min. N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) (100µL) was then added to each vial and incubated at 50°C for 30 min. Samples were transferred to microinserts before running on GC-MS (Agilent 7890A GC, 5975 inert MSD with triple axis detector, 30m DB5-MS column with 10m duraguard column). Samples were run using 1µL injections on scan mode and a solvent delay of 10 min. Run time was 60 min per sample. Each sample was run twice non-consecutively to ensure consistency throughout the sequence.

2.2.5.1 Data analysis for GC-MS

Chromatogram files were converted to ELU format using the AMDIS Mass Spectrometry software (Stein, 1999). Chromatograms were aligned and abundance of metabolites calculated using the Spectconnect software (Styczynski et al., 2007) with the support threshold set to low. In order to determine if differences between week 0 and chemotherapy (weeks 2-16) existed, principle component analysis (PCA) was conducted in SIMCA (Umetrics) using the relative abundance matrix (RA) output from Spectconnect. Data were mean centered and pareto scaled prior to PCA. Independent unpaired t-tests with Bonferroni correction were calculated in Excel to determine metabolites that were significantly altered by chemotherapy (p < 0.05.). Compounds that also contributed to the separation of week 0 from chemotherapy samples according to the PCA loadings plot (compounds in bottom left quadrant) were chosen for further investigation.
2.3 Results

Chemotherapy affected both bacterial diversity and bacterial profiles in human milk. Bacterial diversity within samples was lower in milk collected throughout chemotherapy compared to milk samples collected at week 0 and from healthy lactating women (Figure 2-2). Bacterial profiles at week 0 were similar to those from healthy women, however this changed within 2 weeks of treatment (Figure 2-3A). Samples collected at weeks 4-16 shared similar profiles and differed from week 2 and from week 0/healthy samples (Figure 2-3A). These differences were not due to natural changes over time, as the bacterial community in two milk samples analyzed from a healthy subject did not change over a four-month period (Figure 2-3A, green samples). The bar plot in Figure 2-3B shows the bacterial communities present in these samples with a striking increase in abundances of *Acinetobacter* and *Xanthomonadaceae* in milk collected during chemotherapy. A comparison of relative abundances of *Acinetobacter*, *Xanthomonadaceae* and *Stenotrophomonas* (a genus part of the *Xanthomonadaceae* family) between the chemotherapy (week 4-week 16) and non-treatment (week 0 and healthy samples) groups are displayed in Figure 2-4 and were shown to be significantly higher during chemotherapy. We also examined the differences between three bacteria believed to confer beneficial health effects to the infant, *Bifidobacterium*, *Eubacterium* and *Lactobacillus*. The first two were significantly decreased during chemotherapy whereas no differences were observed for *Lactobacillus* (Figure 2-4). Overall, a total of 22 out of the 49 genera identified were differentially abundant between the two groups (Table 2-1).
Figure 2-2. Changes in bacterial diversity as a result of chemotherapy

Bacterial diversity within a sample (i.e alpha diversity) was measured by calculating Shannon’s diversity index. Each point on the graph represents a subject with the line representing the mean for all samples within a group. The higher the index the greater the bacterial diversity found within a sample. The mean of the “Wk0/H” group (week0 and healthy samples) was 4.3, and that of the “Chemotherapy” group (week 4-16) was 2.8. Groups were statistically different from each other as measured by unpaired student’s t-test (p <0.05).
Figure 2-3. 16S rRNA sequencing analysis of bacteria in human milk

Milk samples were collected from a lactating woman undergoing chemotherapy as described in Figure 2-1 as well as from 8 healthy lactating women. (A) Weighted UniFrac principal coordinates analysis (PCoA) plot. Each milk sample, represented by a coloured circle, is plotted on this 3D, 3 axis plane representing 84% of the variation observed between all samples. Samples that cluster together are similar in biota composition and abundance. Orange circles represent samples collected from weeks 4-16 of chemotherapy, blue circles represent samples collected at week 2 of chemotherapy, purple circles represent samples collected at week 0, red circles represent milk samples from healthy lactating women (only one time point) and green circles represent milk samples collected from a healthy lactating women collected 4 months apart. As shown by the plot, there were 3 distinct groups: (i) week 0 samples and healthy milk samples; (ii) week 2 of chemotherapy and (iii) weeks 4-16 of chemotherapy. Data were rarified to 735 reads/sample. (B) Barplot showing the relative abundance of different genera in each sample. Each bar represents a subject and each coloured box a different genus. The height of the coloured boxes represents the relative abundance of that genus within the sample. Genera that were less than 2% abundant in a given sample were placed in the “Remaining fraction” at the top of the graph (grey boxes).
**A**

Weeks 4-16

Week 0 & Healthy samples

**B**

Healthy

Wk0

Chemotherapy

Remaining fractions

- Firmicutes, Clostridia, Clostridiales, Clostridiales_IncertaeSedisXI, Finegoldia
- Proteobacteria, Alphaproteobacteria, Rhodobacteria, Rhodobacteraceae, Rhodobacter
- Bacteroidetes, Firmicutes, Bacilli, Staphylococcaceae, Staphylococcus
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Moraxellaceae, Acinetobacter
- Actinobacteria, Actinomycetales, Propionibacteriaceae, Propionibacterium
- Proteobacteria, Alphaproteobacteria, Caulobacterales, Caulobacteraceae, Brevundimonas/Caulobacter
- Deinococcus-Thermus, Deinococci, Deinococcales, Truepera
- Firmicutes, Bacilli, Bacillales, Bacillaceae, Anoxybacillus/Bacillus
- Bacteroidetes, Bacteroidia, Bacteroidales, Prevotellaceae, Prevotella
- Firmicutes, Bacilli, Bacillales, Staphylococcaceae, Staphylococcus
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Stenotrophomonas
- Proteobacteria, Gammaproteobacteria, Xanthomonadales, Xanthomonadaceae, Stenotrophomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Moraxellaceae, Acinetobacter
- Actinobacteria, Actinomycetales, Micrococcaceae, Rothia
- Firmicutes, Bacilli, Bacillales, Bacillaceae, Lactobacillus
- Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, Cloacibacterium
- Firmicutes, Bacilli, Bacillales, Staphylococcaceae, Staphylococcus
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
- Proteobacteria, Gammaproteobacteria, Pseudomonadales, Pseudomonas
Figure 2-4. **Comparison of relative proportions of bacterial taxa between groups**

Boxplots comparing 6 bacterial taxa between samples collected during chemotherapy (week 4-16) and those without treatment (week 0 and healthy samples, “Wk0/H”). The box signifies the 75% (upper) and 25% (lower) quartiles and thus shows where 50% of the samples lie. The black line inside the box represents the median. The whiskers represent the lowest datum still within 1.5 IQR of the lower quartile and the highest datum still within 1.5 IQR of the upper quartile (IQR= interquartile range). Outliers are shown with open circles. The value “0” represents the geometric mean abundance, thus values above 0 are more abundant and values less than 0 are less abundant than the geometric mean. Significant differences were observed between the two groups for all taxa graphed (Mann Whitney U test p<0.05, FDR <0.1).
Table 2-1. **Comparison of relative abundances of different genera detected in milk between the chemotherapy and non-treatment groups**

Values in the second and third columns represent the base 2 logarithm of the median abundance in all samples within a group (that is, Wk0/H samples (non-treatment group) or Chemo (treatment group)) relative to the geometric mean abundance, which has a value of 0. Thus, positive values are higher than the geometric mean and are thus more abundant than negative values, which are lower than the geometric mean. Significant differences were based on FDR values of <0.1. Out of the 49 genera identified, 22 were significantly different between the two groups.
<table>
<thead>
<tr>
<th>Taxa</th>
<th>Wk0/H</th>
<th>Chemo</th>
<th>p-value</th>
<th>FDR value</th>
<th>Higher in</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acinetobacter</em></td>
<td>5.542</td>
<td>10.88</td>
<td>0</td>
<td>0</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Gammaproteobacteria</em></td>
<td>4.372</td>
<td>7.901</td>
<td>0</td>
<td>0.001</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Stenotrophomonas</em></td>
<td>-3.181</td>
<td>5.751</td>
<td>0</td>
<td>0.002</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Xanthomonadaceae</em></td>
<td>0.605</td>
<td>8.547</td>
<td>0.001</td>
<td>0.003</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Cloacibacterium</em></td>
<td>5.301</td>
<td>1.981</td>
<td>0.001</td>
<td>0.003</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Rhizobiales</em></td>
<td>-1.947</td>
<td>2.541</td>
<td>0.001</td>
<td>0.003</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Porphyrobacter</em></td>
<td>2.763</td>
<td>0.259</td>
<td>0.001</td>
<td>0.005</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Thermus</em></td>
<td>-2.361</td>
<td>1.733</td>
<td>0.002</td>
<td>0.006</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Streptococcus</em></td>
<td>2.313</td>
<td>4.696</td>
<td>0.001</td>
<td>0.007</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Flavobacterium/Cytophaga</em></td>
<td>0.169</td>
<td>3.062</td>
<td>0.002</td>
<td>0.007</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Bifidobacterium</em></td>
<td>0.74</td>
<td>-3.51</td>
<td>0.002</td>
<td>0.008</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Eubacterium</em></td>
<td>0.224</td>
<td>-3.607</td>
<td>0.003</td>
<td>0.01</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Rothia</em></td>
<td>-1.221</td>
<td>2.325</td>
<td>0.002</td>
<td>0.011</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Brevundimonas/Caulobacter</em></td>
<td>3.105</td>
<td>-0.117</td>
<td>0.003</td>
<td>0.013</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Microbiaceae</em></td>
<td>-1.449</td>
<td>1.322</td>
<td>0.006</td>
<td>0.022</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Staphylococcus</em></td>
<td>4.445</td>
<td>2.306</td>
<td>0.008</td>
<td>0.028</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Anoxybacillus/Bacillus</em></td>
<td>1.616</td>
<td>3.53</td>
<td>0.007</td>
<td>0.029</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Bacillus</em></td>
<td>0.243</td>
<td>2.623</td>
<td>0.008</td>
<td>0.03</td>
<td>Chemo</td>
</tr>
<tr>
<td><em>Gammaproteobacteria OTU11</em></td>
<td>3.729</td>
<td>0.3</td>
<td>0.017</td>
<td>0.052</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Petrobacter</em></td>
<td>0.184</td>
<td>-2.054</td>
<td>0.025</td>
<td>0.076</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Truepera</em></td>
<td>2.9</td>
<td>-0.271</td>
<td>0.026</td>
<td>0.081</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Betaproteobacteria</em></td>
<td>2.579</td>
<td>-0.119</td>
<td>0.026</td>
<td>0.093</td>
<td>Wk0/H</td>
</tr>
<tr>
<td><em>Enterobacteriaceae</em></td>
<td>3.204</td>
<td>5.026</td>
<td>0.036</td>
<td>0.123</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Sphingomonas</em></td>
<td>0.939</td>
<td>2.103</td>
<td>0.047</td>
<td>0.139</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Faecalibacterium</em></td>
<td>-0.361</td>
<td>-2.055</td>
<td>0.055</td>
<td>0.156</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Tepidimonas</em></td>
<td>0.662</td>
<td>-2.512</td>
<td>0.048</td>
<td>0.159</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Comamonadaceae</em></td>
<td>1.21</td>
<td>2.257</td>
<td>0.064</td>
<td>0.195</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Schlegelella</em></td>
<td>1.058</td>
<td>-1.364</td>
<td>0.067</td>
<td>0.251</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Pseudomonas</em></td>
<td>3.494</td>
<td>4.771</td>
<td>0.098</td>
<td>0.359</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Finegoldia</em></td>
<td>-2.234</td>
<td>-4.213</td>
<td>0.186</td>
<td>0.42</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Bacteroidetes</em></td>
<td>-3.188</td>
<td>-3.058</td>
<td>0.248</td>
<td>0.463</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Prevotella</em></td>
<td>-3.061</td>
<td>-3.022</td>
<td>0.227</td>
<td>0.467</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Legionella</em></td>
<td>-2.686</td>
<td>-4.538</td>
<td>0.271</td>
<td>0.554</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Corynebacterium</em></td>
<td>-1.075</td>
<td>-2.789</td>
<td>0.235</td>
<td>0.569</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Sphingobium/Sphingomonas</em></td>
<td>1.136</td>
<td>-1.244</td>
<td>0.21</td>
<td>0.584</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Porphyromonas</em></td>
<td>-3.614</td>
<td>-5.246</td>
<td>0.306</td>
<td>0.611</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Rhizobium/Agrobacterium</em></td>
<td>1.212</td>
<td>-0.374</td>
<td>0.23</td>
<td>0.632</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Ruminococcus</em></td>
<td>-3.541</td>
<td>-4.725</td>
<td>0.36</td>
<td>0.642</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Flavobacteriaceae</em></td>
<td>-3.811</td>
<td>-5.116</td>
<td>0.346</td>
<td>0.642</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Gemella</em></td>
<td>-2.353</td>
<td>-0.957</td>
<td>0.375</td>
<td>0.766</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Peredibacter</em></td>
<td>-2.9</td>
<td>-4.141</td>
<td>0.423</td>
<td>0.768</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Sphingobacterium</em></td>
<td>-2.283</td>
<td>-3.517</td>
<td>0.476</td>
<td>0.781</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Atopobium</em></td>
<td>-3.526</td>
<td>-3.942</td>
<td>0.552</td>
<td>0.867</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Methylphilus</em></td>
<td>-3.242</td>
<td>-1.833</td>
<td>0.517</td>
<td>0.879</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Brevundimonas</em></td>
<td>1.479</td>
<td>0.797</td>
<td>0.483</td>
<td>0.918</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Gardnerella</em></td>
<td>-2.188</td>
<td>-2.203</td>
<td>0.637</td>
<td>0.933</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Propionibacterium</em></td>
<td>2.103</td>
<td>2.182</td>
<td>0.721</td>
<td>0.954</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Lactobacillus</em></td>
<td>0.891</td>
<td>1.273</td>
<td>0.837</td>
<td>0.993</td>
<td>N/A</td>
</tr>
<tr>
<td>unclassified Bacteria</td>
<td>5.004</td>
<td>4.489</td>
<td>0.874</td>
<td>1</td>
<td>N/A</td>
</tr>
</tbody>
</table>
The metabolic profile also changed as a result of chemotherapy (Figure 2-5) and was similar between week 2-16, but different to that observed at week 0. A total of 226 metabolites were detected by our GC-MS method, 12 of which were significantly different between the week 0 and chemotherapy (week 2-16) groups (Table 2-2). Dataset 2-2 shows the relative abundance of all metabolites detected in milk.

Table 2-2. Metabolites significantly altered by chemotherapy

(*) metabolite identity confirmed by authentic standards. Chemotherapy refers to samples collected from week 2-16. Significant differences were determined using the unpaired student’s t-test, p < 0.05.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Elevated in</th>
<th>Bonferroni corrected p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>unknown PUFA</td>
<td>week 0</td>
<td>1.81E-07</td>
</tr>
<tr>
<td>DHA *</td>
<td>week 0</td>
<td>0.000304</td>
</tr>
<tr>
<td>arabinose</td>
<td>chemotherapy</td>
<td>0.000456</td>
</tr>
<tr>
<td>threitol</td>
<td>chemotherapy</td>
<td>0.001342</td>
</tr>
<tr>
<td>unknown</td>
<td>chemotherapy</td>
<td>0.002685</td>
</tr>
<tr>
<td>unknown</td>
<td>chemotherapy</td>
<td>0.002768</td>
</tr>
<tr>
<td>decanoic acid</td>
<td>chemotherapy</td>
<td>0.008458</td>
</tr>
<tr>
<td>myristic acid</td>
<td>chemotherapy</td>
<td>0.008727</td>
</tr>
<tr>
<td>1-monopalmitin</td>
<td>chemotherapy</td>
<td>0.009143</td>
</tr>
<tr>
<td>butanal</td>
<td>chemotherapy</td>
<td>0.012356</td>
</tr>
<tr>
<td>unknown</td>
<td>chemotherapy</td>
<td>0.017961</td>
</tr>
<tr>
<td>Inositol *</td>
<td>week 0</td>
<td>0.037225</td>
</tr>
</tbody>
</table>
Figure 2-5. Principal component analysis of metabolites in breast milk at week 0 and during chemotherapy

(A) Scoreplot displaying the distribution of samples based on metabolites alone, where the distance between samples represents how similar the metabolome of those samples are. Each point represents the average of 2 technical replicates. (B) Loadings plot. Each point represents a metabolite. Metabolites present in a given quadrant of the loadings plot are present in highest abundance in samples present in the same quadrant of the scoreplot (A).
2.4 Discussion

This study has shown that a course of chemotherapy has significant effects on bacterial and metabolic profiles in human milk, moving away from those of healthy lactating women. Of note, the subject did not report any additional drug use, antibiotics, illness or major changes in diet over the course of the study.

The consequences of decreased bacterial diversity in human milk and the implications on the child are still unknown however the decreased milk diversity could impact intestinal diversity and it has been shown that low intestinal diversity in the first weeks of life has been associated with necrotizing enterocolitis (Mshvildadze et al., 2010) and an increased risk of allergy and atopy in school age children (Madan et al., 2012). Lower intestinal diversity has also been observed in children with type 1 diabetes compared to age matched controls (Giongo et al., 2011).

In addition to overall changes in microbial profiles, we observed a significant decrease in the relative abundance of *Bifidobacterium* in the chemotherapy group compared to the non-treatment group. *Bifidobacterium* is the predominant organism in the gut of breast fed infants, attributed to its ability to metabolize the human milk oligosaccharides (HMO) present in large amounts in milk (Sela et al., 2008; Asakuma et al., 2011). Maternal levels have also been shown to impact gut *Bifidobacterium* abundance, with low levels in milk correlating with low levels in the neonatal gut (Grönlund et al., 2007). The potential consequences of decreased numbers of *Bifidobacterium* being passed on from the mother to the neonate could be an increased risk of asthma or obesity later in life. High levels of *Bacteroides* have been reported in the gut of infants with low levels of *Bifidobacterium* (Jost et al., 2012) and early colonization with high *Bacteroides* counts has been attributed
to an increased risk of developing asthma and obesity (Vael et al., 2008, 2011a, 2011b). In addition, depleted levels of *Bifidobacterium* have been shown to promote colonization of opportunistic pathogens such as *Klebsiella* and *Citrobacter* (Jost et al., 2012).

*Staphylococcus*, including coagulase negative species, are one of the predominant organisms in human milk (Hunt et al., 2011; Jost et al., 2013; Ward et al., 2013) and were also significantly reduced as a result of chemotherapy. It has been shown that numerous human milk isolates of *Staphylococcus epidermidis* can inhibit the growth of *Staphylococcus aureus* (Heikkila and Saris, 2003), the main causative agent of mastitis, a painful inflammatory condition of the breast, that often leads to pre-mature cessation of breastfeeding in many women. While we were not able to identify the *Staphylococcus* in our samples down to the species level with 16S rRNA gene sequencing, culture analysis on mannitol salt agar plates did show that the *Staphylococcus* isolates were not *S.aureus* and the select few that were tested were coagulase negative. This reduction of *Staphylococcus* (likely coagulase negative species) as a result of chemotherapy could make lactating women more prone to infections, affecting both themselves and their infants. Like *Bifidobacterium*, *Staphylococcus* is passed from the milk to the neonate, with higher numbers in the intestine of breastfed compared to formula fed infants (Martín et al., 2012). Interestingly, a metagenome analysis revealed the presence of immunosuppressive motifs in bacterial DNA from human milk, with the majority of these belonging to *Staphylococcus* (Ward et al., 2013). The exposure of the neonate to this DNA, either ingested from the milk or through live bacteria that have released their DNA once in the gut, could help to regulate the immune response against a variety of innocuous bacterial, environmental and food antigens.
The utilization of bacterial products by other bacteria is termed metabolic cross-feeding and plays an important role in bacterial selection. For example, the byproducts of bacterial metabolism such as lactate and acetate production are utilized as an energy source by many butyrate producing bacteria (Barcenilla et al., 2000; Belenguer et al., 2006; Duncan et al., 2004a, 2004b) such as *Eubacterium*, which was decreased in our chemotherapy group. Butyrate is important for health, as it reduces inflammation and metabolic diseases, promotes colonic repair and protects against colon cancer (Hamer et al., 2008; Lin et al., 2012). On the other hand some pathogens persist and cause disease only in the presence of certain commensal bacteria (Ramsey et al., 2011), likely due to the metabolites produced. Thus changes in bacterial communities in human milk will inevitably alter the metabolic milieu, selecting for bacteria able to utilize those metabolites. As a result, potential shifts from a healthy and balanced intestinal microbiota can occur, having important consequences on health.

Docosahexaenoic acid (DHA), inositol and an unknown poly-unsaturated fatty acid (PUFA) were among the 12 metabolites that differed between week 0 and weeks 2-16, with reduced levels detected during chemotherapy (weeks 2-16). DHA is the most abundant long chain PUFA in the brain, retina and nerve cells and is supplied mainly through breast milk (Morse, 2012). DHA deficiencies lead to reduced brain, eye and neuronal development (Morse, 2012) and it has been observed that breastfed infants have better visual acuity and neuronal development compared to those fed formula (Hoffman et al., 2009). Reduced levels of DHA and alpha-linolenic acid (a precursor of DHA) have been reported in milk of mothers with atopic children compared to milk from mothers with non-atopic children (Duchén et al., 2000; Yu et al., 1998). Another principal
metabolite in the neonatal brain is inositol and is important for osmoregulation, cellular nutrition and detoxification (Kok et al., 2001). Palmitic acid levels were also reduced during chemotherapy, though the results were not significant. Palmitic acid is the most abundant lipid in human milk and has been shown to increase bone strength in infants (Litmanovitz et al., 2013) and limit intestinal damage and pro-inflammatory immune responses in mice (Lu et al., 2013). While changes in metabolite concentrations do occur over the course of lactation, especially between colostrum and mature milk, no changes in the above metabolites over the course of the first year of life have been detected in mature milk (Jóźwik et al.; López-López et al., 2002; Marangoni et al., 2000; Ribeiro et al., 2008). Due to the high variability in milk metabolites between individuals, we did not have enough samples from healthy women to make substantive claims as to how the Hodgkin’s patient compared, however, there were no obvious differences in the metabolic profiles of control samples taken at early compared to later stages of lactation.

We recognize that the main limitation of the study is its single case study content of a patient undergoing chemotherapy. However, the findings were revealing and warrant further investigation into how more commonly prescribed drugs, such as antibiotics, may effect the milk microbiota and if changes do occur, how long do these changes last.

2.5 Conclusion

The bacterial and metabolic composition in human milk, so critical for infant development, can change significantly by maternal exposure to chemotherapeutic agents. This study emphasizes the importance of conducting further studies to examine changes in the microbiota and metabolomic as a result of drug intake in lactating mothers, and the consequences of these changes to the long term health of the infant.
2.6 References


Jóźwik, M., Jóźwik, M., Teng, C., Jóźwik, M., and Battaglia, F.C. Human breast milk sugars and polyols over the first 10 puerperium days. Am. J. Hum. Biol. 25, 198–204.


protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. PLoS One 7, e35240.


Chapter 3

3 Human milk microbiota profiles in relation to birthing method, gestation and infant gender

This chapter is reproduced with permission (Appendix 8.6) from:


Supporting Datasets 3-1 (clinical metadata), 3-2 (OTU annotations) and 3-3 (ALDEx2 summary) are available for download online as additional excel files

3.1 Background

With the incidence of various non-infectious diseases on the rise, there is much interest in the developmental origins of health and disease and the potential role of early life feeding practices in modulating these outcomes. Breast fed infants have been shown to be better protected than formula fed infants against necrotizing enterocolitis and diarrhea, allergy and asthma, inflammatory bowel disease, type 1 and type II diabetes, obesity and cardiovascular disease (Guaraldi and Salvatori, 2012; Stuebe, 2009). In addition to immune protection and bioactive compounds being conveyed through maternal milk, a possible protective role of bacteria has been suggested. Lower than average levels of \textit{Bifidobacterium} in human milk correlate with low levels of \textit{Bifidobacterium} in the neonatal gut (Grönlund et al., 2007), allowing for higher than normal levels of \textit{Bacteroides} to be established (Jost et al., 2012). These high levels of \textit{Bacteroides} early in life have been associated with an increased risk of asthma and obesity later in life (Vael et al., 2008, 2011a, 2011b). Indeed, efforts to manipulate the microbiota of formula fed
infants through probiotic supplementation have resulted in protection against some of the above diseases, comparable to that observed for breast fed infants (Chouraqui et al., 2008; Dilli et al., 2015; Kankaanpää et al., 2002).

Differences exist in bioactive components, macronutrients, cytokines, enzymes, proteins and immunological factors between Pre-term vs Term milk and milk from mothers giving birth by vaginal vs cesarean deliveries (Dizdar et al., 2014; Koenig et al., 2005; Mehta and Petrova, 2011a, 2011b; Wang et al., 2014; Zanardo et al., 2001). As well, the energy content differs in milk depending on gender of the newborn, with breast milk from mothers who give birth to sons having more fat content than that of daughters (Fujita et al., 2012; Powe et al., 2010).

We hypothesized that physiological or hormonal triggers that influence milk composition might also support different bacterial genera. Thus, we studied human milk samples from women giving birth at different stages of gestation, by vaginal or caesarian delivery, and examined whether gender of the newborn also affected the microbiota profiles.

3.2 Methods

3.2.1 Milk collection and processing

A single milk sample (Day 6 and onwards post partum) was collected from 39 Caucasian Canadian women recruited from London, Ontario and the surrounding area, representing a homogenous community (Dataset 3-1, metadata). Even though the samples collected were from different days after birth, with some considered “transitional” and others “mature,” a Kendall’s Tau correlation test showed no statistically significant correlation between OTU relative abundances over time after a Benjamini-Hochberg correction for
multiple hypothesis tests. This shows that it is acceptable to use all milk samples, regardless of when it was collected, in our analyses.

Ethical approval was obtained from Western Research Ethics Board and Lawson Health Research Institute, London, Ontario, Canada. Subjects provided written consent for sample collection and subsequent analyses. Participants were excluded if they were suffering from mastitis and were/had been on antibiotics during lactation. Cesarean deliveries were classified as either (i) “non-elective”; if there were complications during labour or (ii) “elective”; if they were planned in advance or if the health of the fetus and/or mother was at risk prior to labour.

Wearing sterile gloves the women cleaned their nipple and surrounding area with sterile saline swabs to reduce the presence of skin bacteria. Milk was collected using a sterile HygieniKit Milk Collection System (Ameda, Buffalo Grove, IL, USA) attached to an electric breast pump. Between 5 and 15 ml of milk was pumped into a sterile tube and kept on ice until transfer to the laboratory, which occurred within 1 hour of collection. Samples were aliquoted and stored at -20°C until DNA extraction.

3.2.2 DNA isolation

After thawing on ice, 2 ml of milk were spun down at 20,000 g for 10 minutes and the supernatant discarded. The pellet was then homogenized in 1.4 ml of ASL buffer (QIAamp® DNA Stool Kit, QIAGEN: Valencia, CA, USA) and 400 mg of 0.1 mm diameter zirconium-glass beads (BioSpec Products, Bartlesville, OK, USA). Mechanical and chemical lyses were performed by bead beading at 4,800 rpm for 60 s, then 60 s on ice (repeated twice) using a mini-beadbeater-1 (BioSpec Products) and then incubated at
95°C for 5 minutes. Subsequent procedures were performed using the QIAGEN QIAamp® DNA Stool Kit according to the manufacturer’s protocol, with the exception of the last step in which the column was eluted with 120 µl of elution buffer. DNA was stored at -20°C until further use.

A no template PCR control and a DNA extraction kit reagent control were sequenced alongside the samples. We observed that the taxon abundances in the controls were uncorrelated with the abundances in the experimental samples, and the distance between the controls and samples was large. Thus, we conclude that the controls had different profiles than that of the milk samples (Figure 3-1).
Figure 3-1. Principal components analysis biplot comparing milk samples with controls.

To verify that the microbiota observed in our milk samples was not due to background contamination from reagents in either the DNA extraction kit or from the PCR, a no template PCR control (“NTC1”) and a tube of PBS that was extracted alongside the milk samples (“reagent_control”) were sequenced. Data presented in the biplot are from centered log ratio transformed values. As observed, the controls have a different microbial profile than the milk samples.
3.2.3 PCR amplification

The genomic DNA isolated from the clinical samples was amplified using barcoded primers that amplify the V6 hypervariable region of the 16S rRNA gene which is 70 base pairs long.

V6-Forward:
5’ACACTCTTTCCTACACGACGCTCTTCCGATCTnnnn(8)CWACGCGARGAACC TTACC 3’

V6-Reverse:
5’CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTnnnn(8)ACRACACGAGCT GACGAC 3’

nnnn indicates 4 randomly incorporated nucleotides and “8” represents a specific sample barcode sequence. The PCR was carried out in a 42 µl reaction containing 2 µl of DNA template (or nuclease-free water as a negative control), 0.15 µg/µl bovine serum albumin, 20µl of 2x GoTaq hot start colorless master mix (Promega) and 10µl of each primer (initial concentration: 3.2pmol/µl). Thermal cycling was carried out in an Eppendorf Mastercyler under the following conditions: initial denaturation at 95°C for 2 minutes followed by 25 cycles of 95°C for 1 minute, 55°C for 1 minute and 72°C for 1 minute. After amplification, the DNA concentration was measured with the Qubit® 2.0 Fluorometer (Invitrogen) using the broad range assay. Equimolar amounts of each PCR product were then pooled together and purified using the QIAquick PCR purification kit (QIAGEN). The pooled PCR purified sample was then paired-end sequenced on the Illumina Mi-Seq platform using a 150 cycle kit with a 2x80 run at the London Regional Genomics Center, London, Ontario, Canada following standard operating procedures.
3.2.4 Sequence processing and taxonomic assignment

Custom Perl and Bash scripts were used to de-multiplex the reads and assign barcoded reads to individual samples. Multiple layers of filtering were employed: (i) Paired end sequences were overlapped with Pandaseq, allowing 0 mismatches in the overlapped reads; (ii) Reads were kept if the sequence included a perfect match to the V6 16S rRNA gene primers; (iii) Barcodes were 8mers with an edit distance of >4 and reads were kept if the sequence were a perfect match to the barcode; (iv) Reads were clustered by 97% identity into operational taxonomic units (OTUs) using the Uclust algorithm of USEARCH v7 (Edgar, 2010) which has a de novo chimera filter built into it; (v) All singleton OTUs were discarded and those that represented ≥1% of the reads in at least one sample were kept.

Taxonomic assignments for each OTU were made by extracting the best hits from the SILVA database (Pruesse et al., 2007) and then manually verified using the Ribosomal Database Project (RDP) SeqMatch tool (rdp.cme.msu.edu) and by BLAST against the Green genes database (greengenes.lbl.gov) Taxonomy was assigned based on hits with the highest percentage identities and coverage. If multiple hits fulfilled this criterion, classification was re-assigned to a higher common taxonomy. A summary of each OTU classification and its sequence is shown in Dataset 3-2. The raw sequencing reads generated in this study have been deposited to the NCBI Short Read Archive (SRA) database accession # SRP 064311.

3.2.5 Data analysis

Weighted and un-weighted UniFrac distances (Lozupone and Knight, 2005) were calculated in QIIME (Caporaso et al., 2010) by using a tree of OTU sequences built with
FASTTREE (Price et al., 2009) based on an OTU sequence alignment made with MUSCLE (Edgar, 2004). The QIIME pipeline was also used to generate principal coordinate analysis (PCoA) plots to visualize the Bray-Curtis dissimilarity. Changes in the microbial-community composition were also analyzed by calculating the generalized UniFrac distance (alpha= 0.5, rooted phylogenetic tree) using the GUniFrac package in R, version 3.1.2 (Chen et al., 2012). PERMANOVA was used to test for statistical significance between the groups using 10000 permutations (QIIME package). Barplots, boxplots, stripcharts, PCoA plots and k-means clustering analysis were all generated in R (http://www.R-project.org/).

The ALDEx R package version 2 (Fernandes et al., 2014) was used to compare genera, class, family and phyla between preterm vs term milk; cesarean vs vaginal deliveries and male vs female child. Values reported in this manuscript represent the expected values of 128 Dirichlet Monte-Carlo instances. A value of zero indicates that organism abundance is equal to the geometric mean abundance. Thus, organisms more abundant than the mean will have positive values and those less abundant than the mean will have negative values. Base 2 was used for the logarithm so differences between values represent fold changes. Significance (p<0.05) was based on the Benjamini-Hochberg corrected p-value of the Welch’s t test and the Wilcoxon test.

3.3 Results
Characterization of the milk microbiota in 39 Canadian women showed high variability amongst women (Figure 3-2). Overall, the top 5 most abundant taxa represented in milk were *Staphylococcus* (31%), *Enterobacteriaceae* (10%), *Pseudomonas* (17%), *Streptococcus* (5%) and *Lactobacillus* (3%) (Figure 3-3). While there was variability
amongst samples at the genus level, at higher taxonomic levels, profiles were consistent amongst women with Proteobacteria and Firmicutes making up the highest proportion of phyla in all samples with Actinobacteria and Bacteroidetes also present, but at lower levels (Figure 3-4).
Figure 3-2. Breast milk microbiota in 39 Canadian women identified by 16S rRNA amplicon sequencing.

The relative abundances of bacterial genera in different human milk samples were visualized by bar plots. Each bar represents a subject and each colored box a bacterial taxon. The height of a coloured box represents the relative abundance of that organism within the sample. Taxa present in less than 2% abundance in a given sample are displayed in the “Remaining fraction” at the top of the graph (gray boxes). As shown by the bar plots, a variety of bacteria were detected in breast milk. The legend is read from bottom to top, with the bottom organism on the legend corresponding to the bottom colored box on the bar plot.
Figure 3-3. Five most relatively abundant genera in human milk

Each point on the graph represents a subject, which indicates the percent relative abundance of that genus within the sample. The line represents the mean for all samples within the group.
Figure 3-4. Percent abundance of bacterial phyla in breast milk identified by 16S rRNA amplicon sequencing

(A) The relative abundances of different phyla in different breast milk samples were visualized by bar plots. (B) Box plots of the four phyla identified in breast milk. The box signifies the 75% (upper) and 25% (lower) quartiles and thus shows where 50% of the samples lie. The black line inside the box represents the median. The whiskers represent the lowest datum still within 1.5 interquartile range (IQR) of the lower quartile and the highest datum still within 1.5 IQR of the upper quartile. Outliers are shown with open circles. As shown, 4 phyla are represented in breast milk and present in all samples, with Proteobacteria and Firmicutes being the most abundant.
No differences were detected in microbial profiles based on gestation, mode of delivery or gender of the child, using both weighted (Figure 3-5A) and un-weighted UniFrac distances (data not shown). The above two matrices, however, assign a large emphasis on either rare or highly abundant taxa, so compositional changes that occur in moderately abundant lineages may be overlooked (Chen et al., 2012). For this reason, we also analyzed the data using generalized UniFrac at an alpha of 0.5, which overcomes this problem (Chen et al., 2012). As with weighted UniFrac, no differences were observed based on gestation, mode of delivery or gender of the child (Figure 3-5B), but there was separation of 13 samples, forming 2 distinct clusters, which could not be explained by any of the other metadata that was collected (Figure 3-6). To further confirm the above results, we also utilized the Bray-Curtis dissimilarity metric, which does not take into account phylogenetic relatedness of the biological community, as does UniFrac. As expected, no differences were observed (data not shown).

The dataset was also analyzed using ALDEx2 to examine whether specific taxa were differentially expressed based on gestation, mode of delivery or gender. Again, no differences were detected at the genus (Dataset 3-3), family, class or phyla levels (data not shown).

Of interest, the abundance profiles clearly showed that mothers are transmitting very different bacterial profiles to newborns. For example, the milk of subject 40 contained over 80% abundance of staphylococci, whereas that of subject 25 comprised mostly Gram negative organisms, especially Pseudomonas.
**Figure 3-5. Principal coordinate (PCoA) plots comparing bacterial profiles based on gestation, mode of delivery and gender**

PCoA plots based on (A) weighted UniFrac distances or (B) generalized UniFrac distances at alpha 0.5. Each sample, represented by a coloured circle, is plotted on this two-dimensional, 2 axis plane with the first 2 components plotted. Samples (points) that cluster together are more similar in biota composition and abundance. As shown by the plot, the lack of distinct clustering between groups, for Gestation (1\textsuperscript{st} row), mode of delivery (2\textsuperscript{nd} row) and gender (3\textsuperscript{rd} row), indicate that no bacterial differences exist between pre-term vs term samples, cesarean vs vaginal delivery samples and male vs female samples. PERMANOVA (p<0.5). P\textsubscript{ex}=extremely pre-mature (gestation <28 weeks); P\textsubscript{very}=very pre-mature (gestation 29-32wks); P\textsubscript{late}=late pre-mature (gestation 33-36wks); T=Term (gestation >37wks). “c\_E”= elective cesarean delivery; “c\_NE”= non-elective c-section; “v”=vaginal delivery. “m”= male child; “f”= female child; “m\_tw”= twins both male; “f\_tw”= twins both female; “Twin”= male and female twins.
Figure 3-6. Principal coordinate analysis based on generalized UniFrac distances.

Each sample, represented by a coloured circle, is plotted on this two-dimensional, 2 axis plane with the first 2 components plotted. Samples (points) that cluster together are more similar in biota composition and abundance. GUniFrac, using an alpha of 0.5 (which is more sensitive to changes in moderately abundant taxa), was used to compare microbial profiles based on gestation, mode of delivery and gender. While no differences were seen based on these conditions, there were 3 distinct groups, which could not be explained by any of the metadata collected (Dataset 3-1). The number of clusters was determined using the k-means clustering analysis in R.
3.4 Discussion

This study revealed a range of bacterial genera in human milk, consistent with previous studies (Hunt et al., 2011; Jost et al., 2013; Urbaniak et al., 2012). Interestingly, even when a baby was born extremely prematurely (subject S24), the mother’s milk was similar in composition to that of a woman giving birth at full term (subject S30). It can be speculated that this might be a failsafe mechanism whereby the mother is 'ready' to pass along her bacterial imprint irrespective of when the baby is born. If so, the microbiota would appear to be recalcitrant to late pregnancy hormonal and inflammatory changes, which could indicate an evolutionary pressure protecting this niche for the baby's benefit. Further studies on this concept are warranted, and cases where the milk profiles are very different or the outcome for the baby negative, could be particularly insightful.

It was not surprising that milk from emergency C section deliveries (i.e. non-elective) did not differ from women who gave birth vaginally, as this decision is made at the time of labour, when the hormones and timing for birth have been initiated. We also did not see differences between non-elective C section deliveries and vaginal births which is in contrast to a study published by Cabrera-Rubio et al. in which they concluded that the human milk microbiome is shaped by mode of delivery (Cabrera-Rubio et al., 2012). However, we do not believe that the authors adequately proved this claim. In their analysis the milk samples collected were from both obese and normal weight women with no indication of the proportion in each group. Since it was shown in that same paper that body mass index influences the milk microbiota, the subject’s weight could have been a confounding factor responsible for the observed differences. In addition, the author’s claim that the milk microbiota is influenced by mode of delivery was based
solely on visual observations of barplots (that were not very distinct between the two
groups), with no PCoA plots or statistical analyses to support this claim. These
conflicting results between our group and that of Cabrera-Rubio highlights the need for
future studies with larger sample sizes and inclusion of women from various
demographics.

Because of the multivariate nature of the data, differences, if they are present, may have
been masked by the different variables confounding each other. Thus a larger sample size
allowing for linear regression analysis would strengthen our claims. However there is the
possibility that no matter the sample size, differences will only be apparent when
examined at the level of the individual. Schwarzberg et al. showed that treatment for
periodontitis had no effects on microbial profiles when compared to controls. However,
when pre and post treatment effects for each individual were compared, significant
changes in bacterial composition was observed. The authors thus emphasized the
importance of comparing shifts from a personalized healthy state to a personalized
disease state in order to truly understand biological changes (Schwarzberg et al., 2014).

Proteobacteria and Firmicutes were the dominant phyla, consistent with a previous high-
throughput study performed in Switzerland and analyzed with a different sequencing
platform (454 sequencing) (Jost et al., 2013). The detection of Staphylococcus, as the
most abundant organism, is likewise consistent with other studies (Hunt et al., 2011; Jost
et al., 2013; Urbaniak et al., 2012). Staphylococcus is more abundant in the gut of breast
fed newborns compared to formula fed ones (Balmer and Wharton, 1989; Lundequist et
al., 1985) but their numbers start to decrease after the first week of life when oxygen has
been consumed and an environment favourable for anaerobes is created (Adlerberth,
2008; Jost et al., 2012). Unlike *Staphylococcus*, which is present in high abundance in at least the first week of life, Proteobacteria are never found in high numbers in the feces of newborns, infants or Western adults (Human Microbiome Project Consortium, 2012; Jost et al., 2012; Koenig et al., 2011) Thus, an important question arises, as to what the evolutionary significance is of having such a diverse population of bacteria in milk, if only a select few seem to colonize during the development of the neonate. There are a few possibilities; firstly, persistent colonization is not always needed for beneficial effects, as transient exposure can be just as effective (Gan et al., 2014; Hapfelmeier et al., 2010). Secondly, bacteria in milk may not have to be passed on from mother to child to exert their beneficial effects. Many of the protective factors in milk such as antibodies, immune cells, lactoferrin and beta-defensins originate from the mammary gland and not from the blood (Tunzi et al., 2000). With the ability of bacteria to regulate host gene expression, such as anti-microbials, and their ability to stimulate the immune system, the plethora of bacteria in breast milk could be inducing up regulation of these protective factors in the breast that then get passed on in high concentrations to the neonate via milk.

From another perspective, there is the possibility that some of these milk microbes have either limited or no effects on the offspring but are present for the benefit of the mother, such as in the protection against mastitis. Mastitis is a painful inflammatory condition of the breast with the main causative agent being *Staphylococcus aureus* and it has been shown that some milk commensals have the ability to prevent *S. aureus* growth and infection (Heikkila and Saris, 2003). The same is true for a skin derived strain of *Propionibacterium acnes*, which prevents growth of *S.aureus* via its byproducts of
glycerol fermentation (Shu et al., 2013). With the high abundance of glycerol present in milk, milk derived strains of *Propionibacterium* may have similar growth inhibiting properties, which would account for its presence in every milk sample collected.

*Lactobacillus* was present in high abundance in milk, so for those women who deliver by C section and thus do not transfer lactobacilli from the vagina to the infant, the milk could provide a means for these beneficial organisms to reach the infant gut.

### 3.5 Conclusion

A diverse population of bacteria is present in breast milk dominated by the phyla Proteobacteria and Firmicutes and the taxa *Staphylococcus*, *Pseudomonas*, *Enterobacteriaceae*, *Streptococcus* and *Lactobacillus*. While no differences in microbial profiles were apparent based on gestation, mode of delivery or gender, more studies are still needed on what factors do influence bacterial communities in milk and how these changes impact neonatal and maternal health.
3.6 References


Chapter 4

4 Microbiota of human breast tissue

This chapter is reproduced with permission (Appendix 8.6) from:


Supporting Datasets 4-1 (clinical metadata) and 4-2 (taxonomic classification of OTU sequences) are available for download online as additional excel files

4.1 Introduction

The human body is home to a large and diverse population of bacteria with properties that are both harmful and beneficial to our health (Bäckhed, 2012; Hooper, 2004; Human Microbiome Project Consortium, 2012; Macpherson and Harris, 2004; Serino et al., 2009; Zhu et al., 2013) and for this reason there has been a strong push in recent years to fully characterize the bacteria associated with different parts of the body under different health conditions. These studies have been made possible with the use of deep sequencing technologies and sites once thought of as sterile, such as the stomach, bladder and lungs have now been shown to harbour an indigenous microbiota (Beck et al., 2012; Bik et al., 2006; Wolfe et al., 2012). We hypothesized that microbes may also be present in breast tissue given the known presence of bacteria in human milk (Urbaniak et al., 2012). This is not surprising considering that skin and oral bacteria have access to the mammary ducts through the nipple (Ramsay et al., 2004) with some recent studies suggesting their source to be from the mother’s gastrointestinal tract (Donnet-Hughes et
al., 2010). We rationalized that given the nutrient rich fatty composition of the female breast, the widespread vasculature and lymphatics, and the diffuse location of the lobules and ducts leading from the nipple, bacteria would be widespread within the mammary glands, irrespective of lactation. Thus, the objective of the study was to determine, using culture and non-culture methods, whether breast tissue contains a microbiome. To ensure that the results obtained were not an artifact of a single demographic, tissue was collected and processed from two distant countries, Canada and Ireland.

4.2 Materials and Methods

4.2.1 Clinical samples and study design (Canadian samples)

Ethical approval for this study was obtained from Western Research Ethics Board and Lawson Health Research Institute, London, Ontario, Canada. Patients provided written consent for sample collection and subsequent analyses.

4.2.1.1 Tissue collection and processing

Breast tissue was collected from 43 women (aged 18-90) undergoing breast surgery at St. Joseph’s Hospital in London, Ontario, Canada. Thirty-eight of those women underwent lumpectomies or mastectomies for either benign (n=11) or cancerous (n=27) tumours and 5 women underwent breast reductions and had no history of breast cancer. For those women with tumours, the tissue obtained for analysis was collected outside of the marginal zone, approximately 5cm away from the tumour. After excision, the fresh tissue was immediately placed in a sterile vial on ice and homogenized within 30 min of collection. As an environmental control, a tube filled with sterile phosphate buffered
saline (PBS) was left open for the duration of the surgical procedure and then processed in parallel with the tissue samples.

Tissue samples were homogenized in sterile PBS using a PolyTron 2100 homogenizer at 28000 rpm. The amount of PBS added was based on the weight of the tissue in order to obtain a final concentration of 0.4g/ml. Fresh homogenate and the environmental controls were then plated on different agar plates for culture analysis and the remaining amount aliquoted and stored at -80°C until DNA extraction.

4.2.1.2 DNA isolation

After tissue homogenates were thawed on ice, 400µl (equivalent to 160mg of tissue) were added to tubes containing 1.2ml of ASL buffer (QIAamp® DNA Stool Kit, Qiagen) and 400mg of 0.1mm diameter zirconium-glass beads (BioSpec Products). Mechanical and chemical lyses were performed by bead beading at 4800rpm for 60s, then 60s on ice (repeated twice) (mini-beadbeater-1, BioSpec Products) and then incubated at 95°C for 5min. Subsequent procedures were performed using the Qiagen QIAamp® DNA Stool Kit according to the manufacturer’s protocol, with the exception of the last step in which the column was eluted with 120µl of elution buffer. DNA was stored at -20°C until further use.

4.2.2 Clinical samples and study design (Irish samples)

Ethical approval for this study was obtained from University College Cork Clinical Research Committee. Patients provided written consent for sample collection and subsequent analyses.
4.2.2.1 Tissue collection and DNA isolation

Breast tissue was collected from 38 women (aged 20-85) undergoing breast surgery at Cork University Hospital or South Infirmary Victoria Hospital, Cork, Ireland. Thirty-three women underwent lumpectomies or mastectomies for cancerous tumours (taken at least 5cm away from the primary tumour site) while 5 women underwent breast reductions and had no history of breast cancer. Once collected the specimens were placed in sterile cryotubes and flash frozen in liquid nitrogen within 45 minutes of collection and then stored at -80°C until DNA extraction.

Tissue samples were thawed on ice and the genomic DNA extracted using the Gene-Elute Mammalian Genomic DNA miniprep kit (Sigma-Aldrich) as per the manufacturer’s protocol with the exception of the elution step, where the column was eluted with 70µl of elution buffer.

4.2.3 Ion Torrent V6 16S rRNA sequencing

4.2.3.1 PCR amplification

At the Canadian lab, the genomic DNA isolated from Canadian and Irish clinical samples was amplified using the barcoded primers V6-LT and V6-RT (see section 2.2.4.1 for primer sequence), which amplify the V6 hypervariable region of the 16S rRNA gene. The PCR was carried out in a 40µl reaction containing 5µl of DNA template (or nuclease free water as a negative control), 1.5mM MgCl₂, 0.8µM of each primer, 4µl of 10x PCR Buffer (Invitrogen), 0.2mM dNTPs, 0.05U Taq Polymerase (Invitrogen) and 0.15µg/µl of bovine serum albumin. Thermal cycling was carried out in an Eppendorf Mastercyler under the following conditions: Initial denaturation at 95°C for 2min followed by 25
cycles of 95°C for 1 min, 55°C for 1 min and 72°C for 1 min. After amplification, the DNA concentration was measured with the Qubit® 2.0 Fluorometer (Invitrogen) using the broad range assay. Equimolar amounts of each PCR product were then pooled together and purified using the QIAquick PCR purification kit (Qiagen). The PCR purified sample was then sent to the London Regional Genomics Center, London, Ontario, Canada for V6 16S rRNA sequencing using the Ion Torrent platform as per the Center’s standard operating procedure.

### 4.2.3.2 Sequence processing and taxonomic assignment

Custom Perl and Bash scripts were used to de-multiplex the reads and assign barcoded reads to individual samples. Reads were kept if the sequence included a perfect match to the barcode and the V6 16S rRNA gene primers and were within the length expected for the V6 variable region. Samples with more than 600 reads were kept while those with less than 600 were discarded. Reads were then clustered by 97% identity into Operational Taxonomic Units (OTUs) using UClust v. 3.0.617 (Edgar, 2010). OTUs that represented ≥2% of the reads in at least one sample were kept, while those that did not meet the cut-off were discarded to account for the high error rate intrinsic to Ion Torrent sequencing. Taxonomic assignments for each OTU were made by the Ribosomal database project (RDP) SeqMatch tool (Cole et al., 2009). From the top 20 matches to the RDP named isolates database, the full taxonomy was retained for matches with the highest S_ab score. For multiple top matches with equal scores, the highest common taxonomy was retained (e.g. genus level if multiple species matched equally well). Since the maximum number of matches displayed per sequence is 20, the RDP taxonomic assignments were verified by BLAST against the Greengenes named isolates database with an output of
100 hits (DeSantis et al., 2006). Taxonomy was assigned based on hits with the highest % identities and coverage. If multiple hits fulfilled this criterion, classification was re-assigned to a higher common taxonomy. In instances where the highest % identity/coverage yielded a single match, if this were less than 90% and the S_ab score from RDP was less than 0.7, taxonomy was assigned at the Family level instead of at the Genus level.

4.2.3.3 Data analysis
The QIIME pipeline (Caporaso et al., 2010) was used to (i) calculate weighted UniFrac distances and Shannon’s diversity index (logarithms with base 2); (ii) summarize OTUs by different taxonomic levels and (iii) generate Unweighted Pair Group Method with Arithmetic Mean (UPGMA) trees representing hierarchical clustering of samples. The UniFrac distances were calculated by using a phylogenetic tree of OTU sequences built with FastTree (Price et al., 2009) and based on an OTU sequence alignment with MUSCLE (Edgar, 2004). Weighted UniFrac compares microbial profiles (presence/absence and abundance) between samples (i.e. beta-diversity) (Lozupone and Knight, 2005) while Shannon’s diversity index evaluates the microbial diversity within a sample (i.e. alpha diversity). The higher the Shannon’s diversity index, the more diverse a sample is and a value of zero indicates the presence of only one species (Hill et al., 2003). UPGMA trees allows one to visualize the distance matrix and the robustness of what was observed was tested with jackknifing and bootstrapping. For beta-diversity analyses, the data set was rarified to the lowest read count/sample. Barplots, boxplots and stripcharts were all generated in R (http://www.R-project.org/).
4.2.4 Culture analysis

In an effort to prove that bacteria in the breast were viable, 100µl of Canadian tissue homogenate from each of the 43 samples and 100µl of each environmental control were plated on Columbia blood agar (CBA) plates and incubated aerobically or anaerobically at 37°C. Of note, the facilities to perform culture were not available at the Irish site. DNA from single colonies was extracted using the InstaGene Matrix (Bio-Rad) and then amplified using the eubacterial primers pA/pH, which amplify the 16S rRNA gene:

pA 5’ AGAGTTTGATCCTGGCTCAG 3’ pH 5’ AAGGAGGTGATCCAGCCGCA 3’

The PCR reaction was carried out in 50µl reaction containing 10µl of DNA template (or nuclease free water as a negative control), 1.5mM MgCl₂, 1.0µM of each primer, 0.2mM dNTP, 5µl 10X PCR buffer (Invitrogen), and 0.05 Taq Polymerase (Invitrogen). Thermal cycling was carried out in an Eppendorf Mastercycler under the following conditions: Initial denaturation step at 95°C for 2min, followed by 30 cycles of 94°C for 30s, 55°C for 30s and 72°C for 1min. A final elongation step was performed at 72°C for 10min. After running 10µl of the PCR mixture on a 1% agarose gel to verify the presence of amplicons, 40µl of the PCR mixture was then purified using the QIAquick PCR purification kit (Qiagen). The purified PCR products were then sent for Sanger sequencing to the London Regional Genomics Centre, London, Ontario, Canada. Sequences were analyzed using the GenBank 16S ribosomal RNA sequences database using the BLAST algorithm (Altschul et al., 1990). Taxonomy was assigned based on the highest Max score.
4.2.5 Statistical analysis

Statistical analysis was performed in R using the Kruskal-Wallis one-way analysis of variance followed by the Mann-Whitney U test with Bonferroni correction.

4.3 Results

Tissue was obtained from various locations within the breast, from close to the nipple to as far back as the chest wall (Figure 4-1). Regardless of location sampled within the breast, presence/absence of breast malignancy, country of origin, age, history of pregnancy and method of DNA preparation, a variety of bacteria were detected in breast tissue (Figure 4-2, Figure 4-3 and Dataset 4-1). Bacterial diversity within samples varied between individuals with Shannon’s diversity indices ranging from 0.8-5.2, with an average value of 3.6 (Figure 4-4). To put into perspective, using the same index, oral and gut samples, known for their diverse bacterial communities have values between 3.9-6.5 (Bik et al., 2010; Dethlefsen and Relman, 2011; Dethlefsen et al., 2008; Li et al., 2012), while vaginal samples of low bacterial diversity, generate values between 0.46-2.9 (Kim et al., 2009; Oakley et al., 2008; Ravel et al., 2011).
Figure 4-1. Location within the breast of tissue samples collected for analysis

Tissue samples were collected from 43 women from London, Canada and 38 women from Cork, Ireland undergoing different breast surgeries. (A) and (B) Location of tissue collected from women in Canada undergoing lumpectomies or mastectomies for either malignant “LN” (A) or benign “LNb” (B) tumours. (C) Location of tissue collected from women in Ireland undergoing lumpectomies or mastectomies for malignant tumours “CN”. Blue ovals represent the location of the tumour and purple squares represent the location of the specimen obtained for bacterial analysis. The distance between the blue ovals and purple squares are approximate estimates of the distance between the tumour and the specimen, which was at least 5cm away from the tumour. (D) Location of tissue collected from women in both Canada and Ireland undergoing breast reduction surgery, “LH” and “CH” respectively. Asterisks in the purple boxes underneath the subject number indicate samples from Ireland. All samples were a minimum of 1cm deep to the skin with the surgeons aiming for mid-deep rather than superficial. As shown in the diagram, specimens were obtained from a variety of locations within the breast.
A
“LN” group

B
“LNB” group

C
“CN” group

D
“LH” & “CH” group
Figure 4-2. Breast tissue microbiota in 43 Canadian women identified by 16S rRNA sequencing

The relative abundance of bacterial taxa identified in different tissue samples was visualized by bar plots. Each bar represents a subject and each coloured box, a bacterial taxa. The height of the coloured boxes represents the relative abundance of that organism within the sample. For example, sample L42H is dominated by *Acinetobacter* whereas L19Nb is dominated by *Micrococcus*. Taxa present in less than 2% abundance in a given sample are displayed in the “Remaining fraction” at the top of the graph (grey boxes). As shown by the barplots, a variety of bacteria were detected in breast tissue. NB: The legend is read from bottom to top, as indicated by the red arrow, with the bottom organism on the legend, corresponding to the bottom coloured box on the barplot. “LH”= Tissue collected from women undergoing breast reductions; “LN”= Non-malignant tissue collected adjacent to cancerous tumours; “LNb”= Non-malignant tissue collected adjacent to benign tumours.
Microbiota fraction

Remaining fraction
Bradyrhizobium
Ruminococcus
clostridium
Thermus
Anoxybacillus/Bacillus
Legionella
Bacteroides
Phaeocolarobacterium
Fusobacterium
Roseburia
Bacteroides
trechtmanii
Bifidobacterium
Methylobacterium
Tayubimonas
Butyribrio
Cornebacterium
Truiper
Faecalibacterium
Clostridiales
Shewanella
Ruminococcaceae
Sphingobium/Sphingomonas
Cytophaga/Flavobacterium
Brevundimonas
Porphyrobacter
Schizogallia
unclassified
Lactococcus
Bacillales
Caulobacteraceae
Eubacterium
Microbacteriaceae
Succinivibrionaceae
Rhizobium/Agrobacterium
Lactobacillus
Sphingomonas
Straptonococcus
Betaproteobacteria
Clostridium
Comamonadaceae
Micrococcus
Lysobacter
Bacteria unclassified
Prevotella
Gammaproteobacteria
Pseudomonas
Bacillus
Acinetobacter
Propionibacterium
Staphylococcus
Enterobacteriaceae

Figure 2
Figure 4-3. Breast tissue microbiota in 38 Irish women identified by 16S rRNA sequencing

The relative abundance of bacterial taxa identified in different tissue samples was visualized by bar plots. Each bar represents a subject and each coloured box, a bacterial taxa. The height of the coloured boxes represents the relative abundance of that organism within the sample. For example, sample C40H is dominated by Janibacter whereas C36N is dominated by Enterobacteraceae. Taxa present in less than 2% abundance in a given sample are displayed in the “Remaining fraction” at the top of the graph (grey boxes). As shown by the barplots, a variety of bacteria were detected in breast tissue. NB: The legend is read from bottom to top, as indicated by the red arrow, with the bottom organism on the legend, corresponding to the bottom coloured box on the barplot. “CH”= Tissue collected from women undergoing breast reductions; “CN”= Non-malignant tissue collected adjacent to cancerous tumours
Figure 3
Bacterial diversity within a sample (i.e. alpha diversity) was measured by calculating Shannon’s diversity index. Each point on the graph represents a subject with the line representing the mean for all samples. The higher the index, the greater the bacterial diversity found within a sample. Since Shannon’s diversity index is a logarithmic number with a base of 2, a value of “4” is 2x higher than a value of “3”. The mean value for the Canadian samples was 3.9 (N=43) and for the Irish samples, 3.3 (N=33).

**Figure 4-4. Bacterial diversity of the breast tissue microbiota**

Bacterial diversity within a sample (i.e. alpha diversity) was measured by calculating Shannon’s diversity index. Each point on the graph represents a subject with the line representing the mean for all samples. The higher the index, the greater the bacterial diversity found within a sample. Since Shannon’s diversity index is a logarithmic number with a base of 2, a value of “4” is 2x higher than a value of “3”. The mean value for the Canadian samples was 3.9 (N=43) and for the Irish samples, 3.3 (N=33).
The bacteria identified in tissue were grouped into 121 operational taxonomic units (OTUs) based on 97% sequence similarity (Dataset 4-2). These OTUs belonged to 7 different phyla; *Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, Deinococcus-Thermus, Verrucomicrobia and Fusobacteria* with *Proteobacteria* being the most abundant phylum followed by *Firmicutes* (specifically from the class *Bacilli*) (Figure 4-5). Of the 121 OTUs identified, 57 could be classified at the genus level and 25 at the species level (Dataset 4-2). The most abundant taxa in the Canadian samples were *Bacillus* (11.4%), *Acinetobacter* (10.0%), unclassified *Enterobacteriaceae* (8.3%), *Pseudomonas* (6.5%), *Staphylococcus* (6.5%), *Propionibacterium* (5.8%), unclassified *Comamonadaceae* (5.7%), unclassified *Gammaproteobacteria* (5.0%) and *Prevotella* (5.0%). In the Irish samples the most abundant taxa were unclassified *Enterobacteriaceae* (30.8%), *Staphylococcus* (12.7%), *Listeria welshimeri* (12.1%), *Propionibacterium* (10.1%) and *Pseudomonas* (5.3%).
Figure 4-5. Percent abundance of different bacterial phyla in breast tissue identified by 16S rRNA sequencing

(A) Boxplots of the seven different phyla identified in breast tissue from the Canadian and Irish samples. The box signifies the 75% (upper) and 25% (lower) quartiles and thus shows where 50% of the samples lie. The black line inside the box represents the median. The whiskers represent the lowest datum still within 1.5 IQR of the lower quartile and the highest datum still within 1.5 IQR of the upper quartile (IQR= interquartile range). Outliers are shown with open circles. (B) The least abundant phyla shown in (A) were plotted on another graph with a smaller scale to allow for better visualization of percentage. In samples from both countries *Proteobacteria* was the most abundant phyla followed by *Firmicutes* (Kruskal-Wallis/Mann-Whitney U with Bonferroni correction, p ≤ 0.001).
Of the 30 environmental controls (i.e. PBS tube processed in parallel with tissue sample) sequenced on the Ion Torrent platform, 12 of them were discarded due to low read counts (<600 reads). Of the remaining 18 controls that passed quality control filtering, 13 of them had lower read counts than their respective tissue samples. Of the 6 with higher read counts, weighted UniFrac UPGMA hierarchical clustering showed that microbial profiles differed between the control and its respective tissue sample (Figure 4-6).

Bacteria were able to be cultured from all 43 of the Canadian samples (culture analysis was not performed on the Irish samples) with amounts ranging from 75-2000 cfu/gram of tissue, depending on the sample. Collectively, eight different strains were identified: *Bacillus sp*, *Micrococcus luteus*, *Propionibacterium acnes*, *Propionibacterium granulosum*, *Staphylococcus sp*, *Staphylococcus saprophyticus*, *Streptococcus oralis*, *Streptococcus agalactiae*. No bacteria were isolated from any of the 43 environmental controls.
Figure 4-6. **UPGMA hierarchical clustering comparing Canadian tissue samples with their respective environmental controls**

UPGMA hierarchical clustering translates the UniFrac weighted distances into a rooted tree. Samples that cluster together are more similar in bacterial profiles (presence and abundance) than those that are farther apart on the tree. Jackknifing and bootstrapping were performed to measure the robustness of these observations. The red coloured internal nodes provide strong support (75-100%) for similarities or differences between samples. The horizontal scale bar at the bottom indicates 1% sample divergence. As shown above, there is strong support that the bacterial profile between a tissue sample and its respective environmental control is distinct. Reads were rarified to 475 reads/sample. “E”= environmental control and “N”= tissue sample.
4.4 Discussion

We first suggested the potential for a breast microbiome in 2012 (Urbaniak et al., 2012), and this has now been confirmed in the present study, and recently by Xuan et al who examined breast tumor and normal adjacent tissue also using next-generation sequencing techniques (Xuan et al., 2014). In our study, none of the 81 women recruited had any clinical signs or symptoms of breast infection, yet bacteria were still detected in all Canadian samples analyzed, by both culture and molecular techniques and the Irish samples which were analyzed by just molecular techniques. While only 8 different strains were able to be cultured, we just used Columbia Blood Agar to isolate bacteria, and some fastidious bacteria such as *Prevotella* and *Bacteroides*, require more nutrient rich conditions for growth. Regardless, all those bacteria that were cultured were also detected by 16S rRNA sequencing, supporting the argument that the DNA amplifiers are from viable bacteria and not solely from remnant bacterial DNA.

*Proteobacteria* was the most abundant phylum in breast tissue, unlike in the vagina, oral cavity, bladder, skin and gastrointestinal tract where they make up only a small proportion of the overall bacterial community (Dewhirst et al., 2010; Grice et al., 2009; Human Microbiome Project Consortium, 2012; Hummelen et al., 2010; Wade, 2013; Wolfe et al., 2012). This finding is similar to what was observed in Xuan et al., and suggests that breast tissue may have a unique microbiota, distinct from that found at other body sites, which is consistent with published studies showing that bacterial communities vary across body habitats (Costello et al., 2009; Human Microbiome Project Consortium, 2012). The higher abundance of *Proteobacteria* and *Firmicutes* (specifically the class *Bacilli*) compared with other taxonomic groups may be a result of host-microbial
adaptation to the fatty acid environment in the tissue. A recent study documented a positive correlation between *Proteobacteria* and *Bacilli* and the metabolic byproducts of fatty acid metabolism, as well host derived genes involved in fatty acid biosynthesis (El Aidy et al., 2013). It has also been documented that wild-type mice fed a high fat diet have a gut microbiota enriched in *Firmicutes* (Murphy et al., 2010). Of interest, *Proteobacteria* is also the principal phylum in human milk (Ward et al., 2013) with many of the same bacteria we detected in tissue, raising the possibility that the tissue microbiota could be a source of bacterial inocula for babies.

Some of the organisms detected in breast tissue have been found at other body sites, such as *Lactobacillus iners* and *Prevotella* (vagina), *Enterobacteriaceae* (gastrointestinal tract), *Fusobacterium* and *Streptococcus* (oral cavity), *Propionibacterium* and *Micrococcus* (skin) and *Pseudomonas* (respiratory tract). Species known for their health-conferring properties, such as *Lactobacillus* and *Bifidobacterium* were detected as well as taxa known for pathogenesis at other sites, such as *Enterobacteriaceae*, *Pseudomonas* and *Streptococcus agalactiae*. We can only speculate why potentially pathogenic bacteria do not multiply and induce infections. One concept has been proposed that it is due to a host microbial sensory system aligned with antibacterial gene expression (Xuan et al., 2014).

Of interest, in the Xuan study, differential abundance was observed with the genera *Sphingomonas* and *Methylbacterium* between paired normal and tumour tissue, with the authors suggesting a potential role in tumour development. While we also detected these two genera in our tissue samples, further studies, with a larger sample size of healthy women, would be needed to determine whether differential abundance between healthy
and normal adjacent tissue also exists. While a comparison between tissue from cancer patients and healthy women was not the focus of this paper, we did notice a higher abundance of *E. coli* in women with cancer compared to healthy controls, with this species known for its cancer promoting activity (Davis et al., 1991). However, it is premature to suggest cause and effect until more work is done in this area.

One limitation of the study was the inability to make direct comparisons between the two countries due to the different collection and DNA extraction techniques used. Thus no claims can be made as to whether bacteria detected at one site differ significantly from the other.

This research shows that breast tissue is not sterile but contains a diverse community of bacteria, adding to the literature that body sites once believed to be sterile do indeed have an endogenous microbiome. It has been proposed that the breast microbiome contributes to maintenance of healthy breast tissue by stimulating resident immune cells (Xuan et al., 2014), but the type of bacteria and their metabolic activity, such as ability to degrade carcinogens, may also contribute. Further studies are warranted to determine how this breast microbiome is established, why no infections accompanied colonization, what impact these organisms have on the host and whether external factors such as diet, antibiotics and illness affect this bacterial community and the subsequent consequences for the woman.
4.5 References


Chapter 5

5 The microbiota of breast tissue and its association with breast cancer

This chapter is based on a manuscript under review for publication at mBio. Supporting Datasets 5-1 (clinical metadata) and 5-2 (ALDEx2 summary) are available for download online as additional excel files.

5.1 Introduction

Breast cancer is the leading cause of cancer death in women worldwide and it is expected that within the next 15 years the incidence of breast cancer cases in the U.S. will increase by 50%, resulting in 440,000 women being diagnosed by 2030 (Rosenberg et al., 2015). The etiology of breast cancer is still unknown but is believed to be due to a combination of genetic and environmental factors. Support for environmental factors comes from migration studies showing increased incidence of breast cancer amongst 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). While many environmental factors have been proposed, one not yet considered is the role of the host’s local tissue microbiome in modulating the risk of breast cancer development.

Bacteria inhabit numerous body sites and this collection of bacteria, termed the human microbiota, plays an integral role in human development. Changes in the composition of one’s microbiota could promote disease progression, as individuals with periodontitis (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.,
bacterial vaginosis (Hummelen et al., 2010) and colorectal cancer (Mira-Pascual et al., 2014) have different bacterial communities compared with healthy individuals. While it is still unclear if microbiota changes are a consequence or a cause of the disease, there is evidence in favour of the latter, 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).

The aim of the present study was to assess microbial profiles in breast tissue from women with breast cancer and those who were disease free, and to examine the consequences of this altered bacterial community in terms of breast cancer development. To investigate this, we examined the breast microbiota of 71 women who had either breast cancer, benign tumours or were disease free. Bacteria isolated from cancer patients were characterized and examined for their ability to induce DNA damage.

5.2 Methods

5.2.1 Microbiome analysis

5.2.1.1 Tissue collection and processing

Breast tissue was collected from 71 women (aged 19 to 90) undergoing breast surgery at St. Joseph's Hospital in London, Ontario, Canada. Ethical approval was obtained from Western Research Ethics Board and Lawson Health Research Institute, London, Ontario, Canada. Subjects provided written consent for sample collection and subsequent analyses. Fifty-eight women underwent lumpectomies or mastectomies for either benign (n=13) or cancerous (n=35) tumours, and 23 underwent either breast reductions or enhancements and had no history of breast cancer. For those women with tumours, the
tissue obtained for analysis was collected outside the marginal zone, approximately 5 cm away from the tumour.

After excision, fresh tissue was immediately placed in a sterile vial on ice and homogenized within 30 min of collection. As an environmental control, a tube filled with 1ml of sterile phosphate-buffered saline (PBS) was left open for the duration of the surgical procedure and then processed in parallel with the tissue samples. As an added control, a skin swab was collected of the disinfected breast area prior to surgery. The swab was placed in 1ml of sterile PBS and then vortexed at full speed for 5min to pellet the contents of the swab. The swab was then removed and the liquid stored at −80°C until DNA extraction.

Tissue samples were homogenized in sterile PBS using a PolyTron 2100 homogenizer at 28,000 rpm. The amount of PBS added was based on the weight of the tissue in order to obtain a final concentration of 0.4 g/ml. The homogenate was then stored at −80°C until DNA extraction.

5.2.1.2 DNA isolation

After tissue homogenates, in sealed containers, were thawed on ice, 400µl (equivalent to 160 mg of tissue) was added to tubes containing 1.2 ml of ASL buffer (QIAamp DNA stool kit; Qiagen) and 400 mg of 0.1-mm-diameter zirconium-glass beads (BioSpec Products). 800µl of the PBS control and 800µl of the skin swab control was also added to tubes containing ASL buffer and beads. Mechanical and chemical lyses were performed on all samples by bead beating at 4,800 rpm for 60s at room temperature and then 60s on ice (repeated twice) (Mini-Beadbeater-1; BioSpec Products), after which the suspension
was incubated at 95°C for 5 min. Subsequent procedures were performed using the Qiagen QIAamp DNA stool kit according to the manufacturer's protocol, with the exception of the last step, in which the column was eluted with 120μl of elution buffer. DNA was stored at −20°C until further use.

5.2.1.3 PCR amplification

The genomic DNA isolated from the clinical samples was amplified using barcoded primers that amplify the V6 hypervariable region of the 16S rRNA gene:

V6-Forward:
5’ACACTCTTTCCCTACACGACGCTCTTCCGATCTnnnn(8)CWACCGARGAACC TTACC 3’

V6-Reverse:
5’CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTnnnn(8)ACRACACGAGCT GACGAC 3’

nnnn indicates 4 randomly incorporated nucleotides and “8” represents a specific sample barcode sequence. The PCR was carried out in a 42μl reaction containing 2μl of DNA template (or nuclease-free water as a negative control), 0.15 μg/μl bovine serum albumin, 20μl of 2x GoTaq hot start colorless master mix (Promega) and 10μl of each primer (initial concentration: 3.2pmol/μl). Thermal cycling was carried out in an Eppendorf Mastercyler under the following conditions: initial denaturation at 95°C for 2 minutes followed by 25 cycles of 95°C for 1 minute, 55°C for 1 minute and 72°C for 1 minute. After amplification, the DNA concentration was measured with the Qubit® 2.0 Fluorometer (Invitrogen) using the broad range assay. Equimolar amounts of each PCR product were then pooled together and purified using the QIAquick PCR purification kit.
(QIAGEN). The pooled PCR purified sample was then paired-end sequenced on the Illumina Mi-Seq platform using a 150 cycle kit with a 2x80 run at the London Regional Genomics Center, London, Ontario, Canada following standard operating procedures.

5.2.1.4 Sequence processing and taxonomic assignment
Custom Perl and Bash scripts were used to de-multiplex the reads and assign barcoded reads to individual samples. Multiple layers of filtering were employed: (i) Paired end sequences were overlapped with Pandaseq, allowing 0 mismatches in the overlapped reads; (ii) Reads were kept if the sequence included a perfect match to the V6 16S rRNA gene primers; (iii) Barcodes were 8mers with an edit distance of >4 and reads were kept if the sequence were a perfect match to the barcode; (iv) Reads were clustered by 97% identity into operational taxonomic units (OTUs) using the Uclust algorithm of USEARCH v7 (Edgar, 2010) which has a de novo chimera filter built into it; (v) All singleton OTUs were discarded and those that represented ≥2% of the reads in at least one sample were kept (a filter for PCR and environmental controls and the skin swabs). Taxonomic assignments for each OTU were made by extracting the best hits from the SILVA database (Pruesse et al., 2007) and then manually verified using the Ribosomal Database Project (RDP) SeqMatch tool (rdp.cme.msu.edu) and by BLAST against the Green genes database (greengenes.lbl.gov) Taxonomy was assigned based on hits with the highest percentage identities and coverage. If multiple hits fulfilled this criterion, classification was re-assigned to a higher common taxonomy.

5.2.1.5 Data analysis
PCoA plots of weighted UniFrac distances (Lozupone and Knight, 2005) were generated in QIIME (Caporaso et al., 2010) by using a tree of OTU sequences built with
FASTTREE (Price et al., 2009) based on an OTU sequence alignment made with MUSCLE (Edgar, 2004). PERMANOVA was used to test for statistical significance between groups using 10000 permutations (QIIME package).

Microbiome data is compositional in nature (i.e. proportional distributions that are not independent of each other) and thus has several limitations (Fernandes et al., 2014). A simple example is as follows: If a sample has two organisms A (50%) and B (50%) and after antibiotic treatment organism A is completely killed, the proportion of B in that sample will now be 100% even if its actual abundance has not changed. Transforming the data, using centered log-ratios (CLR) alleviates the constraints inherent with compositional data (Aitchison, 1986) by allowing for subcomposition coherence, linear sample independence and normalization of read counts. CLR transformed data with a uniform prior of 0.5 applied was used when generating the K-means clusterplot and the dendogram of Euclidian distances.

The ALDEx R package version 2 (Fernandes et al., 2014) was used to compare genera the relative abundance of genera. Values reported in this manuscript represent the expected values of 128 Dirichlet Monte-Carlo instances of CLR transformed data. A value of zero indicates that organism abundance is equal to the geometric mean abundance. Thus, organisms more abundant than the mean will have positive values and those less abundant than the mean will have negative values. Base 2 was used for the logarithm so differences between values represent fold changes. Significance was based on the Benjamini-Hochberg corrected p-value of the Wilcoxon rank test (significance threshold of 0.1).
The Microbiome Regression-based Kernel Association Test (MiRKAT) (Zhao et al., 2015) was performed in R using the MiRKAT package. Differences in microbiota profiles were tested using a kernel metric constructed from weighted UniFrac, unweighted UniFrac and GUniFrac (Chen et al., 2012) distances and the Bray-Curtis dissimilarity metric. “Optimal” MiRKAT allows for the simultaneous examination of multiple distance/dissimilarity metrics alleviating the problem of choosing the best one and was performed on the aforementioned metrics. The p-values generated were the mean of 128 Dirichlet Monte-Carlo instances.

The R script of “SourceTracker “(version 0.9.1) was used to assess contamination of the tissue microbiota. Tissue samples were designated as “sink” and PBS controls as “source.” Barplots, boxplots, K-means clusterplots and dendograms were all generated in R (http://www.R-project.org/).

Full details regarding Irish tissue sample collection, patient demographics, DNA extraction protocols and the steps followed to generate the OTU table used for the analysis in Figure 5-6, can be found in our previous publications (Lehouritis et al., 2015; Urbaniak et al., 2014).

5.2.2 DNA damage assay

5.2.2.1 Bacterial strains

Isolates were obtained by plating 100µl of tissue homogenate on Columbia blood (CBA), MacConkey and Beef Heart Infusion (BHI) agar plates and incubating both aerobically or anaerobically at 37°C. DNA from single colonies was extracted using the InstaGene Matrix (Bio-Rad) and then amplified using the eubacterial primers pA/pH, which
amplifies the complete 16S rRNA gene (see section 4.2.4 for the primer sequences). The PCR reaction was carried out in 50µl reaction containing 10µl of DNA template (or nuclease free water as a negative control), 1.5mM MgCl₂, 1.0µM of each primer, 0.2mM dNTP, 5µl 10X PCR buffer (Invitrogen), and 0.05 Taq Polymerase (Invitrogen). Thermal cycling was carried out in an Eppendorf Mastercyler under the following conditions: Initial denaturation step at 95°C for 2min, followed by 30 cycles of 94°C for 30s, 55°C for 30s and 72°C for 1min. A final elongation step was performed at 72°C for 10min. 40µl of the PCR mixture was then purified using the QIAquick PCR purification kit (Qiagen) and the purified products sent for Sanger sequencing to the London Regional Genomics Centre, London, Ontario, Canada. Sequences were analyzed using the GenBank 16S ribosomal RNA sequences database and the Greengenes database. Taxonomy was assigned based on the highest Max score. Because the 16S rRNA gene does not differentiate members of the Enterobacteriaceae family very well, to confirm that our isolates were indeed *E.coli*, we utilized the API® 20E strip to differentiate species that are part of the this family. *E.coli* strain IHE3034 was kindly provided by Jean Philippe Nougayrède (INRA, Toulouse, France).

### 5.2.2.2 Infection assay

HeLa cells were maintained and passaged in DMEM/glutamax media (Invitrogen) supplemented with 10% FBS (Invitrogen). On the day of the experiment a 24 well plate containing sterile cover slips was seeded with 0.5ml of 1x10⁵ cells/ml, resulting in 5x10⁴ HeLa cells/well. The plates were then incubated at 37°C with 5% CO₂ for 24 hours after which times the media was removed and the wells washed with sterile PBS. HeLa cells (2 wells for each organism) were then infected at a MOI 100 cfu of bacterial cells per HeLa
cell for 4 hours with either *Staphylococcus epidermidis* (subject 31), *Micrococcus luteus* (subject 8), *Micrococcus sp* (subject 8), *E.coli* (subject 41 (isolates H and E), subject 34 and strain IHE3034), *Propionibacterium acnes* (subject 20) and *P. granulosum* (subject 20) or at a MOI 1 for 2 hours with *Bacillus cereus* (subject 34 & subject 6). The bacterial cultures for infection were prepared by inoculating 5ml of BHI with 1 colony and incubating aerobically at 37°C for 15 hours, with the exception of *Propionibacterium*, which was incubated anaerobically for 72hr. Bacterial cultures were then spun down at 3500g for 10min, washed and resuspended in PBS. Bacterial cells were then diluted to the appropriate concentration in DMEM media containing 10% FBS and 25mM HEPES. 40µM etoposide (Sigma) was used as a technical positive control. The pH was checked at the end of the experiment to ensure consistency between wells.

### 5.2.2.3 Immunofluorescence

After infection, media was removed and HeLa cells were washed 3x with sterile PBS. Cells were then fixed and permeabilized for 12min at room temperature (RT) with a -20°C solution of 95% methanol and 5% acetic acid. Cells were then blocked for an hour with 0.3% Triton-100/5% goat serum. After blocking, a 1/200 dilution of the primary antibody (rabbit anti phospho-H2AX mAb; Cell Signaling technologies) was added and incubated over night at 4°C. After washing a 1/1000 dilution of the secondary antibody (goat anti-rabbit IgG, Alexa Fluor 647 conjugate; Cell Signaling technologies) was added and incubated at RT for 30min. Cells were then counter stained with 1µg/ml of DAPI (Life Technologies) for 1min. Cover slips were mounted on microscope slides containing a drop of ProLong Gold antifade mountant (Life Technologies). The experiments were performed three times.
Images were captured using the NIKON eclipse TE2000-S digital microscope. Eight fields of view for each replicate were recorded, for a total of 16 fields of view for each condition. Using ImageJ software (version 1.48a), the mean fluorescent intensity of each γH2AX stained cell was measured from the digital images. The digital images were also used to determine the percent of total cells stained positive for γH2AX, which was calculated by dividing the number of red cells (i.e. γH2AX positive) by the number of blue cells (i.e. DAPI stained) and multiplying by one hundred.

5.2.2.4 Statistics
Bar graphs of the mean and standard deviation from the 3 experiments were plotted using Prism (version 5.0a). Significance (p<0.05) was tested by a 1 way ANOVA followed by a 2-sided unpaired student’s t-test (Prism v 5.0a) with the false discovery rate (“FDR”) multiple test correction performed in R.

5.3 Results
5.3.1 Microbiota analysis
16S rRNA amplicon sequencing of the V6 hypervariable region was performed on 70 tissue samples and 38 environmental controls. A full summary of patient demographics can be found in Dataset 5-1. To assess the contribution of environmental contamination towards the overall tissue microbiota, we utilized the contamination predictor tool, “SourceTracker”, which compared the microbial population in the tissue samples to that of the phosphate buffered saline (PBS) environmental controls. Figure 5-1 shows that while there is contamination present, it makes up only a small proportion (average 10%) of the overall microbial community in breast tissue. A dendogram of Euclidian distances
of the centered log-ratio (clr) transformed dataset (Aitchison, 1986) was then constructed to visualize which tissue samples were similar to the PBS controls and to skin swabs collected from the disinfected breast area prior to surgery. As seen in Figure 5-2, skin swabs, PBS controls and the no template PCR control (NTC) formed a single cluster, which was separate from most of the tissue samples, indicating distinct microbial profiles. It is also important to note that no differences were observed within the controls and within tissue samples obtained from different surgeons and surgical rooms. To ensure stringent quality control, we removed those tissue samples that were part of the PBS/skin/NTC cluster from further analysis. In addition, OTUs present in over 2% abundance in the NTC and PBS controls were also removed from further analysis. 16S rRNA sequencing data of the remaining samples and OTUs, showed a diverse population of bacteria consisting of 61 OTUs and 28 taxa (Figure 5-3A) dominated by the phyla *Proteobacteria* and *Firmicutes* (Figure 5-3B).
Figure 5-1. Assessment of DNA contamination towards the breast tissue microbiota

SourceTracker was used to examine if and in what proportion the tissue microbiota was contaminated with environmental DNA by comparing the OTUs detected in the PBS environmental controls to that of the OTUs detected in breast tissue. While there is some contamination present, as would be expected in low biomass samples, this contamination makes up only 10% of the overall bacterial community in tissue. Each pie graph corresponds to one subject. The pink slices represent the proportion of OTUs in the tissue samples that are similar to those found in the PBS controls. The blue slices represent the OTUs unique to the tissue samples.
Figure 5-2. Examining the relatedness of bacterial profiles in tissue samples to that of the environmental controls

Dendogram of Euclidian distances were constructed from centered log-ratio transformed data to compare bacterial profiles between tissue samples, PBS controls, skin swabs and the no template PCR control. Two distinct branches are evident, separated by a red line. One branch consists of tissue samples (right of red line) and the other mainly of environmental controls (left of the red line). Those samples that fell within the environmental control branch (left) were removed from further analysis. “HC”- PBS control during surgery of healthy patients; “BTC”= PBS control during surgery of tumour patients; neg= PCR no template control; skin= skin swab of either healthy patients (“H”) or tumour patients (“BT”)
Figure 5-3. Breast tissue microbiota identified by 16S rRNA gene amplicon sequencing.

(A) The relative abundances of bacterial genera in different breast tissue samples were visualized by bar plots. Each bar represents a subject and each colored box a bacterial taxon. The height of a coloured box represents the relative abundance of that organism within the sample. Taxa present in less than 2% abundance in a given sample are displayed in the “Remaining fraction” at the top of the graph (gray boxes). As shown by the bar plots, a variety of bacteria were detected in breast tissue. The legend is read from bottom to top, with the bottom organism on the legend corresponding to the bottom colored box on the bar plot. (B) Box plots of the six phyla identified in breast tissue. The box signifies the 75% (upper) and 25% (lower) quartiles and thus shows where 50% of the samples lie. The black line inside the box represents the median. The whiskers represent the lowest datum still within 1.5 interquartile range (IQR) of the lower quartile and the highest datum still within 1.5 IQR of the upper quartile. Outliers are shown with open circles
A

Percent abundance of different phyla in breast tissue

B

Microbiota fraction

Remaining fraction

Bradyrhizobium

Kocuria

Micrococcaceae

Propionibacterium

Sphingomonas

Akkermansia

Thermus

Gardnerella

Cytophagales

Flavobacteriaceae

Bacteroidetes

Microbacteriaceae

Bacillales

Pseudomonas

Bradyrhizobiaceae

Thermoanaerobacterium

Lactococcus

Corynebacterium

Thermoanaerobacterium

Lactobacillus

Prevotella

Gammaproteobacteria

Bacillus

Staphylococcus
A comparison of normal adjacent tissue from women with breast cancer with that of tissue from healthy women showed distinctly different bacterial profiles on weighted UniFrac PCoA plots (Figure 5-4A). Unsupervised K-means clustering of the clr transformed dataset indicated two clusters and the PCA plot in Figure 5-4B shows clear separation between the healthy and cancer groups. Differences between the groups were also confirmed using the Microbiome Regression-based Kernel Association Test (MiRKAT) (Table 5-1).

**Table 5-1. Summary of p-values generated by MIRKAT**

<table>
<thead>
<tr>
<th></th>
<th>Bray-Curtis</th>
<th>Weighted UniFrac</th>
<th>Unweighted UniFrac</th>
<th>GUniFrac $\alpha=0.5$</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>4.58E-05</td>
<td>2.64E-06</td>
<td>4.12E-06</td>
<td>2.64E-06</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>0.000129826</td>
<td>1.04E-05</td>
<td>0.004191322</td>
<td>1.15E-05</td>
<td>0</td>
</tr>
<tr>
<td>Average</td>
<td>7.87E-05</td>
<td>4.52E-06</td>
<td>0.000395004</td>
<td>5.33E-06</td>
<td>0</td>
</tr>
<tr>
<td>Median</td>
<td>7.70E-05</td>
<td>4.24E-06</td>
<td>2.05E-04</td>
<td>4.87E-06</td>
<td>0</td>
</tr>
</tbody>
</table>

ALDEx2, which allows for the direct comparison of bacterial taxa between groups showed significantly higher compositional abundances of *Prevotella*, *Lactococcus*, *Streptococcus*, *Corynebacterium* and *Micrococcus* in healthy patients and *Bacillus*, *Staphylococcus*, *Enterobacteriaceae* (unclassified), *Comamonadaceae* (unclassified) and *Bacteroidetes* (unclassified) in cancer patients (Figure 5-5)(Dataset 5-2).
Figure 5-4. Comparison of bacterial profiles between breast cancer patients and healthy controls

(A) Weighted UniFrac principal coordinate (PCoA) plot and (B) K-means clusterplot of centered log-ratio transformed data. Each breast tissue sample, represented by a coloured point is plotted on a three-dimensional, 3-axis plane representing 79% of the variation observed between all samples (A) or 44.85% of the variation on a 2-axis plane (B). Samples (points) that cluster together are similar in biota composition and abundance. The distinct separation between the two groups indicates that bacterial profiles differ between women with and without cancer. The PERMANOVA test performed on the weighted UniFrac distances showed that the observed differences were statistically significant (10000 permutations; pseudo F-statistic=14.4; p-value <0.01).
These two components explain 44.85% of the point variability.
Figure 5-5. Differences in relative abundances of taxa exist between healthy and cancer patients

Summary of taxa that were statistically significantly different between the two groups as determined by ALDEx2. The top panel shows the bacteria that were higher in healthy patients compared to those with cancer (i.e. normal adjacent tissue) and the bottom panel shows the bacteria that were higher in cancer patients compared to healthy controls. The box signifies the 75% (upper) and 25% (lower) quartiles and thus shows where 50% of the samples lie. The black line inside the box represents the median. The whiskers represent the lowest datum still within 1.5 interquartile range (IQR) of the lower quartile and the highest datum still within 1.5 IQR of the upper quartile. Outliers are shown with open circles.
To assess whether bacteria surrounding the tumour microenvironment might be associated with the severity of cancer, we compared bacterial profiles in normal adjacent tissue from women with various stages of breast cancer. No differences were found based on invasiveness or stage (data not shown). However normal adjacent tissue from women with benign tumours had profiles that were more similar to normal adjacent tissue of women with cancerous tumours rather than tissue from healthy subjects (Table 5-2).

**Table 5-2. Comparison of relative abundances of bacteria in breast tissue between women with benign tumours, cancerous tumours and healthy controls.**

ALDEx2 was performed to compare the relative abundances of bacteria in breast tissue between healthy women, and normal adjacent tissue from those with benign tumours and cancerous tumours. Bacterial profiles in normal adjacent tissue from women with benign tumours are more similar to normal adjacent tissue from women with cancerous tumours rather than healthy controls.

<table>
<thead>
<tr>
<th>Cancer vs Healthy</th>
<th>Benign vs Healthy</th>
<th>Benign vs Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher in Healthy</td>
<td>Higher in Healthy</td>
<td>Higher in Benign</td>
</tr>
<tr>
<td>Prevotella</td>
<td>Prevotella</td>
<td>Micrococcus</td>
</tr>
<tr>
<td>Lactococcus</td>
<td>Lactococcus</td>
<td></td>
</tr>
<tr>
<td>Corynebacterium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micrococcus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher in Cancer</td>
<td>Higher in Benign</td>
<td></td>
</tr>
<tr>
<td>Bacillus</td>
<td>Bacillus</td>
<td></td>
</tr>
<tr>
<td>Staphylococcus</td>
<td>Staphylococcus</td>
<td></td>
</tr>
<tr>
<td>Enterobacteriaceae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comamonadaceae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>unclassified</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We have previously published two reports showing which bacteria are present in tumour tissue and normal adjacent tissue of women from Ireland (Lehouritis et al., 2015; Urbaniak et al., 2014). In this report, we now show, using weighted UniFrac distances, that bacterial communities do not differ between tumour tissue and normal adjacent tissue, both at the population level (Figure 5-6A) and within an individual (Figure 5-6B). Thus, when suitably-collected tumour tissue for microbiome analysis is not available, normal adjacent tissue may be a practical alternative.
Figure 5-6. Comparison of bacterial profiles between tumour tissue and normal adjacent tissue

Tissue from malignant tumours and matched normal adjacent (i.e. non-malignant) tissue from 33 women from Cork, Ireland were collected and analyzed. (A) Weighted UniFrac PCoA plots to assess bacterial profiles at the population level. Lack of distinct clustering between tumour and normal adjacent tissue groups indicates that bacterial profiles are similar between the two tissue types. (B) Differences in microbial communities between conditions can sometimes exist within an individual, which may not be evident at the population level. To examine whether this was the case, weighted UniFrac distances were analyzed between matched tumour and normal adjacent tissue (i.e. “paired”). The smaller the weighted UniFrac distance is between two samples, the more related they are. As depicted in the boxplots, the UniFrac distances between paired samples was even smaller than the distances calculated between all normal adjacent tissue (NA vs NA), all tumour tissue (T vs T) and between normal adjacent and tumour tissue (NA vs T) showing that even within an individual bacterial profiles are the same between tumour and normal adjacent tissue. ** denotes p-value <0.01 (1-way ANOVA, followed by a student’s t-test with the FDR multiple test correction).
**Figure S3**

**A**

![PCA plot with PC1 and PC2 axes. Red dots represent normal adjacent tissues, and blue dots represent tumors.](image)

**B**

![Box plot comparing weighted UniFrac distances for different sample pairs. Significant p-value indicated with ** p<0.01.](image)
5.3.2 Assessment of DNA damage ability of breast tissue isolates

*E.coli* strains belonging to the B2 phylotype harbour the *pks* pathogenicity island, which encodes for machinery for the production of the genotoxin, colibactin. These *pks* + strains have been implicated in colon cancer (Arthur et al., 2012; Buc et al., 2013) via their ability to induce DNA double stranded breaks and chromosomal instability (Cuevas-Ramos et al., 2010; Nougayrède et al., 2006). As shown in Figure 5-5, the family *Enterobacteriaceae*, of which *E.coli* is a member of, was relatively more abundant in cancer patients compared to healthy controls. For this reason, we wanted to examine whether *E. coli*, cultured from breast tissue of cancer patients, had the ability to induce DNA double stranded breaks (DSB). Cellular levels of $\gamma$H2AX, a surrogate marker of DSB, were measured in HeLa cells after incubation with various *E.coli* tissue isolates. *E.coli* IHE3034, which contains the *pks* pathogenicity island and induces DSB (Nougayrède et al., 2006) was used for comparison.

HeLa cells exposed to *E.coli* tissue isolates had significantly higher levels of $\gamma$H2AX compared with untreated cells, as measured by mean fluorescent intensity (MFI) and % of cells that stained positive for $\gamma$H2AX, with levels equivalent to that induced by *E.coli* IHE3034 (Figure 5-7). Additional isolates from breast cancer patients were also examined for the ability to induce DNA damage; (i) *Bacillus* and *Staphylococcus* were tested as these genera were more abundant in cancer patients; (ii) *Micrococcus*, as this genus was higher in healthy individuals and (iii) *Propionibacterium* as there were no differences in relative abundances between cancer patients and healthy controls. Neither
Bacillus, Microccoccus nor Propionibacterium isolates induced DSB, whereas Staphylococcus did (Figure 5-8).

γH2AX foci can occur and be resolved very quickly in response to DNA damage, thus a time course was performed, with Bacillus treated cells analyzed every 15min over a 2 hour period. No statistically significant differences at any time point were observed between treated and untreated cells (data not shown).
Figure 5-7. DNA damage ability of *E.coli* breast tissue isolates from cancer patients. *E.coli* was isolated from normal adjacent tissue of 2 patients with breast cancer and tested for its ability to induce DNA double stranded breaks. *E.coli* (isolates H and E) from subject 41, isolate L from subject 34 and strain IHE3034 were incubated with HeLa cells at MOI 100 cfu bacteria per HeLa cell for 4hr and then stained for γH2AX and DAPI. (A) Representative immunofluorescent images of HeLa cells at 1000x magnification. (B) Image J was used to measure the mean fluorescent intensity (MFI) of γH2AX positive cells from the digitally acquired images. (C) Percent of total cells stained for γH2AX calculated from the immunofluorescent images. Data displayed in the bar graphs represent the mean +/- SD of 3 experiments representing a total of 48 fields of view and approximately 300 cells for each treatment group. ** denotes p-value <0.01
**B**

![Graph showing Mean Fluorescent Intensity](image)

**C**

![Graph showing % cells stained γH2AX](image)
Figure 5-8. DNA damage ability of bacteria isolated from breast tissue of cancer patients

Bacteria isolated from normal adjacent tissue of patients with breast cancer were tested for their ability to induce DNA double stranded breaks. *Staphylococcus epidermidis*, *Micrococcus luteus*, *Micrococcus sp*, *Propionibacterium acnes* and *Propionibacterium granulosum* were incubated with HeLa cells at MOI 100 cfu bacteria per HeLa cell for 4hr or with *Bacillus cereus* at MOI 1 for 2hr and then stained for γH2AX and DAPI. (A) Image J was used to measure the mean fluorescent intensity (MFI) of γH2AX positive cells from digitally acquired immunofluorescent images. (B) Percent of total cells stained for γH2AX calculated from the immunofluorescent images. Data displayed in the bar graphs represent the mean +/- SD of 3 experiments (with the exception of *Propionibacterium* which was only done once) representing a total of 48 fields of view and approximately 300 cells for each treatment group. * denotes p-value <0.05.
5.4 Discussion

This study has shown that different bacterial profiles exist in “normal adjacent” breast tissue from women with breast cancer compared with “normal” tissue from healthy controls. In colorectal cancer (CRC) and oral squamous cell carcinoma (OSCC) bacterial profiles in the stool and saliva respectively, also differ between healthy and diseased patients (Mager et al., 2005; Wang et al., 2012; Weir et al., 2013) with evidence suggesting that changes in this community composition and function may be driving cancer progression at these sites (Ahn et al., 2012; Schwabe and Jobin, 2013). This raises the possibility that the differences observed in the breast could also play a role in breast cancer progression.

*Enterobacteriaceae* and *Staphylococcus* are two taxa found in higher abundance in breast cancer patients compared with healthy controls. Examination of three *E.coli* isolates (a member of the *Enterobacteriaceae* family) and one *Staphylococcus epidermidis* isolate, cultured from normal adjacent tissue of breast cancer patients, all displayed the ability to induce DNA double stranded breaks (DSB). DSB are the most detrimental type of DNA damage and are caused by genotoxins, reactive oxygen species, and ionizing radiation (Lees-Miller, 2003). Non-homologous end joining, the mechanism by which DSB are repaired, is extremely error-prone often resulting in missing bases at the site of damage (Lees-Miller, 2003). Accumulation of these mis-repairs within the cell over time leads to genomic instability and eventually cancer (Khanna and Jackson, 2001). DSB caused by bacteria such as *Helicobacter pylori* and certain strains of *E.coli* have been shown to induce chromosomal instability with prolonged exposure ( Cuevas-Ramos et al., 2010; Toller et al., 2011). While the same mechanisms may be involved in the *in vitro* assay
described here (or indeed breast tissue transformation), further tests would need to be done to verify whether chromosomal abnormalities do occur subsequent to the DNA damage induced by these breast isolates. In support of this hypothesis, total cell numbers were consistent between all treated and untreated groups, suggesting no induction of apoptosis. It is important to note that bacterial induced DNA damage may not be sufficient in itself to promote breast cancer development unless it occurs in a genetically susceptible host. All genetic and 3-30% of sporadic cancer cases have mutations in DNA repair or DNA checkpoint machinery (Negrini et al., 2010). Thus women who have impaired DNA repair/checkpoints may be more susceptible to bacterial induced DNA damage and may be at a higher risk of developing breast cancer than women without these mutations, even if they have the same “detrimental” microbes in their mammary glands.

*Bacillus* was also elevated in breast cancer patients compared with healthy controls, confirming our previous findings (Urbaniak et al., 2014). While *Bacillus* did not induce DSB like *E.coli* and *S.epidermidis* it could have other pro-carcinogenic effects. One study has shown that a *Bacillus cereus* strain, isolated from gingival plaque, metabolizes the hormone progesterone into 5 alpha-pregnane-3,20-dione (5αP) (Ojanotko-Harri et al., 1990). 5αP is higher in breast tumours compared with healthy breast tissue (Wiebe et al., 2000) and is believed to promote tumour development by stimulating cell proliferation (Wiebe, 2006; Wiebe et al., 2000). While our molecular analysis did not permit species level identification, all *Bacillus* strains cultured from our breast cancer patients were of the species *B.cereus*. 
An epidemiological study has shown that women who drink fermented milk products have a reduced risk of breast cancer development, irrespective of multivariable risk factors (van’t Veer et al., 1989). This protection could be attributed to the health promoting properties of the various lactic acid bacteria (LAB) present in fermented products. *Lactococcus* and *Streptococcus*, two such bacteria that were higher in healthy women compared with breast cancer patients, exhibit anti-carcinogenic properties and could play a role in prevention. Natural killer (NK) cells are vital in controlling tumour growth with epidemiological studies showing that low NK cell activity from PBMC is associated with an increased incidence of breast cancer (Imai et al., 2000; Strayer et al., 1984). *Lactococcus lactis* has been shown to activate murine splenic NK cells, enhancing cellular immunity (Kosaka et al., 2012). While no studies have yet been published comparing NK cell functionality in the breast between “normal” (i.e. healthy patients) and “normal adjacent” (breast cancer patients) tissue, it could be assumed, based on the PBMC data, that NK functionality is also impaired in the breast of those with cancer. *Lactococcus sp* present in the mammary glands, could be modulating cellular immunity by maintaining the cytotoxic activity of resident NK cells (Carrega et al., 2014) thus helping to prevent cancer development. *Streptococcus thermophilus* on the other hand, better than any other LAB tested, protects against DNA damage caused by reactive oxygen species by producing antioxidant metabolites that neutralize peroxide and superoxide radicals (Koller et al., 2008).

Orally administered *Lactobacillus sp*, has shown to be protective in animal models of breast cancer (de Moreno de LeBlanc et al., 2005). While total numbers did not differ between healthy and diseased patients, those with breast cancer may not experience the
full anti-carcinogenic benefits afforded by *Lactobacillus* due to the decrease in *Lactococcus* and *Streptococcus*, as LAB have been shown to act in synergy with each other (Shiou et al., 2013).

*Prevotella*, which was more abundant in healthy women compared with breast cancer patients, produces the short chain fatty acid (SCFA), propionate. Propionate, like other SCFA, has many beneficial health effects in the gut, one of them being the ability to regulate colorectal tumour growth (Hosseini et al., 2011). In both animal and human studies, higher levels of *Prevotella* were observed in the stool of healthy subjects compared to those with CRC (Weir et al., 2013; Zackular et al., 2013). However in the oral cavity, patients with OSCC have higher levels of *Prevotella* compared with healthy controls and when *Prevotella* presence was used as a diagnostic tool, the authors could predict 80% of the cancer cases (Mager et al., 2005). The conflicting association of *Prevotella* in CRC and OSCC could be due to the fact that metabolites function differently at different body sites. While SCFA are anti-inflammatory in the colon and associated with health (Louis et al., 2014), in the vagina, they are pro-inflammatory and associated with bacterial vaginosis (Aldunate et al., 2015). What role *Prevotella* and/or propionate may be playing in breast health (or disease) remains to be determined.

While we have reported differential abundances of certain organisms between health and diseased states, in reality, it is probably not a single organism driving disease progression or protection but rather an interplay of poly-microbial interactions. To get a better understanding of the microbial influence on breast cancer, the functionality of these microbes should be investigated. Further studies examining bacterial metabolites and
bacterial-induced host metabolites would provide vital information on the role of bacteria in breast health.

5.5 Conclusion

This is the first report, to our knowledge to show that bacterial profiles differ in breast tissue between healthy subjects and normal adjacent tissue of breast cancer patients. Some of the bacteria that were relatively more abundant in breast cancer patients had the ability to induce DNA double stranded breaks. Further studies need to be done to examine whether this DNA damage can lead to chromosomal aberrations and whether the differences in the bacterial profiles are a cause or a consequence of the disease. This study raises important questions as to the role of the breast microbiota in breast cancer development or prevention and whether bacteria could be harnessed for interventions to help prevent disease onset.
5.6 References


Chapter 6

Draft genome sequence of a *Bacillus cereus* strain isolated from non-malignant breast tissue from a woman with breast cancer

6.1 Genome announcement

*Bacillus cereus* was the predominant organism cultured from the breast of cancer patients that we studied (normal adjacent tissue), with it being present in more than half of the samples tested. No isolates were cultured from healthy controls (women undergoing breast reductions or enhancements). 16S rRNA gene targeted sequencing of the V6 hypervariable region showed higher relative abundances of *Bacillus sp* in breast cancer patients compared to healthy controls (Chapter 5). Previously, *B. cereus*, isolated from gingival pockets of a patient with periodontitis, has shown the ability to metabolize progesterone into metabolites that increase the risk of breast cancer development (Ojanotko-Harri et al., 1990; Wiebe, 2006). Various species of *Bacillus* (though not *B. cereus*) are dominant in pulmonary tissue from lung cancer patients (Apostolou et al., 2011). To get a better understanding of what *B. cereus* could be doing in the breast and whether it could have a role in breast cancer development or progression, we sequenced the genome of a *B. cereus* isolate, designated CU-1, cultured from a 70 year-old woman with stage 1 invasive breast cancer.

Genomic DNA was isolated using the PureLink Genomic DNA Mini Kit (Invitrogen). Libraries were then made using the Nextera XT DNA Library Preparation Kit (Illumina) and paired end sequenced (2x150bp) on the Illumina MiSeq platform (London Regional Genomic Center, London, Canada). Reads were assembled into contigs using Velvet
(version 1.2.10) set at the optimal kmer size of 31. 1467 contigs >100bp were generated, with 17x read coverage within contigs and an N50 of 17818bp. Gene calling of the assembled genome was carried out with RAST. A genome comparison between our breast tissue isolate and other fully sequenced \textit{B. cereus} genomes, deposited in NCBI, was performed using JSpecies (version 1.2.1), an \textit{in silico} alternative to DNA-DNA hybridization.

The \textit{B. cereus} CU-1 strain has a genome size of 5.5 Mb, typical of other \textit{B. cereus} strains. RAST analysis revealed 5536 coding sequences, with 41% assigned to a functional subsystem (Figure 6-1).

![Subsystem Category Distribution](image)

**Figure 6-1. Schematic of protein functional groups**

5536 coding sequences of \textit{B. cereus} CU-1 were annotated with RAST and assigned to different functional groups (i.e. “subsystems”). Of the 5536 coding sequences, 41% belonged to a subsystem.
JSpecies comparisons showed that CU-1 was similar to other *B. cereus* strains, being the most similar to *B. cereus* ATCC14579 (Table 6-1). Interestingly, our strain and not any of the others in Table 6-1, has the Lde efflux pump, which is associated with fluoroquinolone resistance. Since fluoroquinolones are synthetic antibiotics, this lends weight to the idea of CU-1 being associated with the human body over a period of time.

Table 6-1. JSpecies analysis comparing *B. cereus* CU-1 to other *B. cereus* isolates with fully sequenced genomes.

A comparison was also performed with a different species from the same genera. The tetra regression coefficient is an indication of how similar two strains are, with values closer to “1” representing more similarity.

<table>
<thead>
<tr>
<th>Name</th>
<th>Tetra regression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. cereus</em> ATCC14579</td>
<td>0.99934</td>
<td>Avirulent strain type strain</td>
</tr>
<tr>
<td><em>B. cereus</em> B4264</td>
<td>0.99908</td>
<td>Isolated from male patient with fatal pneumonia</td>
</tr>
<tr>
<td><em>B. cereus</em> F837/76</td>
<td>0.99845</td>
<td>Isolated from prostate wound</td>
</tr>
<tr>
<td><em>B. cereus</em> ATCC10987</td>
<td>0.99844</td>
<td>Dairy isolate, avirulent strain</td>
</tr>
<tr>
<td><em>B. cereus</em> 03BB102</td>
<td>0.99841</td>
<td>Isolated from blood from a male welder with fatal pneumonia</td>
</tr>
<tr>
<td><em>B. cereus</em> E33L</td>
<td>0.9984</td>
<td>Isolated from a swab of a zebra carcass in Namibia</td>
</tr>
<tr>
<td><em>B. cereus</em> NC7401</td>
<td>0.99834</td>
<td>Produces high levels of the emetic toxin cereulide, isolated from symptomatic patient</td>
</tr>
<tr>
<td><em>B. subtilis</em> subspecies <em>subtilis</em> str.168</td>
<td>0.72452</td>
<td>Strain 168 was isolated after <em>B. subtilis</em> Marburg was mutagenized with X-rays by two Yale University botanists. Nearly all <em>B. subtilis</em> strains used in academia are derivatives of strain 168.</td>
</tr>
</tbody>
</table>
The genome was examined for genes involved in estrogen or progesterone metabolism, as hormone regulation plays a key role in many types of breast cancer. However none of these were present in our strain. Other genes were identified that could play a role in breast cancer development or prevention (Table 6-2) but future *in vitro* studies would need to confirm their functionality.

Overall, although the fully closed genome of CU-1 was not mapped, the analysis shows a genome with the genetic tools for virulence and regulation of cancer development. The well known pathogenicity of this species and the lack of obvious infection in this patient raises an interesting question regarding the activity of CU-1 within the breast and a possible role of the breast tissue environment in down regulating or inhibiting the expression of these virulence genes.
Table 6-2. Potential proteins involved in cancer development or protection.

Analysis of the 5536 coding sequences revealed proteins that could contribute to health and disease. The ones with a possible connection with breast cancer progression or prevention are shown in this table.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toxins: Hemolysin BL, Non-hemolytic enterotoxin, Cytotoxin K</td>
<td>Known to cause disease in humans. Have the potential to induce DNA damage.</td>
</tr>
<tr>
<td>Chitin binding protein and chitinase</td>
<td>Chitin is the second most abundant polysaccharide in nature next to cellulose. Chitinases are found in a variety of prokaryotes even though they do not produce endogenous chitin. It is believed that chitin binding protein and chitinases act as virulence factors in some pathogens by modulating adhesion and/or invasion into host cells. Increased serum levels of chitinase correlate with disease severity, poorer prognosis, and shorter survival in many human cancers such as breast, colon, prostate, ovaries, brain, thyroid, lung, and liver.</td>
</tr>
<tr>
<td>Azurin</td>
<td>Azurin released from <em>Pseudomonas aeruginosa</em> blocks breast cancer cell proliferation and induces apoptosis through the mitochondrial pathway both <em>in vitro</em> and <em>in vivo</em>. Azurin is being considered as a potential chemotherapeutic treatment for breast cancer.</td>
</tr>
<tr>
<td>Bacillolysin</td>
<td>Bacillolysin is an enzyme found in many <em>Bacillus</em> species. It has been shown that bacillolysin from <em>B. megaterium</em> has the ability to convert plasminogen to angiostatin-like fragments. Angiostatin inhibits proliferation of vascular endothelial cells, which is a fundamental process in angiogenesis. Angiostatin has been shown to suppress both <em>in situ</em> and metastatic tumour growth in animal models.</td>
</tr>
<tr>
<td>S layer</td>
<td>S-layers from <em>Bacillaceae</em> function as adhesion sites for cell associated exo-enzymes. Strains containing S-layers are resistant to polymorphonuclear leukocytes in the absence of opsonins.</td>
</tr>
</tbody>
</table>

6.2 References


Chapter 7

7 Discussion

7.1 Verification of a breast tissue microbiome

The tissue microbiota data presented in Chapters 4 and 5 challenge the assumption that breast tissue is sterile unless under active infection. While HTS has allowed scientists to unravel the diversity and functionality of microbial communities across the body, concerns have been raised that sites once thought of as sterile but now purported to have a “microbiome” are really just contaminated DNA, sequenced from reagents and laboratory equipment (Salter et al., 2014). I dispute this suggestion for the work reported herein. In these studies, numerous aliquots of PBS (“environmental” controls) were processed in parallel with the tissue samples. Most of the tissue samples had microbial profiles that were different from the controls. Those with similarities to the controls and any OTUs associated with them were removed from the dataset before further characterization of the microbiota. In addition, the contamination predictor tool, “SourceTracker”, showed that even before implementation of the above, 90% of the microbial reads were derived from the human specimens and not from the environment. Even with these quality control measures in place, some people will still be wary of sequencing results. However, the ability to culture live bacteria from tissue (and not from any of our environmental controls) further suggests that organisms are present within the breast. These results are consistent with a study published in 1988 from the University of Michigan in which bilateral breast tissue from 30 women undergoing either breast reductions or enhancements were analyzed by culture (Thornton et al., 1988). The most abundant organisms identified were coagulase-negative staphylococci,
Propionibacterium acnes, Corynebacterium and Lactobacillus. To a lesser extent, Bacillus, alpha- and beta-streptococci, Clostridium, Peptococcus, Micrococcus, Enterococcus and gram-negative rods were also cultured. After a 12-month follow-up, only 2 of the 30 women developed infections, suggesting to the authors that the bacteria that had been isolated were not contaminants from surgical equipment. All of the organisms detected by Thornton and colleagues were also identified in our studies by either culture or HTS. In short, it may be difficult to overcome skepticism. In dealing with a non-sterile environment of a surgical suite and human body, it may never be possible to prove conclusively that an isolate was not from the air or other sources. The low numbers of bacteria compared to other body sites also make it hard to prepare histological sections that show the organisms within tissues.

7.1.1 Why do these bacteria not induce infections or inflammation

Similar to the Thornton study, Bacillus and Propionibacterium (acnes and granulosum) were isolated from our tissue specimens. Whole genome sequencing performed on one Bacillus cereus tissue isolate revealed the presence of genes that encode for the toxins hemolysin BL, non-hemolytic enterotoxin and cytotoxin K, which are believed to contribute to enteritis caused by this species (Fagerlund et al., 2008). P. acnes and P. granulosum, which are harmless constituents of the skin microbiota, are associated with catheter and prosthetic joint infections. Propionibacterium avidum, another skin commensal, has been implicated in breast abscesses after breast reduction mammoplasty (Kritikos et al., 2015; Levin et al., 2008). Overall, between 75-2000 cfu/ml were cultured from homogenized tissue (~190-5000 cfu/gram of tissue), and qPCR performed on tissue
samples collected from the Irish cohort showed an average of $5 \times 10^5$ 16S rRNA gene copies per gram of tissue. In blood, numbers as low as 10 cfu/ml can induce an inflammatory response with sometimes fatal outcomes (Potgieter et al., 2015). This therefore begs the question as to why these organisms did not cause infections or induce pro-inflammatory responses?

One possibility is that these bacteria have adapted to survive in the breast without being detected and attacked by the immune system. While *B. cereus* has the genes that encode for virulent enterotoxins, they may not be expressed within the mammary environment. While it is hard to ascertain what these factors could be, as many act in concert with each other, it has been shown that enterotoxin production is regulated by energy availability, growth rate, cell numbers, aeration and pH (Ceuppens et al., 2011). For example, the enterotoxin genes are under the control of the quorum sensing regulator, phospholipase C regulator (PlcR), hence enterotoxins are produced during stationary phase (Ceuppens et al., 2011). Thus, the density of cells within the breast may be below the threshold needed to activate PlcR. Furthermore, lactose has been shown to inhibit enterotoxin production, even in the absence of growth impairment (Rowan and Anderson, 1997).

It is important to note that the bacteria identified in tissue were mostly at the genus, and in some cases species, level. Bacterial strains vary greatly in their properties even within a species. A notable example is *E. coli*: strain Nissle 1917 is considered a probiotic while O157:H7 is a deadly gastrointestinal pathogen. Thus, the strains in the breast may be less virulent than other strains of the same species known to cause infections and/or induce pro-inflammatory responses.
Another potential reason for the lack of infection and inflammation is the presence of mechanisms within the mammary gland, similar to those in the GIT, that ensure homeostasis. Two of them, compartmentalization and immune hyporesponsiveness, are discussed below.

### 7.1.1.1 Compartmentalization

It had long been a subject of debate as to why healthy individuals do not mount an overt pro-inflammatory response against their gut commensals even though they share the same molecular patterns as pathogens, which do trigger an inflammatory cascade. It was not until a paper published in 2000 that we began to appreciate the importance of host-microbe compartmentalization in maintaining homeostasis (Macpherson et al. 2000). In this paper, it was shown that specific pathogen free mice lacked serum IgG antibodies specific against *Enterobacter cloacae*, the dominant aerobe in the gut of these mice. Upon systemic intravenous administration of $10^6$ cfu of live bacteria, a strong IgG response against *E. cloacae* was mounted. This showed that the non-responsiveness observed in the serum of un-manipulated mice was due to ignorance rather than tolerance.

The single layer of epithelial cells lining the lumen of the gut creates a physical barrier between the luminal contents and the “inside” of the body and plays an essential role in compartmentalization (Peterson and Artis, 2014). When this barrier is disrupted, microbial translocation occurs resulting in an inflammatory response (Adams et al., 2008; Dubinsky et al., 2006; Fava and Danese, 2011; Pastor Rojo et al., 2007). In a similar fashion, the ducts and the lobules of the mammary glands are made up of a lumen, surrounded by epithelial cells that form a stratified cuboidal epithelium (Chapter 1).
While this is different than the simple columnar epithelium of the GIT and has more of a role in secretion and absorption, it still provides a protective physical barrier against environmental agents. During lactation, milk and the bacteria present in it, are sequestered within the lumen of these ducts and lobules, which could explain the lack of infection and inflammation while the mother is breast-feeding. Even though our tissue samples were collected from non-gravid/non-lactating women, some of whom never had children, there is still an extensive ductal system in the “resting” state and in nulliparous women (Chapter 1). Thus, there is the possibility that the bacteria detected in our tissue samples (composed of a collection of ducts surrounded by adipose tissue) are compartmentalized within the lumen of these ducts and are not found within the surrounding adipose rich stroma. Case reports of breast infections caused by \textit{S. aureus} and \textit{P. avidum} report liponecrosis and adipose edema, with the ability to culture organisms from these areas (Jari et al.; Kritikos et al., 2015; Levin et al., 2008). These observations, to some degree, would support the idea that infections occur when bacteria penetrate into the stroma. To determine whether bacteria were localized within the ducts, histological sections were stained with methylene blue or probed with an antibody against peptidoglycan (Appendix 8.3). Unfortunately the methods used did not allow us to determine where these bacteria localized. Scanning electron microscopy could be informative, but with so few organisms present, it could take many samples to find the organisms.

\subsection*{7.1.1.2 Immune hyporesponsiveness}

Another mechanism that could help maintain homeostasis, regardless of where the bacteria are localized, could be an altered state of immune regulation. Immune
hyporesponsiveness in the GIT is one way in which the host can co-exist with the trillions of bacteria that reside there (Maynard et al., 2012). For example, intestinal macrophages are phenotypically and functionally different than those derived from sterile sites, like the blood. Upon exposure to either Gram negative or Gram positive organisms these intestinal macrophages produce less inflammatory cytokines compared to blood monocytes (Smythies et al., 2005). Although there is no data for human breast macrophages, bovine mammary macrophages are also less responsive against pathogens than blood monocytes and are also less effective at priming antigen-specific inflammatory T cells (Politis et al., 1991, 1992), collectively protecting the host against immune related tissue pathology (Günther et al., 2012).

### 7.2 Factors contributing to the microbial differences between breast cancer patients and healthy controls

Bacterial profiles were shown in Chapter 5 to differ between breast tissue from women who are healthy and normal adjacent tissue from those with breast cancer. We hypothesized that these differences could be contributing to disease development since women with cancer had higher relative abundances of bacteria that induce host DNA damage and lower relative abundances of bacteria that can inhibit tumour growth (properties that have been shown by others).

There is the possibility, however, that these differences are a consequence rather than a cause of the disease. It is well established that the tumour mass creates an environment vastly different than the surrounding tissue. Metabolism differs so much so that metabolic readouts can be used as biomarkers for tumour detection (Balog et al., 2013; Bathen et al., 2013; Budeczies et al., 2012). However, despite these metabolic differences, which
could impart a selective advantage for one organism(s) over another, our microbiota analysis of tumour tissue vs normal adjacent tissue showed no differences in bacterial profiles. Since normal adjacent tissue is more similar to healthy tissue (i.e. from women without disease) than to tumour tissue, I am inclined to believe that the differences observed between healthy and normal adjacent tissue were present before the development of the tumour. However, this is not to say that host metabolites have no role in shaping the breast microbiome. Perhaps due to genetics or lifestyle factors (diet, stress, smoking etc) women who develop breast cancer have different metabolic profiles than those who do not, which consequently influence the bacterial patterns in the breast. For example, in our healthy cohort, higher relative abundances of *Prevotella* were observed, and this genus has the ability to uptake and metabolize estradiol, which in turn enhances bacterial growth (Kornman and Loesche, 1982). Studies in pre- and post-menopausal women have shown that estradiol levels are lower in normal adjacent tissue of breast cancer patients than tissue from healthy controls (Geisler, 2003; Savolainen-Peltonen et al., 2014). Perhaps due to these low levels of estrogen in breast cancer patients, *Prevotella* strains do not have the optimal energy requirements for growth, which could explain the lower levels of *Prevotella* observed in breast cancer patients compared to healthy controls.

A metabolomics analysis is currently being performed on tissue samples and hopefully will provide information on how the host can shape this tissue microbiome, and also on the metabolic activity of the microbes (i.e. bacterial metabolites) and how they interact with the host (i.e. bacterial induced host metabolites). It may even shed light on how these breast tissue bacteria might contribute to, or help prevent, breast cancer
development. Our preliminary analysis does show that there are higher levels of ceramide in healthy tissue compared to normal adjacent tissue from breast cancer patients. Ceramide is a pro-apoptotic signaling molecule that controls aberrant cell growth and there is the possibility that probiotics, like *L. rhamnosus* GR-1, can breakdown lactosylceramide (found in high abundances in milk), into ceramide (Appendix 8.4), possibly protecting against breast cancer development in lactating women. Other strains, found within tissue, could also have similar properties.

It appears that the breast microbiome can be influenced by bacterial composition at distal sites. Women with celiac disease, who are known to have altered gut microbial communities, have different microbial profiles in their milk compared to healthy mothers (Olivares et al., 2015). Thus, there is the possibility that the differences in bacterial communities observed between healthy and breast cancer patients could be a reflection of differences in the GIT. Indeed, the gut microbiota of post-menopausal women with breast cancer is different than that of disease free subjects (Goedert et al., 2015). A possible way to test this concept would be to examine the gut microbiota of women prescribed antibiotics or who have had radical changes in diet and then determine whether these changes alter the breast microbiota, possibly through collection of nipple aspirate fluid. If true, this would have important implications, and may provide a means to manipulate the breast microbiota in order to mitigate disease and promote health.
7.3 The influence of the breast tissue microbiome on the response to chemotherapy

7.3.1 The ability of bacteria to bio-transform drugs

In Chapter 5, I present a rationale for how some bacteria could promote breast cancer development and how others could help prevent it. But the breast microbiota could have a significant impact on women’s health irrespective of its role in breast cancer development, by its ability to modulate the efficacy of chemotherapeutics. As shown in Chapter 2, chemotherapy drastically changed microbial profiles in human milk as early as Week 2 after the start of treatment. To test the reverse, *E. coli*, the most relatively abundant organism within breast tumours, was used as a model to test the impact of bacteria on cancer treatment. It was found that *E. coli* had the ability to bio-transform a variety of drugs used to treat various cancers, leading to decreased efficacy *in vitro* in half of the drugs tested (Lehouritis et al., 2015) (Table 7-1). This was further elucidated *in vivo*, as gemcitabine was less effective at controlling established tumours that had been intra-tumourally injected with *E. coli* compared to PBS. The bio-modulation that occurred in gemcitabine, after incubation with pure bacterial cultures, is consistent with possible acetylation. Xenobiotics are metabolized in two distinct phases, the 2nd of which includes acetylation (Omiecinski et al., 2011). Byproducts of acetylation are typically more hydrophilic than the parent compound and therefore more readily excreted from the cell (Omiecinski et al., 2011). Certain strains of *E. coli* have been shown to possess N-acetyltransferases, the enzymes responsible for acetylation (Chaslus-Dancla et al., 1986; White-Ziegler et al., 2002). So, a possible mechanism explaining our results could be that N-acetyltransferases from *E. coli*, acetylate gemcitabine, which decreases drug efficacy by
either (i) promoting faster removal of the drug from the tumour once it has entered or (ii) inhibiting drug uptake, preventing the drug from entering the tumour in the first place.

Table 7-1. Effects of bacteria on chemotherapeutic drug cytotoxicity in vitro

*E.coli Nissle 1917,* grown to log phase, was incubated for 2 hours with each of the drugs listed below. After incubation, the drug solution was filter sterilized to remove bacteria and then added to plates containing the 4T1 mouse mammary carcinoma cell line. Drug cytotoxicity was measured using the Cell Titre 96 Aqueous One solution Cell Proliferation Assay (Promega).

<table>
<thead>
<tr>
<th>Drug</th>
<th>Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gemcitabine</td>
<td>decreased</td>
</tr>
<tr>
<td>Doxorubicin</td>
<td>decreased</td>
</tr>
<tr>
<td>Daunorubicin</td>
<td>decreased</td>
</tr>
<tr>
<td>Mitoxantrone</td>
<td>decreased</td>
</tr>
<tr>
<td>Idarubicin</td>
<td>decreased</td>
</tr>
<tr>
<td>Capecitabine</td>
<td>no change</td>
</tr>
<tr>
<td>5-fluorouracil</td>
<td>no change</td>
</tr>
<tr>
<td>Vinblastine</td>
<td>no change</td>
</tr>
<tr>
<td>Irinotecan</td>
<td>no change</td>
</tr>
<tr>
<td>Estramustine</td>
<td>no change</td>
</tr>
<tr>
<td>Fludarabine</td>
<td>increased</td>
</tr>
<tr>
<td>Tegafur</td>
<td>increased</td>
</tr>
</tbody>
</table>

7.3.2 The ability of bacteria to modulate the immune system which indirectly effects drug therapy

In addition to the direct effects that bacteria can have on drug efficacy, they can also have indirect effects. Common drugs administered to women to treat breast cancer include doxorubicin (or epirubicin), paclitaxel, gemcitabine and 5-fluorouracil (5-FU). While their primary mode of action is disruption of DNA synthesis or microtubule formation, their efficacy may also depend on the engagement of the immune system. For example, 5-FU
and gemcitabine have been shown to sensitize tumours to CD8\(^+\) T cell and NK cell mediated cytotoxicity and to reduce the levels of myeloid derived suppressor cells, thereby restoring CD8\(^+\) T cell and NK cell functionality (Ghiringhelli and Apetoh, 2015; Weir et al., 2011). Both paclitaxel and doxorubicin upregulate the cation-independent mannose-6-phosphate receptor on the surface of tumor cells, thereby increasing the efficiency of granzyme B mediated killing by cytotoxic T lymphocytes (CTL) and NK cells (Ramakrishnan et al., 2010). Doxorubicin has also been shown to mediate tumour infiltration of a subset of dendritic cells that are highly efficient at presenting tumour antigens to CTLs (Ma et al., 2013). This interaction between the immune system and chemotherapeutics for optimal tumour killing is further highlighted by the fact that (i) synergy has been observed in mouse models of solid tumours between chemotherapy and immunotherapy (Nowak et al., 2003); (ii) immune-deficient mice do not respond as well to certain chemotherapeutics compared to their immune-competent counterparts (Casares et al., 2005; Tesniere et al., 2010) and (iii) patients with genetic immune defects are less responsive to chemotherapy (Ghiringhelli et al., 2009; Tesniere et al., 2010). Therefore if a patient does not have a proper functioning immune system, then the engagement of the immune system by chemotherapeutics will not be fully exploited. This could partly explain why drug therapy is so variable between patients even within the same subtype and stage.

One of the most well studied and established factors that modulate the immune system is our microbiota (Hooper et al., 2012; Macpherson and Harris, 2004). The importance of the microbiota in cancer treatment was recently unraveled by two groups who showed that chemotherapy was not as effective in germ-free and antibiotic treated mice because
of inefficient priming of immune cells (Iida et al., 2013; Viaud et al., 2013). In the Viaud study, *Lactobacillus johnsonii* and *Enterococcus hirae* could reverse these drug defects, while *E. coli* could not. *E. coli* was the most abundant organism observed in our breast tumour specimens while *Lactobacillus* presence was negligible. Thus, antimicrobial or probiotic therapy to try and alter the breast microbiota might be worth considering in order to improve drug efficacy. However, what these strategies may entail would have to be considered in the context of the drug and the immune response elicited as opposite bacterial effects were observed in the Iida study—*Lactobacillus* (*L. murinum, L. intestinalis* and *L. fermentum*) presence was negatively correlated with efficacy while *E. coli* showed a positive correlation.

### 7.4 Future directions

A common thread throughout this thesis had been the idea that one’s microbiota can modulate health and disease. While it is informative to study bacterial composition and function once disease has developed, there are merits to examining how factors that shape one’s microbiome during infancy, such as breast feeding, can influence disease risk later in life.

In the milk study presented in Chapter 3, we showed that bacterial communities are highly variable amongst healthy women. One mother (Subject 32) had a milk microbiota composed of 72% *Enterobacteriaceae* and no *Bifidobacterium* nor *Lactobacillus* while in another mother (Subject 6), who gave birth vaginally and at term like subject 32, *Enterobacteriaceae* made up only 0.8% of her microbiota with *Bifidobacterium* (24.5%) and *Lactobacillus* (8%) dominating. The latter are generally associated with health whereas *Enterobacteriaceae* are more associated with disease at various body sites. An
important question that therefore arises from these results is whether there is a milk microbiota that is “healthier” than another or whether each unique microbiota is adapted for a particular host and is the best profile for that individual and her offspring.

To help address this idea, we had intended on administering human milk from subject 6 (high *Bifidobacterium*/Lactobacillus, low *Enterobacteriaceae*) to neonatal germ free mice and compare the outcome with milk from Subject 32 (low *Bifidobacterium*/Lactobacillus, high *Enterobacteriaceae*) in terms of immune development and function over time. A collaboration was established with Dr. Sven Pettersson of the Karolinska Institute and the studies were planned for his research site in Singapore. However, too few animals were available to provide us with enough meaningful data, so we did not go ahead with the study. Nevertheless, this remains a worthwhile experiment and if differences were to exist, animal models of breast cancer could help elucidate whether the different immune effects induced by different milk microbial communities would put the offspring at a higher risk of breast cancer development.

### 7.5 Concluding remarks

The studies conducted on the breast tissue and the breast milk microbiotas presented in this thesis raise important questions as to what bacterial profiles are considered “healthy” and what factors influence their composition. Longitudinal studies following lactating women over a span of a few months or even years would give us a better appreciation of the factors that shape the milk microbiota (i.e. diet, antibiotics, stress, probiotic intake) and how quickly the microbiota can rebound from these changes. If further studies show that milk and tissue microbial communities are similar, human milk could provide a non-
invasive way in which to study factors that shape the tissue microbiota. Future studies that monitor healthy subjects (from either those from whom we collect milk or tissue) for breast cancer development could provide more insight into the organisms that may drive disease progression. Once diagnosed with cancer, antimicrobial or probiotic therapy might be worth considering to try and alter the breast microbiota in order to improve the efficacy of chemotherapy. Overall, the results presented in this thesis represent a new line of thought that had not been considered before. Hopefully, the data will stimulate new areas of research that helps prevent breast cancer development or improves treatment strategies.
7.6 References


8 Appendix

8.1 Optimization of DNA extraction methods for breast milk and breast tissue

8.1.1 Breast milk

Two different DNA extraction protocols were compared to determine which one would generate higher yields of bacterial DNA. The methods followed were by Hunt; “Characterization of the diversity and temporal stability of bacterial communities in human milk” (Hunt et al., 2011) and Martin; “Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women” (Martín et al., 2007). A schematic of each method is presented in Figure 8-1. Some modifications to the above protocols were also compared. For Hunt’s protocol, 1ml and 2ml of milk were also tested as well as 0.5ml, 1ml and 2ml of milk that had been centrifuged at 20,000 x g for 10min (as per the Martin protocol). For the Martin protocol, DNA was also extracted from 2ml of milk pellet.

Three different breast milk samples were tested with the above methods. After extraction, the DNA was amplified using HDA primers, which amplify the V2-V3 hypervariable region of the 16S rRNA gene (Table 8-1). The PCR was carried out in a 50μl reaction containing 10μl of DNA template (or nuclease free water as a negative control), 2.5mM MgCl₂, 0.64mM of each primer, 5μl of 10X PCR Buffer (Invitrogen), 0.21mM dNTPs, 0.05U Taq Polymerase (Invitrogen) and 0.15μg/μl of bovine serum albumin. Thermal cycling was carried out in an Eppendorf Mastercyler using a step-down approach with an initial denaturation of 94°C for 2min followed by 9 cycles of 94°C for 45s, 61°C for 45s, 72°C for 45s and then another 14 cycles of 94°C for 45s, 61°C for 45s, 72°C for 45s. This
was followed by an elongation step at 72°C for 2 min. After PCR, samples were run on a 1.5% TAE agarose gel for 1 hr at 120 V, stained with 0.5 µg/ml ethidium bromide for 20 min and then viewed under UV light in an AlphaImager (Alpha Innotech Corporation).

8.1.2 Breast tissue

65 µl, 100 µl, 200 µl and 400 µl of homogenized tissue [0.4 g/ml], representing 26 mg, 40 mg, 80 mg and 160 mg respectively were used for extractions following the protocols outlined by Hunt and Martin (Figure 8-1), with the exception of Martin’s initial centrifugation step. These weights were chosen based on the recommendations outlined in the QIAamp DNA Mini Kit (max 25 mg for tissue) and the QIAamp DNA stool Mini Kit (180-220 mg of stool). In addition to the above protocols, a method used to detect bacteria in colonic mucosa of patients with IBD was also tested (Sproule-Willoughby et al., 2010) (Figure 8-2). Three different tissue samples were extracted with these 3 methods and amplified as per the milk samples.
Figure 8-1. Schematic of DNA extraction protocol for bacterial detection in breast milk

These published methods were also used for bacterial detection in breast tissue.
**Sproule-Willoughby protocol**

Based on QIAamp DNA Mini Kit

1. **80µl tissue**
2. **200µl lysozyme solution**
   - 37°C, 30min
3. **200µl Buffer ATL**
   - 20µl proteinase K
4. **Bead beading:**
   - 30s, 4800rpm x3
5. **56°C, 30min; 95°C, 10min**
6. **200µl Buffer AL**
7. **70°C, 10min**
8. **200µl ethanol**
9. **Samples added to Qiagen column and washed as per kit instructions**
10. **DNA eluted in 60µl Buffer AE**

**Figure 8-2. Schematic of DNA extraction for bacterial detection in breast tissue**

This protocol was also compared to the ones presented in Figure 8-1. lysozyme solution= 20mg/ml lysozyme (Sigma) in 20mM Tris-HCl pH 8.0, 2mM EDTA and 1.2% Triton X-100.
8.1.3 Chosen method for extractions

Based on band intensities of the amplicons after running the PCR on the gel, the Martin protocol using 2ml of milk and 400µl of tissue produced the strongest band, indicative of a higher yield. No bands were detected in the no-template negative control.

Numerous studies comparing various DNA extraction methods have shown that bead beading gives a better representation of bacterial diversity in mixed samples (Yuan et al., 2012). Bead beading can, however, shear DNA, causing lower DNA yields after amplification (Lever et al., 2015). To assess whether the bead beading time could be increased without losing DNA integrity, 30s of bead beading (Martin protocol) was compared to 60s, repeated twice. Since longer bead beading did not appear to change amplicon yields, this adjustment was made to the Martin protocol and used for both milk and tissue extractions.

8.2 Denaturing gradient gel electrophoresis

DNA extracted from breast tissue was analyzed by DGGE before being sent for HTS, to get an idea of how diverse the bacterial communities were in the samples. The U968f-GC/L1401r primer pair, which amplifies the V6-V8 region of the 16S rRNA gene, was used (Table 8-1). The PCR was carried out in a 50µl reaction containing 10µl of DNA template (or nuclease free water as a negative control), 3.0mM MgCl₂, 0.2mM of each primer, 5µl of 10X PCR Buffer (Invitrogen), 0.2mM dNTPs, 0.05U Taq Polymerase (Invitrogen) and 0.15µg/µl of bovine serum albumin. Thermal cycling was carried out in an Eppendorf Mastercyler under the following conditions: Initial denaturation at 95°C for
2 min followed by 35 cycles of 95°C for 30 s, annealing at 56°C for 40 s, elongation at 72°C for 60 s and then a final elongation step of 72°C for 5 min.

Following amplification the products were then run on a 30-70% denaturing gradient gel. Preparation of gradients and gel electrophoresis were carried out using the D-code Universal Detection system (Bio-rad) following the manufacturers protocol. A 30% and 70% denaturing solution were made up using the recipe outlined in Table 8-2. 50 µl of TEMED [N,N,N',N'-tetramethylethlenediamine; Sigma] and 90 µl of 10% ammonium persulphate (Bio-Rad) were added to each solution to promote polymerization. Gels were allowed to polymerize overnight. Either 20 µl (Figure 8-3A) or 75 µl (Figure 8-3B) of PCR product were mixed with 2× loading buffer (0.25 ml of bromophenol blue [2%, Sigma], 0.25 ml of xylene cyanol [2%, Sigma], 7 ml of glycerol, and 2.5 ml of ddH₂O) and loaded into the wells. Gels were run at 130 V in 1× TAE at a constant temperature of 60°C until the second dye front (xylene cyanol) approached the end of the gel. After electrophoresis, gels were removed, allowed to cool before the removal of the glass-plate sandwich, stained for 20 min in 5 µg/ml ethidium bromide, and destained for 10 min in 1× TAE. Gels were viewed under UV light in an AlphalImager (Alpha Innotech Corporation).
Table 8-1. Primers used in the studies discussed in the appendix

Bolded sequence refers to the GC clamp. The qPCR primers target genes from *L. rhamnosus* GR-1. *pfk= 6* phosphofructokinase. ¹(Burton and Reid, 2002), ²(Martín et al., 2007).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDA-1</td>
<td>5’ ACTCCTACGGGAGGCAGCAG 3’</td>
<td>V2-V3 region 16S rRNA gene¹</td>
</tr>
<tr>
<td>HDA-2</td>
<td>5’ GTATTACCGCGGTGGCTGCTGCA 3’</td>
<td></td>
</tr>
<tr>
<td>U968F-GC</td>
<td>5’CGCCCCGCGCGGGCCCGGGGGGGGGGGGGGGAAAACGCGAAGACCTTAC 3’</td>
<td>V6-V8 region 16S rRNA gene² (DGGE)</td>
</tr>
<tr>
<td>L1401r</td>
<td>5’ CGGTCGTTGTCAAGACCCA 3’</td>
<td></td>
</tr>
<tr>
<td>orf_2365-F</td>
<td>5’ CTITGGACATTCGCTTGGAG 3’</td>
<td>glucosylceramidase (qPCR)</td>
</tr>
<tr>
<td>orf_2365-R</td>
<td>5’ CTITGGACATTCGCTTGGAG 3’</td>
<td></td>
</tr>
<tr>
<td>orf_1121-F</td>
<td>5’ CTAAATGAAACAGGCAATCCAA 3’</td>
<td>beta-galactosidase (qPCR)</td>
</tr>
<tr>
<td>orf_1121-R</td>
<td>5’ TCAAGACTGCGAACAAGG 3’</td>
<td></td>
</tr>
<tr>
<td>orf_817-F</td>
<td>5’ CATTATCCCCGACACGA 3’</td>
<td>pkf (qPCR-housekeeping gene)</td>
</tr>
<tr>
<td>orf_817-R</td>
<td>5’ GAAATCACCATAAGGCAGCAA 3’</td>
<td></td>
</tr>
</tbody>
</table>

Table 8-2. Recipe for making denaturing polyacrylamide gel solutions

<table>
<thead>
<tr>
<th></th>
<th>30%</th>
<th>70%</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% acrylamide/bis-</td>
<td>5ml</td>
<td>5ml</td>
</tr>
<tr>
<td>acrylamide solution (Bio-rad)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50X TAE</td>
<td>0.5ml</td>
<td>0.5ml</td>
</tr>
<tr>
<td>Urea</td>
<td>3.15g</td>
<td>7.35g</td>
</tr>
<tr>
<td>Formamide</td>
<td>3ml</td>
<td>7ml</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>16.5ml</td>
<td>12.5ml</td>
</tr>
</tbody>
</table>
Figure 8-3. Denaturing gradient gel electrophoresis analysis on breast tissue

DNA was extracted from normal-adjacent tissue from 9 women with tumours and the corresponding PBS environmental controls. DNA was also extracted from 9 healthy tissue samples from women who were disease free. The 16S rRNA gene was amplified using primers that span the V6-V8 hypervariable regions (U968/L1401). The PCR products were then run on a 30-70% denaturing gradient gel and stained with ethidium bromide. (A) DGGE gel with 20µl of PCR product loaded into the wells. “N” = normal adjacent tissue and “E”= PBS environmental controls. (B) DGGE gel with 75µl of PCR product loaded into the wells. With the exception of two faint bands in 17E, no bands were detected in any of the other environmental controls.
A

B

Normal-adjacent

Healthy controls
8.3 Histology

Histology was performed on the breast tissue samples to determine where bacteria localize within the mammary glands. Upon collection, a small piece of fresh tissue was placed in a cryomold filled with OCT (Tissue-Tek), snap frozen in liquid nitrogen and then stored at -80°C. Frozen tissue samples were cut into 5µm sections using a Leica CM 1850 cryostat set at an internal temperature of -30°C. Sections were captured onto Superfrost™ Plus microscope slides (FischerBrand), dried overnight at room temperature and then stored at -80°C until use.

8.3.1 Simple stain

Frozen slides were taken out of the -80°C freezer, allowed to warm up to room temperature, and then heat fixed by passing through the flame of a Bunsen burner. Slides were then covered with 1% methylene blue stain for 1min and excess stain washed off with ddH₂O. Slides were then viewed with a Zeiss AxioSkop microscope using both the 40X and 100X objectives, under oil immersion. As a positive control, an overnight culture of Lactobacillus rhamnosus GR-1 was thinly spread over the slides before being heat fixed.

Unless distinct and large numbers of bacteria are present, as with the positive control, a simple stain is not helpful in detecting bacteria in breast tissue (Figure 8-4). However, this simple stain did show that we had good quality slides with breast architecture intact.
Figure 8-4. Methylene blue stain of breast tissue cryosections

Cryosections of 5µm breast tissue samples were stained with 1% methylene blue solution for 1 min and viewed under oil immersion at 400X magnification. (A) Image of normal adjacent breast tissue from a woman with stage 1 invasive breast cancer. Image is representative of 3 different clinical samples. (B) Image of tissue section with an overnight culture of *Lactobacillus rhamnosus* GR-1 spread across the top, to act as a positive control.
8.3.2 Immunohistochemistry

We next tried to detect bacteria using an antibody specific for peptidoglycan, a molecule present in all bacterial cell walls. This antibody was used by a group at McMaster University to detect bacteria present in the human brain (Branton et al., 2013). First, in order to determine the versatility of the antibody in detecting different bacteria, immunohistochemistry was performed on pure cultures of Bacillus cereus, Staphylococcus epidermidis and L. rhamnosus GR-1.

Overnight cultures of each organism were placed on two microscope slides, left to dry and then heat fixed. Endogenous peroxidases were inhibited by covering the slide with a 0.3% H₂O₂ solution for 20min, which was then washed with TBS for 5min. One slide received acid treatment and the other did not. According to the manufacturer, pre-treatment with a strong acid for Gram-positive bacteria (or a detergent, for Gram-negative bacteria), may be needed to expose the epitopes. For those receiving acid, 25mM of HCl was added to the slides for 5min and then washed 3x with TBS. Slides were then blocked with 5% goat serum for 1 hour. Afterwards, a 1/100 dilution of the primary antibody (mouse anti-bacterial peptidoglycan monoclonal antibody, EMD Millipore) was added to the slides and incubated overnight at 4°C. After washing, the slides were incubated at a 1/500 dilution for 45min with the secondary antibody (HRP conjugated goat anti-mouse IgG (H+L), Jackson Laboratories). The slides were then developed in the dark with AEC (3-amino-9-ethylcarbazole) stain (Table 8-3) for 30min and the reaction stopped with ddH₂O. Slides were viewed with a Zeiss AxioSkop microscope under oil immersion with the 100X objective.
The antibody only bound to *Staphylococcus* (Figure 8-5) and neither *Lactobacillus* nor *Bacillus*, regardless of acid treatment. While different acid concentrations could have been tested, addition of 10mM or 25mM of HCl to the tissue slides ruined the architecture of the breast. Due to the amount of time and money needed for optimization and the fact that this was a minor side project, we decided not to pursue this idea further.

**Table 8-3. Recipe for AEC stain for immunohistochemistry**

<table>
<thead>
<tr>
<th>AEC stain</th>
<th>*Acetate buffer (100ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16mg AEC powder (Sigma)</td>
<td>1.36g sodium acetate trihydrate</td>
</tr>
<tr>
<td>4ml DMF</td>
<td>0.6ml acetic acid</td>
</tr>
<tr>
<td>56ml acetate buffer*</td>
<td>make up to 70ml with ddH₂O</td>
</tr>
<tr>
<td>40µl 30% H₂O₂</td>
<td>adjust pH to 5.2</td>
</tr>
<tr>
<td>Dissolve DMF and AEC together, then add acetate buffer, filter into designated bottle, add in H₂O₂</td>
<td>top up to 100ml</td>
</tr>
</tbody>
</table>

Figure 8-5. Immuno-staining of breast tissue cryosections.

An overnight culture of *Staphylococcus* was immunolabeled with anti-peptidoglycan antibody and developed with AEC. Cultures were treated either with 25mM HCl or without (image displayed). Positive staining was observed with and without acid treatment.
8.4 Examining the ability of the probiotic *Lactobacillus rhamnosus* GR-1 to break down lactosylceramide into ceramide

8.4.1 Background

Lactosylceramide is a lipid molecule ubiquitous on all eukaryotic cells and is a major component in human and bovine milk (Newburg and Chaturvedi, 1992). Surprisingly, considering the health benefits of milk, lactosylceramide has properties that could promote cancer development, such as the ability to stimulate reactive oxygen species from various types of cells (Arai et al., 1998), the ability to promote VEGF (vascular endothelial growth factor) induced angiogenesis (Kolmakova et al., 2009) and the ability to upregulate the expression of PECAM-1 (platelet/endothelial cell adhesion molecule) (Gong et al., 2004), which is associated with a poor prognosis in breast cancer patients (Martin et al., 2005). Lactosylceramide, however, can be broken down into ceramide, a pro-apoptotic signaling molecule that suppresses tumour growth (Obeid et al., 1993; Ogretmen and Hannun, 2004). Increased ceramide levels also make multi drug resistant tumours sensitive to therapy (Liu et al., 2004, 2001). Certain strains of probiotic bacteria have been shown to possess the ability to increase endogenous ceramide levels on tumour cells, leading to cellular death (Angulo et al., 2011; Di Marzio et al., 2001).

Lactosylceramide is first broken down into glucosylceramide by beta-galactosidase (EC 3.2.1.23), with this intermediate then broken down to ceramide by glucosylceramidase (EC 3.2.1.45). An RNA-seq experiment performed by a former graduate student in the lab, examining the effects of different growth conditions on *Lactobacillus rhamnosus* GR-1, revealed higher abundance of mRNA transcripts for glucosylceramidase when GR-1 was grown to stationary phase in milk compared to broth.
Considering the importance of ceramide in breast cancer risk reduction, I wanted to explore whether GR-1 did have the ability to break down lactosylceramide to ceramide and whether the ceramide generated could inhibit breast cancer cell proliferation in vitro. First, a quantitative real-time PCR (qPCR) experiment was performed to confirm the RNA-seq results and to determine at what growth phase these enzymes were most active. The methods for RNA extraction, primer design and qPCR are outlined below.

8.4.2 RNA isolation from bacterial cultures grown in milk or broth

A fresh overnight culture of *L. rhamnosus* GR-1 was prepared in 12ml de Man-Rogosa-Sharpe (MRS) media (Difco Laboratories). The culture was diluted in a 96-well plate and scanned using a Multiskan Ascent (Thermo Scientific) to select a dilution with an absorbance of approximately 1.09 at 600nm. A culture of this dilution was made from the fresh overnight culture in sterile PBS. Next 3.3ml of the diluted culture was transferred into 2 sterile culture tubes and centrifuged for 4 minutes at 12000 x g. The supernatant was discarded, the pellet washed twice with sterile PBS, and then re-pelleted for 4 minutes at 12000 x g. Following the washes, the pellet from one tube was resuspended in 3.3ml skim milk that had been pre-warmed to 37°C and the other in 3.3ml MRS. Milk was prepared by suspending skim milk powder (Sigma) in deionized water to a concentration of 10% w/v followed by a 5 minute autoclave cycle at 121°C. Next, 2.5ml of the milk or MRS suspension was added to 47.5ml pre-warmed skim milk or 47.5ml MRS respectively in a 50ml conical tube, mixed and loosely capped. Immediately, the culture was placed in the interchange of an anaerobic chamber and following de-oxygenation, placed into the chamber’s incubator at 37°C. Milk cultures were grown for
10 or 22 hours, corresponding to mid-log and stationary phases of growth respectively. MRS cultures were grown for either 5 or 14 hours, corresponding to mid-log and stationary phases of growth respectively.

8.4.2.1 RNA stabilization

Upon completion of the incubation period, 20ml of milk or MRS culture was immediately added to 40ml of Bacterial RNAProtect (Qiagen) containing 50mg/ml Rifampicin (Sigma) inside the anaerobic chamber. This mixture was thoroughly vortexed and incubated at room temperature for 10min to allow for full RNA stabilization and then removed from the chamber. Next, the cultures were centrifuged at 6000 \( \times \) g at 4\( ^\circ \)C for 20min. The supernatant was then discarded and the pellet immediately frozen in liquid nitrogen and stored at -80\( ^\circ \)C until extraction.

8.4.2.2 Optimized \( L.rhamnosus \) GR-1 RNA extraction

The frozen pellet was immediately suspended and vortexed in 20ml of lysis solution made from 20ml DEPC treated water; 20mg/ml lysozyme (Sigma), 50U/ml mutanolysin (Sigma) and 100 \( \mu \)g/ml rifampicin (Sigma). The lysis mixture was then incubated at 37\( ^\circ \)C for 45min with vortexing every 5min. After the incubation period, the mixture was centrifuged at 6000 \( \times \) g for 20min at 4\( ^\circ \)C. The supernatant was discarded and the pellet suspended in 20ml TRIzol reagent (Invitrogen) and vortexed for 2min. The sample was then passed through an RNase-free 21 gauge needle and syringe to aid in homogenization of the pellet followed by incubation at room temperature for 10min. It was then centrifuged at 12000 \( \times \) g for 10min to remove excess contaminant proteins and the supernatant then transferred to a fresh tube. Next, 4ml chloroform was added and the sample mixed for 15s by vortexing, incubated at room temperature for 5min and then
centrifuged at 12000 \text{x} g for 20min. Next 10ml of the upper aqueous phase was
transferred into 500µl aliquots into fresh RNase-free microcentrifuge tubes. RNA was
precipitated with the addition of 5µg nuclease-free glycogen (Ambion), 50µl 3M sodium
acetate and 550µl isopropanol, inverted 40 times and allowed to precipitate at room
temperature for 10min. RNA was then pelleted by centrifugation at 12000 \text{x} g for 20min
at 4°C. The resulting pellet was rinsed twice by adding 70% ethanol (non-denatured)
followed by vortexing and centrifugation at 12000 \text{x} g for 5min. The pellet was then air
dried for 10min on ice and resuspended in 100µl RNase-free water (Qiagen). Samples
were stored at -80°C until further analysis.

8.4.2.3 Total RNA cleanup and DNA digest

RNA cleanup was carried out using the RNeasy mini-columns (Qiagen) according to the
manufacturer’s instructions. The optimal on-column DNA digest was applied using
RNase-free DNaseI (Qiagen). Clean, DNA-free, total RNA was eluted from the column
with 2 aliquots of 50µl RNase-free water. The RNA was then concentrated via ethanol
precipitation using 5µg glycogen, 2.5 volumes of 100% ethanol and 0.1 volumes of 3M
sodium acetate, incubated overnight at -80°C and centrifuged at 12000 \text{x} g for 30min.
After discarding the supernatant, the pellet was rinsed twice with 75% ethanol, air dried
and then resuspended in 15µl RNase-free water.

8.4.2.4 Analysis of RNA quality

RNA quality was analyzed by loading 1µl total RNA solution on a 1% agarose gel made
with tris-borate (TBE) buffer. Gels were run at 6V/cm for 1 hour, stained with 0.5µg/ml
ethidium bromide in TBE buffer for 20min and then viewed under UV light in an
AlphaImager (Alpha Innotech Corporation). Quality was determined by the presence of crisp rRNA bands. High quality samples were then analyzed using a nanodrop (Thermo Scientific) to ensure a 260nm/280nm ratio between 1.8-2.0 and a 260nm/230nm ratio above 1.4. Samples meeting these criteria were then used for qPCR.

8.4.3 qPCR primer design

Primers were designed against beta-galactosidase (EC 3.2.1.23), glucosylceramidase (EC 3.2.1.45) and the housekeeping gene 6-phosphofructokinase (pfk) using the GR-1 genome sequence generated in our lab. Primer3plus (http://biotools.umassmed.edu/cgi-bin/primer3plus/primer3plus.cgi) was used to design the primers, taking into account the following considerations:

- avoid repeats of G or C longer than 3 bases
- have G and C on the ends of the primers as this helps in specificity
- length of primer should be between 18-22 nucleotides long
- avoid nucleotide runs of more than 4bp (so 4 or more of the same nucleotide beside each other)
- GC content should be between 50-60%
- primer annealing temperature should be between 55-60C
- melting temperature of primers should not differ by more than 2C
- the amplicon should have a GC content of 40-60%
- amplicon length for qPCR should be approximately 100bp

Once adequate primers were chosen, the primer sequences were inputted into Beacon Designer (www.premierbiosoft.com/molecular_beacons/index.html) to verify that the primer pairs would not form hairpins or cross-dimerize. mfold (http://unafold.rna.albany.edu/?q=mfold) was used to make sure that the chosen amplicon would not form secondary structures above the annealing temperature. Finally
blastn (http://blast.ncbi.nlm.nih.gov) was used to verify the specificity of the primers and to ensure that it would only bind the gene of interest. The primer sequences used are summarized in Table 8-1.

### 8.4.4 Quantitative real time PCR

Synthesis of cDNA from the RNA isolated above was performed using random hexamers and SuperScript® II Reverse Transcriptase (Invitrogen) according to the manufacturer’s instructions. Amplification of cDNA by qPCR was performed using the iQ™ SYBR® Green Supermix (Bio-Rad) using 10μl iQ SyberGreen, 1μl of forward primer (10μM), 1μl of reverse primer (10μM), 1μl of cDNA (or DNA standard) and 7μl of ddH₂O. Reactions were run in a Rotor-Gene 6000 thermocycler (Corbett) under the following program: an initial melting ramp from 72°C to 95°C, followed by 35 cycles of 95°C for 45s, 60°C for 45s, and 72°C for 60s, ending with a hold at 95°C for 5 min. Data was analyzed using the Rotor-Gene 6000 series software (version 1.7; Corbett). A series of standards was created to quantify gene concentrations during data analysis. Total DNA was isolated from an overnight culture of GR-1 using InstaGene matrix (Bio-Rad). PCR was carried out in 50-μl reaction mixtures consisting of the following components: 1× PCR buffer (Invitrogen), 2.5mM MgCl₂ (Invitrogen), 0.2mM dNTP mix, 0.5μM each primer pair used for qPCR (Table 8-1), 0.05U Taq polymerase (Invitrogen), 2μl template, and ddH₂O to reach 50μl. Amplification was performed in a Mastercycler (Eppendorf) using the following program: 95°C for 1 min, followed by 30 cycles of 95°C for 45 s, 60°C for 45 s, and 72°C for 60s with a final elongation step of 72°C for 2 min. Products were viewed via electrophoresis on a 1% agarose gel using 1× TBE, stained with ethidium bromide, and viewed under UV light in an Alphalmager (Alpha Innotech Corporation). The
concentration of DNA in each sample was determined using a nanodrop, and a series of standards with concentrations ranging from 5,000 pg/µl to $5 \times 10^{-3}$ pg/µl were made. These standards were used as templates in the qPCR reactions to construct standard curves with known concentrations of DNA. Genes of interest were normalized to the housekeeping gene 6-phosphofructokinase (pfk) (EC 2.7.1.11) before determining the relative fold change compared to log phase growth in broth.

The qPCR confirmed the results of the RNA-seq experiment in that glucosylceramidase mRNA transcripts are more abundant at stationary phase when GR-1 is grown in milk compared to broth. Transcripts were also more abundant at log phase when grown in milk compared to broth. Beta-galactosidase mRNA transcripts were also significantly higher when grown in milk compared to broth at log phase but not so when grown to stationary phase. These results are displayed in Figure 8-6.
Figure 8-6. qPCR results showing the relative fold change of the genes involved in the breakdown of lactosylceramide to ceramide

Concentrations of each gene were determined using a standard curve and then normalized to the housekeeping gene, pfk. Abundances are displayed as fold changes compared to log phase growth in broth ("log_broth"). Log phase= 5hr broth and 10 hour milk and stationary phase=14 hour broth and 22hr milk. For all variables with the same letter ("a", "b", "c", "d", "e", "f", "h") the difference between the means is statistically significant following the Bonferroni multiple test correction (p<0.05). Bars represent the mean +/- standard deviation of 3 separate experiments.

8.4.5 Verifying the in vitro ability of GR-1 to degrade lactosylceramide

As qPCR confirmed that the genes involved in lactosylceramide breakdown to ceramide were upregulated when grown in milk compared to broth, the next step was to show that GR-1 did have the ability to breakdown lactosylceramide. To test this, GR-1 was grown to log and stationary phase of growth in nutrient broth supplemented with 17µM of purified lactosylceramide (Matreya LLC), the concentration present in bovine milk (Newburg and Chaturvedi, 1992).
To test the amount of lactosylceramide remaining over time, an ELISA was performed using an antibody specific against lactosylceramide. While HPLC is the common method to detect sphingolipids, I opted to try an ELISA. Unfortunately, the antibody purchased did not recognize sphingolipids and this method had to be abandoned. In the future, HPLC or LC-MS could be used to measure both lactosylceramide breakdown and the production of ceramide by GR-1. Bacterial supernatants containing ceramide could then be used in *in vitro* experiments to test the ability of the ceramide generated to inhibit breast cancer cell proliferation.

### 8.5 References


8.6 Permission to reprint published manuscripts

Copyright for Microbiome/BioMed Central

BioMed Central Open Access Charter

Every peer-reviewed research article appearing in any journal published by BioMed Central is 'open access', meaning that:

1. The article is universally and freely accessible via the Internet, in an easily readable format and deposited immediately upon publication, without embargo, in an agreed format - current preference is XML with a declared DTD - in at least one widely and internationally recognized open access repository (such as PubMed Central).

2. The author(s) or copyright owner(s) irrevocably grant(s) to any third party, in advance and in perpetuity, the right to use, reproduce or disseminate the research article in its entirety or in part, in any format or medium, provided that no substantive errors are introduced in the process, proper attribution of authorship and correct citation details are given, and that the bibliographic details are not changed. If the article is reproduced or disseminated in part, this must be clearly and unequivocally indicated.

BioMed Central is committed permanently to maintaining this open access publishing policy, retrospectively and prospectively, in all eventualities, including any future changes in ownership.
Copyright for Applied and Environmental Microbiology/American Society of Microbiology

Dear Camilla Urbaniak,

Please note that as stated in the copyright form signed by the article’s corresponding author, an author of an ASM journal article retains the right to republish discrete portions of that article in any other publication (including print, CD-ROM, and other electronic formats) for which he/she is author or editor, provided that proper credit is given to the original ASM publication. An ASM author also retains the right to reuse the full article in his/her dissertation or thesis. For more information, please see the Instructions to Authors section on copyright: http://aac.asm.org/site/misc/journal-ita_edi.xhtml#06.

ASM no longer generates licensing agreements for this sort of permission request through RightsLink due to the fees that RightsLink has levied on publishers for free permission licenses. The language in the copyright form, in the Statement of Author Rights, and in the Instructions to Authors should be sufficient permission for authors to confidently reuse their work in the above-mentioned formats.

Please contact us if you have any questions.

Thank you.

ASM Journals

journals@asmusa.org
8.7 Ethics approval

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Gregor Reid
File Number: 100388
Review Level: Delegated
Approved Local Adult Participants: 150
Approved Local Minor Participants: 0
Protocol Title: Investigating the bacterial microbiota and nitrosamine content in breast, breast discharge and breast milk (REB 181333)
Department & Institution: Schulich School of Medicine and Dentistry/Microbiology & Immunology, Western University
Sponsor: Ethics Approval Date: September 04, 2012 Expiry Date: August 31, 2014
Documents Reviewed & Approved & Documents Received for information:

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Comments</th>
<th>Version Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in number of local Participants</td>
<td>Increased to 150.</td>
<td></td>
</tr>
<tr>
<td>Revised Western University Protocol</td>
<td>Revised study objectives and methodology</td>
<td></td>
</tr>
<tr>
<td>Revised Letter of Information &amp; Consent</td>
<td>Breast Tissue, Breast Milk and Breast Discharge</td>
<td>2012/07/20</td>
</tr>
</tbody>
</table>

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines, and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Ethics Officer to Contact for Further Information

This is an official document. Please retain the original in your files.
Western University Health Science Research Ethics Board
HSREB Annual Continuing Ethics Approval Notice

Date: July 13, 2015
Principal Investigator: Dr. Gregor Reid
Department & Institution: Schulich School of Medicine and Dentistry/Microbiology & Immunology, Western University

Review Type: Full Board
HSREB File Number: 100985
Study Title: Investigating the bacterial microbiota and nitrosamine content in breast, breast discharge and breast milk (REB 18133).

HSREB Renewal Due Date & HSREB Expiry Date:
Renewal Due -2016/07/31
Expiry Date -2016/08/31

The Western University Health Science Research Ethics Board (HSREB) has reviewed the Continuing Ethics Review (CER) Form and is re-issuing approval for the above noted study.

The Western University HSREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guideline for Good Clinical Practice (ICH E6 R1), the Ontario Freedom of Information and Protection of Privacy Act (FIPPA, 1990), the Ontario Personal Health Information Protection Act (PHIPA, 2004), Part 4 of the Natural Health Product Regulations, Health Canada Medical Device Regulations and Part C, Division 5, of the Food and Drug Regulations of Health Canada.

Members of the HSREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.
# Curriculum Vitae

**Name:** Camilla Teresa Urbaniak

**Education:**

- University of Western Ontario  
  London, Ontario, Canada  
  Ph.D, Microbiology and Immunology  
  Supervisor: Dr Gregor Reid  
  2011-2016

- McMaster University  
  Hamilton, Ontario, Canada  
  M.Sc., Medical Sciences  
  Supervisor: Dr. Andrew Macpherson  
  2005-2008

- University of Toronto  
  Toronto, Ontario, Canada  
  Hons B.Sc.,  
  2001-2005

**Scholarships and Awards**

- Translational Breast Cancer Studentship  
  Doctoral fellowship (*Breast Cancer Society of Canada*)  
  2011-2016

- CIHR Strategic Training Program in Cancer Research and Technology Transfer (CaRTT)  
  Doctoral fellowship (*London Regional Cancer Program*)  
  2012-2013

- Western Graduate Research Award  
  Tuition scholarship (*University of Western Ontario*)  
  2011-2015

- Graduate Research Scholarship in Medical Sciences  
  M.Sc. fellowship (*McMaster University*)  
  2005-2007

- The Duke of Edinburgh’s Award  
  2008

- Fr. Robert Madden Leadership Award  
  University of Toronto Graduation Award  
  2005

- University of Toronto Entrance Scholarship  
  2001

**Teaching Assistantships**

- “Advanced Medical Sciences Laboratory”, Department of Medical Sciences, University of Western Ontario (Jan 2014-Dec 2015)
“Biology of Prokaryotes”, Department of Microbiology and Immunology, University of Western Ontario (Sept 2012-Dec 2015)

“Human Biochemistry”, Faculty of Health Sciences, McMaster University (Sept 2005-May 2007)

**Relevant Work Experience**

Clinical Research Coordinator
Department of Pediatric Gastroenterology, McMaster University

2008-2011

**Invited Speaker**

Women and their Microbes, 2nd annual conference
Amsterdam, Netherlands

June 2015

Microbiome Program, Mayo Clinic, Rochester, Minneasota, USA

“The microbiota of breast milk and tissue: what could changes mean for health.”

Apr 2015

8th International Symposium on the breast, Dr. Susan Love Research Foundation, Santa Monica, USA

“Microbes in the breast-important implications”

Feb 2015

Translational Breast Cancer Research Meeting;
Research Success and Prospects for the Future
London, Canada

“Microbes and Health: Could bacteria play a role in breast cancer development or treatment. A translational approach to a growing problem.”

Jan 2014

South Western Ontario Lactation Consultants Meeting
London, Canada

“The potential role of bacteria in breast cancer development.”

Sept 2012

Crohn’s and Colitis Foundation of Canada, Education Event
Waterloo, Canada

“The interactions between the intestinal microbiota and the immune system.”

May 2010

**Conferences** (*denotes oral presentation)

International Union of Microbiological Societies, Montreal, Canada (July 2014)*

Canadian Society of Microbiology Student Oral Symposium, Montreal, Canada (July 2014)*

Rowett-INRA Conference “Microbiology: from sequence to function”, Aberdeen, Scotland (June 2014)
Infection and Immunity Research Forum, London, Canada (Nov 2013)


Oncology Education and Research Day, London, Canada (June 2013)

London Health Research Day, London, Canada (March 2013)

Infection and Immunity Research Forum, London, Canada (Nov 2012)

21st annual Canadian Society for Immunology Conference, Mont Tremblant, Quebec (April 2008)

Publications

Camilla Urbaniak, Michelle Angelini, Gregory Gloor and Gregor Reid. Human milk microbiota profiles in relation to birthing method, gestation and infant gender. Microbiome. 2016, 4:1


Camilla Urbaniak and Gregor Reid. How the microbiota might affect women during long Spaceflight missions. Journal of Women’s Health (to be published March 2016)

Submitted manuscripts

Camilla Urbaniak, Gregory Gloor, Muriel Brackstone, Leslie Scott, Mark Tangney and Gregor Reid. The microbiota of breast tissue and its association with breast cancer (mBio)