8-6-2016

Investigating the Role of Anaerobic Bacteria in Colorectal Carcinoma

Oluwadara Coker
ooc9075@rit.edu

R·I·T

Investigating the Role of Anaerobic Bacteria in Colorectal Carcinoma

Oluwadara Coker

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Bioinformatics

Thomas H. Gosnell School of Life Sciences
College of Science

Rochester Institute of Technology
Rochester, NY 14623-5603
August 6th, 2016

ROCHESTER INSTITUTE OF TECHNOLOGY
THOMAS H. GOSNELL SCHOOL OF LIFE SCIENCES
BIOINFORMATICS PROGRAM

To: Head, Thomas H. Gosnell of Life Sciences

The undersigned state that Oluwadara Coker, a candidate for the Master of Science degree in Bioinformatics, has submitted his thesis and has satisfactorily defended it.

This completes the requirements for the Master of Science degree in Bioinformatics at Rochester Institute of Technology.

Committee Approval:

Robert C. Osgood, Ph.D. Date
Head of Committee, Thesis Advisor

Gary R. Skuse, Ph.D. Date
Committee Member

Maureen Ferran, Ph.D. Date
Committee Member

Table of Contents

List of Figures

ACKNOWLEDGEMENTS

I thank God, first and foremost, for His guidance, wisdom, and strength throughout my journey on the hard lessons learned and conquered challenges of my life and academic career.

I want to extend my heartfelt gratitude to my parents for their continual advice and wisdom during tough times. I thank them for their support, perpetual love and understanding. I also thank my siblings, Ebun and Anu, for cheering and encouraging me along the way. I thank my parents once more for the sacrifices they made for us, told and untold, as they continued to guide us whenever necessary. Your sacrifices and values have made me into the individual I am today and I can only hope to do right by you both as I continue onward. I thank my Auntie Funmi and Uncle Robert for their eternal support, care, and love as they acted as my anchors, grounding me in the unwavering faith that I can do anything I set my mind to do. I thank them for their invaluable wisdom and counsel on the numerous discussions we had over the course of my life. I wish to thank the rest of my family for showering me with their encouragement and love throughout this journey. I love you all very much.

I want to also give my thanks to Dr. Robert Osgood, my advisor, and for whom I have profound respect and admiration, for his continual support in not just my graduate career but also as I continue to learn the necessary skills to navigate through new challenges. I appreciated how he allowed me enough room to learn the things I needed to learn and the mistakes I made doing it but always ready to pull me back before I come to close the edge of the cliff. I appreciated the many times we debated about the principles and realities of scientific research, the hilarious situations that one could hypothetically end up in and how they would get out of it, and the general lessons of life. I appreciate you taking me on as a student of yours and am proud to have been your protégé.

I also wish to thank my other committee members, Dr. Gary Skuse and Dr. Maureen Ferran, for their extensive knowledge of tissue culture and cancer as well as their unique insights on the problems that I faced. I have immense respect for them and their advice on how to best approach certain aspects of the project and hope to continue to receive their input as I move forward in my career.

I would also like to thank the GSOLS and CHST faculty and support members who have done tremendous work in providing me the necessary resources I needed to complete the project. Particularly, I thank Ms. Faye Modeste and Ms. Kim Corbett for their tireless efforts to make sure the orders went through and the frequent checking to make sure that I had everything I needed. I also thank Ms. Amanda Dolan and Ms. Nicki Bruno for their efforts to make sure course registrations and orders worked out smoothly as well as Ms. Allison Healy for helping me get the cells I needed. I would also like to thank Mike whose help in preparing the materials and autoclaving reagents I appreciate very much.

I wish to also thank Dr. Hyla Sweet and her lab assistants, Lauren Hesse and Teresa Zgoda, for their help and advice in the numerous confocal sessions I had with them. Without your help, I am confident I would not have been able to complete a core part of this project. I look forward to receiving Dr. Sweet's advice in future confocal sessions.

I would like to extend my eternal gratitude to AccessComputing for funding the final stages of my project. Without these funds I would not have been able to obtain the last set of materials I needed to complete project.

Finally, I would like to thank my friends, whose encouragement, intellect, wit, care and comedic moments I treasure daily. Thank you, Pete, Dusty, Erin, Brandon, Matt, Nate, Daron, and many others.

ABSTRACT

Colorectal cancer (CRC), with multifactorial influences of genetic, molecular, inflammatory, and environmental factors, is a leading cause of cancer-related deaths. Due to the high diversity and concentration of microbes found in the human colon, the microbiome has become a prime suspect of being a biological contributor. This prompts the need to investigate the relationship between microorganisms and CRC. Recent studies have shown that *Fusobacterium nucleatum*, *Campylobacter showae*, *Leptrotrichia buccalis*, and *Selenomonas sputigena* are overrepresented in colon tumors. These anaerobes have known associations to the oral microbiome and cause infections. Here, these findings are further expanded in an experimental attempt to investigate a possible etiological relationship between the four microbes and CRC. This was done by examining bacterial biofilm impact on the viability of CRC. Crystal violet biofilm assays showed that each of the four anaerobes are capable of producing biofilms, which is a known contributor to disease and has been proven to alter host tissue microenvironment of the human colon. The images of bacterial biofilms grown in presence and absence of CRC cells were generated through confocal microscopy studies. Co-cultured bacterial and cellular formation was visualized in 2D and 3D perspectives using ImageJ and Icy, respectively. Comstat2 was used to quantitatively analyze 3D biofilm characteristics. The results showed that *L. buccalis* is a prime suspect of possessing key "driver" genes that not only encouraged the proliferation of CRC cells but also minimized cell deaths. Furthermore, the relationship between bacterial biofilms and CRC is not only statistically significant but suggests that one factor influences the other inasmuch as the opposite holds true.

INTRODUCTION

The History of Cancer: Past to Present

Siddhartha Mukherjee's *The Emperor of all Maladies: A Biography of Cancer* explains that the first discovery of cancer was made by the Egyptian physician Imhotep in 2625 BC, in which the Egyptian described the disease as "bulging masses on the breast" (Mukherjee, 2011, p.39) that is "…large, spreading, and hard" (Mukherjee, 2011, p.39). Scientists today know this disease as breast cancer. However, Mukherjee further explains that after this discovery, there was no mention of such cases until 440 BC. The records, written by Greek historian Herodotus, foretell a story of how the queen of Persia, Atossa, noticed "a bleeding lump in her breast" (Mukherjee, 2011. p.39) that is today described as inflammatory breast cancer where malignant cells invade the lymph glands of the breast, the infection made manifest as a red, swollen mass (Mukherjee, 2011).

However, despite first discoveries being recorded in mid-2000 BC, there is evidence that cancer existed before it was documented. In 1914, an excavation in southeastern Africa done by a team of archaeologists uncovered a jaw bone that had lymphoma, which is a type of cancer that initiates in cells whose origins comes from the body's immune system (American Cancer Society, 2016), and was dated from 4000 BC. These findings indicate that cancer has been around for a very long time. Yet, despite its ancient existence, its occurrence was notably rare throughout human history.

The author contributes this rarity to the fact that cancer is a disease dependent upon the age of individuals (Mukherjee, 2011). He explains that in ancient times, people simply did not live long enough for cancer to occur. For example, the statistics given by the Centers of Disease Control explain that 0.44% of women who are now 30 will get breast cancer sometime during the next 10 years (Centers for Disease Control and Prevention [CDC], 2015). For women who are sixty years old today, the chances of getting breast cancer within the next ten years increases to 3.46% (i.e., 3 or 4 out of every 100 women who are 60 years old will get breast cancer by the age of 70) (CDC, 2015). Mukherjee explains that human beings of ancient past were done in by other diseases long before cancer could emerge. The author goes on to explain that 19th-century physicians thought that cancer was caused by civilization, by the "rush and whirl of modern life,

which somehow incited pathological growth in the body" (Mukherjee, 2011, p.41). While the physicians were correct to link cancer to civilization, Mukherjee argues that cancer was not caused by civilization; instead, though the extension of human life, "civilization *unveiled* it" (Mukherjee, 2011, p.41).

Since then, cancer has expanded into many types – its nomenclature contingent on the *in situ* origin of diseased tissues. These cases have elicited traditional paradigms used to study cancer, such as surgery, biopsy, and autopsy techniques (Mukherjee, 2011). Mukherjee compares the diagnosis of cancer in the past to today's time, explaining that the death of a child with leukemia – a type of cancer that impacts cells originating in the bone marrow such as white blood cells (National Cancer Institute, 2016) – in the 1850s would have been diagnosed as an abscess or infection. However, with the introduction of techniques such as the mammography to detect breast cancer in its early stages and many others, the rate at which cancer has been recognized sharply increased (Mukherjee, 2011). In 1900, tuberculosis was the most common reason for mortality in America followed closely by pneumonia, diarrhea, and gastroenteritis. Cancer was placed seventh. By the early 1940s, cancer skyrocketed to the top of the list – second only to heart disease (Mukherjee, 2011). This was due not only to the development of treatments for other diseases but also because of the increasing awareness of cancer.

The rise of cancer as a dominant force that took the lives of many prompted the urgent need to understand its roots and causes. Early experiments accumulated the evidence that cancer resulted from fixed changes which broke off the cellular chains that held them down, allowing them to grow uncontrollably (Varmus, 1989). It was observed that daughter cancer cells resembled a phenotypic morphology that characterized the cancerous state its predecessor exhibited. Many of the experiments performed included the use of physical and chemical agents that acted as mutagens to induce cancerous growth (Pierotti, Sozzi, and Croce, 2003; Varmus, 1989) and epidemiological studies that linked an individual's lifestyle to cancer, suggesting that cancer could also rise through environmental factors (Varmus, 1989; Weissman, 1979). Together, they formed the basis that cancer is the result of alterations that occur in the DNA, or more specifically, of the structural and/or functional alterations that occur in specific genes whose job is to control the life cycle of cells (Yokota, 2000; Bernards and Weinberg, 2002).

These genes are classified into two key categories that have revolutionized our understanding of how cancer works at the molecular level: oncogenes and tumor suppressor genes (Fig. 1).

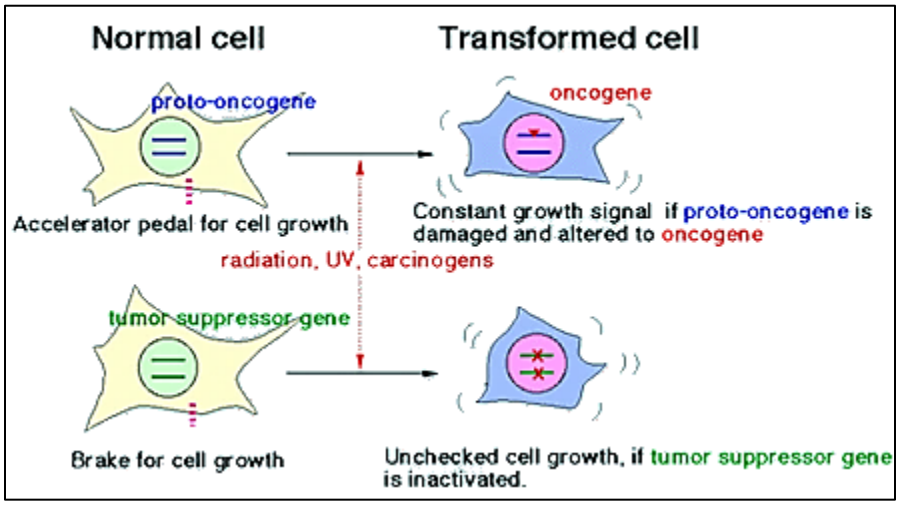

Figure 1. Impact of oncogenes and tumor suppressor genes on cell growth control. Proto-oncogenes perform normal cellular functions where a change in the gene sequence could lead to gross proliferation of cells. In contrast, tumor suppressor genes regulate the cell cycle, where its inactivation could disrupt cell behavior, allowing cells to proceed through the cell cycle virtually unchecked. As shown above, in a normal cell, proto-oncogenes act as the gas, promoting cell growth and division. Tumor suppressor genes are analogous to the brakes, preventing the cell from growing through division at inappropriate times. Adapted from http://www.rerf.jp/dept/radi/eng/oncog.html.

Although the term 'oncogene' was first coined by George Todaro and Robert Huebner in 1969 (Mukherjee, 2011), the work began in earnest with the virologist Francis Peyton Rous's discovery of the Rous sarcoma virus (RSV). RSV, found in excised fibrosarcomas (i.e., connective tissue tumors) (Lodish, Berk, and Zipursky *et al*., 2000), could induce solid tumors in chickens infected by the virus (Bister, 2015). These findings were then built upon with the discovery of RSV as a transforming principle by Peter Duesberg and Peter Vogt (1970) where the RNA of transforming derivatives of sarcoma viruses were shown to have two RNA subunits, denoted *a* and *b*, while the RNA of non-transforming derivatives were shown to have only the *b* subunit. This suggested that the *a* subunit present in transforming derivatives was responsible for inducing the oncogenic phenotype (Duesberg and Vogt, 1970). Their hypothesis was later confirmed in a temperature-based experiment by Steve Martin (1970), revealing the existence of

the viral *src* gene, which was capable of inducing oncogenic properties. The *src* gene, specifically named v-*src*, was defined to be the first oncogene discovered while its cellular homolog was given the name c-*src* (Martin, 2001). Many cellular oncogenes are, in fact, proto-oncogenes, or genes that perform normal cellular functions such as the promotion of cell growth and release of growth hormones, transcription factors, cell signaling molecules, etc. (Lodish, Berk, and Zipursky *et al*., 2000). Multiple studies conducted since have revealed numerous proto-oncogenes.

The discoveries of the 1970s continued over the next two decades, laying the groundwork for not just the findings of oncogenes but also to the other vital class of genes that defines cancer genetics: tumor suppressor genes. This discovery started with the observation that normal cells fused with tumor cells, termed hybrid cells, grown in animals did not display any malignant behavior (Ephrussi *et al*., 1969; Harris, 1988; Ho Park and Vogelstein, 2003). This observation prompted the hypothesis that the genetics of nonmalignant cells somehow suppressed the tumorigenicity of malignant cells. When propagating hybrid cells for longer periods of time in culture, the malignant phenotype returned. Karyotypic studies revealed hybrid cells that reverted back to malignant phenotype had lost certain chromosomes associated with normal cells and thus supported the hypothesis that tumorigenicity could be suppressed even in the presence of active oncogenic activities in hybrid cells (Ho Park and Vogelstein, 2003; Geiser *et al*., 1986).

While the somatic hybridization experiments did not directly discover tumor suppressor genes, they convinced cancer researchers that there were genes whose sole duty was to regulate pathological growth of malignant cells. Subsequent experiments began with two unique cases of retinoblastoma (RB) (Knudson, 1971), a type of cancer that initiates in the far backend of the retina (American Cancer Society, 2016). In both cases, the disease occurred in children; however, in one case, the children's parents also had the disease, while in another case, the parents of the diseased also had RB (Knudson, 1971).

These cases of RB prompted Knudson (1971) to provide an explanation for the development of the disease. He proposed that RB was the result of two mutational events in which biallelic inactivation of the RB gene, called *RB1,* was needed for the eye cancer to occur (Ho Park and Vogelstein, 2003). In expanding his hypothesis, Knudson (1971) explained that children whose parents had retinoblastoma only had one functional copy of the *RB1* gene

because they inherited a non-functional copy from the diseased parent (i.e., familial). Therefore, only one mutational event was needed in order for the disease to occur. In the second, the children were born with two functional copies where two mutational events were needed to initiate the disease (i.e., sporadic) (Knudson, 1971). Karyotypic analysis and cytogenetic experiments (Orye, Delbeke, Vandenabeele, 1974; Francke, 1976; Benedict *et al*., 1983) revealed that Knudson's proposal was consistent, prompting RB to be the first tumor suppressor gene discovered.

The discovery of oncogenes and tumor suppressor genes fundamentally altered the way cancer research is done today because they have cemented the idea that cancer is a genetic disease. Tumor suppressor genes act as regulators by controlling cell division, repairing DNA damage, or controlling apoptosis (i.e., programmed cell death) (American Cancer Society, 2016) where its biallelic inactivation events (i.e., both alleles must be mutated in order for the disease to occur) prompts the development of tumors. Oncogenes, its polar opposite, are the result of activating proto-oncogenes (American Cancer Society, 2016). When comparing the initiation of tumor suppressor genes to oncogenes, one sees that the latter is usually dominant while the former is usually recessive. There are, of course, exceptions to this hypothesis; the most notable one being the p53 tumor suppressor gene in which it was shown that the suppression of wild-type p53 gene is not necessary for tumorigenesis (Baker *et al*., 1990).

These two concepts have also paved the way for a relatively recent field that has today become a critical aspect of cancer biology: cancer epigenetics. Epigenetics, a term coined by Conrad Waddington (1939), was originally defined as "the casual interactions between genes and their products, which bring the phenotype into being" (Esteller, 2008, p.1148; Waddington, 1939). Following the works of Holliday (1987), the modern term has altered to mean changes in gene expression that are not the result of alterations in the DNA sequence and thus can also be passed down from parent to child.

In the field of cancer biology, epigenetics is generally studied to understand how external factors are used to influence disease progression from non-malignant to malignant cells such as how certain drugs can be used to activate tumor suppressor genes by targeting the DNA methylation and histone modification of proteins produced by these genes (Esteller, 2008). The reason why these observations are of immense interest is because, unlike mutations, DNA

methylation and histone modification are reversible changes. It is possible to reverse the changes that influence cancer cell behavior by re-expressing the DNA-methylated genes in cancer cell lines (Esteller, 2008). This is typically done through de-methylating agents that can restore the original functionality (Esteller, 2007; Herman and Baylin, 2003; Yoo and Jones, 2006).

Recent seminal works have described hallmark characteristics that all cancer types share due to being driven by the buildup of genetic mutations (Hanahan and Weinberg, 2000, 2011) as well as how epigenetic regulatory mechanisms are disrupted in cancer cells (Baylin and Jones, 2011; Sandoval and Esteller, 2012). Collectively, these works have firmly established cancer as a polygenic disease (Banwait and Bastola, 2016): it is the result of genetic abnormalities (i.e., mutations) that lead to the upregulation of oncogenes and the inactivation of tumor suppressor genes but can also be the result of epigenetic factors that may stem from activities that lie outside the genetic realm (e.g., drugs, viruses, etc.). Among the numerous cancer tissues that have been extensively studied since its inception, this paper examines the pathology of colorectal carcinoma.

The Pathology of Colorectal Carcinoma

Colorectal carcinoma (CRC) is the second leading cancer killer and the third most common cancer diagnosed in the United States (CDC, 2014; National Cancer Institute, 2016). Diagnosis of CRC is usually done by performing an endoscopic biopsy or polypectomy followed by microscopic examination of neoplastic cells, after which invasive carcinoma can sometimes be recognized (Fleming *et al*., 2012). The definition of invasive carcinoma is restricted to the submucosal invasion of the colorectum (Fleming *et al*., 2012). Analysis of submucosal invasion reveals that most colorectal adenocarcinomas are derived from precursor lesions (i.e., abnormal damage or change in tissue). Common precursor lesions are adenomas, dysplasia, and serrated polyps (Fleming *et al*., 2012).

It is well documented that CRC initially starts out as a polyp, called an adenoma, which is a benign tumor composed of epithelial cells that can develop into CRC (Vogelstein *et al*., 2013; Sears and Garrett, 2014) and may form on the inner wall of the intestine (Cooper *et al*., 2010). Endoscopic studies show that adenomas can be either pedunculated or sessile (Fleming *et al*., 2012). Adenomatous cells are characterized by their enlarged, hyperchromatic, and elongated nuclei. Under regular conditions, they are classified into three categories based on their structural

components: tubular, tubulovillous, and villous (Fleming *et al*., 2012). Tubular adenomas are made up of crypt-like dysplastic glands. Villous adenomas resemble finger-like projections. Tubulovillous adenomas are intermediate lesions (Fleming *et al*., 2012).

Serrated polyps describe any polyp that exhibits a saw tooth or star-shaped structure in epithelial cells (Fleming *et al*., 2012). Four types of lesions fall into this category: hyperplastic polyp (HP), sessile serrated adenoma/polyp (SSA/P), and traditional serrated adenoma (Fleming *et al*., 2012). Among all four, HPs are the most common. They are found in the distal colon and are roughly < 5 mm in size (Fleming *et al*., 2012). SSA/Ps are often seen in the proximal colon and are generally larger than HPs (Fleming *et al*., 2012). Although traditional serrated adenomas are unique and exhibit low grade nuclear dysplasia, their structure is similar to HPs and SSA/Ps (Fleming *et al*., 2012). Furthermore, polyps that become cancerous are called adenocarcinomas. More than 90% of adenocarcinomas are known to originate from epithelial cells of the colorectal mucosa while the minority of CRC types include neuroendocrine, squamous cell, adenosquamous, spindle cell, and undifferentiated carcinomas (Fleming *et al*., 2012).

The quest to understand the molecular pathogenesis of CRC revealed a step-by-step explanation of how normal colonic tissues progresses to CRC based on key mutations in oncogenes and tumor suppressor genes that possess critical regulatory and/or repair functions was proposed (Fearon and Vogelstein, 1990; Hisamuddin and Yang, 2006). In this model, it was suggested that alterations in either category as well as those resulting from epigenetic mechanisms (e.g., methylation) drives the tumorigenesis of colonic tissues, pushing it from one stage to the next (Fig. 2). These genes are called "driver" genes where mutations that occur in these genes are called "driver" mutations. The term, "driver", is used to define changes in genes that either directly or indirectly contribute to the proliferative potential of cells (Vogelstein *et al*., 2013). Subsequent studies have identified certain driver genes that are consistently correlated with the progression of CRC. Among these genes, the most common ones are adenomatous polyposis coli (*APC*), a tumor suppressor gene, and *KRAS*, a proto-oncogene (Fearon, 2011; The Cancer Genome Atlas Network, 2012; Seshagiri *et al*., 2012). These studies have suggested that certain genes may participate in very important processes that make them more valuable to cellular stability than other genes when evaluating the genetic pathogenesis of CRC. Thus,

researchers have developed models of CRC to elucidate the mechanisms that encourages the driver potential of key genes.

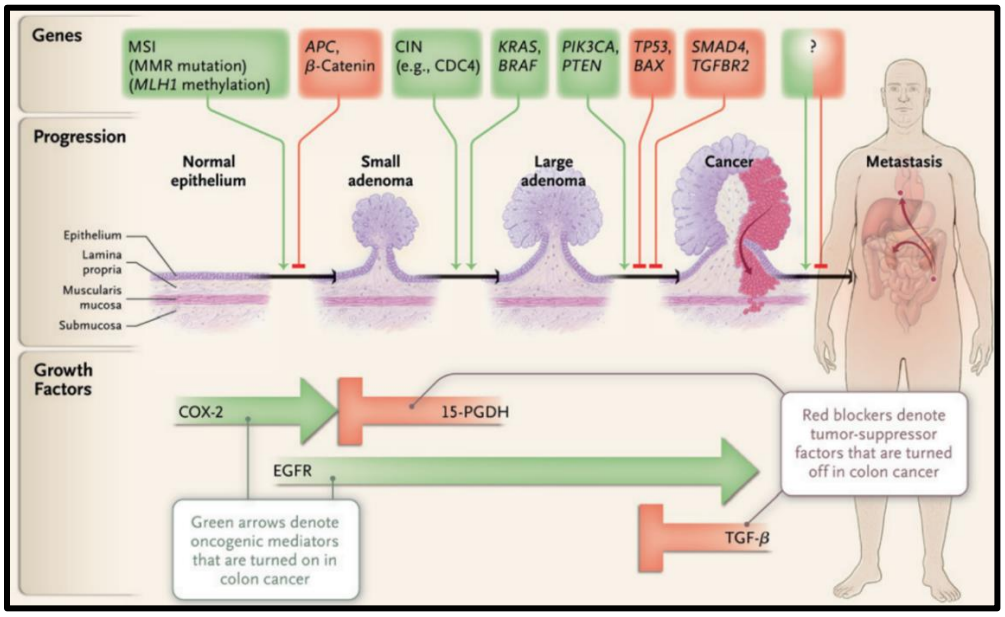

Figure 2. The molecular paradigm of colorectal cancer. The first step in CRC progression is believed to be the formation of aberrant crypt foci (ACF) caused by mutations in the *APC* and *β-catenin* gene that lead to the inactivation of *APC*. The eventual progression to adenoma and carcinoma stages are typically caused by mutations that activate oncogenic properties in the *KRAS* gene and the loss of *p53* (i.e., *TP53*), respectively. Other genes, such as *MSH2*, *MLH1*, *PMS2*, etc., may contribute to the developmental stages of ACF to advanced adenomas through epigenetic mechanisms (Hisamuddin and Yang, 2006; Roper and Hung, 2013; Markowitz and Bertagnolli, 2009).

Human genetic-based models of CRC explain how normal hyperplastic epithelium cells undergo molecular alterations in multiple genes which cause the development of these cells to progress onto an adenoma and then toward adenocarcinoma (Sears and Garrett, 2014). The growth of colonic epithelial cells (CECs) may be determined by the mutations in genes that influence adenoma and adenocarcinoma (i.e., driver mutations) (Sears and Garrett, 2014). These mutations may also reduce their vulnerability to apoptosis, causing them become more specialized in their metabolic processes, and gain control over immunological functions to further promote metastasis (i.e., spread of cancer from one organ and/or tissue to another) (Sears and Garrett, 2014).

A common anatomical location for adenocarcinomas is the human colon, which houses the largest number of microbes (Sears and Garrett, 2014). Therefore, understanding the role of microbes, specifically bacteria, in CRC has generated immense interest in the scientific community. As a derivative of the generic cancer paradigm, colorectal cancer is, at its core, a heterogeneous group of diseases that encompass unique genetic and epigenetic backgrounds (Fleming *et al*., 2012). It is therefore crucial to explore the molecular pathology that underscores its progression in conjunction with the impact of microbiota conceptual frameworks to properly evaluate its carcinogenesis (Sears and Garrett, 2014).

Intestinal Microbial Influence on Colorectal Tumorigenesis

Although it is known that buildup of oncogenic mutations over time causes CECs to replicate uncontrolled – a process that is said to take 10-40 years – it is not known what exactly causes this gradual change. However, the microbiome is a top suspect for triggering the initiation and/or progression of colorectal carcinogenesis due to the fact that colonic tissues are the repositories of the largest and most complex community of microorganisms (Sears and Garrett, 2014). The microbiome is a vast, complex, and dynamic conglomerate of microorganisms that colonizes the human body, constituting roughly 90% of all the cells (Qin *et al*., 2010). Furthermore, it is suspected that the number of microbial genes far supersedes the number of human genes by 100 times or more. However, many of these microbes that make up the microbiome do not all contribute to disease but are instead there for the benefit of the individuals they inhabit.

Healthy microbiomes are directly tied to host benefits while disturbances, natural and/or artificial, may lead to diseases. Thus, interactions between microbes and their hosts play crucial roles in maintaining human health (Costello *et al*., 2012). These interactions are classified into three general categories: symbiosis, commensalism, and pathogenicity (Hooper and Gordon, 2001). Symbiosis corresponds to a relationship between two different species where one of them benefits without harming the other (Hooper and Gordon, 2001). Commensalism corresponds to a relationship where both species coexist without any harm but also without any obvious benefit as well (Hooper and Gordon, 2001). Pathogenicity refers to a pathogenic relationship whereby the host is harmed by the host-microbe interaction (Hooper and Gordon, 2001).

From an ecological perspective, a healthy microbiome is contextualized within ecologic stability (Bäckhed *et al*., 2012) that comes with multiple benefits such as extracting nutrients that are normally inaccessible from dietary substances, encouraging differentiation of host tissues, stimulation of the immune system, and host protection from invasion by pathogens (Costello *et al*., 2012). Furthermore, the balance between beneficial and harmful chemical conversion reactions that take place in the gut microbiota is determined by its specific composition (Bäckhed *et al*., 2012). The reverse is also true where secretion of cytokines (i.e., cell signaling molecules) and defense effector molecules by host immune system shapes the microbiota community and promote mucosal immunity (Bäckhed *et al*., 2012).

Natural disturbances of the microbiome include age, geographical location, host genotype, probiotics (Bäckhed *et al*., 2012), environmental selection, and demographic stochasticity (i.e., ecological drift) (Costello *et al*., 2012). Artificial disturbances include extrinsic factors such as inflammation (Costello *et al*., 2012), intake of food supplements and drugs (e.g., antibiotics), human diet, and stress (Bäckhed *et al*., 2012). For example, excess exposure to antibiotics can disrupt the host-microbe interactions that contribute to human health, ultimately leading to various diseases such as obesity, type I diabetes, inflammatory bowel disease, asthma (Bäckhed *et al*., 2012), as well as acute and chronic disorders such as malnutrition, necrotizing enterocolitis, and antibiotic-associated diarrhea (Costello *et al*., 2012).

These disturbances, whether natural or artificial, have elicited the attempt of associating individual bacterial microbes to human disease. For example, *Streptococcus mitis*, *Staphylococcus epidermis*, *Bacillus* sp., *Mycoplasma* sp., and *Chlamydophila pneumonia* have all been identified in lung cancer cells (Cummins and Tangney, 2013). *Robinsoniella peoriensis*, *Pedioccoccus acidilactici*, *Leuconostoc lactis*, and *L. mensenteroides* have all been identified in pancreatic cancer cells (Cummins and Tangney, 2013). *Staphylococcus epidermis* and *Mycoplasma* sp. have been identified in breast cancer cells (Cummins and Tangney, 2013). *Ralstonia insidiosa, Fusobacterium naviforme,* and *Prevotella* sp. have all been identified in oral cancer cells (Cummins and Tangney, 2013). *Salmonella typhi*, *Helicobacter pylori*, *H. hepaticus,* and *H. bilis* have all been identified in gall-bladder cancer cells (Cummins and Tangney, 2013). *Chlamydia trachomatis* and *Mycoplasma* sp. have been identified in ovarian cancer cells (Cummins and Tangney, 2013). These bacteria species have been associated or defined as being

causative agents of cancers, further underlining the importance of studying bacterial microbes and its role in human diseases.

Researchers have developed murine disease models that support the idea that the microbiota contributes to colon carcinogenesis, but such models weakly illustrate human disease development (Sears and Garrett, 2014). Three models have been proposed in an attempt to provide a framework that not only strongly illustrates human disease development but does so within the framework of microbiota and certain members as either primary (i.e., initiators) or secondary (i.e., promoting growth) indicators that influence human CRC pathogenesis (Fig. 3) (Sears and Garrett, 2014). The first model suggests that individual microbes initiate or promote the growth of CRC. The second model theorizes that there is a collective microbial community that together initiate or promote the growth of CRC. The last model adopts the idea that single microbes interact with the microbial community which in turn drives the initiation or promotion of CRC growth (Sears and Garrett, 2014).

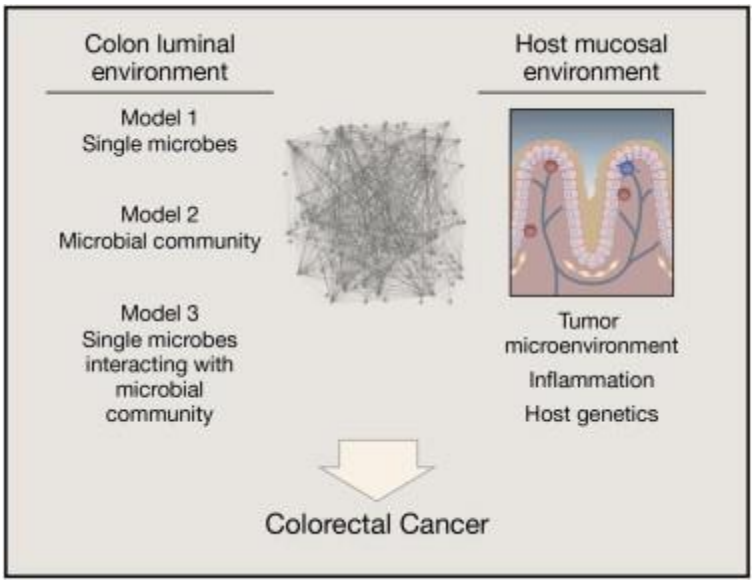

Figure 3. Microbial contributions to the pathogenesis of CRC. The three models summarize the emerging school of thought that the microbial community is a prime suspect for the underlying influence of CRC over long periods of time (Sears and Garrett, 2014).

Numerous suspected bacteria species have warranted possible identification of being causative agents or bacterial drivers of colorectal cancer, where they may possess carcinogenic features that may promote or initiate the disease. Many of these organisms have been identified

through metagenomics analysis where the 16S rRNA amplicons (i.e., a segment of DNA or RNA to be amplified) obtained from tumor and matching non-tumor samples in patients with CRC have been used to generate microbiome maps in order to identify microbes (Tjalsma *et al*., 2012; Marchesi *et al*., 2011; Kostic *et al*., 2012; Castellarin *et al*., 2012). A recent study took advantage of metagenomics methods to identify bacteria organisms from 130 matching tumor and non-tumor samples in 65 CRC patients. The results reveal that organisms within three genera have been found to demonstrate significant co-occurrence within individual colorectal tumor cells where they collectively illustrate a metagenomic signature of CRC: *Fusobacterium*, *Campylobacter*, *Leptotrichia*, and *Selenomonas* sp. (Warren *et al*., 2013). This was done by performing a read-pair alignment analysis where each of the bacterial sequences ran against sequence databases of human rRNA, bacterial, and viral RefSeq genome sequences using the Burrows-Wheeler Aligner for human sequences and Novoalign for bacterial and viral sequences (Warren *et al*., 2013). This allowed the determination of which species were significantly over-represented in colorectal tumor cells. Metagenomics analysis revealed that *F. nucleatum*, *C. showae*, *L. hofstadii*, *L. buccalis*, and *S. sputigena* had an over-representation of mapped read pairs that were of tumor origin (Fig. 4) (Warren *et al*., 2013).

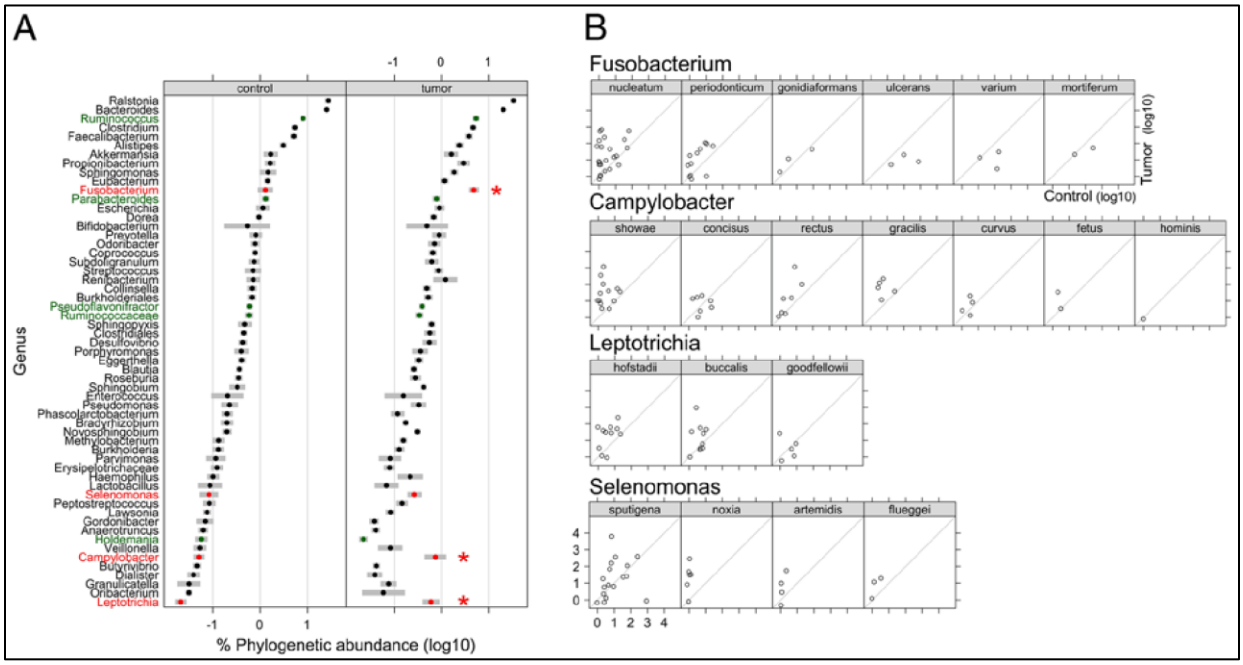

Figure 4. Microbial abundance in CRC and normal gut mucosa tissue via RNA-seq. A) Phylogenetic abundance gathered from unique metatranscriptomics read pair mapping. B)

Distribution of species uniquely mapped to *Fusobacterium*, *Campylobacter*, *Leptotrichia*, and *Selenomonas* normalized sequence pairs (Warren *et al*., 2013).

Between the two *Leptotrichia* species, it can be seen that *L. hofstadii* has the higher read-pair alignments. However, since both strains were isolated from colorectal tumor cells with very similar numbers of unique read-pair alignments, it is not known which one of the two species can facilitate stronger interactions with *F. nucleatum, C. showae,* and *S. sputigena* in the development of colorectal tumors. This is because it is also possible that all five strains work together to promote tumorigenesis. All five bacteria are gram-negative anaerobes that participate in a commensalism relationship with the oral cavity, specifically, the subgingival plaque (Etoh *et al*., 1993; Macuch and Tanner, 2000). Furthermore, the study shows that *Fusobacterium*, *Campylobacter*, and *Leptotrichia* sp., when isolated from tumor tissue, co-aggregated in culture (Warren *et al*., 2013). While co-aggregation is not uncommon in microbial co-cultures given the fact that they occupy the same niche, prior studies have shown that co-aggregation of *F. nucelatum* and *Streptococcus cristatus* increases *F. nucleatum*'s ability to invade into cultured host cells as well as altering the host response to it (Edwards, Grossman, and Rudney, 2006; Zhang, Chen, Rudney, 2011). Additionally, prior studies have shown that co-aggregation of *F. nucleatum* with other species facilitated the survival of obligate anaerobes (i.e., an organism that cannot tolerate any oxygen) in aerated environments (Bradshaw *et al*., 1998).

F. nucleatum has been repeatedly linked with CRC in multiple studies (Casellarin *et al*., 2012; Kostic *et al.*, 2012; McCoy *et al*., 2013). It is a pro-inflammatory anaerobe that is invasive, adherent, and known to be associated with other diseases such as periodontitis (Han *et al*., 2000; Signat *et al*., 2011; Swidsinski *et al*., 2009). Morphological studies have revealed that the organism is a small spindle-shaped rod and, in addition to being gram-negative, is non-spore forming and non-motile. Most of the cells are between 5 to 10 μm with sharply pointed ends (Bolstad, Jensen, and Bakken, 1996). While *F. nucleatum* is indeed an anaerobe, studies show that it can grow in an environment of up to 6% oxygen (i.e., a facultative anaerobe) (Moore *et al*., 1984).

Although a member of the order *Fusobacteriales* by virtue of being part of the *Leptotrichiaceae* family, little is known of the pathogenic potential of *L. buccalis* (Warren *et al*., 2013). Previous studies have shown that this gram-negative, rod-shaped (Bernard *et al*., 1991)

anaerobe was found in patients with lymphoma and leukemia. However, it is not known whether *L. buccalis* played a causative role in the development of cancer as these patients are immunocompromised and therefore more susceptible to bacterial infection (Warren *et al*., 2013; Eribe and Olsen, 2008; Weinberger *et al*., 1991). Although classified as an obligate anaerobe (Grollier *et al*., 1990), it has been proven that some strains of *L. buccalis* are capable of being aero-tolerant having grown well in 5% CO_2 atmosphere (Clark *et al*., 1984, Bernard *et al*., 1991). However, to what extent the bacterium is aero-tolerant remains unclear. Like its fellow *Fusobacterium* brethren, it is also non-motile as evidenced by the absence of flagella (Warren *et al*., 2013).

Like many of its fellow siblings in the family, *C. showae* is known to play pivotal roles in intestinal diseases such as Crohn's disease (Allos and Blaser, 1995; Maher *et al*., 2003; Tay *et al*., 2013), as well as being associated with gingivitis, periodontitis, and cholangitis (Macuch and Tanner, 2000; Etoh *et al*., 1993; Suzuki *et al*., 2013). It is a gram-negative, straight rod organism that possesses multiple unipolar flagella, making it a motile anaerobe. Although the bacterium prefers to grow in an anaerobic environment, it can grow under microaerophilic conditions (Etoh *et al*., 1993). Size of bacterium cells are 0.5 to 0.8 µm wide and 2 to 5 µm long with round ends. It has two to five unipolar unsheathed flagella as well (Etoh *et al*., 1993). Although *C. showae* has been suspected of having a causative link to diseases, its pathogenicity is unknown (Etoh *et al*., 1993; Suzuki *et al*., 2013).

Traditionally found in the upper respiratory tract, *S. sputigena* is an anaerobic gram-negative curved rod organism. However, like *C. showae*, the presence of a flagella points the fact the bacterium is motile (McCarthy and Carlson, 1981). Although studies have shown that *S. sputigena* may be implicated in the pathogenesis of generalized aggressive periodontitis, the exact causative link remains to be seen (Goncalves *et al*., 2012). Furthermore, while *S. sputigena* was found to be the least significant out of the four overrepresented species, previous studies have shown co-aggregation between the bacterium and *F. nucleatum* (Kolenbrander, Andersen, and Moore, 1989) backed by a high correlation value (Warren *et al*., 2013). This warrants the need to investigate the causative link between *S. sputigena* and CRC.

To this end, the possibility that *Fusobacterium*, *Campylobacter*, *Leptotrichia*, and *Selenomonas* sp. may have an etiological role in the development of CRC where the interaction

between the species could promote or induce significant molecular changes of colorectal tumor cells is the subject of this paper. This proposition will be explored using cancer cell lines to explore the relationship between CRC and bacterial biofilm growth of *F. nucleatum, C. showae, L. buccalis,* and *S. sputigena*.

Bioinformatics Approaches in Evaluating CRC

Although cancer research has relied on traditional methods for decades, the sole use of these methods have become woefully insufficient to solve and understand the ever increasing complexities of cancer biology. Bioinformatics approaches have played prominent roles in the identification and validation of biomarkers, developing clinical phenotype profiles that are patient-specific, and providing ways to measure disease progression as well as response to therapy (Wu, Rice, Wang, 2012). For example, by using databases such as Gene Ontology (GO) (Ashburner *et al*., 2000) to extract data on biological processes or Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto *et al*., 2000) to obtain information on biological pathways, one can better understand the biological function of various molecules (e.g. microRNAs (miRNAs)) and their targets (e.g. mRNAs). GO is a database that encompasses a wealth of information on the roles of genes and gene products in many organisms (Ashburner *et al*., 2000). KEGG is a database that is often used to link information between the genes present in genomes and a network of interacting cellular components to produce pathways that represent higher order biological functions (Kanehisa and Goto *et al*., 2000). The information contained within the two databases can then be used to construct a model that represents the interplay between a cell's molecular state and its response to anti-cancer therapy for a specific cancer type (Wu, Rice, Wang, 2012).

In addition to the use of databases, there are numerous algorithms and software programs designed with the intent of analyzing complex cancer data. These techniques take advantage of a variety of things such as sequence homology to identify miRNA families that play a role in cancer and then use these miRNAs to identify gene targets such as transcription factors, secreted factors, receptors, and transporters by tracking what these miRNAs bind to (Lim *et al*., 2003, 2005; Grosshans *et al*., 2005, Krek *et al*., 2005). Other bioinformatic approaches take advantage of pooling in multiple tools that individually analyze different aspects of a complex cancer dataset and then provide results that are used to understand how a particular process is frequently

perturbed in cancer progression (Beck *et al*., 2014). It can also be used to discover novel models that lead to more efficient ways to understand a perturbed process (Roca *et al*., 2014). These tools serve as a massive framework that provides a powerful way to infer conclusions and implications on how perturbed processes work and develop strategies to counteract it.

Thesis Approach

This framework serves as an integral part in the investigation of microbial impact on CRC. The intent is to use software programs visualize how the bacteria species interact with CRC cells to study the dynamic interplay between bacterial biofilms and CRC cells which explain how one affects the other and vice versa. In evaluating this interplay, the paper focuses on three key software programs that were used in the project: ImageJ, Comstat2, and Icy.

ImageJ (Fig. 5a) is an open-sourced imaging program written in Java that is capable of reading many image formats commonly used in the biomedical sciences (Abramoff, Magalhaes, and Ram, 2004). It has many different operations such as reading and writing image files, convolution, edge detection, Fourier transform, histogram, and particle analyses (Abramoff, Magalhaes, and Ram, 2004) – just to name a few. It also allows users to write macros and plugins in different languages although most of them are written in Java. Macros are scripts meant to expand a single task, making it easier to automate it for repeated tasks while plugins are external programs that arm ImageJ with unique capabilities the program does not have (Abramoff, Magalhaes, and Ram, 2004). Finally, the program can run on any operating system and can be integrated into other software programs.

Comstat2 (Fig. 5b) is a graphical user interface derivative of the original Comstat. This is a novel, open-sourced, computer program that analyses three-dimensional biofilm structures using a host of parameters designed to quantify it (Heydorn *et al*., 2000). 3D images of bacterial biofilm are typically attained through confocal microscopy experiments where they can be opened up by an imaging software, such as ImageJ, and then analyzed using the ten parameters. Comstat2 can exist within the ImageJ platform as a plugin to be used whenever the user desires (Fig. 5c). The program can also exist as a standalone program via the utilization of a JAR file (i.e., a runnable software by aggregating different Java class files). This allows users to analyze quantifiable data of biofilms using Comstat2 without the need to have or understand the underlying computer language skills.

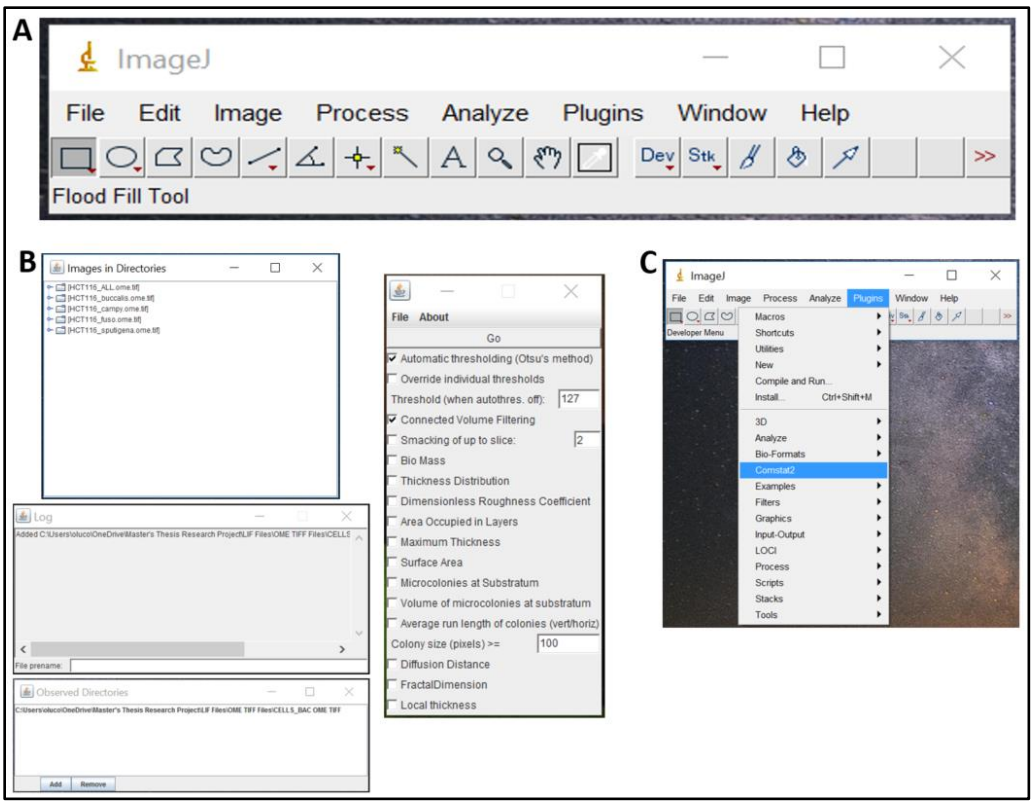

Figure 5. Screenshots of ImageJ and Comstat2 software. (a) ImageJ main window opened using Microsoft Windows OS. (b) Comstat2 opens up three pop-up windows. The top-left opens the window that lists all the folders that contains the correct image files Comstat2 looks for in the directories. The bottom-left show the log screen which records every action taken in the software (top) and the PATH location to each directory opened up (bottom). The window on the right contains the list of parameters that users can choice to analyze 3D biofilm structures. (c) ImageJ plugin menu to open the Comstat2 software.

Icy is an open-source platform for bioimage informatics that is used analyze biological images (de Chaumont *et al*., 2013). Similar to ImageJ, it has an extensive plugin library. One of the most prominent features is the visualization of 3D data, which is performed using Visualization Toolkit (VTK) (Fig. 6). VTK is an open-source software routinely used in 3D computer graphics and is written in several combined programming languages such as C++, Java, and Python (http://vtk.org). Icy is completely written in Java although some of the plugins can be written using JavaScript and Python (de Chaumont *et al*., 2012). Furthermore, Icy natively integrates the ImageJ platform where Comstat2 can be used as an internal ImageJ plugin.

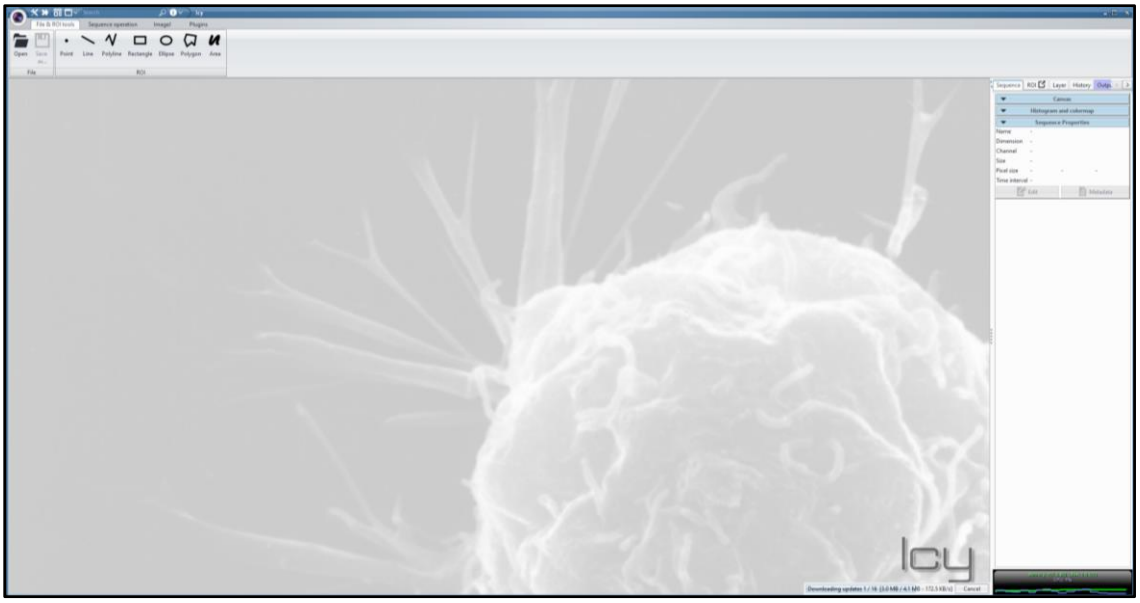

Figure 6. Screenshot of Icy interface. The above images show what Icy looks like when the user opens it up. The ImageJ platform is accessible by virtue of a tab with its own name and contains every plugin of its own. Users can also add their own plugins by creating a JAR file and then embedding it into the software. The same goes for Icy-based plugins as well.

Thesis Goals

In evaluating the impact of anaerobic bacteria on colorectal cancer, this thesis project has four distinct goals. The first is to firmly establish the ability of *F. nucleatum, C. showae, L. buccalis*, and *S. sputigena* to produce biofilms – a known contributor to disease. The second is to explore how bacterial biofilm production impacts the growth of CRC cells by assessing cellular viability. The third is to explore how bacterial biofilm production and establishment is influenced when growing in presence of CRC cells. The last goal of this project is to further explore the relationship between anaerobes and CRC by evaluating the progression of normal epithelial colorectal cells through miRNA sequencing analysis.

METHODS

Reconstitution of lyophilized bacteria cultures and growth conditions

Lyophilized cultures of *Leptotrichia buccalis* (ATCC 14201), *Fusobacterium nucleatum* (ATCC 25586), *Campylobacter showae* (ATCC 51146) (ATCC, Manassas, VA), and *Selenomonas sputigena* (ATCC 35185) (ATCC, Manassas, VA) were reconstituted precisely

according to the manufacturer's instructions and subsequently inoculated aseptically onto the surface of CDC-formulated blood agar plates (VWR, Pittsburg, PA) using the quadrant streak method. The plates were then incubated at 37°C for 48 hours in an AS-500 anaerobic chamber (Anaerobe Systems, Santa Clara, CA) under anaerobic conditions (90% N_2, 5% H_2, and 5% CO_2). Multiple liquid pure cultures of each bacterial species was prepared by aseptically inoculating a single isolated colony into a screw-capped Hungate glass tube which contained 10 mL of sterile, pre-reduced anaerobic broth composed of brain-heart infusion (BHI) medium reconstituted within liquid dental transport medium (LDT; Anaerobe Systems, Santa Clara, CA). All tubes were incubated for 48 h at 37°C under anaerobic conditions.

Preparation of bacteria -80°C glycerol stocks

Cryopreservation of bacteria isolates used in this study was achieved by aseptically preparing and subsequently storing multiple tubes of 20% glycerol bacterial stocks at -80°C. Briefly, a day prior to preparing the glycerol stocks, a rack of sterile internally threaded cryopreservation tubes (Corning, Corning, NY) were placed into the anaerobic chamber (Anaerobe Systems, Santa Clara, CA) and the caps were slightly but carefully loosened. A bottle of glycerol was heat sterilized under standard autoclaving conditions and was immediately placed into the anaerobic chamber with the top slightly loosened to allow the gaseous anaerobic environment to permeate the glycerol in the tubes as they cooled. The following day, 200 µL of sterile pre-reduced glycerol was aseptically pipetted into several sterile pre-labelled cryopreservation tubes, followed by the addition of 800 µL of bacteria culture. Multiple tubes of each bacteria isolates was similarly prepared with subsequent complete mixing by inversion while in the AS 500 anaerobic chamber before being stored at -80°C.

Bacterial biofilm viability

A 1:200 dilution of the bacterial isolates was prepared in LDT + BHI medium and 200 µL of this dilution was seeded into the wells of a 96-well plate and incubated in 5% CO_2 humidified anaerobic and microaerophilic environments at 37°C for 3 days using GasPak™ EZ Anaerobe Container System (BD Diagnostics; Sparks, MD) and GasPak™ EZ Campy Container System (BD Diagnostics; Sparks, MD), respectively. Following the incubation period, 100 µL of a 0.01% crystal violet solution was added to the wells and the wells of each plate were allowed to sit for 20 min with gentle agitation every few minutes. After excess crystal violet was discarded by grasping the plate and applying an almost simultaneous yet quick inversion and

sharp/abrupt stoppage of a forward movement, each plate was gently washed two times using 300 µL of water and subsequently allowed to air-dry. The crystal violet-stained biofilm was subsequently immersed in 200 µL of absolute ethanol and allowed to sit for 20 min. Indirect biofilm formation data was acquired by reading the absorbance values of each well using a 96-well spectrophotometer at 595 nm. Softmax Pro (Molecular Devices, Sunnyvale, CA) software was used to capture the data, however, the data was visualized using the graphing features present in Microsoft Excel (Microsoft; Seattle, WA).

Cell culture of HCT116 cells

Upon receipt, the colon cancer cell line HCT116 (ATCC CCL-247) (ATCC, Manassas, VA) was immediately thawed with gentle agitation in a 37ºC water bath according to exact instructions provided by the ATCC. Thereafter, the cell line was cultured in McCoy's 5A medium (Iwakata and Grace Modification) with L-glutamine (Catalogue #45000-374; VWR, Randor, PA) supplemented with 5% fetal bovine serum (FBS) (Catalogue #10437010, Gibco by Thermo Fisher Scientific Inc., Waltham, MA) in 75 cm^2 Corning T-75 flasks (Product #43725U; Corning, NY) as described by the manufacturer. The cell line was grown at 37ºC in a humidified environment of 5% CO_2 in air atmosphere until cells reached 70-80% confluency with medium renewal every 2-3 d. Cryopreservation of HCT116 cells was achieved by preparing liquid nitrogen stocks in cryopreservation medium that contained a mixture of 50% FBS (Catalogue #10437010, Gibco by Thermo Fisher Scientific Inc., Waltham, MA), 40% culture medium, and 10% Synth-a-Freeze (Catalogue #A13713-01; Cell Therapy Systems by Thermo Fisher Scientific Inc., Waltham, MA). This was then used to produce 1 mL aliquots of cell culture and was stored at -80ºC.

Confocal analysis of bacteria biofilms grown under anaerobic and microaerophilic conditions

Prior to confocal image analysis, biofilm from each bacteria isolate was grown in a 24-mm glass dish with 2 mL of LDT + BHI medium over a period of four days to one week in anaerobic and microaerophilic environments using GasPak™ EZ Anaerobe Container System (BD Diagnostics; Sparks, MD) and GasPak™ EZ Campy Container System (BD Diagnostics; Sparks, MD), respectively. Both groups of samples were maintained at 37ºC in a humidified 5% CO_2 atmosphere. The medium was not replaced during the growth cycle. After aseptically removing the medium, biofilm formation from the bacterial isolates was stained with

Invitrogen's LIVE/DEAD Biofilm Viability Kit (Thermo Fisher Scientific Inc., Waltham, MA) as directed by the manufacturer. Specific areas of each sample was scanned using the Leica TCS SP5 II (Leica Microsystems; Wetzlar, Germany) software for biofilm structures using water immersion lens at 40X magnification. The images were obtained through the generation and image capture of z-stacks. The number of z-stacks obtained and the thickness of each slice was dependent upon biofilm thickness, composition and the type of organisms used. Each image slice was acquired with a resolution of 1024 x 1024 pixels.

Confocal analysis of HCT116 cells

Prior to confocal microscopic analysis, each sample was first grown in a 24-mm glass dish with 2 mL of appropriate medium over a period of one week in a microaerophilic environment using the GasPak™ EZ Campy Container System (BD Diagnostics; Sparks, MD). The medium was not replaced during the growth period. After aseptically removing the medium, the samples were each stained with Invitrogen's LIVE/DEAD Biofilm Viability Kit (Thermo Fisher Scientific Inc., Waltham, MA) and Life Technologie's NucBlue Live ReadyProbes Reagent (Thermo Fisher Scientific Inc., Waltham, MA) according to exact instructions by manufacturer. Briefly, one to two drops of the NucBlue stain was added to each well of cell growth followed by confocal microscopy. Specific areas of each sample was scanned using the Leica TCS SP5 II (Leica Microsystems; Wetzlar, Germany) software for biofilm structures using water immersion lens at 40X magnification. Images were obtained via z-stacks with slices taken at 3.8 µm thickness per slice for a total of 43 slices and a resolution of 1024 x 1024 pixels.

Confocal analysis of co-cultured HCT116 cells and bacteria

Prior to confocal microscopic scanning, HCT116 cells were grown in 24-mm glass dish wells with 2 mL of appropriate medium. Cells were allowed to grow over a period of one week in a microaerophilic environment using the GasPak™ EZ Campy Container System (BD Diagnostics; Sparks, MD) with no medium renewal during growth period. After one week, 1 mL of bacteria samples was added to each well, with each well receiving one of the four bacteria species and subsequently allowed to grow for 24 h. Following 24 h growth period, the medium was aseptically removed and the samples were stained with Invitrogen's FilmTracer Biofilm staining Kit (Thermo Fisher Scientific Inc., Waltham, MA) and NucBlue Live ReadyProbes Reagent (Life Technologies by Thermo Fisher Scientific Inc., Waltham, MA) as directed by the manufacturer. Specific areas of each sample was scanned using the Leica TCS SP5 II (Leica

Microsystems; Wetzlar, Germany) software for biofilm structures using water immersion lens at 40X magnification. The images were obtained via z-stacks. Number of z-stacks obtained and size of each slice was dependent on the composition and type of organisms used. Each image was acquired with a resolution of 1024 x 1024 pixels.

3D construction of images using Icy

Images generated by confocal microscopy were visualized using a collaborative bioimage informatics platform called Icy (Pasteur Institute; Paris, France). Images were extracted from ".LIF" files and opened up in the navigator. 3D reconstruction of images was done by switching visualization mode to 3D *VTK* option. After adjusting parameters to get desired images, they were then saved by taking a screenshot of the screen and then cropping regions of interest.

Computational analysis of biofilm images using Comstat2

Prior to Comstat2 (v2.1; SEAS-NVE A/S, Denmark) analysis, TIFF image stacks of biofilms produced by Leica TCS SP5 II (Leica Microsystems; Wetzlar, Germany) were converted to single OME-TIFF files using the LOCI tools built in Comstat2 (v2.1; SEAS-NVE A/S, Denmark) in greyscale according to the exact instructions provided by the COMSTAT 2.1 Manual. Predetermination of threshold values for each OME-TIFF file was done by using automatic (Otsu's method) thresholding in the Comstat2 (v2.1; SEAS-NVE A/S, Denmark) control panel. Computational analysis of OME-TIFF image files was performed by Comstat2 (v2.1; SEAS-NVE A/S, Denmark) according to the following parameters: Biomass, Thickness Distribution, and Surface Area. From the surface area parameter, average surface area and surface to volume ratio was calculated. Numerical data was produced in the form of a text file, which was opened using Microsoft Excel 2016 (Microsoft; Seattle, WA).

Viability of HCT116 cells and bacteria using ImageJ

Images obtained from confocal microscopy were used to assess viability of cells and bacteria. This was done by first using ImageJ's split channels function under *Image* → *Color* tab to split the images into two channels for samples stained using Invitrogen's FilmTracer LIVE/DEAD Biofilm Viability Kit (Thermo Fisher Scientific Inc., Waltham, MA) to distinguish between image channels stained green (LIVE) and red (DEAD). Images stained with the additional NucBlue Live ReadyProbes Reagent (Thermo Fisher Scientific Inc., Waltham, MA) for co-cultured growth of HCT116 cells and bacteria were split into three channels where the first one was stained blue (nucleus) with the second and third channels stained red and green,

respectively. To figure out which stain was used for each channel, ImageJ's *Subtract Background* function under *Process → Subtract Background* was used. Before calculating cellular and bacteria viability, the images were cleaned up by first changing the colors to black & white under *Image → Adjust → Threshold*. This was followed by changing the color to *B&W* and then clicking *Apply*. Subsequently, the background was filled by clicking the *Fill Holes* option under *Process → Binary → Fill Holes*. The background colors were then inverted by using the *Convert to Mask* option under *Process → Binary → Convert to Mask*. Finally, using the *Watershed* option under *Process → Binary → Watershed*, the images were automatically segmented to cut particles that were touching each other. Cellular viability was counted using the *Analyze Particles* under *Analyze → Analyze Particles*. To separate cells from bacteria species in the images, the size and circularity was adjusted.

Statistical analysis

Examination of the differences between HCT116 cells grown in the presence of vs. the absence of bacterial biofilms was tested with the nonparametric Mann-Whitney *U* test using the R statistical language (https://www.r-project.org/). Statistical significance was determined by comparing the viability of HCT116 cells grown in the absence of bacteria vs. presence of bacteria.

RESULTS

Confocal microscopy was performed on samples composing of individual and collective bacteria species. Each sample was grown in anaerobic and microaerophilic environments. The 2D images portray a snapshot of the live (green) / dead (red) organisms that are present in the biofilm. In addition to viability composition, the general organization from the top of the biofilm can be deduced as well. Images were taken at 40X magnification.

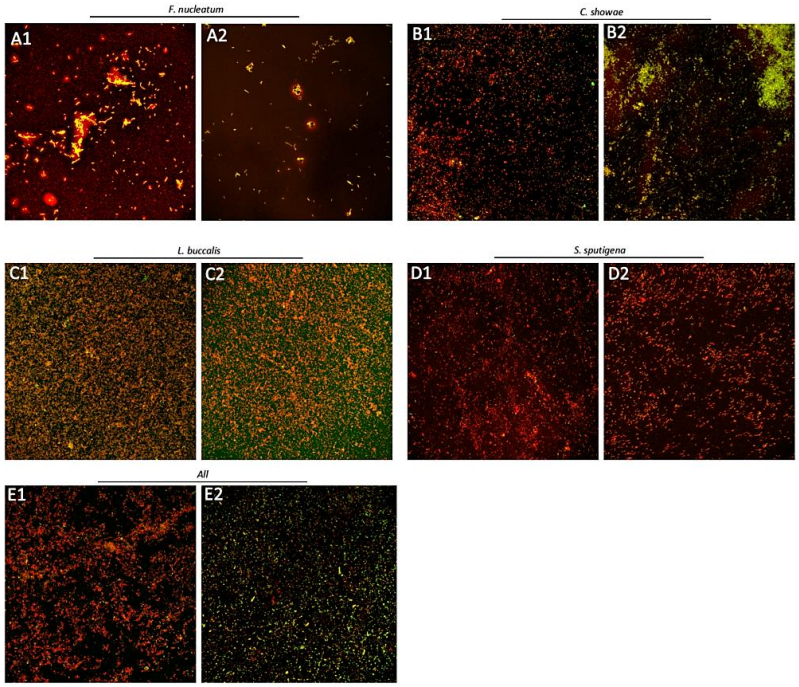

Figure 7. 2D confocal images of bacterial cultures in (1) anaerobic vs. (2) microaerophilic environments. (A) *F. nucleatum*. (B) *C. showae*. (C) *L. buccalis*. (D) *S. sputigena*. (E) All four together. These images were obtained using the ImageJ software. What is notable in each image is the difference in organization and overall morphological formation of biofilm structures grown in the two environments.

Each image generated from confocal studies was then extrapolated in a 3D structure using Icy. Once opened up in the software, the color parameters were adjusted to generate optimal 3D images. The 3D perspective provides more information about the overall biofilm structure, viability, as well as much more prominent composition of clusters and voids.

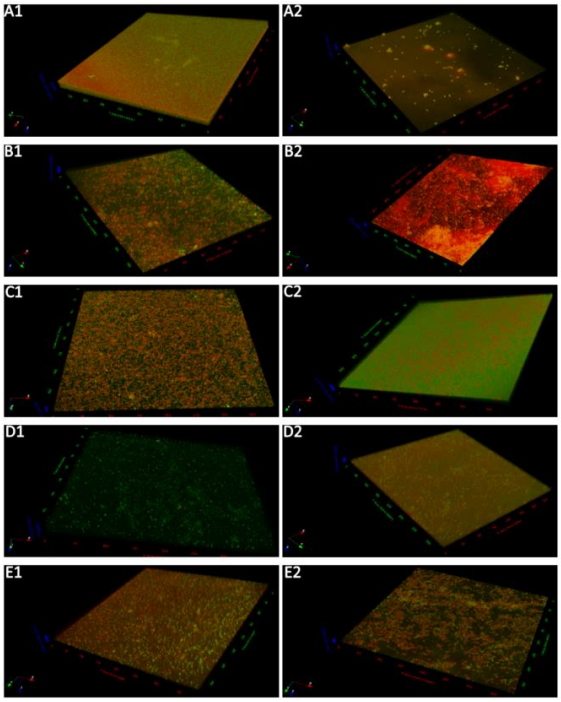

Figure 8. 3D biofilm images of anaerobic microbes grown in (1) anaerobic and (2) microaerophilic environments. (A) *F. nucleatum*. (B) *C. showae*. (C) *L. buccalis*. (D) *S. sputigena*. (E) All four together. These images were obtained using the Icy software. Similar to the results obtained in Figure 7, the overall biofilm structure can be seen to determine how the microbes organized themselves by either clustering around certain areas as well as present a visual ratio of viable (green) vs. non-viable (red) anaerobes.

Crystal violet biofilm assay performed on each organism grown in anaerobic and microaerophilic environments revealed that each anaerobe can produce biofilms.

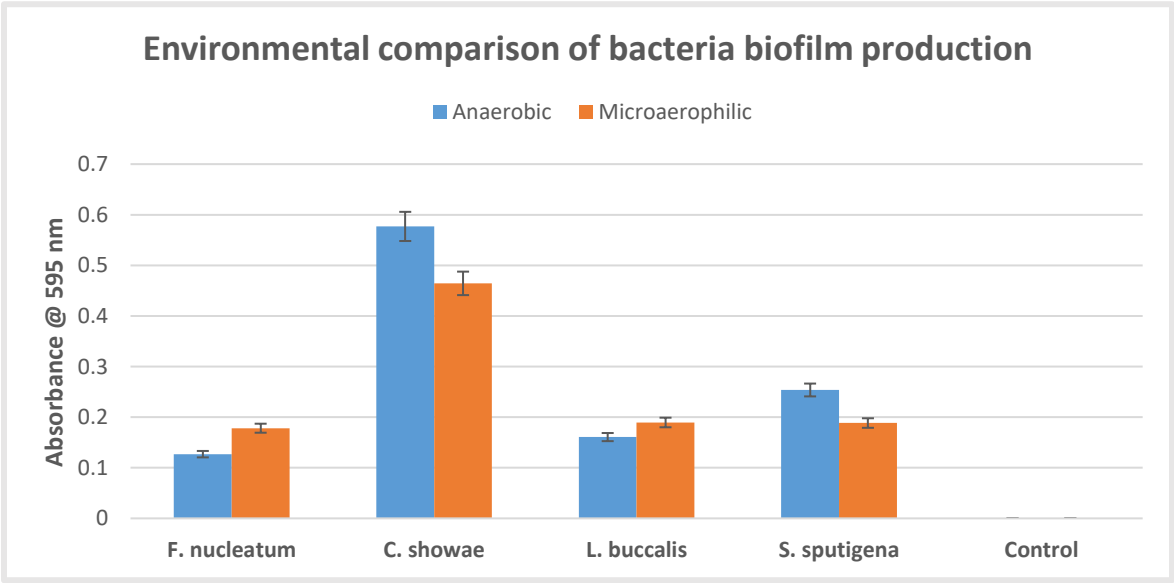

Figure 9. Biofilm growth of anaerobes in anaerobic and microaerophilic environments. This

confirms the ability of each microbe to grow well-established biofilms.

Each confocal image was quantitatively analyzed using Comstat2 using biomass, average, thickness, surface to volume ratio, and average surface area. The four parameters are used to assess how the biofilm adapts over time in both environments.

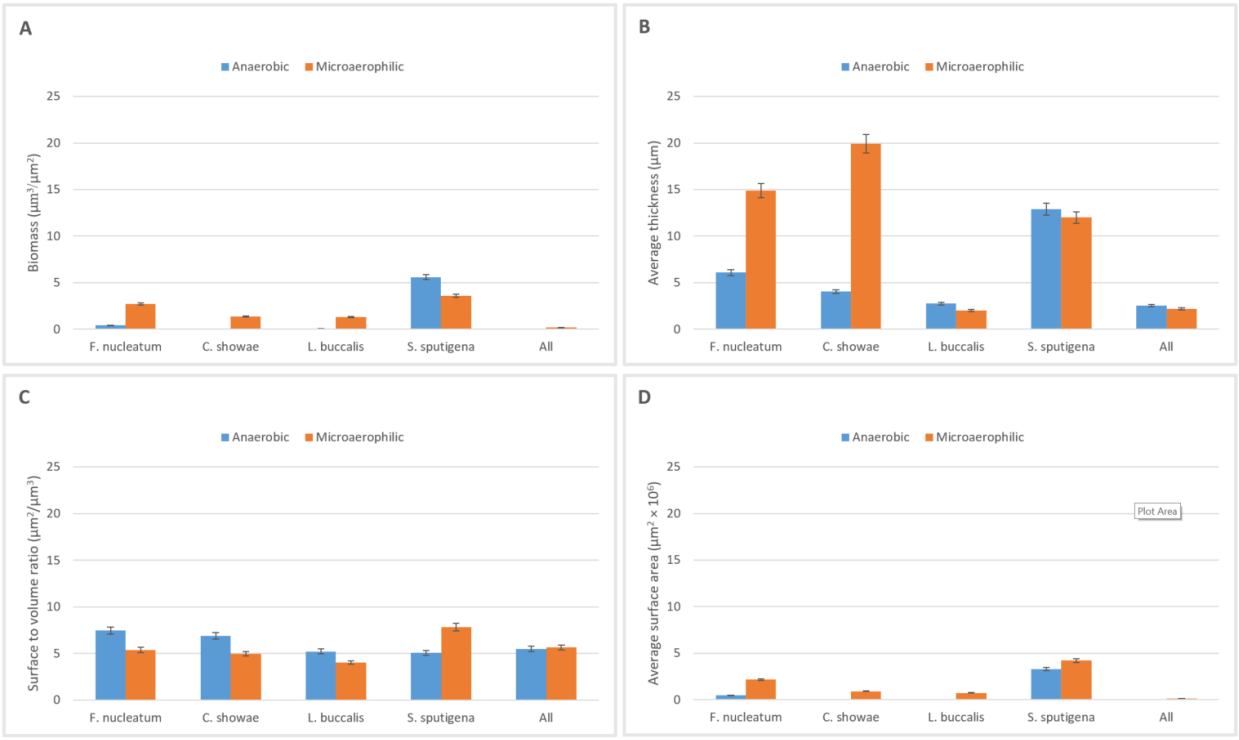

Figure 10. Biofilm parameters of *F. nucleatum*, *C. showae*, *L. buccalis*, *S. sputigena*, and all four together grown in anaerobic and microaerophilic environments. (A) Biomass. (B) Average thickness. (C) Surface to volume ratio. (D) Average surface area. Results were obtained using Comstat2.

Viability of each organism was directly assessed using ImageJ through the *Analyze Particles* option to see how well each organism survived in both environments over a long period of time.

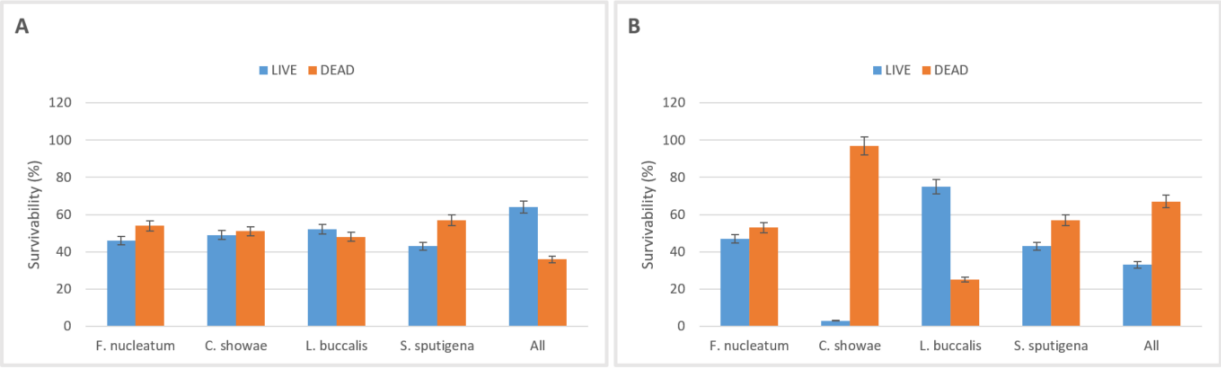

Figure 11. Comparison of live and dead anaerobes in (A) anaerobic and (B) microaerophilic

environments. Viability results were obtained using ImageJ.

HCT116 samples were scanned by confocal microscopy to generate 2D images. HCT116 adopts cancerous phenotypic qualities as shown by the elongated fibroblastic-like shapes and rapid proliferation of cells within a short time period. Images were taken at 40X magnification.

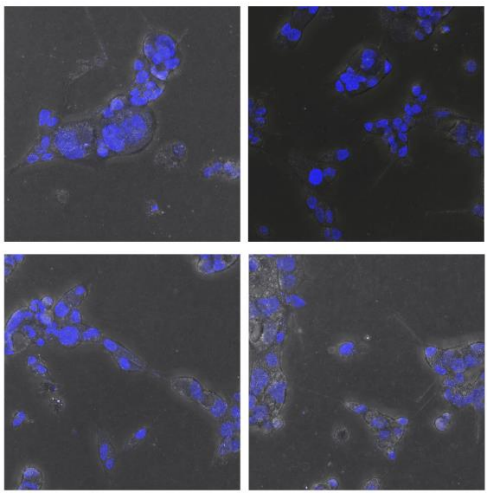

Figure 12. 2D confocal images of HCT116 colorectal carcinoma cells.

2D confocal images of HCT116 samples co-cultured with individual and collective bacteria species was generated. Composition and limited assessment of bacterial biofilm viability can be gleaned from each images. Also prominent in images B and D is the apparent struggle of bacterial biofilm establishment. Images were taken at 40X magnification.

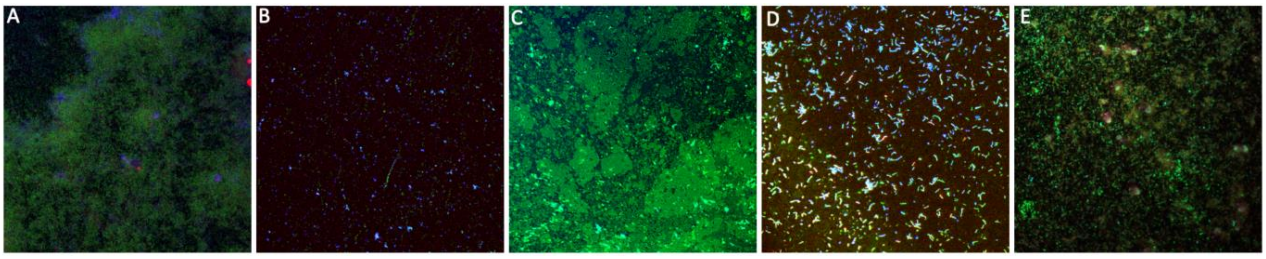

Figure 13. 2D confocal images of HCT116 colorectal carcinoma cells co-cultured with anaerobic microbes in microaerophilic environment. (A) *F. nucleatum*. (B) *C. showae*. (C) *L. buccalis*. (D) *S. sputigena*. (E) All four together.

3D images of each co-cultured sample was generated in Icy and adjusted using the color parameters to produce optimal images. The breakage of biofilm structures points to the struggle of bacterial organisms struggling to generate well-established structures when growing in the presence of HCT116 cells.

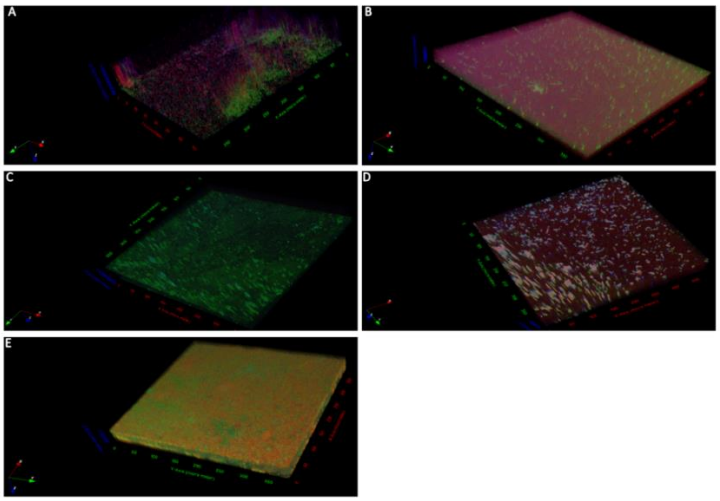

Figure 14. 3D images of HCT116 colorectal carcinoma cells co-cultured with anaerobic microbes in microaerophilic environment. (A) *F. nucleatum*. (B) *C. showae*. (C) *L. buccalis*. (D) *S. sputigena*. (E) All four together.

Each co-cultured sample was quantitatively analyzed using ImageJ's *Analyze Particles* option to provide a direct assessment of live/dead cells. In addition, HCT116 cells grown in the absence of any bacteria species was assessed as well. Statistical significance was done between viability of HCT116 cells grown in the absence and presence of individual and collective anaerobes using R.

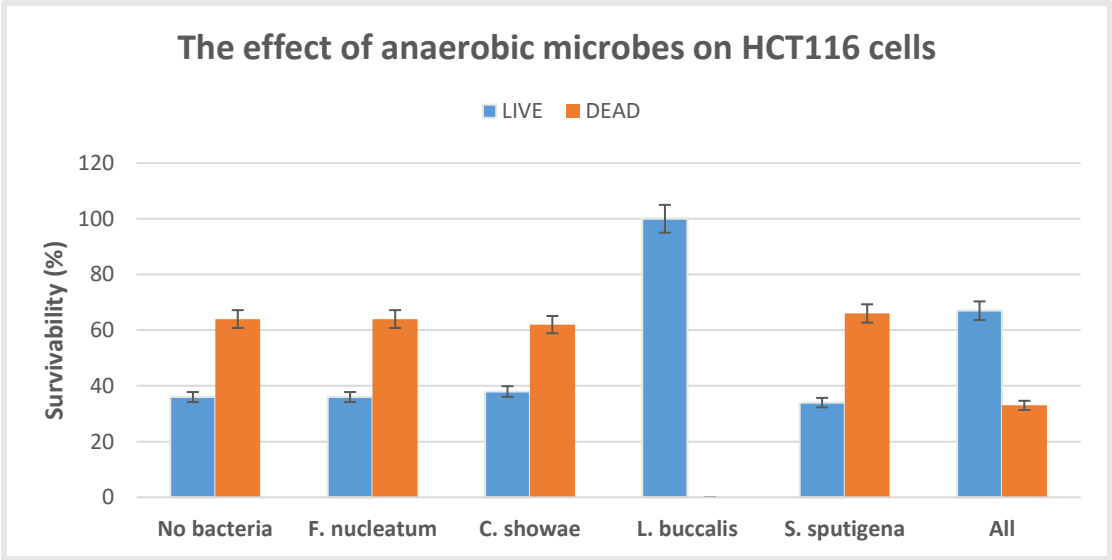

Figure 15. Viability of HCT116 cells grown in the absence and presence of anaerobes. The Mann-Whitney *U* test was performed for each co-cultured growth compared to single cell cultured growth experiments to evaluate statistical significance using a *p*-value cutoff of $p < 0.0001$. *p* values for co-cultured experiments of *F. nucleatum*, *C. showae*, *L. buccalis*, and *S. sputigena* was $p < 2.2 \times 10^{-16}$ while $p = 1.386 \times 10^{-14}$ for all four together.

Comstat2 was used to characterized biofilm structures of individual and collective anaerobes grown in the presence of HCT116 cells.

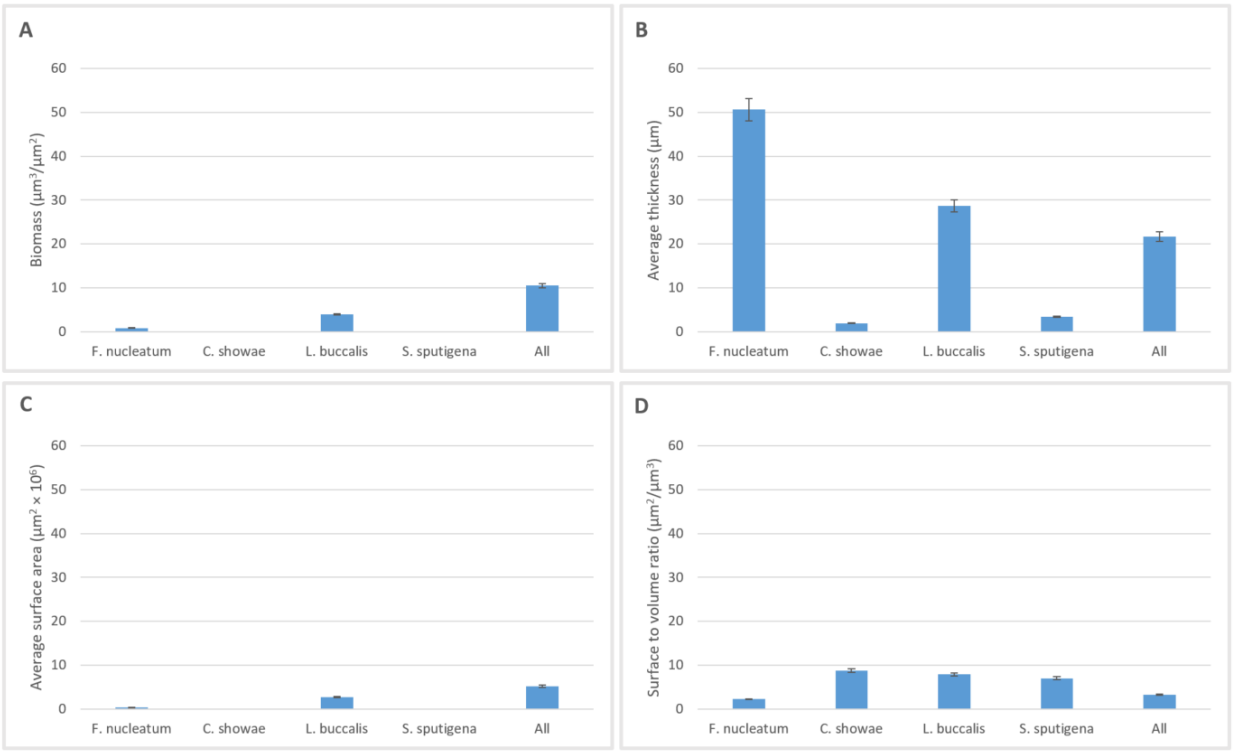

Figure 16. Biofilm parameters of *F. nucleatum*, *C. showae*, *L. buccalis*, *S. sputigena*, and all four together co-cultured with HCT116 cells. (A) Biomass. (B) Average thickness. (C) Surface to volume ratio. (D) Average surface area.

DISCUSSION

2D (Fig. 7) and 3D (Fig. 8) images of *F. nucleatum*, *C. showae*, *L. buccalis*, and *S. sputigena* revealed stark morphological differences of biofilm growth between anaerobic and microaerophilic environments. Nevertheless, the organisms were capable of producing well-structured biofilms as evidenced in Figure 9. This was even more prominent upon in-depth analysis of biofilm characteristics on each microbe grown in both environments (Fig. 10). It is interesting to note that *C. showae* initially produced the most biofilm but over time, the other species catch up and, in the case of *F. nucleatum* and *S. sputigena*, surpassed its growth rate. When comparing the two environments each organism grew in, the initial stage revealed that *C. showae* and *S. sputigena* were more comfortable growing in anaerobic than microaerophilic while the opposite held true for *F. nucleatum* and *L. buccalis*. Upon closer inspection on individual biofilm growth over long periods of time, *S. sputigena* was the only organism to grow

very well in an anaerobic environment compared to microaerophilic. This analysis can be seen in comparing the biomass parameter across all four individual organisms.

Furthermore, when analyzing other biofilm characteristics such as average thickness and average surface area, one can use it to understand the overall spatial size of the biofilm and what percentage of the overall biofilm was exposed to nutrient flow, respectively. Here, it seemed that for both cases *F. nucleatum* and *S. sputigena* did well compared to *C. showae* and *L. buccalis* in both environments. However, it is important to note *C. showae* produced the thickest biofilm in microaerophilic environment while *S. sputigena* produced the thickest biofilm under anaerobic conditions. These results are interesting when put into the perspective of how well the overall biofilm was able to extract the nutrients from its environment, one sees that *C. showae* did very poorly while *S. sputigena* did very well. This is surprising because one would reasonably assume that large biofilms are capable of extracting lots of nutrients but this was only true for *S. sputigena* and (to a lesser extent) *F. nucleatum*. This could mean that *C. showae* and *L. buccalis* were extremely inefficient at extracting nutrients when growing in both environments but further studies are needed to fortify such conclusions.

Finally, when evaluating the surface to volume ratio, one gets a sense of how the biofilm adapts to its environment. When comparing all four individual organisms in anaerobic environments, it becomes clear that *F. nucleatum* had a hard time adapting to the environment and thus needed to contract while spreading over a larger area of the substratum (i.e., bottommost layer) in order to optimize its access to scarce resources while *S. sputigena* was the most efficient in optimizing access to resources. When comparing the anaerobes grown in microaerophilic environment, *S. sputigena* was the one that had the hardest time adapting the environment while *L. buccalis* had the least burden for environmental adaptability. These results are interesting because, for one, they almost contradict each other. For example, *S. sputigena* had the hardest time adapting to a microaerophilic environment yet it had one of the highest biofilm mass, overall average thickness, and an extraordinary ability to extract resources from its surroundings as the biofilm grew larger and larger. Now, one could attribute this to the fact that it possesses flagella which gives it a significant advantage over organisms that do not but the fact that it grew very well despite having difficulty adapting to its environment warrants further

investigation because *C. showae* not only has flagella but possess multiple flagella as explained in the introduction yet it did not produce similar consistent results as *S. sputigena*.

A breakdown of individual organisms grown in anaerobic vs. microaerophilic to their viability rates revealed a much more precise numerical picture of how well each organism did within well-established biofilms (Fig. 11). In spite of *S. sputigena* doing very well in producing biofilm growth, far more organisms died when grown in either environment. To a lesser extent, the same was true for the rest of the organisms with the exception of *L. buccalis*. In fact, when grown under microaerophilic conditions, *L. buccalis* had the highest survivability rate compared to the other three anaerobes. Upon acute analysis, this makes sense given that the organism had the lowest surface to volume ratio compared to the other organisms.

Further analysis on the survivability of all four anaerobes grown together revealed that they seem to fare better in anaerobic vs. microaerophilic. From an objective biological standpoint, this may contribute to the availability of resources that may suddenly seem scarce in microaerophilic compared to anaerobic environments. The almost complete reversal in survivability ratios between two environments points to competition for resources between the species. Another factor to be considered here is the interaction that may occur between the four organisms. As mentioned earlier in the introduction, co-aggregation of bacteria species has been known to facilitate survival of organisms that prefer to grow in one environment over another. Since it is clear from individual bacterial biofilm growth that *F. nucleatum* and *L. buccalis* did very well in microaerophilic over anaerobic – both in the initial stage and in the long term – it would not be surprising that *C. showae* and *S. sputigena* rely on these two species for exchange of communication and genes to promote their own survival. The reciprocal can also be true where *F. nucelatum* and *L. buccalis* rely on the other two species for movement in the environment for search of food and optimal areas to maximize growth and survival rates.

However, when analyzing the biofilm characteristics of all four grown together, it is revealed that they did very poorly across the board. Comparing all four anaerobes grown together to their individual growth reveals that they had the lowest biomass, average thickness, and average surface area while having one of the highest surface to volume ratio. This suggests that although bacteria species may be interacting amongst themselves, these interactions may not

necessarily be one of a beneficial relationship. It seems that competition among organisms is intense even in the presence of resources that would ordinarily be abundant.

This competition is further explored when the anaerobes are individually and collectively grown in the presence of HCT116 cells (Fig. 12). Here, co-cultured experiments in an anaerobic environment was not necessary because in reality, human beings are aerobic creatures and while the intestinal tract is generally considered to be anaerobic, the influx of oxygen in and out of the human body renders the environment microaerophilic due to the consistently basal levels of oxygen present. Confocal 2D (Fig. 13) and 3D (Fig. 14) images of individual and collective species grown in microaerophilic environments reveals tremendous information about how the anaerobes organized themselves as they grew in presence of another entity that was fairly large. In both dimensional images, one sees a "breakage" among biofilm structures of individual organisms while there is a uniform and well-establish biofilm structure when all four grow together. This "breakage" that seems to be prevalent among only individual organisms may suggest that there was a loss of communication as they struggled to grow fortified biofilms among resource-hoarding cancer cells that were already vicious and malignant in their behavior and possess no logical sense of organization.

The microbial impact on CRC is curiously evaluated in Figure 15. To be sure, HCT116 cells co-cultured with each organism – both individual and collective – was deemed statistically significant when compared to samples that grew in the absence of any bacteria growth. The most surprising result was the effect of *L. buccalis* on HCT116 cells. Upon first glance, it would appear that the growth of *L. buccalis* on the cells did not have any effect. However, from a different perspective, one could ponder if *L. buccalis* possess significant driver genes that not only promotes the growth of CRC but also increases the viability of growing cells. While one should indeed take this with a grain of salt given that the organisms were only grown on the cells for 24 h, these results do indeed warrant further investigation into the specific role that *L. buccalis* may play in the progression, proliferation, and viability of CRC. It is unlikely that this is a fault of the algorithm as it caught a substantial number of dead cells for the other organisms.

The other rather surprising result was the effect that all four bacterial organisms grown together had on HCT116 cells. It is the only other co-cultured system in which the number of live cells was greater than the number of dead cells. This is immensely surprising given that the

number of dead HCT116 cells grown in the presence of individual bacterial organisms – with the exception of *L. buccalis* – was higher than the number of live cells. One hypothesis that could explain these results is that in an environment where HCT116 cells was much more efficient at getting resources compared to individual bacteria species, the anaerobes could have been struggling and were thus locked in a deadly battle with these cancer cells. This competition enhances an already intense and harsh environment, causing the CRC cells to die as time goes on. But, in an environment where all four organisms are growing together, it could be that all four organisms have put aside their petty differences and worked together to establish an effective means of communication that not only creates conditions that are favorable to them but also encourages CRC cells to grow and survive. This conclusion should not come as a surprise given the revealing relationship between *L. buccalis* and HCT116 cells.

These results are further underlined when closely evaluating the relationship between bacterial biofilms and HCT116 cells (Fig. 16) in which all four of them grown together in the presence of CRC cells produced the highest amount of biofilm. *L. buccalis*, quite interestingly, has the second highest biofilm production, average thickness and average surface area yet the second highest surface to volume ratio. This suggests that while it struggled to adapt to its environment, it did so while still being capable of growing well-established biofilms and was quite efficient at manipulating its biofilm to access available nutrients. Furthermore, *F. nucelatum* did fairly well as it had the highest average thickness and the lowest surface to volume ratio. While this was not surprising given the growing amount of evidence that points to the relationship between *F. nucleatum* and CRC, these results do indicate the need to understand how *F. nucleatum* affects the pathogenicity of the other three organisms when grown in the presence of CRC cells inasmuch as the mysterious role of *L. buccalis* ought to be further investigated. Alongside *F. nucleatum*, all four species grown together have the next least amount of burden when adapting to a rapidly shifting environment. The collective bacterial growth also had the highest average surface area, suggesting that the four organisms grown together were extremely savvy in exposing a well-establish biofilm structure to resources while growing among cells that were already capable of being resource-efficient.

CONCLUDING REMARKS

This project has attempted to evaluate the impact of anaerobic bacteria on colorectal carcinoma by first exploring the question of whether or not these bacteria species were capable of producing biofilms. Upon finding that they could indeed produce biofilms, the question then turned to the exploration of how bacterial biofilms influence cellular viability of CRC cells. This was done by establishing several co-cultured experiments and then using software programs and algorithms to effectively analyze the relationship between the two entities. Finally, the project explored how bacterial biofilm production and establishment is developed when co-cultured with CRC cells.

Together these results have indicated that all four organisms are capable of producing well-established biofilms, which is a known contributor to disease and alters the host tissue microenvironment of the human colon. They also question the exact nature of the microbiome when growing in presence of host cells and tissues as well as point to the obvious impact of the microbial community on colorectal carcinoma. Particularly notable is the absence of non-viable cells when co-cultured with *L. buccalis*, which prompts further investigation. In the future, it would do well to test the growth of these species among different CRC cell lines and for longer periods of time to effectively evaluate the relationship between bacterial biofilms and CRC. Furthermore, later experiments should focus on monitoring the growth of normal colorectal epithelial cells in the presence of these bacteria species via miRNA sequencing to truly examine the microbial impact of anaerobic bacteria on CRC.

On the last note, this project did not manage to perform miRNA sequencing for differential expression analysis to genetically evaluate the impact of bacterial biofilms on normal colorectal epithelial cells. In the future, this option should be explored and monitored at set time intervals in order to properly capture the progression of non-malignant cells as it goes from one stage to the next. Furthermore, this project only did one of the fifteen different combinations that resulted from testing four anaerobes. To truly evaluate the "driver" potential of bacterial biofilms as well as the relationships they may exhibit, future experiments should focus on a combination of different organisms – both in presence of and in the absence of normal colorectal epithelial cells. While previous studies have identified all four organisms being significantly over-represented in colorectal tumors and forming a unique microbial signature, some of the species – such as *F. nucleatum* and *C. showae* – co-aggregated in culture. It could be that relationships

between certain species over another could be more effective in encouraging these cells to adopt carcinogenic characteristics.

Lastly, any meticulous observer will raise the question of why this project did not use *L. hofstadii* instead of *L. buccalis* when the results clearly show that, although close, the former has higher read-pair alignments associated with CRC microbial signature than the latter. To be sure, the acquisition of *L. hofstadii* was relentlessly pursued but was met with obstacles and dead-ends each time. Future experiments should undoubtedly include *L. hofstadii* to properly and thoroughly examine the relationship between bacterial biofilms and CRC as well as assessing the individual and collective role of anaerobic microbes in the carcinogenesis of the colorectal mucosa.

REFERENCES

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics international*, *11*(7), 36-42.

Allos, B. M., & Blaser, M. J. (1995). Campylobacter jejuni and the expanding spectrum of related infections. *Clinical Infectious Diseases*, 1092-1099.

Allos, B. M., & Blaser, M. J. (1995). Campylobacter jejuni and the expanding spectrum of related infections. *Clinical Infectious Diseases*, 1092-1099.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25-29.

Bäckhed, F., Fraser, C., Ringel, Y., Sanders, M., Sartor, R., Sherman, P., . . . Finlay, B. (2012). Defining a Healthy Human Gut Microbiome: Current Concepts, Future Directions, and Clinical Applications. Cell Host & Microbe, 12, 611-622.

Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K., & Vogelstein, B. (1990). Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science*, *249*(4971), 912-915.

Banwait, J. K., Goettsch, K. A., & Bastola, D. R. (2016). Proteins localised to human mitochondria show structural element enrichment in their corresponding mRNA. *International Journal of Computational Biology and Drug Design*, *9*(1-2), 120-134.

Baylin, S. B., & Jones, P. A. (2011). A decade of exploring the cancer epigenome—biological and translational implications. *Nature Reviews Cancer*, *11*(10), 726-734.

Beck, T. N., Chikwem, A. J., Solanki, N. R., & Golemis, E. A. (2014). Bioinformatic approaches to augment study of epithelial-to-mesenchymal transition in lung cancer. *Physiological genomics*, *46*(19), 699-724.

Benedict, W. F., Murphree, A. L., Banerjee, A., Spina, C. A., Sparkes, M. C., & Sparkes, R. S. (1983). Patient with 13 chromosome deletion: evidence that the retinoblastoma gene is a recessive cancer gene. *Science*, *219*(4587), 973-975.

Bernard, K. A. T. H. R. Y. N., Cooper, C. H. R. I. S. T. I. A. N. E., Tessier, S. U. S. A. N., & Ewan, E. P. (1991). Use of chemotaxonomy as an aid to differentiate among Capnocytophaga species, CDC group DF-3, and aerotolerant strains of Leptotrichia buccalis. *Journal of clinical microbiology*, *29*(10), 2263-2265.

Bernards, R., & Weinberg, R. A. (2002). Metastasis genes: a progression puzzle. *Nature*, *418*(6900), 823-823.

Bister, K. (2015). Discovery of oncogenes: The advent of molecular cancer research. *Proceedings of the National Academy of Sciences*, *112*(50), 15259-15260.

Bolstad, A. I., Jensen, H. B., & Bakken, V. (1996). Taxonomy, biology, and periodontal aspects of Fusobacterium nucleatum. *Clinical microbiology reviews*, *9*(1), 55-71.

Bradshaw, D. J., Marsh, P. D., Watson, G. K., & Allison, C. (1998). Role of Fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. *Infection and immunity*, *66*(10), 4729-4732.

Cai, J., Xu, X., Qian, J., Deng, Z., Cai, Z., Tang, T., . . . Zhang, K. (2012). Expression of miR-21, miR-31, miR-96 and miR-135b is correlated with the clinical parameters of colorectal cancer. Oncology Letters Oncol Lett, 4, 339-345.

Cancer Genome Atlas Network. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*(7407), 330-337.

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., ... & Holt, R. A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome research*, *22*(2), 299-306.

Centers for Disease Control and Prevention. (2015). Colorectal (Colon) Cancer. Retrieved June 3, 2015, from http://www.cdc.gov/cancer/colorectal/

Chen, X., Guo, X., Zhang, H., Xiang, Y., Chen, J., Yin, Y., . . . Zhang, C. (2009). Role of miR-143 targeting KRAS in colorectal tumorigenesis. Oncogene, 28, 1385-1392.

Clark, W. B., Wheeler, T. T., & Cisar, J. O. (1984). Specific inhibition of adsorption of Actinomyces viscosus T14V to saliva-treated hydroxyapatite by antibody against type 1 fimbriae. *Infection and immunity*, *43*(2), 497-501.

Costello, E., Stagaman, K., Dethlefsen, L., Bohannan, B., & Relman, D. (2012). The Application of Ecological Theory Toward an Understanding of the Human Microbiome. Science, 336(6086), 1255-1262.

Cummins, J., & Tangney, M. (2013). Bacteria and tumours: Causative agents or opportunistic inhabitants? Infectious Agents and Cancer, 8, 11-11.

De Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., ... & Lagache, T. (2012). Icy: an open bioimage informatics platform for extended reproducible research. *Nature methods*, *9*(7), 690-696.

Duesberg, P. H., & Vogt, P. K. (1970). Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses. *Proceedings of the National Academy of Sciences*, *67*(4), 1673-1680.

Edwards, A. M., Grossman, T. J., & Rudney, J. D. (2006). Fusobacterium nucleatum transports noninvasive Streptococcus cristatus into human epithelial cells. *Infection and immunity*, *74*(1), 654-662.

Ephrussi, B., Davidson, R. L., & Weiss, M. C. (1969). Malignancy of somatic cell hybrids. *Nature*, *224*, 1314-1315.

Eribe, E. R. K., & Olsen, I. (2008). Leptotrichia species in human infections. *Anaerobe*, *14*(3), 131-137.

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics*, *8*(4), 286-298.

Esteller, M. (2008). Epigenetics in cancer. *New England Journal of Medicine*, *358*(11), 1148-1159.

ETOH, Y., DEWHIRST, F. E., PASTER, B. J., YAMAMOTO, A., & GOTO, N. (1993). Campylobacter showae sp. nov., isolated from the human oral cavity. *International Journal of Systematic and Evolutionary Microbiology*, *43*(4), 631-639.

Fearon, E. R. (2011). Molecular genetics of colorectal cancer. *Annual Review of Pathology: Mechanisms of Disease*, *6*, 479-507.

Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, *61*(5), 759-767.

Fleming, M., Ravula, S., Tatishchev, S., & Wang, H. (2012). Colorectal carcinoma: Pathologic aspects. Journal of Gastrointestinal Oncology, 3(3), 153-173.

Francke, U. (1976). Retinoblastoma and chromosome 13. *Cytogenetic and Genome Research*, *16*(1-5), 131-134.

Geiser, A. G., Der, C. J., Marshall, C. J., & Stanbridge, E. J. (1986). Suppression of tumorigenicity with continued expression of the c-Ha-ras oncogene in EJ bladder carcinoma-human fibroblast hybrid cells. *Proceedings of the National Academy of Sciences*, *83*(14), 5209-5213.

Gonçalves, L. F., Fermiano, D., Feres, M., Figueiredo, L. C., Teles, F. R., Mayer, M. P., & Faveri, M. (2012). Levels of Selenomonas species in generalized aggressive periodontitis. *Journal of periodontal research*, *47*(6), 711-718.

Großhans, H., Johnson, T., Reinert, K. L., Gerstein, M., & Slack, F. J. (2005). The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in C. elegans. *Developmental cell*, *8*(3), 321-330.

Han, Y. W., Shi, W., Huang, G. T. J., Haake, S. K., Park, N. H., Kuramitsu, H., & Genco, R. J. (2000). Interactions between periodontal bacteria and human oral epithelial cells: Fusobacterium nucleatum adheres to and invades epithelial cells. *Infection and immunity*, *68*(6), 3140-3146.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *cell*, *100*(1), 57-70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell*, *144*(5), 646-674.

Harris, H. (1988). The analysis of malignancy by cell fusion. *Cancer Res*, *48*, 3302-3306.

Herman, J. G., & Baylin, S. B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *New England Journal of Medicine*, *349*(21), 2042-2054.

Heydorn, A., Nielsen, A. T., Hentzer, M., Sternberg, C., Givskov, M., Ersbøll, B. K., & Molin, S. (2000). Quantification of biofilm structures by the novel computer program COMSTAT. *Microbiology*, *146*(10), 2395-2407.

Hisamuddin, I. M., & Yang, V. W. (2006). Molecular genetics of colorectal cancer: an overview. *Current colorectal cancer reports*, *2*(2), 53-59.

Holliday, R. (1987). The inheritance of epigenetic defects. *Science*, *238*(4824), 163-170.

Hooper, L. (2001). Commensal Host-Bacterial Relationships in the Gut. Science, 292, 1115-1118.

Jančík, S., Drábek, J., Radzioch, D., & Hajdúch, M. (9). Clinical Relevance of KRAS in Human Cancers. *Journal of Biomedicine and Biotechnology*, *2010*, 1-13.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.

Knudson, A. G. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences*, *68*(4), 820-823.

Kolenbrander, P. E., Andersen, R. N., & Moore, L. V. (1989). Coaggregation of Fusobacterium nucleatum, Selenomonas flueggei, Selenomonas infelix, Selenomonas noxia, and Selenomonas sputigena with strains from 11 genera of oral bacteria. *Infection and immunity*, *57*(10), 3194-3203.

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., ... & Baselga, J. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome research*, *22*(2), 292-298.

Kraus, S., & Arber, N. (2009). Inflammation and colorectal cancer. Current Opinion in Pharmacology, 405-410.

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., ... & Rajewsky, N. (2005). Combinatorial microRNA target predictions. *Nature genetics*, *37*(5), 495-500.

Kulda, V., Pesta, M., Topolcan, O., Liska, V., Treska, V., Sutnar, A., . . . Cerny, R. (2010). Relevance of miR-21 and miR-143 expression in tissue samples of colorectal carcinoma and its liver metastases. Cancer Genetics and Cytogenetics, 200, 154-160.

Kusaba, T., Nakayama, T., Yamazumi, K., Yakata, Y., Yoshizaki, A., Nagayasu, T., & Sekine, I. (2005). Expression of p-STAT3 in human colorectal adenocarcinoma and adenoma; correlation with clinicopathological factors. Journal of Clinical Pathology, 56, 833-838.

Kwong, A., Shin, V., & Ho, J. (2014). MIR143 (MicroRNA 143). *Atlas of Genetics and Cytogenetics in Oncology and Haematology*, *18*(10), 724-727.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., ... & Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, *433*(7027), 769-773.

Lim, Lee P., Nelson C. Lau, Earl G. Weinstein, Aliaa Abdelhakim, Soraya Yekta, Matthew W. Rhoades, Christopher B. Burge, and David P. Bartel. "The microRNAs of Caenorhabditis elegans." *Genes & development* 17, no. 8 (2003): 991-1008.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Proto-oncogenes and tumor-suppressor genes.

Lustig, B., Jerchow, B., Sachs, M., Weiler, S., Pietsch, T., Karsten, U., . . . Behrens, J. (2002). Negative Feedback Loop of Wnt Signaling through Upregulation of Conductin/Axin2 in Colorectal and Liver Tumors. Molecular and Cellular Biology, 22(4), 1184-1193.

Lymphoma. (n.d.). Retrieved August 10, 2016, from http://www.cancer.org/cancer/lymphoma/

Macuch, P. J., & Tanner, A. C. R. (2000). Campylobacter species in health, gingivitis, and periodontitis. *Journal of Dental Research*, *79*(2), 785-792.

Maher, M., Finnegan, C., Collins, E., Ward, B., Carroll, C., & Cormican, M. (2003). Evaluation of culture methods and a DNA probe-based PCR assay for detection of Campylobacter species in clinical specimens of feces. *Journal of clinical microbiology*, *41*(7), 2980-2986.

Marchesi, J. R., Dutilh, B. E., Hall, N., Peters, W. H., Roelofs, R., Boleij, A., & Tjalsma, H. (2011). Towards the human colorectal cancer microbiome. *PloS one*, *6*(5), e20447.

Markowitz, S. D., & Bertagnolli, M. M. (2009). Molecular basis of colorectal cancer. *New England Journal of Medicine*, *361*(25), 2449-2460.

Martin, G. S. (2001). The hunting of the Src. *Nature reviews Molecular cell biology*, *2*(6), 467-475.

McCarthy, L. R., & Carlson, J. R. (1981). Selenomonas sputigena septicemia. *Journal of clinical microbiology*, *14*(6), 684-685.

McCoy, A. N., Araujo-Perez, F., Azcarate-Peril, A., Yeh, J. J., Sandler, R. S., & Keku, T. O. (2013). Fusobacterium is associated with colorectal adenomas. *PloS one*, *8*(1), e53653.

Moore, W. E., Holdeman, L. V., Smibert, R. M., Cato, E. P., Burmeister, J. A., Palcanis, K. G., & Ranney, R. R. (1984). Bacteriology of experimental gingivitis in children. *Infection and immunity*, *46*(1), 1-6.

Mukherjee, S. (2011). The emperor of all maladies. *London: Fourth Estate*.

Nagel, R., Sage, C., Diosdado, B., Waal, M., Vrielink, J., Bolijn, A., . . . Agami, R. (2008). Regulation of the Adenomatous Polyposis Coli Gene by the miR-135 Family in Colorectal Cancer. Cancer Research, 68(14), 5795-5802.

Oncogenes and tumor suppressor genes. (n.d.). Retrieved August 10, 2016, from http://www.cancer.org/cancer/cancercauses/geneticsandcancer/genesandcancer/genes-and-cancer-oncogenes-tumor-suppressor-genes

Orye, E., Delbeke, M. J., & Vandenabeele, B. (1974). Retinoblastoma and long arm deletion of chromosome 13. Attempts to define the deleted segment. *Clinical genetics*, *5*(5), 457-464.

Park, B. H., & Vogelstein, B. (2003). Tumor-suppressor genes. *Cancer medicine*, *6*, 87-102.

Pierotti, M. A., Sozzi, G., & Croce, C. M. (2003). Oncogenes.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., ... & Peng, Y. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, *490*(7418), 55-60.

Roca, H., Pande, M., Huo, J. S., Hernandez, J., Cavalcoli, J. D., Pienta, K. J., & McEachin, R. C. (2014). A bioinformatics approach reveals novel interactions of the OVOL transcription factors in the regulation of epithelial–mesenchymal cell reprogramming and cancer progression. *BMC systems biology*, *8*(1), 1.

Roper, J., & Hung, K. E. (2013). Molecular mechanisms of colorectal carcinogenesis. In *Molecular Pathogenesis of Colorectal Cancer* (pp. 25-65). Springer New York.

Sandoval, J., & Esteller, M. (2012). Cancer epigenomics: beyond genomics. *Current opinion in genetics & development*, *22*(1), 50-55.

Sears, C., & Garrett, W. (2014). Microbes, Microbiota, and Colon Cancer. Cell Host & Microbe, 15, 317-328.

Seshagiri, S., Stawiski, E. W., Durinck, S., Modrusan, Z., Storm, E. E., Conboy, C. B., ... & Guillory, J. (2012). Recurrent R-spondin fusions in colon cancer. *Nature*, *488*(7413), 660-664.

Signat, B., Roques, C., Poulet, P., & Duffaut, D. (2011). Role of Fusobacterium nucleatum in periodontal health and disease. *Curr Issues Mol Biol*, *13*(2), 25-36.

Suzuki, J., Sugiyama, T., Ito, K., Hadano, Y., Kawamura, I., Okinaka, K., ... & Ohmagari, N. (2013). Campylobacter showae bacteremia with cholangitis. *Journal of Infection and Chemotherapy*, *19*(5), 960-963.

Swidsinski, A., Dörffel, Y., Loening-Baucke, V., Theissig, F., Rückert, J. C., Ismail, M., ... & Schilling, J. (2009). Acute appendicitis is characterized by local invasion with Fusobacterium nucleatum/necrophorum. *Gut*, gut-2009.

Tay, A. P., Kaakoush, N. O., Deshpande, N. P., Chen, Z., Mitchell, H., & Wilkins, M. R. (2013). Genome sequence of Campylobacter showae UNSWCD, isolated from a patient with Crohn's disease. *Genome announcements*, *1*(1).

Team, R. C. (2013). R: A language and environment for statistical computing.

Tjalsma, H., Boleij, A., Marchesi, J. R., & Dutilh, B. E. (2012). A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. *Nature Reviews Microbiology*, *10*(8), 575-582.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *science*, *339*(6127), 1546-1558.

Waddington, C. H. (1939). An introduction to modern genetics. *An Introduction to Modern Genetics.*

Warren, R., Freeman, D., Pleasance, S., Watson, P., Moore, R., Cochrane, K., . . . Holt, R. (2013). Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome, 1, 16-16.

Weinberger, M., Wu, T., Rubin, M., Gill, V. J., & Pizzo, P. A. (1991). Leptotrichia buccalis bacteremia in patients with cancer: report of four cases and review. *Review of Infectious Diseases*, *13*(2), 201-206.

Wu, D., Rice, C. M., & Wang, X. (2012). Cancer bioinformatics: A new approach to systems clinical medicine. *BMC bioinformatics*, *13*(1), 71.

Yoo, C. B., & Jones, P. A. (2006). Epigenetic therapy of cancer: past, present and future. *Nature reviews Drug discovery*, *5*(1), 37-50.

Zhang, G., Chen, R., & Rudney, J. D. (2011). Streptococcus cristatus modulates the Fusobacterium nucleatum-induced epithelial interleukin-8 response through the nuclear factor-kappa B pathway. *Journal of periodontal research*, *46*(5), 558-567.