Spring 2018

First Do No Harm: Caution Against the Promises of Whole Genome Sequencing in Medicine

Sarah Seiwald
Regis University

FIRST DO NO HARM: CAUTION AGAINST THE PROMISES OF WHOLE GENOME SEQUENCING IN MEDICINE

A thesis submitted to
Regis College
The Honors Program
in partial fulfillment of the requirements
for Graduation with Honors

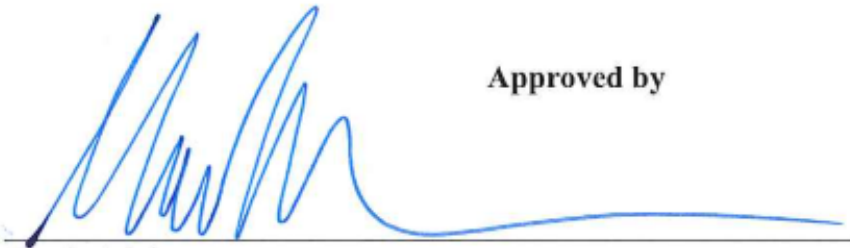
by

Sarah Seiwald
May 2018

Thesis written by

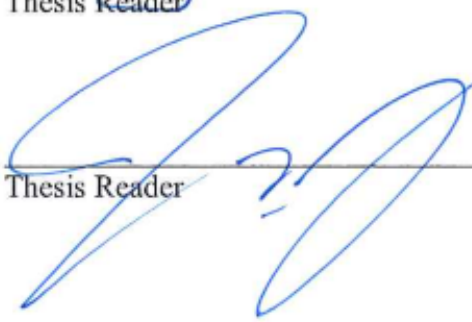
Sarah Seiwald

Approved by



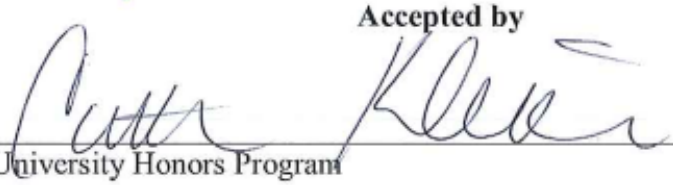
Thesis Advisor

Thesis Reader



Thesis Reader

Accepted by



Director, University Honors Program

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I thank my advisors, Dr. Max Boeck, Dr. Amy Schreier, and Dr. Marie-dominique Franco for their continuous encouragement, support, and guidance during the writing of this thesis as well as in my life pursuits. I am honored to be a member of their academic research as well as their student. Thank you to my research advisor, Dr. Boeck for always inspiring me to ask deeper questions and keeping me on track through the thesis writing process. Thank you to Dr. Marie-dominique Franco and my reader, Dr. Amy Schreier for stimulating my thinking and challenging me in the laboratory. I will never forget the fun conversations and the valuable experiences that we shared.

I would also like to thank my second thesis reader, Dr. Jason Taylor, for his added philosophical perspective and continued support during this thesis endeavor. He challenged me to address deeper insights and develop my own point of view. I thank my Honors Program advisors, Dr. Catherine Kleier, Dr. Lara Narcisi, and Dr. Thomas Howe, for always being available in my times of need and encouraging me to pursue the *Magis* while studying at Regis University.

A huge thank you to my family, without whose love and support, I would not be the woman and scholar I am today. Thank you for your love, kindness, compassion, and encouragement that lead me to apply and finish the Honors Program.

Finally, I thank Regis University's Biology Department for use of their laboratory space, materials, and equipment; along with Brittney Truong, Matt Barton, Emma Kite, and Jibin Abraham for their contribution to the genetics research project; and

TriBeta National Biological Honor Society and the Regis University Research and Scholarship Council for funding.

I. Introduction

I was first introduced to DNA during my first high school Biology course. After learning about the simple, yet intricate structure of the nucleic acid, I held the molecule in my hands. In a simple experiment with strawberries, I lysed DNA from strawberry cells using detergent. It fascinated me to see the isolated product, the substance that is known to carry the instructions for all living organisms. Today, DNA is commonplace in research laboratories and school classrooms and amazingly only 40 years ago we began deciphering its hidden messages.

In every high school and university Biology classroom, DNA's influence is evident. Isolating, cutting, and observing DNA are commonplace laboratory techniques. Starting my undergraduate education as a biology major, I quickly became acquainted with the molecule and was fascinated by the vast array of knowledge already known about its purpose in living organisms. After completing my introductory Biology classes, I devoted myself to studying biology and particularly focusing on DNA and its role in organisms. This fascination led me to inquire about genetics based research opportunities at Regis University during my sophomore year. I asked my genetics professor, Dr. Marie-dominique Franco, if I could join her genetics based research project. Dr. Franco welcomed me as part of her team to research the behavioral changes and genetic diversity of Costa Rican Howler monkeys (*Alouatta palliata*) living with habitat fragmentation. To understand whether a smaller habitat changes the inbreeding levels of howlers, Dr.

Franco introduced me to utilizing sequenced DNA to assess the population's genetic diversity. The project aims and methods captivated me and I worked to understand exactly how the white stringy substance I had once isolated would now offer insight into the behavior of another species.

DNA is the code and blueprint for the construction of living beings. Knowing its importance for all organisms, I now marvel at how easy it is to gain information quickly from any genome. Genetics class and my research with Dr. Franco introduced me to the vast information contained in the sequence of a genome and its applicability to solving questions on behavior, physical attributes, and disease. Humans have 23 human chromosomes containing DNA, tightly packaged around histone proteins. As diploid organisms, a complete set of DNA requires both a maternal and paternal copy of genes. Therefore each somatic, bodily cell contains 46 linear DNA molecules that can be 50 million to 250 million nucleotide base pairs long (National Research Council, 1988) which is roughly two meters in length when stretched out. That is certainly a large amount of information and genetic material contained in each cell.

The nucleotide base pairs that form DNA are in a specified sequence, like letters in words. When using the term "sequence", I refer to the exact order of nucleotide base pairs on a DNA strand. The nucleic acids that make up the structure of DNA are indeed complicated molecules in themselves, but there are regularities that allow for the copying and reading of the DNA "code". First, there are only four nucleotides, or DNA components. Those components fall into two categories: Purines and Pyrimidines. The purine bases, adenine and guanine, pair up with the pyrimidines, thymine and cytosine,

respectively. Differences in physical characteristics that comprise diversity are attributed to small variations in the genetic code or DNA sequence of each individual. Astonishingly, all humans share 99.9% of their genome while the remaining 0.1% is responsible for the variation that makes us each unique (Genome Research Institute, 2015; Culliton, 2003; Collins and McKusick, 2001). Mutations or alterations in the sequence of nucleotide base pairs that comprise DNA are the source for diversity as well the culprits for disease. Baffling to scientists in early DNA research was the importance of the nucleotide sequence and how its messages could be read. The order of nucleotides created the unique nucleic acid (DNA) and therefore a unique individual (Watson and crick, 1953). Almost like building blocks or the letters of a language, science soon discovered that nucleotides code for messages or the instructions required for organism function.

Due to the work of the Human Genome Project, the sequenced human genome is available for researchers and medicine to understand the function of each protein coding DNA segment, or gene. As genes and their associated proteins are mapped on the genome, their functions as well as associated diseases are being cataloged in vast online databases. All of this information is at our fingertips allowing vast progress in research and the application of the genetic sequence in medicine. As sequencing technologies improve, more and more individual genomes will be sequenced. The information obtained however is both exciting and worrisome.

The latest sequencing technology now allows sequencing on the bench top of a classroom. Such advancements are also accompanied by an extended influence of

sequencing in medical practice. Specifically at my own University, my Molecular Biology laboratory recently used the latest in rapid sequencing technology, the MinION by Oxford Nanopore. This machine allows collection of rapid sequences of any DNA molecule. I was excited for the opportunity to use this new technology to sequence hundreds of individual howler monkey DNA samples that I would prepare for my own research in Dr. Franco's laboratory. Almost futuristic in nature, rapid sequencing technology could be the tool I need to enhance my research.

Making new strides in my personal research project, the novel ability to sequence small segments of DNA in a laboratory and also for a large scale research endeavor inspired me to see other fields where sequencing technology exerts its influence. Due to the connection of DNA to disease, my inquiry led me to medicine. A hot topic of debate among doctors and ethicists is the application of whole genome sequencing for the treatment of disease. Whereas my research sequenced only a small segment of DNA, medical professionals are analyzing the full disease risk of individuals present in their entire genetic sequence. Although the promises of knowing one's disease risk are certainty tempting, I find myself questioning whether such information should be known. Here, I present the importance of DNA and its use in my personal genetics research on a Costa Rican Howler population; then, I investigate the promises behind whole genome sequencing in medicine. Considering both the positives and negatives behind knowing one's genetic disease risk, I struggle with the life altering effects of unearthing the secrets kept within one's DNA. Perhaps health providers should worry about whole genome sequencing for diagnosis and treatment as well.

II. The Code of Life

The Discovery of DNA

Human beings continually show an innate interest in the mechanisms that constitute life and configure organisms. Widely accepted in the scientific community is the paramount importance of DNA, the molecule that carries information about each individual's genetic makeup, their characteristics, and their overall state of health. DNA allows genetic traits to be passed down through generations and in doing so, that information is both retained and altered. When first discovered, society initially disregarded DNA as a structural component of cells. However, the discovery of DNA's function as a heritable information carrying molecule was a pivotal moment in Biology. Subsequently determining its structure was a momentous occasion in the scientific community that triggered an explosion of interest and research. With scientific interest, DNA took its place as cornerstone of life, the hereditary molecule. The scientific community would forever be altered as molecular biology and genetics would take hold. Looking into the history behind DNA and genetics, the scientific community's fascination with DNA after its discovery quickly shifts from a desire to determine its structure to a desire to decode its vast information. These decoding endeavors are the stepping stones towards transforming society's views on standard medical practice and using the information to serve the human condition.

Before scientists knew about DNA, they discovered that traits could be passed down from parents to offspring. The proposed mechanism of heredity began with the concept of genes and much later to large packages called chromosomes. Thomas Hunt Morgan, a geneticist who won the Nobel Prize for Physiology and Medicine in 1933, discovered that chromosomes held the genes responsible for phenotypic expression. He observed “crossing over” of chromosomes and deduced from fruit fly models that some genes were “x-linked” meaning they only appeared on the X chromosome. Morgan stated, “Medical science will here take the lead - but I hope that genetics can at times offer a helping hand.” In the early 1930s, Morgan presumed some importance of genetics in disease but denied that genetics would have a significant impact on medical treatment. During his acceptance speech for the Nobel Prize, Morgan demonstrated his reluctance to claim genes and genetic material as the driving force for disease. For treatment, Morgan stated that Doctors will take “the lead” in treatment of genetic diseases rather than knowledge of genes and gene function. In the early 1900s, when families sought to know the likelihood of their offspring acquiring a particular disease, Doctors turned to family history. This game of odds is still predominant in today’s medical practice even after the genetic code was deciphered. While Morgan’s experiments demonstrated that chromosomes have a role in heredity, he was completely unaware of the decipherability of the code that lies within those chromosomes. He fell back on the use of pedigrees and observable phenotypic expression as the best means to address disease (Morgan, 1934). His observations of chromosome separation and x- linked genes solidified the pedigree approach to medicine at least for a time.

The role of genetics in medicine beyond the pedigree approach could not have been possible without the discovery of DNA. Almost exactly 150 years ago, the allusive DNA or deoxyribonucleic acid was discovered. Unbeknownst to the discoverer, Friedrich Miescher, DNA would serve as the building block for molecular biology, genetics, and medical practice for years to come. Almost by accident, Miescher discovered DNA as a peculiar precipitate while working with the nuclei of leukocytes (white blood cells). I myself had isolated this stringy white substance in a simple high school experiment; however, Miescher was unaware of its particular importance. In his laboratory at the University of Tubingen, Miescher deduced that the unknown substance was not a protein because it could not be degraded by protease enzymes and lacked a sulfur component. Instead, the substance contained large amounts of phosphorus further indicating that it was not a protein. In these moments, Miescher was the first to give a preliminary characterization of the novel molecule. Due to its isolation from cell nuclei; he named it “nuclein”. Since these experiments in 1868, the “nuclein” name is preserved with the widely accepted term, deoxyribonucleic acid. Miescher’s discovery lives on as the first description of DNA. Miescher made attempts to describe the functional purposes of his “nuclein” and came to the conclusion that one molecule could not possibly be responsible for genetic diversity. However his later experiments demonstrated that sperm had large amounts of “nuclein”. Miescher surmised that perhaps his “nuclein” played a large role in fertilization. Little did he know that DNA would soon be discovered as the primary molecule responsible for the transfer of genetic information and thus become the source of disease.

The Molecule Responsible for Heredity

Seventy six years would pass until scientists viewed DNA as the information molecule of life. In 1944, Oswald Avery published a landmark paper describing the role of DNA in bacterial transformation experiments. Transformation is the event when DNA is taken from the environment into a bacterium. Avery's experiments with mice and heat killed virulent bacteria demonstrated that DNA could be taken up by other non-virulent bacteria strains. In doing so, those bacteria acquired new virulent characteristics. These experiments supported the claim that DNA, not proteins, was the molecule that carried genetic information for phenotypes (Avery et al., 1944). Genes are transferred from parents to offspring via chromosomes that Morgan described 10 years earlier. These genes are now defined as stretches of DNA that contain genetic information. Genes contain instructions for functional protein molecules and are interspersed throughout the genome. Avery's results showed the ability of DNA to facilitate expression of acquired phenotypes in another organism. The transformation of genetic material alone was astonishing, however the ability of DNA to hold the information required for expressed phenotypes was a momentous discovery in DNA research. Avery et al. discuss in their paper the possible implications of their results stating, "If, ...(DNA) actually proves to be the transforming principle..., then nucleic acids of this type must be regarded not merely as structurally important but as functionally active in determining the biochemical activities and specific characteristics of pneumococcal cells" (Avery et al., 1944). Avery realized the importance of his results yet would not say with full confidence that DNA

was indeed the sole responsible molecule carrying genes. A cautious Avery knew he unlocked a whole new understanding of heredity and the function of DNA.

The conclusion that DNA was functionally active in heredity required the additional viral work of Alfred Hershey and Martha Chase in 1952. To this day, the virus is a helpful model organism due to its simplicity; being solely composed of genetic information (DNA) encapsulated in a protein coat. Hershey and Chase's radioactive labeling of these virus components fortified DNA as the necessary molecule for transfer of genetic information as well as the molecule coding for genes. Well established in their experiments, the protein coat of a virus was left on the exterior of a host bacterium, while the viral DNA was released inside to hijack the inner replicative machinery (Hershey and Chase, 1952). Left on the outside, the protein coat no longer could be responsible for holding genetic information. Once again, DNA demonstrated an important role in gene expression and function. The mid-1900s saw the beginning of DNA as the genetic molecule and with it, a growing interest from the scientific community (Dahm, 2008).

The scientific community wondered exactly how did DNA retain and allow for the expression of genetic information. Surely, there must be some mechanistic explanation for DNA's functions? What structure and mechanisms would allow such a relatively simple molecule to code for all of life? With the help of X- Ray crystallography photographs taken by female scientist, Rosalind Franklin, James Watson and Francis Crick unraveled the conundrum behind the structure of DNA in 1953. Watson and Crick demonstrated in their models that DNA is a polymer made up of nucleic acid subunits, containing a phosphate backbone, 5-carbon sugar, and four bases. Intricately and with

some definable order, the bases of Adenine and Tyrosine always pair up along with the bases Guanine and Cytosine doing the same. Trial and error by Watson and Crick proved beneficial as they finally devised a structure for DNA that could explain key elements of its function (Figure 1). Watson and Crick's model depicted DNA as a double helix with phosphates on the outside, allowing the specific bases to pair in the interior portion. Like a zipper, the DNA strand could be taken apart to allow for DNA replication to take place (Watson and Crick, 1953). Watson and Crick note within their paper, "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material". Watson and Crick were not under any illusion that their structure would shake the very core of biology, fueling future molecular biology. Further research identifies that the consistent base pairing in DNA allows both strands to become "templates" for synthesis of new strands. Due to this reason, DNA proves to be a relatively simple molecule with a significant role for all organisms.

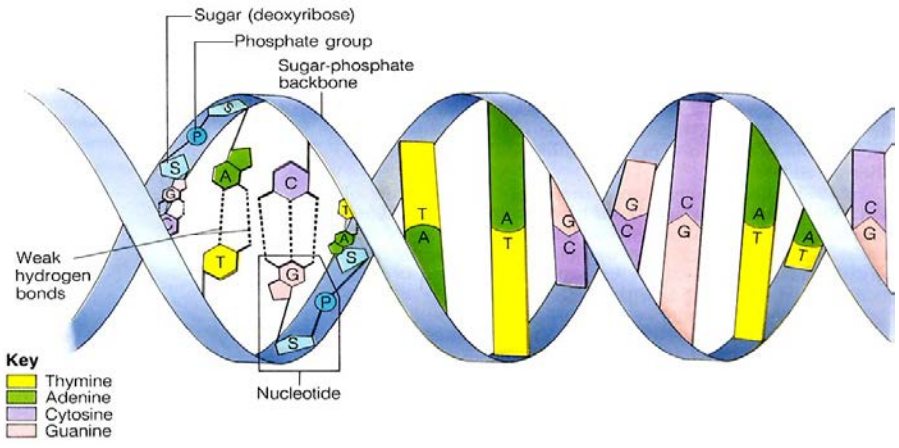


Figure 1. The helical structure of DNA. Contained in the DNA double helix is nitrogenous bases that pair in the center of the molecule. The backbone (depicted as blue ribbon) is comprised of the phosphate groups and the 5- carbon sugar of each nucleotide. Hydrogen bonding between pairs hold the bases together but are weak enough to allow the structure to unzip. Figure retrieved from Yan, 2009.

DNA is at times described as the “Code of Life”, the molecule whose unique sequences allow organisms to exist (Reichard, 1968). In the mid-1900s, the question remained as to how DNA was translated into protein, the functional component of the cell. The 1950’s established that DNA served as a template for RNA, another nucleic acid (Mukherjee, 2016). The exact mechanism behind the RNA to protein phenomena was not known. Researchers knew that somehow RNA coded for protein synthesis outside of the nucleus. However, not until 1968 did scientists decipher how the genetic code leads to the production of proteins. Three scientists are credited with the Nobel Prize in Medicine or Physiology for their work connecting protein synthesis to initial DNA sequences. Marshall W. Nirenberg, Robert Holley, and Gobind Khorana established the link between DNA and proteins by deciphering the genetic code. Nirenberg contributed by successfully synthesizing a strand of RNA and subsequently making a protein from that code. His work demonstrated that the inner machinery of cells uses translated genetic information (RNA) to form functional gene products (Nirenberg, 1968). Also looking at a sequence of ribonucleotides, Robert Holley derived the structure of tRNA. In essence, from Holley’s work, the chemical structure of DNA served as a precursor for the chemical structure of downstream proteins (Holley, 1968). Alongside Holly’s accomplishments, Gobind Khorana greatly elucidated the genetic code by synthesizing his own predetermined DNA sequence. Using a similar method to Nirenberg, Khorana analyzed resulting protein amino acid sequences to discover the triplet codon of messenger RNA and thus the code behind protein synthesis (Khorana, 1968). All three scientists demonstrated that proteins were derived from an original nucleotide sequence.

Therefore any deviation from the correct bases could result in malformed proteins and thus disease.

In 1968, at the Nobel Prize Award Ceremony at the Karolinska Institutet in Sweden, Professor P. Reichard stated in his award address that, “So far the work can be described as basic research. However, through this work we can now begin to understand the causes of many diseases in which heredity plays an important role”. Reichard was not wrong in his statement of the importance of inheritance research. Roughly thirty years prior, Thomas Hunt Morgan mentioned in his Nobel Prize speech that genetics would only supplement understanding of disease. However with the progression of science, the importance of DNA and understanding its messages became essential to understanding disease. Researchers now had the connection between DNA and phenotype; RNA messages transcribed from DNA were translated into proteins. The scientific community knew the implications of decoding the mechanism behind the human genome and they foresaw its medical implications. The knowledge gained in the mid-1900s laid the foundation for molecular biology and its continued focus on gene regulation and function to create advancements in medical practice (Reichard, 1968).

Sequencing the Code of Life

While the connection between DNA and protein products was a large stepping stone in DNA analyses, researchers still needed a way to read the nucleotide sequence and decipher the messages hidden in the genome. Only a few years later, a novel method to sequence DNA segments was devised. In the early 1970s, Frederick Sanger introduced the first “rapid” sequencing technology. Sanger’s first sequencing technique is described

as “rapid” because researchers could go to the actual DNA itself to determine the nucleotide sequence rather than taking a protein and going backwards to deduce its sequence. Rapid for Sanger was a mere 50 bp in a few days. At this rate, sequencing an entire genome would require centuries. Sanger coined his first method as the “plus and minus” technique and used DNA with radiolabeled nucleotides (Sanger et al., 1973). Utilizing this approach, Sanger sequenced the first full genome of a bacteriophage which was a momentous achievement for its time. After this success, Sanger continued to improve his technique giving rise to new approaches to understanding molecular biology and genetics (Heather and Chain, 2016). In 1977, Sanger published an additional paper on DNA sequencing and called his new method “chain termination”. This became the standard method for DNA sequencing for the next 30 years (Sanger et al., 1977). The first sequencing generation was born. Despite seemingly low rates of nucleotide sequencing, the process was fast enough to inspire a multi-billion dollar race to sequence the entire human genome.

In 1976, it took months to sequence 50 base pairs; a mere 14 years later scientists declared they wanted to sequence all 3 billion. This endeavor, a multi-billion dollar effort to sequence the entirety of the human genome, was labeled The Human Genome Project (HGP). The project formally began in 1990 and its primary mission was to determine the nucleotide sequence contained in all 23 human chromosomes, an estimated 3 billion base pairs. The National Institute of Health and the U.S. Department of Energy organized the project’s development and researchers across the globe showed interest. The sequencing techniques devised by Frederick Sanger inspired the project and with continued

improvements to sequencing rate, the project was launched with an estimated completion period of 15 years and cost of roughly 3 billion dollars (Hutchison, 2007). Contained in the human genome are thousands of genes that code for a particular protein product. The project sought to use the vast letter information of DNA sequencing and translate that into defined protein products. The project also aimed to map the genes on human chromosomes. To do this, the Human Genome Project required the DNA of twenty-one people found in Buffalo, New York to create a comprehensive view of the total human genome. Researchers wanted readable information that could be used to diagnose genetic diseases. They wanted to know what bits of insight the human genome could provide about a person's life and overall health (Kelavkar, 2006). For accurate analyses and mapping of gene locations, 3 billion base pairs would be enough to gain a complete representative human genome (National Research Council, 1988).

Upon completion of the Human Genome Project in 2003, scientists were excited for the downstream applications of a complete set of human genetic instructions. Not only would a blueprint of our genetic code be mapped out, but knowing the accurate sequence placed science on a new playing field. Innovations previously not thought of were now possible by looking at what some still call the "Book of Life" (Collins, 1988). In a 1988 lecture before the start of the project, the director of the HGP, Francis S. Collins, M.D. Ph. D., discussed that a "map of human genetic terrain…would lead [scientists] to previously unimaginable insights, and from there to the common good". The common good to which Collins refers to is partly the benefits that would come to medicine and the future treatment of human disease. By knowing the genetic code,

deviation from its sequence could be stepping stones to disease treatment. Knowing the correct nucleotide sequence could prove helpful in identifying the source of disease.

The HGP sequencing results ushered in a new form of research that allowed genes to be found and studied for their interactions and products. Researchers could develop logical strategies to address human disease and find its source in the genome. In order for scientific progress to be made, the sequencing results needed to be made free and public. Society's need for free genomic information was especially evident in research surrounding BRCA1, the gene that causes breast cancer. Myriad, a genetic testing company patented BRCA1 and BRCA2, the two genes largely connected to breast and ovarian cancer in 1988, thus limiting any further research surround those genes. The two genes were untouchable by researchers unless they wanted to pay money to Myriad (Goldacre, 2010; Rettner, 2013). Thankfully in 2013, the Supreme Court ruled that genes could no longer be patented, therefore freeing researchers to investigate genes further (Stordahl, 2013). The HGP made free and public access a paramount goal. Collin described in his 1988 lecture that "The DNA sequence arms scientists seeking to understand disease with new information and techniques to unravel the mysteries of human biology". In 1988, Collin recognized that DNA was the key to tackling diseases and knowing its sequence would dramatically help in diagnosis, prevention, and treatment options (Collins, 1988). To this day, the vast mysteries of DNA are still being unraveled with the help of the immense contribution from the Human Genome Project (Hamdoun et al., 2017).

The Genome Era

The Human Genome Project officially ended in 2003, ahead of schedule and under budget, costing 2.7 billion dollars (Davies, 2010). Currently, the entire sequence of DNA contained in human chromosomes is made freely available to the public for research use. The National Research Council, assigned to process the feasibility of the sequencing project, reported a preliminary estimate for the total number of human genes to be around 100,000 (National Research Council, 1988; A Brief History of the Human Genome, 2012). Today, the accepted gene number is around 19,000 and more genes and their possible variants are being mapped and cataloged (Ezkurdia, 2014).

From the results of the HGP, more than 1,800 diseases and their associated genes are currently known. As a result, 200 quick DNA analyses tests are used to screen for various genetic risk factors (van El et al., 2013). A "risk factor" is defined as a genetic sequence or fault in a gene that disables its intended function. Without the correct function of genes, various diseases arise and place individuals at risk. The Genome Era began upon the completion of the human genome. Mapping the location of genes, since the turn of the century has thus far allowed quick diagnosis of various diseases and has spurred the use of sequencing in diagnosis methods. After the completion of the project, many scientists predicted that sequencing and genomics would dramatically change health care (Guttmacher et al., 2005). Treatment options would look to molecular solutions and fully understanding the source of an individual's condition. A. Alan Guttmacher, Ph.D. noted that collecting an individual's unique sequence would likely be standard medical practice in the next ten years (2005). Although this is not yet the case,

the medical field has implemented this technology in oncology practice. Researchers and physicians are realizing the extreme benefit that sequencing can play in diagnosis and treatment for patients (Nagarajan, et al., 2017)

While the Human Genome Project required the DNA of twenty-one people, some sequencing efforts today focus on retrieving the full genetic code individually for thousands of people. Held in an individual genome are small nucleotide variations that give rise to diverse phenotypic expression. These variations can take the form of mutations, or polymorphisms. Mutations are changes to the nucleotide sequence at the DNA level and can lead to deleterious problems in protein expression. One such disease is Retinoblastoma (RB), a recessive disease of the eye in which malignant tumors form in the retina tissue due to random and/ or heredity mutations in a gene on chromosome 13 (Matea et al. 1997). Retinoblastoma forms from two mutated copies of the RB1 gene. The RB1 gene codes for the RB protein that regulates the cell cycle by arresting cells in the G1 phase or growth phase. Inheritance of the disorder requires two mutated copies of the gene to display malignant eye tumors. Patients who inherit a paternal mutated gene largely exhibit bilateral eye tumors while patients with sporadic copies commonly only show unilateral incidence (Mateu et al. 1997).

Due to the hereditary nature of retinoblastoma, families with incidence of the disease are advised to have their children come in for regular examinations from their ophthalmologist in the first few years of life. However, prescreening genome analysis can be done to determine whether a child indeed has two mutated RB1 gene copies. Common techniques utilize Polymerase Chain Reaction (PCR) to amplify the q14.2 locus and

sequence the DNA specifically looking for mutations (Richter et al. 2003). A linkage analysis of a child's DNA could also look for polymorphisms associated with mutated copies which serve as precursors to the malignant tumor formation (Mateu et al. 1997). Prescreening genetic analysis offers patients a definitive way to know whether they will develop the malignant eye tumors. When families seek answers, genetics and sequencing can offer some answers. PCR methods and sequencing offer families peace of mind and save them costly doctor visits (Richter et al. 2003).

Current research on retinoblastoma is yielding exciting prescreening and treatment possibilities for individuals with early signs of the cancer. Recent prescreening research uses next generation sequencing (NGS) with an in-house analysis pipeline to detect RB in patients. The in-house bioinformatics pipeline is created from careful analysis of SNVs, insertion/ deletions, as well as differentiating between somatic and germline RB. A more targeted approach using NGS more efficiently determines mutations in the RB gene and further the nature of the mutation. Conventional methods such as Sanger sequencing require that 27 exons are separately sequenced while NGS uses fewer assays of DNA and is sensitive to specific mutations. NGS can be used to enhance risk assessment to future generations (Devarajan et al. 2015).

The polymorphisms that NGS seeks to identify in patient samples are either from a single nucleotide (SNP's), a variation in copy number (CNP's) of repeated bases, or indels which are base deletions or insertions. However, polymorphisms can also be in the noncoding regions of DNA and thus do not cause downstream problems. These polymorphisms prove useful in determining lineage relationships (Sebat et al., 2004). To

be concise, laboratory researchers focus on these variable portions of individual genomes to quickly answer genetic based questions in a variety of species. These highly variable regions allow assessments of familial relationships and determination of particular allele presence. While the Human Genome Project was a momentous undertaking in itself, sequencing has since then improved dramatically allowing sequencing machinery and the necessary manpower to decrease in size. Faster computing capabilities allow more information to be stored and derived from sequencing efforts, thus allowing research with next-generation sequencing to take hold in the scientific community (Schuster, 2008). The biological questions that can now be asked are practically limitless with the invention of rapid, relatively cheap sequencing technologies. The question remains however, how will these technologies revolutionize the understanding and treatment of disease for researchers and medical professionals alike.

III. Comparison of genetic structure of a mantled howler monkey (*Alouatta palliata*) population at the edge and interior of a Costa Rican forest fragment

In my previous chapter, I emphasized the discovery of DNA as the hereditary molecule of life and argued that humanity's preoccupation with understanding our genetic makeup and seeking an understanding of disease are primary driving forces for the use of sequencing technologies. Since the discovery of DNA, sequencing technologies allowed vast achievements in the understanding of organisms and how they function. Geneticists and molecular biologists use direct sequences of DNA to understand how specific genes are transcribed and also how genetic mutation and protein malfunction can lead to disease. Interestingly, both the non-coding portions of DNA and the coding elements allow scientists to study molecular interactions and the presence of disease risk factors in human and non-human populations. Still to this day, scientists are unlocking the vast social and medical potential that DNA sequences hold.

DNA molds not only molecular biology and genetics but also extends into other scientific disciplines. Considering non-coding segments of DNA, sequencing is instrumental for a variety of scientific questions. Specifically, scientists can sequence these portions to assess kin relationships and inbreeding levels. Sequencing polymorphic DNA regions, or sequences of DNA that have varied numbers of nucleotide repeats,

allows a quick determination of heterozygosity versus homozygosity. For diploid organisms which contain a paternal and maternal copy of their genes, a gene can have multiple alleles or forms. An individual with two copies of the same allele of a gene on both chromosomes is considered to be homozygous. Conversely, if there are two different alleles present, the individual is heterozygous. The proportion of heterozygous individuals in a population is an indicator of genetic diversity and overall population health. High genetic diversity is associated with higher chances of populations being able to fight off disease and other environmental challenges (Woodruff, 1989). Thus, by evaluating the proportion of heterozygous individuals in a population, the level of inbreeding in a population can be indirectly assessed with simple DNA sequencing analyses.

Sequencing the DNA of individual organisms gives insight into the health of individuals as well as into a population's overall genetic diversity. Molecular biologists commonly use sequencing to determine molecular interactions in organisms. Recently, sequencing has also been used to investigate the negative consequences of environmental stress on community stability in non-human primate populations (Winkler et al., 1999). By peering into the genome of a species, scientists can interpret how a population copes with environmental stress. Sequencing proves useful in studying non-human primate populations because we can answer behavioral questions by looking at the DNA code and employing statistical analyses. Environmental pressures can influence social cohesion, migration and other activities of inhabitant species. These changes in activity and behavior can ultimately lead to increased inbreeding in populations, thus causing a

change in the overall genetic diversity of a species over time. The genetic sequences of individuals prove useful for scientists to draw conclusions concerning the effects of environmental factors on populations.

My own interest in DNA began when I started undergraduate research in a Genetics laboratory working with non-human primate DNA. As a student researcher at Regis University, I personally worked with DNA acquired non-invasively from the non-human primate species, mantled howler (*Alouatta palliata*). My research on *A. palliata* utilizes DNA obtained from fecal samples to understand how a changing environment is influencing their behavior and ultimately altering their population genetic diversity. Fecal samples allow researchers to study the monkeys' genetics to aid in conservation efforts and understand how this particular primate species responds to habitat fragmentation. Luckily genetic analyses utilizing next generation sequencing allows a thorough, non-invasive investigation.

My project looks specifically at the level of inbreeding in groups of *A. palliata* in the edge versus the interior of a forest fragment. Entering into the project, I knew that I would be working with DNA samples obtained from monkey feces, but I was unsure of exactly what information would be gained from the DNA sequences. I soon learned that sequencing highly variable DNA regions allows a quick determination of heterozygous versus homozygous monkey individuals. By comparing the number of nucleotide repeats in segments of DNA, the proportion of heterozygous individuals and thus the level of inbreeding in a population can be investigated. In this chapter, I present my genetics

research using next generation sequencing technologies to assess the level of inbreeding in a Costa Rican howler population that faces severe fragmentation of forest habitat. Sequencing thus proves to be a useful tool not only for molecular genetics but for ecology as well.

Introduction

The human impact of deforestation on non-human primate populations is a center of concern for a variety of primate communities in the New World (Pope, 1996; Winkler et al., 2004; Winkler et al., 1999). Over the last few decades, deforestation contributing to habitat loss and fragmentation is a leading threat to biodiversity (Clarke, Zucker, & Scott, 1986; Arroyo-Rodriguez & Diaz, 2010). Destruction to habitat reduces the resources available to native species and can result in animals migrating to other areas (Oklander, Kowalewski, & Corach, 2010). The fragmentation of forest into disconnected patches increases the ratio of forest edge to interior. The edge effects from fragmentation present new ecological challenges for primate species that potentially can alter population density, social cohesion, and group sizes (Broadbent et al., 2008; Schwitzer et al., 2011). There is less ability for primates to disperse to new groups and a potential decrease in gene flow may ensue. As a result, the genetic diversity of a species could be detrimentally affected from increased inbreeding. With increased infringement of human activities on primate habitats, it is imperative to understand how monkey populations are being affected by increased edge effects.

Edge effects are the structural changes of ecosystems resulting from newly formed boundaries that allow exposure to new biotic and abiotic conditions (Schwitzer, 2011; Murcia, 1995; Lidicker, 1999). The varied microclimate near the boundary of two different habitat types results from altered amounts of light, wind, and moisture that potentially produce a varied microclimate (Chen et al., 1999; Murcia, 1995; Laurance et al., 1998). This difference in abiotic factors can result in fewer tall trees and less canopy cover at the forest edge. Further, Chapman (1998) reports that the average tree diameter at breast height (DBH), an indicator of food availability, is lower at the forest edge (Wilcove et al., 1986). For these reasons, subpopulations of primate species at the forest edge would likely need to travel farther distances to acquire their food. At the forest edge, primates are more susceptible to foreign species of plants and animals that can more easily invade and further create ecological shifts (Laurance et al., 1998). With a larger edge to interior forest ratio, inhabitants of fragments are particularly vulnerable to these ecological changes from edge effects and would likely need to adapt to these habitat changes (Laurence et al., 1998).

In particular, forest fragmentation and the associated edge effects pose threats to the survival and overall population numbers of mantled howler monkey (*Alouatta palliata*) populations (Estrada and Coates-Estrada, 1996). *A. palliata* is a large neotropical monkey species that is recognized as being resilient to changes in habitat that lead to smaller habitat size (Estrada, 1982; Bicca-Marques, 2003; Emmons and Feer, 1997). Howler monkey groups are commonly large (>10 individuals) with multiple males and females (Clarke, Zucker, & Scott, 1986). The howler monkeys' resilience to smaller

habitat is due in part to the species' dispersal patterns. Both sexes disperse from their natal groups which decreases the chances of inbreeding in howler groups (Pope, 2000). They also experience limited contest competition (Wang & Milton, 2002) as this species is mainly folivorous and seasonally supplements its diet with fruit (Crockett & Eisenberg, 1987; Milton, 1981).

Howler monkeys that live in forest fragments with increased edge effects have decreased habitat size, resource availability, and ideal plant life (Arroyo-Rodríguez and Dias, 2010). Furthermore, howler monkeys that inhabit fragmented or isolated forests have decreased opportunities for dispersal to other groups. Sometimes monkeys can find corridors to other fragments in order to relocate to other howler groups (Arroyo-Rodríguez and Dias, 2010). However, in these narrow corridors there is higher threat of predation. Therefore, there is still cause for concern for the long-term impact of decreased habitat size on the survival of the species (James, 1992). While both male and female howler monkeys disperse to other howler groups once they are sexually mature, in smaller and disconnected habitats howler individuals from the same initial group often migrate together to another group contributing to less gene flow within populations (Pope, 2000; Oklander et al., 2010).

Forest fragmentation and its resulting edge effects have the potential to negatively impact group numbers and composition, social cohesion, overall activity budgets, and dietary composition (Schwitzer et al., 2011). Although the exact consequences of fragmentation are not known, Irwin et al. (2010) observe that different subgroups of a

population can experience varied responses in different areas of a forest fragment. Due to the decrease in food resource availability and varied microclimate at the forest edge, primate subgroups will likely have dissimilar behavior to groups in the forest interior. Stoner (1996) specifically reports that the howler monkey selects its habitat based upon the density of preferred trees present for food. At the forest edge, lower food availability will likely have negative impacts on group size and social cohesion. Chapman (1990) reports that *A. palliata* groups organize based on food availability and recorded that in fragments, smaller, more dispersed groups were at the forest edge with larger groups in the interior. These changes among subgroups can lead to differences in genetic composition over time.

Amidst severe environmental changes, it is important for species to maintain genetic diversity. Genetic diversity increases the likelihood of species surviving against stressors which include fighting off disease (Acevedo-Whitehouse et al., 2003; Woodruff, 1989). This diversity is the result of gene flow, or the transfer of alleles from one population to another. As mentioned previously, alleles are the variations of a gene that are present in a population (Oklander, 2010). When a large proportion of individuals carry two different alleles of a gene, the population has a better chance of surviving environmental stressors such as disease. However, habitat fragmentation is a major threat to gene flow and thus genetic diversity. A larger forest edge to interior ratio can lead to higher levels of inbreeding among howler groups on the edge, thereby decreasing population diversity. At the forest edge, *A. palliata* groups are smaller and more spread out, which likely will increase the incidence of inbreeding in edge subpopulations

(Chapman, 1990). With more inbreeding, the survivability of the population will decrease. Diversity is important to maintain because it increases the likelihood that some individuals will have certain allele variations that can help them better survive and pass on their genetics (Oklander et al., 2010).

Increased fragmentation of forest habitat coincides with higher incidences of inbreeding among howler monkeys (Oklander, 2010). Specifically, past studies assessed the genetic structure of *A. palliata* to determine the species' behavioral response to human infringement on its habitat due to farming and other practices such as logging (Winkler et al., 1999; Bastos et al., 2010; James, 1992; Pope, 1996). Winkler et al. (1999) report that two separate howler populations showed high levels of microsatellite diversity despite being in areas of heavy land use and deforestation. Often, howler monkeys are associated with being resilient to changes in habitat structure (Bicca-Marques, 2003; Emmons and Feer, 1997). Regardless, there is conflicting evidence on whether human activities and consequent change in habitat structure detrimentally affect howler monkey genetic diversity (Estrada and Coates-Estrada, 1996; Milton et al., 2008; Van Belle et al. 2012). Dispersal of groups are limited in fragmented habitat and different responses of subgroups within the same fragment can occur, potentially leading to different levels of inbreeding between subgroups (Irwin et al., 2010). Thus, genetic structure can be an indicator of the effect of habitat fragmentation and specifically edge effects on the stability of monkey populations.

Analysis of microsatellite loci is commonly used to assess inbreeding levels (Bastos et al., 2010; James, 1992; Pope, 1996; Winkler et al., 2004). Microsatellites are short DNA sequences with varied numbers of nucleotide repeats across individuals, groups, and populations. These loci can be amplified using Polymerase Chain Reaction (PCR) which exponentially replicates the specific segment of DNA containing the genetic marker, or microsatellite (Morin & Woodruff, 1996; Pruetz & Leasor, 2002). By comparing variation in microsatellite repeats and numbers, the level of genetic variation can be quantified (Morin & Woodruff, 1996; Pruetz & Leasor, 2002). For each microsatellite, individuals are homozygous if they carry two of the same allele (with same number of repeats) on both chromosomal copies. Individuals are heterozygous for a particular microsatellite locus if they carry two different alleles (with different number of repeats). Different microsatellite polymorphisms are thus used as genetic markers for heterozygosity and homozygosity (Oklander, 2010). We can measure the genetic health of a population Using Hardy-Weinberg equilibrium. Hardy Weinberg equilibrium utilizes the frequencies of alleles in a population to characterize the genetic structure of populations without outside evolutionary pressures (Woodruff, 1989). A population is considered genetically healthy if it displays high levels of genetic variation at or above Hardy Weinberg equilibrium.

In Costa Rica, tropical rainforests have experienced human-induced deforestation for farming and logging purposes since the 1970s (Garber, Molina, & Molina, 2010; Arroyo-Rodriguez & Mandujano, 2006). The boundary at forest edge is prominent due to nearby cattle ranches and fruit plantations (Garber et al., 2010). As a result, there is

reduced forest habitat conducive for providing high-quality food resources and migration opportunities for wildlife (Garber, Molina, & Molina, 2010; Arroyo-Rodriguez & Mandujano, 2006). The La Suerte Biological Research Station (LSBRS) in Costa Rica serves as "home" for native *A. palliata* groups and has experienced this fragmentation. The forest at La Suerte comprises a large and small forest fragment that is connected by a corridor. Here I investigate the consequences of edge effects on the genetic diversity of mantled howler monkeys in the forest fragments at the La Suerte Biological Research Station (LSBRS) in Costa Rica by analyzing subgroups at the edge and interior of forest fragments. In 2016, Schreier et al. (2016) reported that the La Suerte forest fragments have higher species richness and diameter at breast height (DBH), which signifies higher fruit production, in the forest interior as compared to the edges (Chapman et al., 1992). The forest edge is defined as within 100m from the forest boundary. I analyzed the genetic diversity of the howler monkeys at the forest edge versus interior via next generation sequencing of microsatellite loci to see whether there is a difference in the level of heterozygosity between the howler monkeys at the edge and in the interior of the forest.

The goal of my research is to directly assess the differences in genetic structure of the La Suerte mantled howler monkey (*A. palliata*) groups between the forest edge and interior using microsatellite analysis. Due to the detrimental effects of increased forest edge, resulting in less food availability and habitat (Estrada and Coates-Estrada 1996) as well as decreased dispersal ability to other howler groups (Oklander et al., 2010; Winkler, 2004), I predict that *A. palliata* groups that live in the interior portion of the forest

fragments will have higher levels of heterozygosity compared to groups that live near the edge. This research will be compiled with previous studies on howler monkey response to forest fragmentation to determine what extent howler populations are affected by human infringement and fragmentation of their habitat. Further, the data collected from this study can aid in analysis of gene flow of fragmented populations in general.

Methods

Study Site

Genetic samples were collected at the La Suerte Biological Research Station (LSBRS) in Costa Rica (10° 26'N, 83°46'W). The forest surrounding LSBRS contains a small and large forest fragment connected by a corridor which together comprise 300 hectares (ha) of primary and secondary forests and regenerating pastures (Garber et al., 2010). La Suerte is home to three primate species: the mantled howler (*A. palliata*), Central American spider monkey (*A. geoffroyi*), and the white-faced capuchin (*C. capucinus*). Previous studies report that there are 8-1 howler groups with approximately 15 individuals each (Pruetz & Leasor, 2002; Garber et al. 2010). The boundary at forest edge is prominent due to nearby cattle ranches and banana and pineapple plantations (Garber et al., 2010). Schreier et al. (2016) report that the La Suerte forest fragments have higher species richness and DBH, or diameter at breast height, in the forest interior compared to the edges. The forest edge is defined as within 100m from the forest boundary.

Sample Collection

During the summers of 2014, 2015, and 2016, Regis University students and Dr. Amy Schreier carefully collected and cataloged fecal samples from *A. palliata* at LSBRS. To collect fecal samples, researchers searched for howler individuals and waited for defecation. Researchers used wood spatulas to transfer approximately 5.0 g of fresh feces into 5-ml collection vials containing 2.5 ml of RNAlaterTM. GPS locations were used to determine whether each sample was from the forest edge or interior. Sample vials were labeled with the date, sex, and age-class and stored at room temperature. At the end of the sample collection periods in Costa Rica, all fecal samples were shipped to the Biology Department at Regis University and stored at -20^{o}C.

186 fecal samples were collected for DNA analyses. Of these, samples were designated as edge or interior based on the location collected and distance in relation to the forest edge. Edge was first defined as 50m resulting in 37 edge population samples and 149 interior population samples. Defining the edge as 100m, there are 91 edge and 95 interior samples.

DNA Extraction

To analyze the difference in genetic diversity of howler monkey groups living in the forest edge compared to the interior of the two forest fragments of LSBRS, I extracted and purified the DNA of howler individuals from the fecal samples using a QIAamp DNA Stool Mini Kit (Qiagen). The DNA purification procedure has been developed to preferentially extract endogenous DNA from the individual's GI tract as opposed to the

microbiome and plant materials found in monkeys' fecal matter. The procedure allows for high amounts of endogenous DNA to be purified that can be subsequently amplified using PCR (Nechvatal et al., 2008).

Polymerase Chain Reaction

To assess edge effects from fragmentation on genetic diversity of subpopulations, microsatellite primers specific to *A. palliata* were used with PCR to amplify specific and diverse microsatellite sequences of DNA (Ellsworth and Hoelzer, 1998). I specifically used primers designed to amplify the Ab06 and Amp1 microsatellite regions (Ellsworth & Hoezler, 1998; Milton et al., 2008). The PCR cycling conditions for these primers are: Initial denaturation at 94°C for 4 minutes, 30 cycles at 94°C for one minute, 1 minute at a primer specific annealing temperature. The primer annealing temperatures are 50°C for Ab06 and 64°C for Amp1. This cycle condition is followed by 1 minute at 72°C, ending with extension for 10 minutes (Van Belle et al., 2012). Each PCR tube will contain 20µl total volume comprising 10mM Tris HCl (pH 8.3), 50 mM KCl, 0.25mM MgCl2, 0.5mM of the respective primer (forward and reverse), 0.25 mM of a complete mixture of all dNTPs, 300ng of purified DNA, and 5U of Taq polymerase (Fermentas). The PCR amplification products were visualized through gel electrophoresis to ensure the proper amplification of each polymorphism during PCR.

Next Generation Sequencing

Barcode Primers: To sequence the DNA, PCR reactions were repeated with barcoded primers specific to each sample. Each forward and reverse primer contained a four-nucleotide barcode. By combining forward and reverse primers with different barcode sequence combinations, each DNA PCR sample can be specifically marked with a preceding and ending nucleotide sequence. In next generation sequencing this aids in individual determination.

Magnetic Bead PCR Purification: In a PCR tube, 65ul of DNA was combined with 2X volume of Ampure XP beads, mixed with pipetting and incubated at room temperature (5 mins). Samples were placed on a magnetic stand (1 min) and supernatant was removed. Beads were washed twice with 75% Ethanol (20 ul). DNA was eluted from the beads with 50ul of H_2O.

End Repair/ A- tailing: Library preparation was performed according to methods outlined by Oxford Nanopore (Potter, n.d.). First, I completed end repair of the PCR products. The following components listed in Table 1 were mixed gently and incubated at 20^{o}C (5mins) and at 65^{o}C (5mins).

Table 1. End Repair Components

Reagents	**Volumes (ul)**
~ 1 ug DNA	45
Rxn buffer ~ end prep	7
Ultra II End Prep enzyme	3
Nuclease Free H2O	3
Total Volume	60 total

Second Magnetic Bead Purification: The second purification was performed with the following differences: 60 ul of Magnetic bead mixture was added and the final elution was with 31ul nuclease free water.

Adapter Ligation: The components in Table 2 were mixed together and incubated at room temperature for ten minutes.

Table 2. Adapter Ligation to DNA

Reagents	**Volume (ul)**
End-prepped DNA	30
Adapter Mix	20
Blunt/TA Ligation Master Mix	50
Total	100 total

Final Magnetic Bead Purification: A final magnetic bead step was completed with the following differences: 40 µl AMPure XP beads was added to the adapter ligation reaction, the beads were washed twice with 140 µl of the Adapter Bead Binding buffer to remove excess adapters and final elution in 15 µl of Elution Buffer with a 10 minutes incubation period at room temperature.

Loading the MinIon: MinIon was primed according to the recommended methods obtained from Oxford Nanopore (Potter, n.d.). A total of 1000 µl of Flow Cell priming mix (Table 3a.) was placed in the priming port in a two-step process. The DNA library (Table 3b.) was loaded drop wise onto the flow cell.

Table 3. a. Flow Cell Priming Mix. b. DNA Library Preparation

a.

Reagent	**Volume (ul)**
RBF	480
Nuclease-free water	520
Total	1000 total

b.

Reagent	**Volume (ul)**
RBF	35
Nuclease- free water	2.5
LLB	25.5
DNA library	12.0
Total	75.0total

Statistical Analyses

Using population genetics statistical analysis, I assessed the null hypothesis that there is genetic equilibrium in the howler populations through Hardy-Weinberg Equilibrium (HWE) and indirectly assessed the level of inbreeding in the population using a chi-square goodness of fit test to determine which population genotype ratios were significantly different from expected values (Table 4). I subsequently analyzed the simulated allele frequencies via F-statistics to assess the heterozygosity to homozygosity ratio across the subpopulations at the forest edge versus interior and in the population as a whole (Wright, 1951). Using formulas derived by Peakall and Smouse (2009) and outlined by Barton (2016) and Truong (2017), I first determined the observed (Ho) and expected (H_e) heterozygosities and then derived the fixation indices, F_{IS} and F_{ST}, which are direct measures of population genetic variability (Weir & Cockerham, 1984) (Box 1 and 2). The fixation indices for the subpopulations were then compared to determine the level of inbreeding occurring between groups (Oklander et al., 2010) (Table 5).

Box 1: Subpopulation and mean heterozygosity measurements (Peakall and Smouse, 2009, Barton, 2016; Truong, 2017). The subpopulations fall into two categories: edge and interior (50m) and edge and interior (100m).

H_o is observed heterozygosity of a subpopulation and is given as the proportion of *N* samples that are heterozygous at a given locus.

$$\text{Ho} = \frac{\text{No. of Heterozygotes}}{N}$$

H_e is expected heterozygosity of a subpopulation and is the sum of the squares of allele frequencies subtracted from 1.

$$H_e = 1 - \sum p_i^2$$

$\overline{H}_o$ is mean observed heterozygosity and is the sum of both subpopulation's observed heterozygosity, divided by total number of subpopulations, *k*.

$$\overline{H}_o = \sum_{i=1}^{k} H_o/k$$

$\overline{H}_e$ is mean expected heterozygosity and is the sum of both subpopulation's expected heterozygosity, divided by the total number of sub-populations, *k*.

$$\overline{H}_e = \sum_{i=1}^{k} H_e/k$$

H_T is total expected heterozygosity and is the sum of the squares of allele frequencies for the entire population subtracted from one.

$$\overline{H}_e = \sum_{i=1}^{k} {p_{Ti}}^2$$

Box 2: Fixation indices. $F_{IS}$and F_{IT} values are between -1 and +1. Values near zero indicate non-random mating, positive values indicate inbreeding and an excess of homozygosity, and negative values indicate excess heterozygosity.
To compare the genetic differentiation between subpopulations, values for F_{ST} range from 0 to +1 and higher positive values designate differentiation among subpopulations (Peakall & Smouse, 2009; Barton, 2016, Truong, 2017).

F_{IS} is the inbreeding coefficient within individuals relative to the subpopulation, measuring the reduction in heterozygosity of an individual due to non-random mating within its subpopulation.

$$F_{IS} = \frac{\overline{H}_e - \overline{H}_o}{\overline{H}_e}$$

F_{IT} is the inbreeding coefficient within individuals relative to the total population.

$$F_{IT} = \frac{H_T - \overline{H}_o}{H_T}$$

F_{ST} is the inbreeding coefficient within subpopulations relative to the Total population. This statistic measures the genetic differentiation between subpopulations. $F_{!"}$ values less than 0.05 were interpreted to indicate no evidence of genetic differentiation between subpopulations

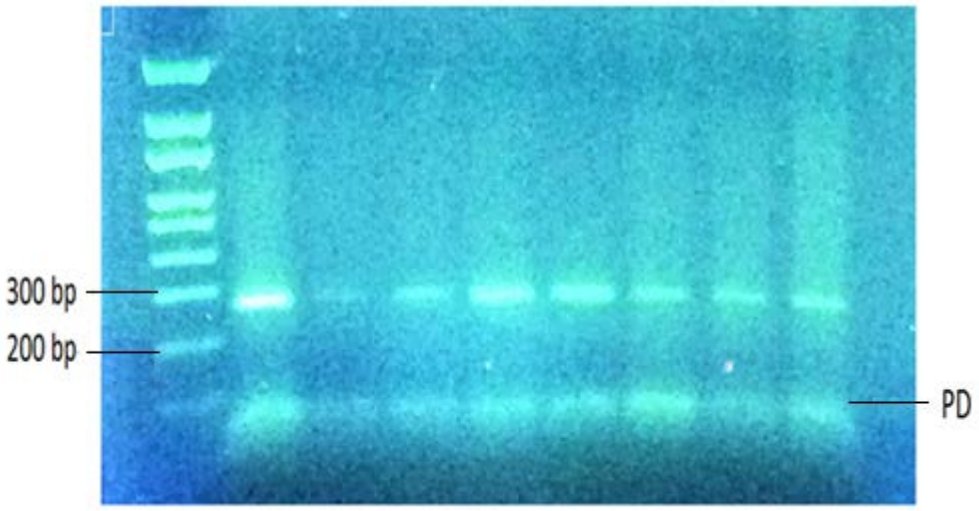
$$F_{ST} = \frac{H_T - \overline{H}_e}{H_T}$$

Results

DNA was successfully isolated from all samples and the presence of microsatellite alleles (Apm01 and Ab06) were confirmed through PCR and gel electrophoresis (Figure 2); The Apm01 locus size ranges between 180-208 bp while Ab06 DNA segments are within the range of 270-280 bp (Figure 2A-B). Due to small differences in bp size that can occur from additional nucleotide repeated motifs, gel electrophoresis did not clearly indicate heterozygotes versus homozygotes in the subpopulations samples. PCR of all samples were pooled for genetic analyses via sequencing by the Oxford Nanopore. Real time computer input of data shows reads were recognized by the pores in the MinION sequencing platform (Figure 3).

Due to the unclear nature of the sequencing results provided by the MinION, simulations of varied genotype ratios in the howler subpopulations were completed to offer insight into what future results might provide. Simulation of randomized heterozygous and homozygous ratios yielded theoretic F-statistic values that could be obtained upon clear sequencing results (Table 5). To understand what population numbers and genotype ratios would result in genetic differentiation between subpopulations, simulations with randomly assigned genotype ratios were statistically analyzed. As expected, the populations with the most genetic differentiation signified through Chi-Square and F_{st} values were those with divergent high/ low heterozygosity ratios. For example, Chi-squared values of 228.9463 and 154.4526 were calculated for low heterozygosity (10/90) in both 50m and 100m interior subgroups. These numbers, are higher than the 0.05 significance threshold (Chi-Square = 5.99) thus indicating divergence from Hardy Weinberg Equilibrium and expected genotype ratios. Particularly, edge populations with 10/90 heterozygosity for both 50m and 100m edge groups were largely differentiated from interior subpopulations with 50/50 heterozygosity with 0.16944 and 0.16423 F_{st} values respectively. With future sequencing results, the actual genotype ratios can be calculated and treated in the same manner.

A



B

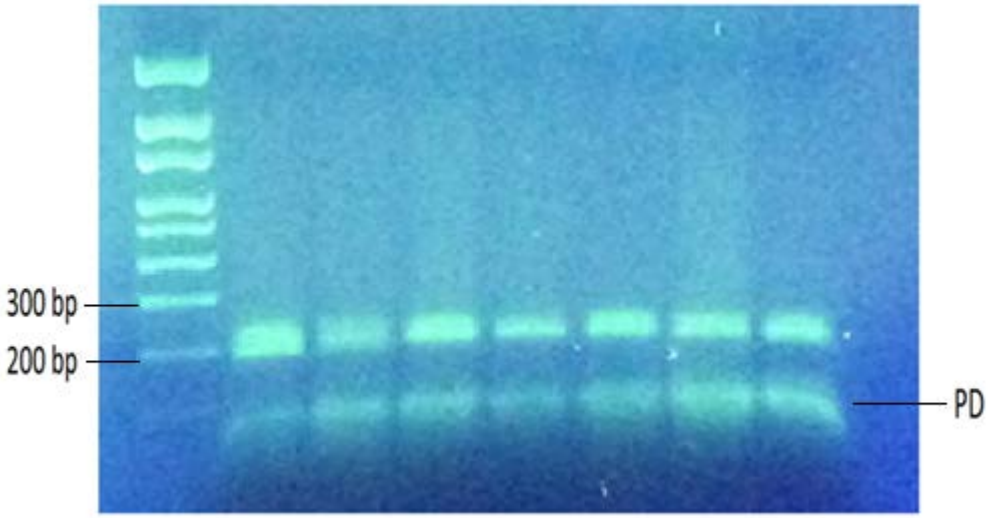


Figure 2. A-B. Gel electrophoresis of *Alouatta palliata* DNA samples. The DNA ladder is in the leftmost lane of both gels. Gel A. shows the presence of the polymorphism amplified using Ab06 primers. The size of the PCR product is around 270 bp. Gel B. shows the presence of the smaller microsatellite region amplified from Apm1 primers. The Apm1 PCR product is around 220 bp. The lower bands indicated on both gels are primer dimers (PD).

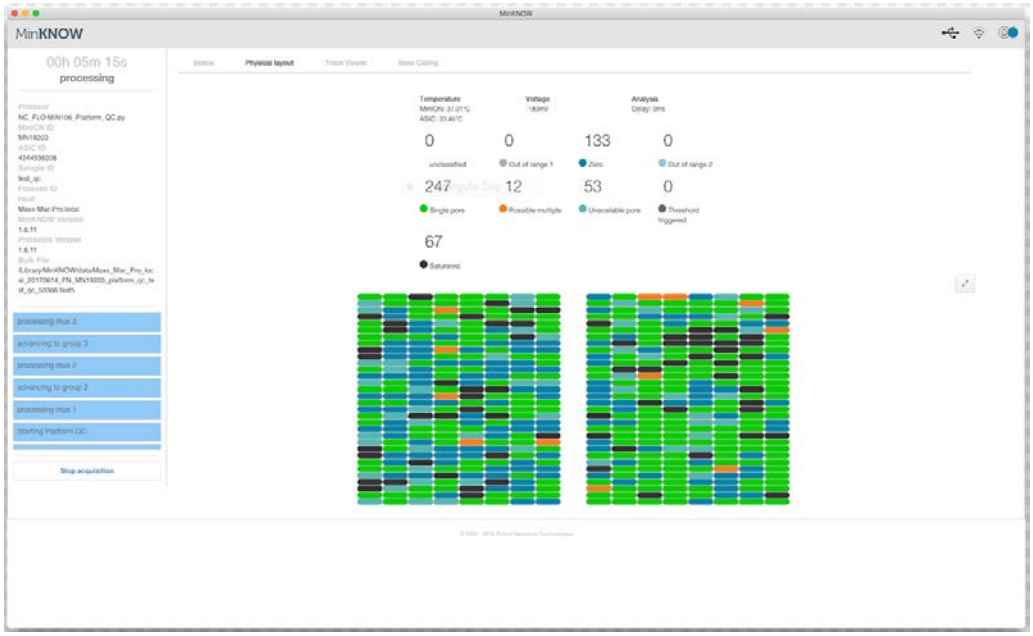


Figure 3. MinION sequencing. Each colored cell represents real time sequencing reads recognized by the nanopores of the MinION. A computer screen displays the incoming data from thousands of DNA strands recognized by the nanopore proteins. Green cells represent single, good reads while the remaining colors signify a sequencing issue or misread.

Table 4. Chi-Square Results for simulations. Groupings based on edge versus interior classification are designated (edge (E) or interior forest (I)). The ratios represent the proportion of heterozygotes/ homozygotes in the subpopulation. The Chi-square values are reported for the comparison of simulated "observed" values compared to the expected values. Values with an asterisk (*) are higher than 5.99 which signifies a significance p-value of 0.05.

	Edge/Interior	**50/50**	**70/30**	**30/70**	**90/10**	**10/90**
50m	E	0.72973	13.7027*	4.783784	37.21622*	68.35135*
	I	1.939597	76.03356*	49.13423*	264.9866*	228.9463*
100m	E	4.164835	32.73626*	19.26374*	144.033*	131.4396*
	I	0.789474	57.27368*	34.53684*	154.4526*	166.6*

Table 5: Fixation Indices. The ratios represent the proportion of heterozygotes/ homozygotes in the subpopulation. Values above 0.05 indicate genetic differentiation between the two subpopulations: interior and exterior. Values in light gray represent F_{ST} values > 0.05 and < 0.1 and values in dark gray are F_{ST} values above 0.1.

50m

		Edge				
		50/50	70/30	30/70	90/10	10/90
	50/50	0.00046	0.0119	0.01404	0.05343	0.16423
	70/30	0.0392	0.0035	0.08843	0.00594	0.31575
Interior	30/70	0.03752	0.10507	0.00853	0.19423	0.08197
	90/10	0.17237	0.07033	0.26707	0.01085	0.5872
	10/90.	0.15008	0.27052	0.07827	0.4026	0.09161

100m

		Edge				
		50/50	70/30	30/70	90/10	10/90
	50/50	0.03898	0.02888	0.08763	0.1588	0.16944
	70/30	0.08579	0.00634	0.22484	0.02793	0.39436
Interior	30/70	0.11049	0.16232	0.0669	0.3709	0.03425
	90/10	0.17664	0.05918	0.35906	1E-06	0.56646
	10/90.	0.279	0.3761	0.16671	0.63298	0.00444

Discussion

To test the hypothesis that the howler subpopulation at the forest edge (50m and 100m) would show signs of inbreeding, I sought to gather discernible sequencing information from the Oxford Nanopore MinION. However, Nanopore individual sequence results could not be confidently identified as homozygous or heterozygous. The sequencing read lengths are not clear enough to decide whether both copies of alleles are the same size or have different numbers of nucleotide repeats. While the results could not be used to discern genotype, this first run with the MinION is a starting point for the use of novel sequencing technologies on these Costa Rican howlers. Perhaps future sequencing work could be done with another rapid sequencing technology, such as the Illumina, to cross reference the reads from the Nanopore. Future improvements on the sequencing technology and/ or the preparation of endogenous DNA will result in clearer sequenced reads.

Without verified genotype from the individual howler samples, investigation on possible allele frequencies of the howler subgroups at the forest edge and interior was completed randomly following established genotype ratios. For the interior and edge howler samples, each sample was randomly assigned as homozygous for two different alleles or heterozygous. For the purposes of statistical inquiry, random genotype assignment to all individuals allowed simulations with the data to randomly generate allele frequency numbers to be used with a Chi-square test and F-statistics. F-statistics of theoretical situations or genotype ratios following the genetics of the inhabitant howler

population at La Suerte allows insight into possible scenarios that could result from accurate sequencing reads in the future. The random assignment ratios were used following Table 4 and 5 and show which genotype ratio comparisons allow for genetic differentiation between subgroups. Therefore, these simulations give a picture of what real sequencing results might yield.

From the simulated genotype ratios and comparisons between edge and interior samples, significant F-Statistic and Chi-square values, indicating genetic difference between the groups and significant divergence of individual groups from Hardy Weinberg Equilibrium respectively were noted. Specifically considering the hypothesis that the edge subgroups at 50m and 100m would have higher homozygosity, the genotype ratios of 30/70 and 10/90 simulate increased inbreeding scenarios. As expected, the Chi-squared values of 68.35135 and 131.4396 were calculated for low heterozygosity (10/90) in both 50m and 100m edge subgroups. However, there was not enough deviation from Hardy Weinberg for the 50m edge subgroup (30/70) with a lower Chi- Square value of 4.783784 (< 5.99) meaning it is not significantly different from a population with no evolution occurring. Therefore if the edge (100m) subpopulation had a majority of homozygous individuals, they would not follow Hardy Weinberg Equilibrium and would also be genetically different than another population experiencing no evolutionary pressures.

In order to simulate possible edge effect scenarios and compare the genetic structures of edge to interior populations, F-statistics offer a direct comparison of the

edge and interior subgroup genotype frequencies. Edge effects, or the increase of the forest edge to interior ratio is shown to result in less food availability and habitat (Estrada and Coates-Estrada 1996) as well as decreased dispersal ability to other howler groups (Oklander et al., 2010; Winkler, 2004). Therefore, to see which genotype ratio comparisons yield significant genetic differentiation, both the 30/70 and 10/90 edge groups were compared to the 50/50 interior groups. When comparing the 10/90, low heterozygosity edge group to the 50/50 for both 50m and 100m subgroups, the edge 10/90 heterozygous groups were significantly different with F_{st} values of 0.05343 and 0.1588 respectively (Significance >0.05). Both F-stat values are above 0.05 meaning that the simulation genotype composition of both edge/ interior classifications (50m and 100m) to interior subgroups are significantly genetically different from each other. Interestingly, the 30/70 heterozygous edge group was only significant from the 50/50 interior subgroup when the edge was increased from 50m to 100m resulting in an F-stat value of 0.08763. These simulations allow determination of what possible genotype ratios in the subpopulations would result in significant genetic differences between subgroups. In the future, improved sequencing reads could lend a helping hand to determine the actually allele frequencies of the inhabitant howler monkeys at La Suerte Forest.

Accurate sequencing results are important for this study to understand the behavioral and genetic consequences of increased edge effects on howler monkeys from habitat fragmentation. The edge effects from fragmentation present new ecological challenges that potentially can alter population density, social cohesion, and group sizes (Broadbent et al., 2008; Schwitzer et al., 2011). Such challenges can result in limited

dispersal of howlers to new groups, thus resulting in a potential decrease in gene flow for the howler population. Amidst severe environmental changes, it is important for species to maintain genetic diversity. Genetic diversity increases the likelihood of species surviving against stressors which include fighting off disease (Acevedo-Whitehouse et al., 2003; Woodruff, 1989). As such, there is still cause for concern for the long-term impact of decreased habitat size on the survival of the species (James, 1992). With increased infringement of human activities on primate habitats, it is imperative to understand how monkey populations are being affected by increased edge effects.

With continued emphasis on accurate and effective sequencing technology, perhaps the MinION technology could one day be regularly used for genetic applications. Accurate sequencing should allow quick determination of inbreeding levels in the howler population. As for now, the lack of clear sequencing results for two small microsatellite DNA regions by the new MinION sequencing technology poses a concern for applied research. These very technologies are being used to sequence human genomes (Jain et al., 2018) and will likely become more commonplace in medical practice due to the established link of genetics to disease. However we must ask ourselves if the sequencing technology is reliable enough to be used with human lives. While there is concern due to the lack of accurate sequencing results as seen in my personal research, the technology will undoubtedly continue to improve. When sequencing error no longer becomes an issue, society must ask themselves how such technology and information can impact human lives especially in regard to personal health and lives.

In the following chapter, I will turn my attention from sequencing technologies of small non-human DNA sequences to the application of advanced sequencing platforms to sequence the entirety of human genomes. With the established genetic link to disease, whole genetic sequences are now easily attained and analyzed for specific mutations known to commonly occur with a particular disease. However instead of focusing on the sequence itself, I turn my attention to the potentially life altering health information inferred from that sequence and its imminent use in the medical field. Knowing that genetic risk assessment for particular diseases can be given to patients, there are more challenges and questions that medical ethicists and society must address in the coming era of personalized medicine.

IV. The New Face of Medicine

I previously discussed the use of genetic sequencing technologies, namely the Oxford Nanopore, on evaluating inbreeding levels in a Costa Rican howler monkey population. The information gained from analyzing millions of DNA sequence reads has the potential to give insight into the howler monkey population's behavior. Although the sequencing information is difficult to decipher, the results are a starting point for using rapid sequencing technologies for the genetic assessment of the Costa Rican howler monkeys at La Suerte. I explicitly used the Nanopore for all sequencing reported here. Initially however, I used Sanger Sequencing, first developed in 1977, to sequence the howler DNA samples and assess inbreeding levels (Sanger et al., 1977). The Nanopore and Sanger methods successfully sequenced amplified DNA samples. However, accurate analyses of the results proved difficult from both methods. Considering my experience with two different sequencing mediums, I was inspired to investigate how sequencing advancements are refining not only scientific research but also the understanding and treatment of genetic-based diseases.

As sequencing technologies improve, the cost and analyses of DNA reads must improve as well. The Sanger method, which is part of the first generation of sequencing, is relatively cheap per sample at 7 dollars but relatively expensive per base when compared to next generation or Nanopore sequencing. For added perspective, it costs around $7,000 today to sequence 1 million bp using Sanger sequencing as compared with

Oxford Nanopore that costs less than a cent. For my sequencing needs, Functional Biosciences (WI), a DNA sequencing company, sequenced the howler DNA samples and displayed the results as electropherograms. Electropherograms are DNA sequencing reads with prominent peaks and colors designating the recorded base pair. I analyzed electropherograms of the specified polymorphisms present in the howler DNA and manually counted the number of nucleotide repeats to assess heterozygosity versus homozygosity. While many of the samples resulted in clear reads, some could not be confidently designated as heterozygous or homozygous. Rather than a strong nucleotide peak at the end of the polymorphism repeats, I had considerable doubt on the identity of the last few nucleotides with homozygous individuals being the trickiest to assess. As such, my research advisers and I sought out a new sequencing technique that could sequence all samples in one run. The search lead to the Oxford Nanopore. The Oxford MinION claims reasonably accurate sequencing results (95.74%) and also sequencing results in real time (Jain et al., 2016; Cherf et al., 2012; Manrao et al., 2012). The MinION technology fascinated me and I was more than excited to try the novel sequencing technology.

Instead of sequencing by synthesis, the Oxford Nanopore technology directly sequences DNA molecules as they pass through 512 membrane-embedded Nanopore proteins. As the DNA goes through the protein pore, characteristic disturbances in an electric signal across the membrane are picked up by a computer system, marking them as an A, G, T, or C (Jain et al., 2016; Cherf et al., 2012; Manrao et al., 2012). Once loaded with samples, the MinION directly sequences the DNA in a span of 48 hours (Lu

et al., 2016). During the sequencing process, a computer screen displays the incoming data and thousands of DNA strands are recognized by the Nanopore proteins (Figure 3). I could not help but recall that such advancements were not possible only 3 years ago, when the Nanopore was first introduced. As a relatively new sequencing method, the Nanopore shows vast potential for applications beyond sequencing small DNA segments (Lu et al., 2016). The application of accurate and rapid sequencing technology is already extending its influence beyond scientific research into health assessment and medical practice. As advancements in read length and accuracy continue, the Nanopore and other technologies are increasingly being used to understand genetic-based disease by rapidly sequencing whole genomes (Ku et al., 2013; Jain et al., 2018).

The discovery of DNA and the invention of sequencing technologies vastly altered not only scientific research but our understanding of humanity and our health. Upon the completion of the Human Genome Project (HGP) in 2001, the HGP Director, Francis S. Collins, and Victor A. McKusick, MD predicted that the entirety of the human genome would vastly change the face of medical practice. They stated, "…this is a time of dramatic change in medicine...We must commit ourselves to exploring the application of these powerful tools to the alleviation of human suffering" (Collins and McKusick, 2011). Collins and McKusick know that science and medicine would be crossing a "threshold" into a new way to tackle human disease. In order to fully explore the life-altering benefits of sequencing the human genome, Collins and McKusick recognize that as doctors and fellow researchers, they have a duty to utilize all of these new "powerful tools" (Collins and McKusick, 2011). Amongst these tools are rapid sequencing

technologies which allow assessment of disease-causing genes and in some cases targeted treatment options.

The HGP's wanted to obtaining a complete human genome sequence to aid medical understanding of disease. This legacy continues to take shape and mold patient diagnosis and treatment with the explosion of sequencing based research (Collins, 1999). Just as the HGP houses the first human genome on a computer system, the continually growing research behind the genetic basis for disease is recorded on various databases. One such database is the *Online Mendelian Inheritance in Man* (OMIM®) which is a compilation of recorded human genes and genetic disorders. The OMIM compliments the aims of the Human Genome Project to understand the function of each human gene to aid medical practice (National Center for Biotechnology Information). Sequencing technologies combined with online databases of mapped disease causing genes are a powerful arsenal in the hands of medical professionals.

To examine the transforming role of Next Generation Sequencing technologies on medicine, this chapter specifically explores the use of Whole Genome Sequencing (WGS), sequencing all 6 billion bases, for signs of disease and ultimately how genetic results can positively and negatively impact patients and life outcomes. Largely, medical professionals will need to become true experts as patients now can seek out and live with this new genetic information. Health risk not previously available to patients is now commercially available. The question remains, what will patients do with this new information? We must consider how patients will react and how can medical

professionals help. Sequencing of whole genomes allows the discovery of disease genes early and accurate diagnosis of current symptoms allowing for individualized patient treatment and prognosis. As such, WGS and diagnosis tests fall under the umbrella of the recently coined, “personalized or precision medicine.” Defined broadly, personalized medicine is the utilization of genome test results in order to tailor disease treatments to the specific needs of the patient (Lesko and Schmidt, 2014). This can include tailored drug therapies to the exact mutation present or even drug dosage amounts. Despite its proliferation as a diagnostic tool, there are a number of serious ethical pitfalls that remain to be addressed by scientists and doctors, and importantly by individuals.

Whole Genome Sequencing: Our Complete Book

While there is inherent beauty in the gene sequences that make up each individual’s characteristics, that diversity is accompanied with life’s maladies. The maladies that I refer to are the inescapable fact of genetic disease. Sequencing allows researchers to understand the genes that make up a diverse array of physical attributes and the myriad of intricate molecular functions of the human body. However, one of the main purposes of the HGP being overseen by the National Institutes of Health (NIH) was to improve medical treatment of disease (Stankiewicz and Lupski, 2010). Reading the DNA sequence brings humanity closer to understanding a large source of life’s suffering. Variation or deviation from a normal gene sequence can occur during conception with the mixing of maternal and paternal DNA copies or it can happen due to abiotic factors such as UV radiation that can create breaks in DNA strands (Guttmacher et al., 2005). While not all mutations lead to abnormal gene products, the very mechanisms which create

humanity's immense diversity also allow disease. The genetic nature of disease is therefore a tool for medicine to potentially improve patient outcome and treatment.

Rapid sequencing technologies are continually transforming the diagnosis of disease. Disease diagnosis mainly uses patterns of familial inheritance to identify current disease and predict future risk. With the progression of sequencing technology in the early 2000s, genetic tests mostly used targeted DNA approaches to identify genes connected to specific disorders including conditions such as Tay Sach's and phenylketonuria (PKU) (NIH, 2010). Practitioners can order targeted sequencing of amplified regions containing disease gene variants to combine with phylogenetic analyses for a comprehensive profile of their patients (Van El et al., 2013). With this method, genetic counselors focus on high risk disease probabilities for that specific patient and limit their genetic inquiry to the "driver" mutations or risk alleles most common in the population (Rajendran and Deng, 2017). In doing so, genetic counselors and patients disregard the information contained in the rest of the genome.

Targeted sequencing methods are useful to determine the cause of particular diseases, yet these targets are often chosen based on familial history and medical diagnosis. This targeted approach poses a concern for individuals without any prior family incidence of disease. For example, in a 2012 study on the applicability of targeted approaches, Calvo et al. sequenced the coding exons regions of ~1,000 nuclear mitochondrial genes in 42 patients with a condition called human oxidative phosphorylation. While 10 cases could be accurately diagnosed, 19 patients did not carry known variants in the targeted regions. Therefore, Calvo et al. (2012) concluded that the

variants must be in other genomic sites. While targeted approaches can offer insight into some cases, WGS offers a complete view of the genome, allowing thorough assessment of faulty genes and other mutated DNA regions. Due to the complex nature of disease, deleterious mutations can also occur in non-coding DNA and potentially alter the expression of important nearby genes. Therefore, in complicated cases, sequencing the entire 6 billion base pairs of DNA could prove useful. WGS can serve as an overall health profile beyond the scope of targeted sequencing or multigene panels. As such, WGS is increasingly used for health applications (Wheeler et al., 2008; Welch et al., 2011). With WGS and subsequent analyses alongside known disease variants, the genetic basis for current conditions as well as a complete profile of future disease risk becomes available to the patient.

WGS is a common practice in the realm of scientific inquiry and its price tag is increasingly becoming cheaper. Do a Google search of companies that offer whole genome sequencing and the first company, Genome One based in Sweden, offers to completely sequence a human genome for $1,000. This is certainly a drop from the HGP's 3 billion dollar budget. Genome One will provide a complete sequence but the analyses or Whole Genome Analyses (WGA) of what that sequence means is omitted. However, if a person does not want to pay $1,000, they can volunteer their genome sequence to science. 2005 marked the beginning of a volunteer, non-anonymous project to sequence 100,000 human genomes. This grandiose effort is the ongoing Personal Genome Project (PGP). At over 5,000 participants thus far, PGP seeks to expand an online database where researchers can seek out common genetic variants associated with

disease similarly to the OMIM online database (Harvard PGP, 2018). These exciting insights into disease and phenotype are possible due to the advancements in sequencing technologies. The PGP amongst other large sequencing plans will help usher in future changes to medical practice, leading to an emphasis in prevention and therapeutic strategies (Ku et al., 2013).

WGS and WGA of patient DNA create a comprehensive disease profile that includes all SNP's, recognized risk alleles, and whether individuals are homozygous or heterozygous for particular variants. This information is then used to create a risk assessment profile. The variants increase an individual's basal level risk of disease for the general population by an order of magnitude. The risk assessment comes from the work completed in genome wide association studies (GWAS) that seek to catalog common disease variants or SNP's co-existing with disease phenotype (Hindorff et al., 2009). For example, in Iceland a comprehensive study sequenced the whole genomes of 2,636 people and discovered rare allele variants only present in an isolated Icelandic population. By "rare allele variants," Gudbjartsson et al. (2015) describe various single nucleotide polymorphisms discovered in the study population coexisting with particular diseases. Some allele variants were linked to diseases including liver disease and early-onset atrial fibrillation. Studies such as this allow comprehensive association of these variants with expressed phenotypes including both physical traits and health (Gudbjartsson et al., 2015). The knowledge gained from WGS studies further improves the scope of next generation sequencing diagnosis. With continued discovery of disease-

causing genes, WGS could prove to be more useful in medicine and not just for gene research.

While not yet a primary method for diagnosis in the medical field, WGS is becoming a reality for some disease tests. However, the clinical utility of WGS is in question due to the ambiguous nature of most risk allele results. There are few diseases whose genes increase disease risk for a patient in a meaningful way. For instance, the presence of a mutation in BRCA1 can increase a woman's risk of breast cancer to 80%. These odds surely cannot be ignored; however, for a majority of diseases, a 2010 study on the clinical utility of GWAS reports that risk assessment only provides "scant evidence" for a possible disease (Manolio, 2010). In essence, individual risk results are hard to assign due to difference in lifestyle and environment. Optimal risk assessment metrics have yet to be found and there is a lack of accurate interpretation of results.

Based on the percentage of the population with a particular disease phenotype, mutations marginally increase the risk for a particular disorder. While WGS with analyses can scan for multiple conditions at one time, there are few factors that can separate individuals enough to be taken from low to high risk (Manolio, 2010). By separate, I refer to low, medium, or high disease risk as defined by the area under a curve (AUC) for disease risk assessment (Figure 4). The curve is created based on the proportion of risk alleles an individual has as well as assessment of family history. A shift in position for an individual under the curve is due to particular risk allele presence.

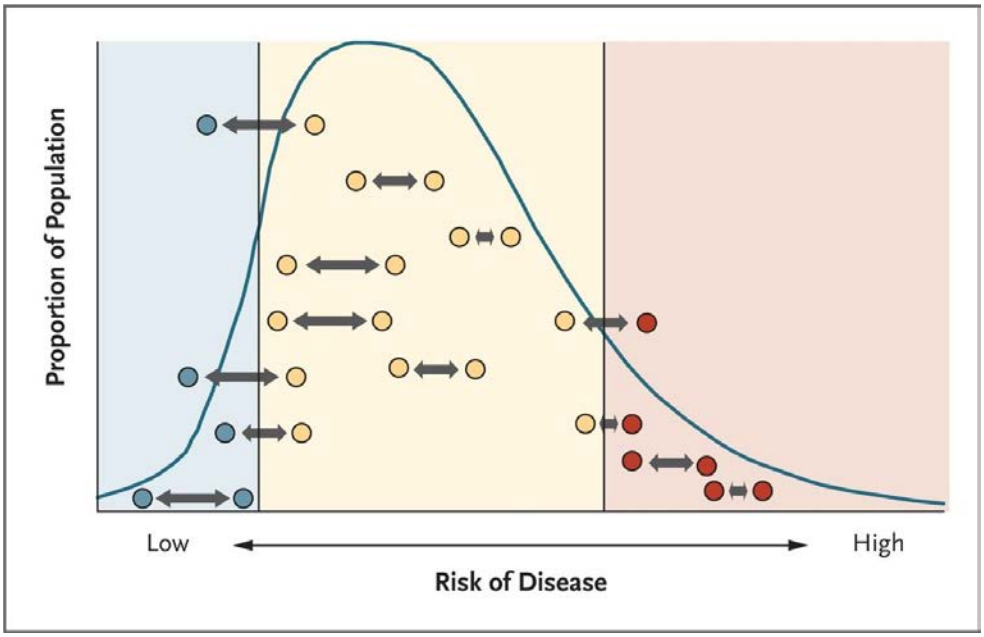


Figure 4. Risk Threshold classification for disease in the population. The basal risk of disease for most individuals lies in the median (yellow) area of the curve. The arrows between individual points signify a shift in risk due to new information specific to that person. Such shifts can be enough to change a risk category. Figure retrieved from Manolio, 2010.

The ability to peer inside one's genome and know your personal disease risk is an attractive idea in our present society. Starting in 2004, sequencing efforts became commercially available or direct to consumer (Niemiec et al., 2016). A popular company is 23andMe which allows the consumer to choose the health information shared with them after targeted sequencing. Shortly, the commercial genetic testing company will launch WGS as well (Hughes, 2013). With their services, an individual can specifically test for over a 1,000 allele variants associated with disease. For instance, a person could want the results for the late onset Alzheimer's gene variant or risk alleles for Parkinson's disease. Once known genes are sequenced with 23andMe, consumers often take the

health data to medical professionals to verify any current phenotype or higher risk. These companies allow untrained laymen to peer inside their genome and simply tell them if a gene variant is present or not. There is no risk assessment or one-on-one conversation. Individuals simply are left with a surface level assessment of their possible future health.

What Your Genome Can Do For You

Scientists and medical practitioners are both excited and worried about the imminent influence of WGS in common medical practice. While some doctors anxiously await the imminent changes to healthcare, others recognize that the benefits also come with negative consequences that should not be ignored (Kawamoto et al., 2009; Johnson, 2016; Parens, 2015; Lesko and Schmidt, 2014). Ideally, genomic DNA is a necessary change to medicine that will reduce treatment time, overall number of deaths, and total health costs (Kawmoto et al., 2009). However, ethicists are still grappling with the possible psychological trauma from knowing the chance of illness (Cacioppo et al., 2016). Doctors and researchers alike recognize that sequencing of individual patient genomes gives a direct blueprint to the genetic abnormalities responsible for deleterious phenotypes and thus allows doctors to better serve their patient's unique health needs. Personalized genomic data changes the approach to medicine. Whether this change is positive is left up to the practitioners and patients to decide.

The ability to read the entirety of one's genome allows the discovery of possible risk alleles associated with disease (Wheeler et al., 2008; Lindor et al., 2016). As WGS moves outside of large scale research efforts, researchers are discovering its profound applicability to solving medical conundrums including cases with rare allele variants as

well as variants previously unknown to cause a particular disease. In 2011, Welch et al. (2011) investigated a medical mystery surrounding a patient with acute promyelocytic leukemia via WGS. Through traditional methods of analysis including metaphase cytogenetic and interphase fluorescence in situ hybridization (FISH), no evident cause for the condition could be identified. The patient agreed to WGS and within 7 weeks, DNA was analyzed against common mutations and the information concerning a cryptic fusion oncogene was verified with PCR (Welch et al., 2011). With only 10ng of patient DNA, WGS identified a disease's genetic cause and also improved the speed at which doctors could administer a treatment regimen (Welch et al. 2011). Today the sequencing speed is drastically improved with times ranging from as low as 26 hours to roughly two weeks; time after the sequencing is reserved for WGA or analyses to determine variant alleles (Miller et al. 2015). Sequencing results received in a timely manner enable a physician to present the outcomes immediately and start a new course of treatment.

As the cost and error rate of WGS are reduced, healthy people desire the results to preemptively detect potential signs of disease. Tests on healthy adults help prepare individuals for future health challenges. Despite not having current phenotypic expression, individuals recognize that WGS can detect all common variants including rare single-nucleotide variants, copy number variants, and insertions and deletions that can later be mapped against known risk alleles (Lindor et al., 2017). Regarded as the "ultimate genetic test," WGS offers a glimpse into the instructions that make up one's physical characteristics as well as the possible fate of their health (Drmanac et al., 2014). Knowing the contained information is certainly an enticing offer considering the genome

remains constant starting from conception to the end of life. Perhaps genes can offer a helping hand when their messages are read correctly.

While the risk results of WGA are never 100% certain, people can still prepare themselves for the worst outcome and in some cases take considerable effort to alter their fate. WGS offers an error rate of roughly 1 in 100 kb or one in 200–500 SNVs discovered, for some technologies (Wall et al., 2014; Drmanac et al., 2015). The analyses' results are accompanied by this high error rate, which although rare, can result in false or missed diagnoses. Despite the imperfection of the technology, people still strongly believe that by knowing the possible outcomes from inherited genes, they negate the possibility of being blindsided later in life. They can seek the optimal treatment and begin making better life choices now rather than later. Put simply, an increased risk of heart disease from a particular risk allele could be decreased by eating non-processed foods and going for a walk every day. Changes such as these could be made from the information contained in a genome. Perhaps in the case of genetics, knowledge is indeed power.

V. Taking Control Over Your Genes

In my previous chapter, I outlined some of the major arguments for Whole Genome Sequencing (WGS) to aid medical practice. Specifically looking at the information gained from non-targeted sequencing, patients do not have to waste time completing grueling diagnostic tests but can simply give a small sample of blood and have an answer in days to weeks (Welch et al. 2011). The results remove the guessing game for current disease pathology and can give patients peace of mind. In a timely manner, patients can know the genetic basis of their condition and begin a tailored treatment regimen that in some cases can drastically improve treatment outcome (Miller et al. 2015; Welch et al., 2011). However, with the commercialization of WGS for diagnosis of current and possible disease, the possibility for inadvertent risk allele discoveries also poses a concern for the wellbeing of the patient. While looking for the specific genetic cause of a particular disease, the genetic risk factors for another ailment can also be discovered. Sharing this information without full explanation and understanding likely places added anxiety on the patient and their family (Cacioppo et al., 2016). Therefore, if WGS is implemented into standard medical practice, practitioners are faced with a new challenge to their oath, “First, do no harm” (Shmerling, 2015).

To “First, do no harm,” every new medical practice must go through a comprehensive evaluation before it can be accepted in healthcare. While I cannot share personal experience of being given WGS sequencing results, I can imagine my response

to knowing my future health risk. I envision being afraid of the information, knowing that I have no control over the messages my genes carry. Results received as a healthy individual would undoubtedly be received differently than someone who already has a disease with an unknown cause. Knowing my personal risk for one, two, or maybe three diseases has the potential to vastly alter my outlook on life. The discovery of a significant increase in risk would become all-consuming with days spent anticipating its arrival. Thankfully not all risk alleles give a high probability.

Reading Your Book of Life

Not only is the sharing of incidental disease variants a problem for doctors but also whether discoveries have enough concrete evidence to be deemed legitimate findings. By legitimate, I allude to the uncertainty caused by one's unique environment, family history, and the even more rare chance of sequencing error. What individuals must realize is that health risk assessment is not diagnosis. There is considerable chance that a disease will not be present later in life. Not all individuals with a common disease variant acquire the disease and not all with the disease have one or more copies of a driver variant. This phenomenon can lead some individuals to disregard the information from WGS and others to continue their lives constantly worrying about the timing of a condition. Doctors are unable to know the severity of the condition as well as when it will appear (Mukherjee, 2016). Therefore, the uncertainty of acquiring the disease is still present both before and after WGS. Looking at the example of breast cancer, not all individuals with a mutation in BRCA1 will develop the cancer and some mutations in

BRCA1 only marginally increase a patient's risk. Therefore, we can recognize that not every gene is fully predictive of risk. Environment and chance play a part. Disease manifestation is not always that simple either with various diseases involving hundreds of genes. Reading our book of life is proving difficult as we seek to interpret its hidden messages. While one is inclined to say that more knowledge is power, we have to keep in mind the anxiety that a new form of uncertainty can bring. We must ask the question whether this difficulty is worth the professed reward or if we should leave our genetic secrets unearthed.

Commercially available, 23andMe, a genetic sequencing company, seeks to give sequencing results and chosen health risk information with little hassle to its consumers. Proudly displayed on their website is the slogan, "To be your best self. Look inside yourself." The language implies that individuals should look inside their genome to know their destiny and ultimately to understand themselves. The essence of their advertising is such that people feel that reading their genes will unlock the key to a life of happiness or their "best self." A simple sample of spit is enough for the company to extract viable DNA, and when the spit vial arrives at your doorstep, the words, "Welcome to you" are written on the cover. While 23andMe intends for its users to know their health assessment and make positive life changes, they instill this idea that one's genes are the key to knowing ourselves both in the present and future.

The publicity surrounding WGS sounds too good to be true. Sequencing companies advertise an opportunity to prepare oneself for possible diseases by making

beneficial life changes. I wonder, though, if I would be one of the thousands of people who would purchase their genetic code and disease risk. Considering a current disorder, of course I would want to know its genetic cause. The anxiety of a current disease would already be present and knowing the exact cause might offer some comfort (Sapp et al., 2014). But when it comes to whether I would get sequenced by a company such as 23andMe and be checked for more than 1,000 disease mutations, I told myself, "No." I anticipate having unnecessary anxiety and unexpected, potentially life altering results with particular risk allele presence. I cannot imagine bringing myself to opening the results with the words, "Welcome to you" boldly displayed.

Not surprisingly, my fear of stress caused by WGS results is a current hot debate among geneticists and medical professionals. High on the list of ethical concerns around WGS is the release of sensitive information to the patient which includes incidental discoveries of other strong genetic risk factors beyond the desired result (Cacioppo et al., 2016). While the handling of this sensitive health information is an entirely separate issue than the use of whole genomes for health assessment, I am reluctant to say I want that information provided to me. Given the scope of this concern among ethicists, the American Society of Human Genetics Board of Directors and The American College of Medical Genetics Board of Disorders (ASHG/ASMG) gave a detailed report in 1995, 8 years before the completion of the first human genome, in which they predicted that people would have intense psychological stress and anxiety from future disease prognosis. In the report, they advise that such distress should be prevented while still understanding that genetic testing has a valuable place in the medical field. In a

formalized statement, the ASHG/ASMG seek to avoid undue early stress by recommending that future professionals only report possible diseases that can have effective treatment, prevention, or delayed onset, whereas other disease risks should not be reported. My response to this statement was rather positive and I saw myself accepting these limits on testing results. Analyses simply would neglect those definite disease-causing genes. Perhaps with strict protocols on handling information, WGS could become a benefit rather than potentially harmful reality.

Although strict handling policy could help circumvent anxiety, there are various investigations on patient response to genomic results that suggest patient anxiety and life practice are not as significantly altered as commonly believed. According to a study published in *The New England Journal of Medicine*, 90.3% of study applicants did not react with significantly higher levels of anxiety after receiving WGA results (Bloss et al., 2011; Heshka et al., 2008). In a review study compiling the results of more than 30 scientific papers on post retrieval of disease risk, Heska et al. (2008) discovered that there are no psychological problems associated with receiving risk assessments. They report that anxiety is at first present but does not last in high levels in the following months. Participants were largely happy knowing the results rather than not knowing. While sequencing can reveal higher disease risk, it can also ease the mystery behind the inheritance of disease causing genes.

Another analyses of anxiety from sequencing results shows that despite the presence of strong risk factors, individuals feel less anxiety than worry regarding their

results. This study, also from the *New England Journal of Medicine* follows the reactions of 160 adult children of an Alzheimer's parent. The disclosure of sequencing results specifically focusing on the driver mutation in APOE, showed no significant psychological trauma. Rather, participants largely felt relieved due to removed uncertainty and also recognizing the chance of not acquiring the disease is higher than their risk. As an added perspective, individuals with a negative assessment of Alzheimer's risk had pre-result stress relieved. Here, the information gained was power in their hands. While there is still uncertainty due to the combination of environment with familial inheritance, participants gained peace of mind (Green et al., 2009). Participants in WGS seek an answer to or even avoidance of life's uncertainty. If sequencing analyses results offer no deleterious anxiety or after effects, then perhaps by gaining their code, there is an aspect of their humanity that could be detrimentally affected.

Continued Uncertainty

While I predicted that sequencing results would be received with fear and long-term concern, I was surprised to discover that this is not a significant consequence of WGS results. Perhaps there is something else that is lost by knowing personal health risk. Therefore, I turned my investigation to possible life changes as a result of sequencing information.

Given the exciting promises of WGS accompanied by the worries of the ethical community, an American physician and oncologist wrote the book, *The Gene: An Intimate History* to share his opinions on the coming societal changes from genomic

technology. The author, Siddhartha Mukherjee, details the discovery of DNA and shares his worry about DNA's future trajectory. Mukherjee describes that screening for all possible diseases will create a world inhabited by "previvors who have been screened for genetic vulnerabilities" and thus use their genetic knowledge to avoid disease where possible (2016). Tied to this need to be prepared and know one's intimate health future appears to be a desire to change one's destiny, to avoid life's uncertainties, to ultimately be 23andMe's "best self." However, Mukherjee continues his argument describing that while, "grief might be diminished…" so "might tenderness… Infirmities might disappear, but so might vulnerability" (2016). Just as individuals seek to be prepared and alleviate their health vulnerabilities they also miss out on the lessons learned from confronting that vulnerability. While Mukherjee's argument is mainly directed to purposeful editing of personal genomes to fix mutated genes, his assessment is still applicable for the sharing of a total disease risk assessment. WGS exemplifies humanity's constant need to alleviate all forms of life suffering and we appear unafraid to read our own instruction manuals and unearth its secrets.

Life's suffering and unknown challenges undoubtedly leave individuals vulnerable and at times helpless, especially in regard to their bodily health. I wrestle with understanding the purpose behind all of life's suffering but perhaps there is a value in accepting life's unpredictable nature. Recently I came across a book that offered some help, entitled *Man's Search for Meaning*. Within its pages the author, Victor Frankl, describes his experience in a concentration camp during World War II. The miraculous part of the autobiography is Frankl's determination to respond well to his life's

misfortune. Frankl proposes a psychological analysis of human suffering where he asserts that human beings possess a spiritual freedom or ability to choose their response to pain and uncertainty. If humans have the ability to choose their response to pain, perhaps times of struggle can be converted to moments of triumph. Frankl argues that, "The way in which a man accepts his fate… the way in which he takes up his cross, gives him ample opportunity… to add a deeper meaning to his life" (2006). The "opportunity" is a chance to unearth personal moral values amidst suffering. While reading Frankl's argument, I envisioned that a deleterious allele discovered through WGS could be considered a moment of suffering. Therefore, posed with this circumstance, all individuals have an opportunity to accept their fate and continue living life as well as they can.

Of course, life choices after receiving WGS analyses are left up to each individual to decide. Many would argue that individuals could not easily find a sense of joy or purpose amidst psychological stress but Frankl says that all humans do have a choice to respond well. To respond well suggests that personal integrity is upheld. Frankl asserts that "Man *can* preserve a vestige of spiritual freedom, of independence of mind, even in such times of psychic and physical stress" (2006). This spiritual freedom is the choice to go one's own way rather than allowing suffering to change overall life perception. In the case of genes, we all can choose to define our own lives rather than letting our genes lead our decisions. Life is full of uncertainty; it is our sole responsibility to respond well to that ambiguity. We can either attempt to know what our genes have in store for our overall life or we can live with their secrets hidden. In either case, we still have

uncertainty and we are vulnerable to the fate our genes grant us. If we choose to have our disease assessment, perhaps we can be prepared with purposeful life choices to hopefully reduce our risk. One could argue that both choices are valid and we still have the opportunity to respond well to life's circumstances.

Knowing that disease risk results can offer peace of mind, and enable individuals to take control of their health, there is still much ambiguity in regard to future health. WGS results give at least some form of autonomy to its users, asking them to decide how they will handle the information. Oftentimes, when armed with this genetic "book", the insights may not make sense to its readers. That is why it is crucial for doctors to understand genetics and be able to explain options and future directions to their patients. Medicine did not previously know our gene risk and doctors have always had a moral obligation to tell their patients of a current disease. However, the role of physician changes when the information is a risk assessment and carries uncertainty. With the commercialization of genetic sequencing, more and more individuals will be seeking their doctor's expertise. Doctors must be prepared with the knowledge to serve their patients' best interests. Knowing that WGS is a coming change to medical practice, practitioners must still continue to live by their creed to "First, do no harm". Although this is a negative duty for doctors, there is also the implied positive duty to alleviate suffering. The future will determine to what extent WGS will influence the full range of a doctor's responsibility to their patients.

While there is much to be learned from the unknown challenges of life, WGS does not eliminate the opportunity to respond well to life's circumstances. Instead, it can

remind people of their autonomy and agency to decide for themselves how their life will develop. Most importantly, WGS results might serve as a reminder to take care of our bodies but also to live life well. The sequencing future of medicine must allow patient choice and ensure their informed consent regarding receiving their sequencing results. It will be up to the patient to decide whether they want to have a translation of their life code. While there is uncertainty in not knowing, I acknowledge the value of being prepared. After all, a person could become more vulnerable if disease is entirely unexpected. Health professionals must ask themselves, "By denying WGS, are we robbing individuals of their agency?" Although I previously thought that vulnerability and life's lessons could be altered or lessoned from gaining sequencing results, I now realize that vulnerability is still present from the uncertainty of risk as well. Perhaps Frankl is right when he says that we can respond well to suffering in all of its forms.

WGS is more complex than some people may know and cannot offer complete answers. Although sequencing is not always perfect, as seen in my previous genetics experiment, our society and medical professionals are faced with this new medium of information. Personal genomes, previously untapped, provide our doctors with new tools to serve their patients and they are charged with educating those who choose to read their book of life. Although enticing, the translation is not yet easy to understand. It is up to each individual to choose how they act on this knowledge. The knowledge can encourage people to decide how they live the rest of their life or it can become a looming fear of what may not come. Individuals must be careful though to understand that risk assessment from sequencing is not set in stone and cannot predict the future. If WGS is

imminent, health professionals can protect the interests of their patients by emphasizing just how complex one's genetic code really is and give them options for their next steps.

Perhaps individuals armed with the information, no matter how uncertain it may be, might be reminded to appreciate life for all of its uncertainties. The genetic information might give people an extra nudge to live life well. When getting sequenced, individuals should not have the mindset that they can predict their health future and therein avoid life's challenges. Instead, if we choose to purchase our sequence, we all should take the power that sequencing may provide and appreciate its knowledge, while fully knowing that there is still much uncertainty and much still to be deciphered. Rather than have your genes inadvertently tell you how your life might develop, what you should eat or whether you should have children, we should dictate our own life course and its numbered moments. When asked whether we want the analyses of our sequence results, the choice we make should come from knowing ourselves and how we respond to risk. Risk places our life in numbered form so that we can attempt to act upon its meaning. Therefore, knowing oneself is important as we now enter into a medical profession that seeks to "personalize" its approach. Our genes are not the key to knowing ourselves, it is our response and acceptance of our vulnerability. Only when we value the bodies and life that we are given, will we being to appreciate our moments of vulnerability for revealing who we truly are as individuals.

VI. Conclusion

As I look back on my time in a genetics laboratory and as an honors student at Regis University, I am continually reminded of a phrase that my writing Professor, Dr. Bruhn, used to tell the class, “At Regis, we like to cultivate ignorance.” At the time, Dr. Bruhn sought to push back on the first year student perception of college. Considering the endeavor to write this thesis, it showed me that the more knowledge we gain, the more we realize how much we do not know. There is a lot to be said about the uncertainties of life, being ignorant of when life will throw you a curveball or not knowing how to accurately discern sequencing results. Yet as we dive deeper into the mysteries of our own genetic code, we ultimately find more uncertainty. However this time, that uncertainty can be accompanied by grief and fear for what may never occur. Not all risk assessments end in disease, but they do have a lasting impact on one’s definition of their own life. One thing that Honors has taught me is that we are all on a pursuit for life’s meaning. Intrinsically tied to the definition of life is suffering and ultimately death, yet we inadvertently spend many moments of our life trying to evade death and eliminate our suffering. As a future physician, that will undoubtedly be my life pursuit, to alleviate the suffering of my fellow human beings.

When we peek inside our genome, we seek to redefine our fate, hoping to change the course alluded to in our genes. However, genes are not the ultimate road map to our life’s happiness or sorrow. We are. We decide how to respond to the uncertainties of life and all of life’s suffering. We can either choose to let ourselves be consumed by our

worries and fears or we can choose to learn from our moments of suffering. We can still triumph over the genes that we were dealt by chance. Because after all, is not life left up to chance? At conception with the random combination of maternal and paternal genes to the random people that we encounter every day, our life is left up to us to discern our response.

Although knowing the possible health problems that we may have in the future can offer hope and maybe even relief at not having any diseased alleles, we still seek to be prepared. Of course lives can be saved by knowing disease risk, yet as the use of WGS increases in medicine, it is up to each individual whether they want to know all of their DNA secrets, however uncertain they may be. Outlook on life can be altered for good or worse, but one's response reveals who we are as individuals. There is value in cherishing life with all of its uncertainties. Regardless of receiving WGS results, we are still faced with life's lessons. We can make conscious decisions to improve our lifestyles whether that is with more exercise or with moments of joy. I strongly believe that individuals can live without having a glimpse into their future and WGS cannot offer that guarantee. With the imminent use of WGS both commercially and in medicine, health providers must make a conscious effort to treat sensitive genetic information with care. Individuals who choose to have WGS need to understand their risk assessment beyond just the presence or absence of certain alleles. Before receiving sequencing results, we must recognize that our genes are not the key to knowing ourselves.

References

A Brief History of the Human Genome Project. (2012, November 8). Retrieved September 28, 2017, from https://www.genome.gov/12011239/a-brief-history-of-the-human-genomeproject/.

About – The Harvard Personal Genome Project (PGP). (n.d.). Retrieved January 24, 2018, from https://pgp.med.harvard.edu/about/

Acevedo-Whitehouse, K., Gulland, F., Greig, D., & Amos, W. (2003). Inbreeding: disease susceptibility in California sea lions. *Nature*, *422*(6927), 35.

Altmann, J., Alberts, S. C., Haines, S. A., Dubach, J., Muruthi, P., Coote, T., ... & Wayne, R. K. (1996). Behavior predicts genes structure in a wild primate group. *Proceedings of the National Academy of Sciences*, *93*(12), 5797-5801.

Arroyo-Rodríguez, V., & Dias, P. A. D. (2010). Effects of habitat fragmentation and disturbance on howler monkeys: a review. *American Journal of Primatology*, *72*(1), 1-16.

Arroyo-Rodríguez, V., & Mandujano, S. (2006). Forest fragmentation modifies habitat quality for *Alouatta palliata*. *International Journal of Primatology*, *27*(4), 1079-1096. doi:10.1007/s10764-006-9061-0

Avery, O. T., MacLeod, A. O., & McCarty, M. (1944). Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type iii. *Journal of Experimental Medicine*, *79*, 137-158.

Barton, M. (2016). *Vegetation structure of a Costa Rican rainforest fragment and the genetic structure of the inhabitant Alouatta palliata population* (Undergraduate honors thesis). Retrieved from http://epublications.regis.edu/cgi/viewcontent.cgi?article=1702&context=theses

Bastos, H. B., Gonçalves, E. C., Ferrari, S. F., Silva, A., & Schneider, M. P. C. (2010). Genetic structure of red-handed howler monkey populations in the fragmented landscape of Eastern Brazilian Amazonia. *Genetics and Molecular Biology*, *33*(4), 774–780.

Bicca-Marques, J. C. (2003). How do howler monkeys cope with habitat fragmentation? In *Primates in fragments* (pp. 283-303). Springer US.

Bloss, C. S., Schork, N. J., & Topol, E. J. (2011). Effect of direct-to-consumer genomewide profiling to assess disease risk. *New England Journal of Medicine*, 364(6), 524-534. doi:10.1056/NEJMoa1011893

Broadbent, E. N., Asner, G. P., Keller, M., Knapp, D. E., Oliveira, P. J., & Silva, J. N. (2008). Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. *Biological Conservation*, *141*(7), 1745-1757.

Brooks, L. D., Ph.D. (2015, June 3). Genetic Variation Program. Retrieved October 17, 2017, from https://www.genome.gov/10001551/genetic-variation-program/

Cacioppo, C. N., Chandler, A. E., Towne, M. C., Beggs, A. H., & Holm, I. A. (2016). Expectation versus reality: The impact of utility on emotional outcomes after returning individualized genetic research results in pediatric rare disease research,

a qualitative interview study. *Plos ONE*, *11*(4), 1-12.doi:10.1371/journal.pone.0153597

Calvo, S. E., Compton, A. G., Hershman, S. G., Lim, S. C., Lieber, D. S., Tucker, E. J., ... & Christodoulou, J. (2012). Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Science Translational Medicine*, 4(118), 118ra10-118ra10.

Chapman, C. A. (1990). Ecological constraints on group size in three species of neotropical primates. *Folia Primatologica*, *55*(1), 1-9.

Chapman, C. A., & Balcomb, S. R. (1998). Population characteristics of howlers: ecological conditions or group history. *International Journal of Primatology*, *19*(3), 385-403.

Chen, J., Saunders, S. C., Crow, T. R., Naiman, R. J., Brosofske, K. D., Mroz, G. D., ... & Franklin, J. F. (1999). Microclimate in forest ecosystem and landscape ecology: variations in local climate can be used to monitor and compare the effects of different management regimes. *BioScience*, *49*(4), 288-297.

Cherf, G. M., Lieberman, K. R., Rashid, H., Lam, C. E., Karplus, K., & Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. *Nature Biotechnology*, *30*(4), 344-348.

Clarke, M. R., Zucker, E. L., & Scott, N. J. (1986). Population trends of the mantled howler groups of La Pacifica, Guanacaste, Costa Rica. *American Journal of Primatology*, *11*(1), 79-88.

Collinge, S. K. (1996). Ecological consequences of habitat fragmentation: implications for landscape architecture and planning. *Landscape and Urban Planning*, *36*(1), 59-77.

Collins, F. S. (1999). Medical and societal consequences of the Human Genome Project. *New England Journal of Medicine, 341*(1), 28-37.

Collins, F. S., & McKusick, V. A. (2001). Implications of the Human Genome Project for medical science. *Journal of the American Medical Association, 285*(5), 540-544.

*C*ulliton, B. J. (Ed.). (2003, January 15). Genome Variations. Retrieved October 17, 2017, from http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp4_1.shtml

Dahm, R. (2008). Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Human Genetics*, *122*(6), 565-581.

Drmanac, R., Peters, B. A., Church, G. M., Reid, C. A., & Xu, X. (2015). Accurate whole genome sequencing as the ultimate genetic test. *Clinical Chemistry*, *61*(1), 305-306.

Ellsworth, J. A., & Hoelzer, G. A. (1998). Characterization of microsatellite loci in a New World primate, the mantled howler monkey (*Alouatta palliata*). *Molecular Ecology*, *7*(5), 657.

Emmons, L., & Feer, F. (1990). *Neotropical rainforest mammals: A field guide*. Chicago: University of Chicago Press.

Estrada, A., & Coates-Estrada, R. (1996). Tropical rain forest fragmentation and wild populations of primates at Los Tuxtlas, Mexico. *International Journal of Primatology*, *17*(5), 759-783.

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., ... & Tress, M. L. (2014). Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Human Molecular Genetics*, 23(22), 5866-5878.

Garber, P.A., Molina, A., and Molina, R. (2010). Putting the community back in community ecology and education: the role of field schools and private reserves in the ethical training of primatologists. *American Journal of Primatology, 72,* 785-793.

Goldacre, B. (2010, April 03). The absurdity of patenting genes: Bad Science. Retrieved October 19, 2017, from https://www.theguardian.com/commentisfree/2010/apr/03/ben-goldacre- gene-patents

Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., Martin, C. L., ... & Rehm, H. L. (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in Medicine*, *15*(7), 565.

Green, R. C., Roberts, J. S., Cupples, L. A., Relkin, N. R., Whitehouse, P. J., Brown, T., ... & Chen, C. (2009). Disclosure of APOE genotype for risk of Alzheimer's disease. *New England Journal of Medicine*, *361*(3), 245-254.

Gudbjartsson, D. F., Helgason, H., Gudbjartsson, H., Kong, A., Bjornsson, E. S., Olafsson, S., & ... Oddson, A. (2015). Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics*, 47(5), 435-444. doi:10.1038/ng.3247

Guttmacher, A. E., & Collins, F. S. (2005). Realizing the promise of genomics in biomedical research. *Journal of the American Medical Association*, 294(11), 1399-1402.

Hamdoun, Z., and Ehsan, H. (2017). Aftermath of the Human Genome Project: an era of struggle and discovery. *Turkish Journal of Biology*, 41(3), 403-418. doi:10.3906/biy-1609-77.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: the history of sequencing DNA. *Genomics*, *107*(1), 1-8.

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology*, 36, 39-56.

Heshka, J. T., Palleschi, C., Howley, H., Wilson, B., & Wells, P. S. (2008). A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. *Genetics in Medicine*, *10*(1), 19.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, *106*(23), 9362-9367.

Holley, R. W. (1968, November 12). *The genetic code*. Lecture presented at The Nobel Prize in Physiology or Medicine 1968, Stockholm. In *Nobel Lectures*. (1972). Amsterdam: Elsevier Publishing Company.

Hughes, V. (2013, January 07). Are You Scared of What's in Your Genome? Retrieved January 27, 2018, from http://www.slate.com/articles/health_and_science/medical_examiner/2013/01/ethics_of_genetic_information_whole_genome_sequencing_is_here_and_we_need.html

Hutchison III, C. A. (2007). DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research*, *35*(18), 6227-6237.

Irwin, M. T., Wright, P. C., Birkinshaw, C., Fisher, B. L., Gardner, C. J., Glos, J., & Raherilalao, M. J. (2010). Patterns of species change in anthropogenically disturbed forests of Madagascar. *Biological Conservation*, *143*(10), 2351-2362.

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., ... & Malla, S. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology*. doi:10.1038/nbt.4060

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*(1), 239.

James, R. (1992). Genetic Variation in Belizian Howler Monkeys (*Alouatta pigra*), Doctoral Dissertation, Rutgers University, New Brunswick, NJ.

Johnson, A. (2016). An impediment to personalized medicine commercialization is the lack of understanding of the value of the testing. *Journal of Commercial Biotechnology*, *22*(1), 3-5. Doi: 10.5912/jcb732.

Kawamoto, K., Lobach, D. F., Willard, H. F., & Ginsburg, G. S. (2009). A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Medical Informatics & Decision Making*, *9*(1), 1-14. doi:10.1186/1472-6947-9-17

Kelavkar, U. P. (2001). Human Genome Project. *Encyclopedia of Life Sciences*. doi:10.1038/npg.els.0001899Khorana, H. G. (1968, November 12). *Nucleic acid synthesis in the study of the genetic code*. Lecture presented at The Nobel Prize in Physiology or Medicine 1968, Stockholm. In *Nobel Lectures*. (1972). Amsterdam: Elsevier Publishing Company.

Ku, C. S., & Roukos, D. H. (2013). From next-generation sequencing to nanopore sequencing technology: paving the way to personalized genomic medicine. *Expert Review of Medical Devices*, *10*(1), 1-6.

Laurance, W. F., Ferreira, L. V., Merona, J. M., Laurance, S. G., Hutchings, R. W., & Lovejoy, T. E. (1998). Effects of forest fragmentation on recruitment patterns in Amazonian tree communities. *Conservation Biology*, *12*(2), 460-464.

Lesko, L. J., & Schmidt, S. (2014). Clinical implementation of genetic testing in medicine: a US regulatory science perspective. *British Journal of Clinical Pharmacology*, *77*(4), 606-611. doi:10.1111/bcp.12299.

Lidicker, W. Z. (1999). Responses of mammals to habitat edges: an overview. *Landscape Ecology*, *14*(4), 333-343.

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265-279.

Manolio, T. A. (2010). Genomewide association studies and assessment of the risk of disease. *New England Journal of Medicine*, *363*(2), 166-176.

Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., ... & Gundlach, J. H. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature Biotechnology*, *30*(4), 349-353.

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470(7333), 198-203.

Marian, A. J. (2012). Challenges in medical applications of whole exome/genome sequencing discoveries. *Trends in Cardiovascular Medicine*, 22(8), 219-223. doi:10.1016/j.tcm.2012.08.001

Marshall ,M. (1968, November 12). *Alanine Transfer RNA.* Lecture presented at The Nobel Prize in Physiology or Medicine 1968, Stockholm. In *Nobel Lectures*. (1972). Amsterdam: Elsevier Publishing Company.

Material presented at the graduate workshop Introduction to Genetic Analysis for

Meder, B., Haas, J., Keller, A., Heid, C., Just, S., Borries, A., ... & Weichenhan, D. (2011). Targeted next-generation sequencing for the molecular genetic

diagnostics of cardiomyopathies. *Circulation: Cardiovascular Genetics*, 11(3). CIRCGENETICS 110.958322.

Miller, N. A., Farrow, E. G., Gibson, M., Willig, L. K., Twist, G., Yoo, B., ... & Petrikin, J. E. (2015). A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Medicine*, 7(1), 100.

Milton, K., Lozier, J. D., & Lacey, E. A. (2008). Genetic structure of an isolated population of mantled howler monkeys *(Alouatta palliata)* on Barro Colorado Island, Panama. *Conservation Genetics*, *10*(2), 347-358.

Morgan, T. H. (1934, June 4). *The relation of genetics to physiology and medicine.* Lecture presented at The Nobel Prize in Physiology or Medicine 1934, Stockholm. *In Nobel Lectures.* (2014). Amsterdam: Elsevier Publishing Company.

Morin, P. A., & Woodruff, D. S. (1996). Non-invasive genotyping for vertebrate conservation. In: Molecular Genetic Approaches in Conservation. T. B. Smith & R. K. Wayne, eds. Oxford University Press. New York, 298-313.

Murcia, C. (1995). Edge effects in fragmented forests: implications for conservation. *Trends in Ecology & Evolution*, *10*(2), 58-62.

Nagarajan, R., Bartley, A. N., Bridge, J. A., Jennings, L. J., Kamel-Reid, S., Kim, A., & ... Merker, J. D. (2017). A window into clinical Next-Generation Sequencing-based oncology Testing Practices. *Archives of Pathology & Laboratory Medicine*, 141(12), 1679-1685. doi:10.5858/arpa.2016-0542-CP

National Center for Biotechnology Information. Online Mendelian Inheritance in Man, OMIM..McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), Accessed January 19th, 2018.

National Research Council. (1988). *Mapping and sequencing the human genome*. National Academies Press.

Nechvatal, J.M., Ran, J.L., Basson M.D., Nampracham, P., Niec, S.R., Badsha, K.Z., Matherly, L.H, Majumdar, A.P., and Kato, I. (2008). Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *Journal of Microbiological Methods, 72*, 124-132.

Niemiec, E., Borry, P., Pinxten, W., & Howard, H. C. (2016). Content Analysis of Informed Consent for Whole Genome Sequencing Offered by Direct-to-Consumer Genetic Testing Companies. *Human mutation*, *37*(12), 1248-1256.

NIH Fact Sheets - Genetic Testing: How it is Used for Healthcare. (2010, March 29). Retrieved January 24, 2018, from https://report.nih.gov/NIHfactsheets/ViewFactSheet.aspx?csid=43

Oklander, L. I., Kowalewski, M. M., & Corach, D. (2010). Genetic consequences of habitat fragmentation in black-and-gold howler *(Alouatta caraya)* populations from northern Argentina. *International Journal of Primatology, 31*(5), 813-832. doi:10.1007/s10764-010-9430-6

Peakall, R. & Smouse, P. (July 2009). Introduction to population genetic analysis. Populations Studies, Canberra, Australia.

Pope, T. (2000). Reproductive success increases with degree of kinship in cooperative coalitions of female red howler monkeys *(Alouatta seniculus)*. *Behavioral Ecology Sociobiology, 48*, 253-267.

Pope, T. R. (1992). The influence of dispersal patterns and mating system on genetic differentiation within and between populations of the red howler monkey (*Alouatta seniculus*). *Evolution,* 46, 1112–1128.

Pope, T. R. (1996). Socioecology, population fragmentation, and patterns of genetic loss in endangered primates. In Avise, J., and Hamrick, J., (eds.), *Conservation Genetics: Case Histories from Nature*, Kluwer Academic, Norwell, MA, pp. 119–159.

Pope, T. R. (1998). Effects of demographic change on group kin structure and gene dynamics of populations of red howling monkeys. *Journal of Mammalogy*, 79, 692–712.

Potter, B. (n.d.). MinION library preparation. Retrieved March 22, 2018, from http://bedford.io/projects/zika-seq/protocols/minion-library-preparation.html

Pruetz, J.D., and Leasor, H.C. (2002). Densities of primate species in forest fragments at La Suerte Biological Field Station, Costa Rica. *Neotropical Primates,* 10, 4-9.

Rajendran, B. K., & Deng, C. X. (2017). Characterization of potential driver mutations involved in human breast cancer by computational approaches. *Oncotarget*, *8*(30), 50252.

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology*, *17*(1), 230-237.

Reichard, P. (1968). *Award Ceremony Speech*. Speech presented at The Nobel Prize in Physiology or Medicine 1968, Sweden.

Rettner, M. R. (2013, April 16). 4 Ways the Gene Patent Ruling Affects You. Retrieved October 19, 2017, from https://www.scientificamerican.com/article/4-ways-the-gene-patient-r/

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., ... & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*(4839), 487-491.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463-5467.

Sapp, J. C., Dong, D., Stark, C., Ivey, L. E., Hooker, G., Biesecker, L. G., & Biesecker, B. B. (2014). Parental attitudes, values, and beliefs toward the return of results from exome sequencing in children. *Clinical Genetics*, *85*(2), 120-126.

Schreier, A. L., Franco, M. D., Barton, M., Pryor, N. P., & Barrickman, N. L. (2016). A preliminary examination of genetic diversity in mantled howler monkeys (*Alouatta palliata*) in a fragmented forest in Costa Rica. *American Journal of Physical Anthropology*, 159(Supp), 283-284.

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nature Methods*, *5*(1), 16-18.

Schwitzer, C., Glatt, L., Nekaris, K. A. I., & Ganzhorn, J. U. (2011). Responses of animals to habitat alteration: an overview focusing on primates. *Endangered Species Research, 14*(1), 31-38.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., ... & Navin, N. (2004). Large-scale copy number polymorphism in the human genome. *Science*, *305*(5683), 525-528.

Shmerling, M. R. (2015, October 14). First, do no harm. Retrieved January 28, 2018, from https://www.health.harvard.edu/blog/first-do-no-harm-201510138421

Skirton, H., Cordier, C., Ingvoldstad, C., Taris, N., & Benjamin, C. (2015). The role of the genetic counsellor: a systematic review of research evidence. *European Journal of Human Genetics*, *23*(4), 452-458. doi:10.1038/ejhg.2014.116

Stancu, M. C., van Roosmalen, M. J., Renkens, I., Nieboer, M., Middelkamp, S., de Ligt, J., ... & Korzelius, J. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *bioRxiv*, 129379.

Stankiewicz, P., & Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annual Review of Medicine*, *61*, 437-455.

Stoner, K. E. (1996). Habitat selection and seasonal patterns of activity and foraging of mantled howling monkeys (*Alouatta palliata*) in northeastern Costa Rica. *International Journal of Primatology*, *17*(1), 1-30.

Stordahl, N. (2013, April 08). BRCA1 and BRCA2 Gene Patent Debate Reaches the U.S. Supreme Court: Why Everyone Should Care. Retrieved October 19, 2017, from

https://www.huffingtonpost.com/nancy-stordahl/brca1-and-brca2-gene-patents_b_3015595.html

The American Society of Human Genetics Board of Directors and The American College of Medical Genetics Board of Disorders. ASHG/ASMG REPORT points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. *American Journal of Human Genetics*.;57, 1233–1241

Truong, B. (2017). *Yours, Mine, and Ours: The Ethics of Using Human and Non-Human Primate DNA in Genetic Research* (Undergraduate honors thesis). Retrieved from https://epublications.regis.edu/cgi/viewcontent.cgi?referer=https://www.bing.com/&httpsredir=1&article=1829&context=theses

Van Belle, S., Estrada, A., Strier, K.B., and Di Fiore, A. (2012). Genetic structure and kinship patterns in a population of black howler monkeys, *Alouatta pigra*, at Palenque National Park, Mexico. *American Journal of Primatology, 74,* 984-957.

Van El, C. G., Cornel, M. C., Borry, P., Hastings, R. J., Fellmann, F., Hodgson, S. V., ... & Scheffer, H. (2013). Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. *European Journal of Human Genetics*, *21*(6), 580.

Wall, J. D., Tang, L. F., Zerbe, B., Kvale, M. N., Kwok, P. Y., Schaefer, C., & Risch, N. (2014). Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Research*, *24*(11), 1734-1739.

Watson, J. D., & Crick, F. H. (1953). A structure for deoxyribose nucleic acid. *Nature*, 171, 737–738.

Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370.

Welch, J.S., Westerville, P., Ding, L., Larson, D.E., Klco, J.M., Kulkarni, S., Wallis, J., Chen, K., Payton, J.E., Fulton, R.S. and Veizer, J., (2011). Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. *Journal of the American Medical Association*, 305(15) 1577-1584.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. and Gomes, X., (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452(7189), 872-876.

Wilcove, D. S., McLellan, C. H., & Dobson, A. P. (1986). Habitat fragmentation in the temperate zone. *Conservation Biology*, *6*, 237-256.

Winkler, L. A., Peter, G., and Sohn, R. (1999). Monitoring health, genetic diversity, movement, and fission-fusion social patterns in a New World monkey (*Alouatta palliata*): an interdisciplinary project. In Bowman, J., and Huettmann, F. (eds.), *Investigations of Animal Movements: Analytical Techniques Across the Animal Kingdom*, University of New Brunswick press, Canada, pp. 79–84.

Winkler, L. A., Xichen Zhang, L. A., Ferrell, R., Wagner, R., Dahl, J., Peter, G., & Sohn, R. (2004). Geographic microsatellite variability in Central American howling monkeys. *International Journal of Primatology*, *25*(1), 197-210.

Woodruff, D. S. (1989). The problems of conserving genes and species. *Conservation for the twenty-first century. Oxford University Press, New York*, 76-88.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics, 15, 323-354.*

Yan, Y. (2009, August 26). DNA, RNA and Chromosomes. Retrieved March 25, 2018, from http://blog.nus.edu.sg/yiuyan/about/