DRUG USE AND AGE ALTERS THE DRUG METABOLISM POTENTIAL OF THE HUMAN GUT MICROBIOME

By

Travis D. D. LeBlanc

Submitted in partial fulfilment of the requirements for the degree of Master of Science

At

Dalhousie University

Halifax, Nova Scotia

August 2016

For my family,

who have always had my back.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The human microbiome is the collection of bacteria living in and on the human body. Several studies have demonstrated that the gut microbiome can alter the metabolism of commonly used medications. However, there has yet to be a broad-scope analysis to determine the presence and variance of microbial genes with drug metabolizing potential. The aim of this study is to survey the drug metabolising potential of the human microbiome, particularly in the gut. We examine what drug metabolising functions are present in these microbes, and how the abundance of these genes changes based on various factors.

A collection of 225 microbial genes with the potential to metabolize drugs were manually curated from the Kyoto Encyclopedia of Genes and Genomes. Two previously published and one unpublished study were used to relate these possible microbial drug metabolizing genes (MDMGs) in relation to several factors including body site, sex, age, frailty and medication use.

Nearly all MDMGs were found to be present in the human microbiome, with large amounts of interpersonal variation. Although there were no significant differences based on sex, we identified 146 functions that were different in proportion within the gut compared to other body sites, 63 functions that were significantly correlated with an individual's age, 24 functions correlated with frailty index score, and 40 functions that had different abundance based on certain medication use.

The ability of the human microbiome to alter drug metabolism is diverse and variable across the population. Several factors have been identified that influence the abundance of these functions, but this only represents a small proportion of the observed variation. Future studies will test other factors that could alter the drug metabolising potential of the gut microbiome, as well as the biological significance of these functions. This knowledge could then be incorporated into personalized medicine treatment.

LIST OF ABBREVIATIONS USED

ANOVA	Analysis of variance
BBR	Berberine
BP	Base pair(s)
BVU	(E)-5-(2-bromovinyl)uracil
CGR	Cardiac glycoside reductase
CRT	Conditionally rare taxa
CYP450	Cytochrome P450
CV	Coefficient of variation
dhBBR	Dihydroberberine
DNA	Deoxyribonucleic acid
FDR	False discovery rate
FI	Frailty index
GI	Gastrointestinal
GWAS	Genome-wide association studies
HMP	Human Microbiome Project
HUMAnN	HMP Unified Metabolic Analysis Network
IBD	Inflammatory bowel disease
IV	Intravenous
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO(s)	KEGG Orthology/Ortholog(s)
MDMG(s)	Microbial Drug Metabolizing Gene(s)

mRNA	Messenger ribonucleic acid
MTC	Multiple test(ing) correction
OTU	Operational taxonomic units
PC	Principle component
PCA	Principle component analysis
PCoA	Principle coordinate analysis
PEAR	Paired-End reAd mergeR
PICRUSt	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
QIIME	Quantitative Insights Into Microbial Ecology
rRNA	Ribosomal ribonucleic acid
SD	Standard deviation
SHMT	Serine hydroxymethyltransferase
SOP	Standard operating procedure
STAMP	Statistical Analysis of Taxonomic and Functional Profiles
T2D	Type-2 diabetes
Y/O	Years old

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor Dr. Morgan Langille for dedicating countless hours towards teaching me skills required to conduct my research as well as guiding me through the development of my Masters project.

I would like to thank all the members of the Langille lab, who helped me expand ideas regarding my research and provided valuable feedback throughout the years. A special thanks goes out to Dr. André Comeau and Gavin Douglas, without whom I would be lost attempting to analyze and understand bioinformatic data.

Dr. Ryan Pelis and Dr. Jana Sawynok, I am grateful for your support as members of my advisory committee. You were always available to answer my questions and to provide guidance regarding the future directions of my work.

Furthermore, I would like to thank all members of the Department of Pharmacology. Your questions throughout research day talks and individual conversations have been extremely beneficial in shaping my thesis. Additionally I would like to thank all departmental lecturers in the Intro and Advance Pharmacology courses. Your teaching provided me with the necessary foundation of knowledge to progress through my research.

To Luisa Vaughan and Sandi Leaf, I apologize for the hundreds of times I came to you with questions throughout the year. Thank you for always helping me out, and thank you for all the administrative work you do.

I would also like to thank Dr. Ryan Pelis, Dr. Joe Bielawski, and Dr. Kishore Pasumarthi for their roles as members of my defense committee.

Finally, I would like to thank my friends and family whose continuous support helped motivate me not only through this degree, but through my education as a whole.

CHAPTER 1: INTRODUCTION

1.1 PERSONALIZED MEDICINE

The term 'personalized medicine' refers to a medical model of tailoring a treatment regimen to a patient's specific need (Bates, 2010). Conceptually, this is an appropriate way of treating patients since every individual is different, and it does not make sense for everyone to receive the same treatment in every situation. The sequencing of the human genome represented a major turning point, advancing personalized medicine out of just theoretical possibilities into using an individual's genetic make up for clinical applications such as modified drug treatments and identification of disease susceptibility (Burke & Psaty, 2007; J. M. Meyer & Ginsburg, 2002; Venter, 2001).

There has been skepticism regarding personalized medicine, in both the biological significance it will have as well as our ability to analyse the vast amount of information in the human genome (Fernald, Capriotti, Daneshjou, Karczewski, & Altman, 2011; Lesko, 2007). Nonetheless, the use of personalized medicine has been crucial in our understanding of certain diseases and medical conditions. The identification of mutations in the BRCA gene have been linked to certain types of breast and ovarian cancers (Olopade, Grushko, Nanda, & Huo, 2008). Several genetic markers have been found related to the diagnosis and pathogenesis of Alzheimer's disease via genome-wide association studies (GWAS) (Kim et al., 2014). Further GWAS have found genetic relationships to several other diseases such as inflammatory bowel disease (IBD), schizophrenia, bipolar disorder, and coronary heart disease (Cardno & Owen, 2014; Chen et al., 2014; den Hoed et al., 2015). These studies highlight the fact that as our abilities in genetic analysis grow, a deeper understanding of several diseases and conditions on the molecular level will be established.

The term personalized medicine can be expanded beyond just human genome markers to include other aspects that make an individual unique. In the case of obesity and nutrition, a personalized treatment would not only look at genetics, but take into

consideration the dietary patterns of an individual, as well as anthropometry and levels of physical activity (Gibney & Walsh, 2013).

Advances in medical technology also aid in personalized medical treatments. This is especially true in complex conditions like Crohn's disease. The exact cause of this disease is unknown, and those suffering from Crohn's experience a wide and variable range of symptoms that may change as the disease progresses (Baumgart & Sandborn, 2012). In cases such as this, identifying the stage and nature of the disease is crucial for effective treatment. Development and improvement of techniques such as ultrasonography, computed tomography, and molecular endoscopy allow physicians to get a better idea of what aspects of Crohn's are present in a particular individual (Neurath, 2015). From there, a treatment can be prescribed that best fits the patient's current need.

Identifying genetic markers and other factors related to disease susceptibility is an important aspect of personalized medicine. However, identification of these factors is only the first step. In order for this information to be helpful for a patient, an effective medical treatment must be initiated, which often requires some type of drug regimen. This brings up another broad area of personalized medicine. Drugs are not all 'one-size-fits-all', and our uniqueness may actually change the way our bodies process different drugs.

1.2 DRUG METABOLISM

A compound moving through the human body does so in four major steps: absorption, distribution, metabolism and excretion**.** All steps are essential and interconnected, but the step most interesting in the field of personalized medicine is metabolism. Drug metabolism is the process by which drugs (or xenobiotics) are modified by the body. These modifications often make a compound more hydrophilic, making it more readily eliminated from the body, thus preventing potentially toxic buildup (Remmer, 1970). Drugs are a critical aspect of modern medicine, and detailed

understanding of what is happening to drugs as they move through our bodies is essential for the health and wellbeing of the population.

Our understanding of drug metabolism starts by taking a step back and examining differences based off the route of administration. Major routes include inhalation, injection, oral, topical, and sublingual, each with their own pros and cons (Rowland, 1972). The most common route of administration is oral, as it is convenient and production of drugs usually costs less than with other routes. One major issue with oral formulation of drugs is their susceptibility to the effect known as 'first-pass metabolism' (Pond & Tozer, 1984). This is a process by which a drug is metabolised between its site of entry into the body, and its target site (or the site at which concentration is measured). With oral drugs, the liver is the major site for this first pass effect, but other sites such as the gastrointestinal tract also contribute. As a result the bioavailability of oral drugs, the proportion reaching systemic circulation, is reduced below 100%. On the other hand, drugs administered via intravenous (IV) injections have 100% bioavailability as they are injected directly into circulation and show no first-pass effect. First-pass metabolism is closely related to drug dose, as a treatment is beneficial only when the intended amount of drug is entering the system.

As mentioned previously, the liver is crucial organ for drug metabolism. This is because the liver contains many diverse enzymes utilized in xenobiotic metabolism (Remmer, 1970). These metabolism enzymes are broken up into two main categories, phase I and phase II, although some phase III modifications are sometimes included when discussing xenobiotic metabolism. In phase I metabolism, a compound is modified via oxidation, reduction, or hydrolysis, making the target more hydrophilic and easier to excrete. In phase II metabolism, enzymes such as methyltransferase, sulfotransferase, acetyltransferase, glucuronosyltransferase, glutathione S-transferase, and acetyl CoA are used to further reduce activity and increase water solubility of a xenobiotic compound. (Liston, Markowitz, & DeVane, 2001; Sheweita, 2000). Phase III includes modifications not present in the other two classes, including beta-glucuronidase, beta-glucosidase, and azoreductase (Haiser & Turnbaugh, 2013).

A major family of enzymes involved in phase I metabolism are the cytochrome P450s (CYPs). These enzymes are broken down into families, subfamilies and individual genes, all of which play different roles in drug metabolism. The primary CYP enzyme in humans is CYP3A4, where some studies predict this enzyme metabolizes over half of all drugs (Smith, 2009). Interestingly, it has been found that aspects of our diet such as grapefruit juice contain active compounds known as furanocoumarins which strongly inhibit the CYP3A4 enzyme, leading to increased bioavailability for drugs normally metabolized by CYP3A4 (He et al., 1998; Paine et al., 2006). This highlights the fact that interactions with the metabolic systems in our bodies must be taken into account, as unknowingly altering them may disrupt a medical treatment and result in harm.

Other enzymes in the CYP family have important clinical significance. CYP2D6 and CYP2C9 are abundant enzymes in the liver (Zhou, Liu, & Chowbay, 2009; Zhou, Zhou, Yang, & Cai, 2009). What is interesting amount these particular enzymes is the amount of polymorphism, the presence of multiple different forms, and gene copy number across the population. While many individuals contain the normal number of genes for these enzymes, some individuals may lose one or two alleles for a CYP, causing them to be classified as poor metabolisers. Along the same line, an individual may have a duplication in the gene for one of these CYPs, classifying them as an ultrarapid metabolizer. Race is one of the principle factors behind these polymorphisms with, for example, 5-14% of Caucasians lacking CYP2D6 activity, and less than 1% of Asians classified as poor metabolizers for this enzyme (Zhou, Liu, et al., 2009). This has enormous biological consequences. If an individual with an unknown genotype is taking a drug metabolized primarily by one of these CYPs the drug could be cleared too quickly to have an effect (in the case of an ultrarapid metaboliser), or the drug could build up in the system leading to harmful side effects (in the case of a poor metaboliser) (Teh & Bertilsson, 2012). These genes illustrate the importance of identifying internal differences in drug metabolism functions, and adjusting treatment so the patient receives the intended level of drug in their system.

These metabolizing enzymes, along with many others, are present to protect the body against foreign toxins. However, they also serve as a roadblock for effectively

getting a drug, particularly through the oral route, into systemic circulation. The dose of a drug must be large enough so that it is not completely metabolized and eliminated before causing its desired effect, but no so large that an excessive amount can result in toxicity. The ratio of the dose that would cause toxicity over the dose that produces an effect is called the therapeutic index (Muller & Milton, 2012). By understanding how a drug is metabolized, a dose can be prescribed so that the amount in systemic circulation remains within the window of being effective without hitting toxicity.

In some situations, body's metabolism enzymes can be used advantageously. This is the case for drugs like codeine, where the metabolites of the drug (in this case, morphine) are actually more pharmacologically active than the parent form (Haffen et al., 2000). Drugs in this category are defined as prodrugs. The relationship between prodrugs and metabolic polymorphisms is essentially opposite to that of normal drugs. If a prodrug is given to a poor metabolizer, the prodrug will not be readily converted to its active form, therefore it may not hit effective therapeutic levels. However, a prodrug given to an ultrarapid metabolizer would be rapidly converted to its active form, potentially hitting toxic levels faster than a normal metabolizer taking the same dose.

The circulation systems adds an increased layer of complexity to drug metabolism. The liver's main role in drug metabolism is to modify drugs for elimination. However, in a process known as enterohepatic circulation, certain drugs may be moved from the liver to the bile, deposited in the intestine, and then transported back to the liver instead of being excreted (M. S. Roberts, Magnusson, Burczynski, & Weiss, 2002). As a result, these drugs may have a longer duration of action, increased pharmacological effect, or result in toxicity especially in the liver.

Taking a broader look, several factors are connected to the process of drug metabolism in the body. Lifestyle factors such as diet were previously mentioned to impact drug metabolism, but several other physiological factors are also important. In elderly individuals, physiological functions begin to deteriorate resulting in the inability to clear drugs as effectively. This could be tied to a reduction of functioning liver cells, decreased cardiovascular function and therefore a lower circulation rate of compounds through the body, or even decrease in nephrons in the kidney impacting the clearance of

xenobiotics. As a result a lower dosage is often needed to avoid toxicity (Turnheim, 2004). The same is true for patients with diseases effecting the liver or blood flow, as the rate of drug metabolism clearance will be reduced which may result in increasing concentrations of an un-metabolized drug (Thomson, 1973). Obesity has been linked to direct changes in enzyme activity, as well as changes in hepatic and renal function that could alter the metabolism and elimination of a drug (Hanley, Abernethy, & Greenblatt, 2010). Taking all of these factors into consideration is essential in providing proper medical treatment that suits an individual's unique situation.

1.3 MICROBES

Breaking down xenobiotics is not an ability unique to humans. In fact, these functions were around long before humans ever evolved on this planet. In the infancy of Earth when life was just emerging, there exists microbes. These tiny creatures have spent the last 3 billion years learning how to survive and spread across the planet to countless niches (Visscher et al., 2000). Although the vast majority of these organisms are not visible with the naked eye, it has been estimated that collectively, the biomass of all microbes is larger than that of plants and animals combined (Hogan, 2011). The term 'microbes' encompasses many things, including archaea, bacteria, eukaryota, and in some sense viruses. However, within this thesis I will be using the term microbe primarily in reference to the domain of bacteria unless otherwise stated.

These microbes are tiny, but they are also extraordinarily complex. Like humans, their genetic information is encoded into DNA, and that DNA is used for their development and to carry out various functions. When humans reproduce, their progeny in most cases gets 23 chromosomes of DNA from each parent. Alternatively microbes often undergo a process called binary fission, where a parent cell replicates its DNA and splits into two daughter cells (although other reproductive methods are known) (Angert, 2005). Some of the tools microbes have to process their DNA are remarkably similar to humans. For example, the ribosome, a cellular organelle responsible for protein synthesis, is found in both humans and microbes. In both groups, the ribosome acts in very similar

fashion; reading mRNA and using that information to join together amino acids. On the other hand, the different subunits of the ribosome are slightly different between humans and microbes. In the small subunit of a human ribosome there is the 18S rRNA (the S standing for Svedberg units, a type of 'size' measurement), whereas the equivalent match in bacteria and archaea the small subunit is 16S rRNA (Janda & Abbott, 2007).

In certain situations, microbes appear to be even more complex than humans. One such example of this is the operon. In human's, every gene is regulated by its own promotor and terminator. However, many microbes make use of an operon, a cluster of genes regulated by a single promotor (Blumenthal, 2004). There is an obvious functional benefit to this type of organization. If one particular pathway requires several different proteins, it would make sense to produce all those proteins at the same time in one 'assembly line' rather than having multiple different processing centers for each required piece.

Microbes are also more complex than humans in the way their genes are inherited. Outside of gene therapy and viral infections, every human acquires all of their DNA from their parents (Flotte & Carter, 1995). This straightforward path is why it is so easy to use DNA in order to trace family trees. Additionally, this means that, in general, humans will have many features and functions similar to their parents. Microbes on the other hand, are able to undergo a process called lateral gene transfer (LGT). In addition to getting DNA from the division of their 'parent' cell, microbes are able to take in genes from other microbes (Ochman, Lawrence, & Groisman, 2000). The consequences of this process are significant, as depending on what genes are transferred, a microbe could have many functional capabilities their parent did not possess leading to an adaptive advantage. For example, LGT of particular genes are known to allow microbes to gain resistance to particular antibiotics (Stokes & Gillings, 2011). Bacteria with this newly acquired function would survive and be selected for in an environment with antibiotics, whereas other cells with identical DNA except the portion that was transferred would not survive. Gene transfer is an incredible adaptive advantage as it does not require microbes to wait a generation to acquire different DNA, and critical functions may be able to quickly spread to an entire microbial population.

Microbes have an enormous array of functions they can use for their own benefit, as well as inadvertently benefitting every other organism on this planet. Over three billion years ago in the early years of life on planet Earth, microbes were responsible for a crucial processes called nitrogen fixation, where normally unreactive nitrogen in the atmosphere (N_2) was converted to ammonia (NH_3) (Dixon & Kahn, 2004). This allowed nitrogen to be available for other organism, and since it is an essential compound for building DNA and proteins, many living organisms would never have existed if it weren't for these microbes. Microbes are able to break down previously-living plants and animals, utilizing some dissolved organic matter for themselves, but also releasing carbon compounds back into the environment (Amon & Benner, 1996). They are able to degrade compounds that are often viewed as harmful to the environment such as plastics and oils (Leahy & Colwell, 1990; Mergaert, Webb, Anderson, Wouters, & Swings, 1993). Other environmental threats, such as aromatic pollutants, are not safe from microbes. Due to the speed of adaptation, many microbes are able to acquire genes allowing them to break down these toxins and use them for sustenance (Díaz, 2004). It has also been found that microbes are able to break down other human-made compounds such as therapeutic drugs (Goldman, 1978).

These microbes and their corresponding functions make them a valuable tool for human utilization. Microbes with specific genes can be enriched in areas of environmental concern to harness their ability to degrade toxins as mentioned above. In addition, microbes have been used in sewage treatment to break down particular compounds such as ammonia (Wagner, Rath, Koops, Flood, & Amann, 1996), while nitrogen fixation by microbes is often used to increase production of agriculture by providing valuable nitrogen to growing plants (Kahindi et al., 1997). Manipulation of microbe metabolism has been conducted using microbes in order to produce ethanol fuel (Ingram et al., 1999). Microbes also play a fundamental role in the fermentation process for various foods such as bread, cheese, wine, and beer (Hutkins, 2006).

In addition to the diverse range of functions, microbes have a variable range of habitats. In the soil, and in the ocean, billions of microbes can be found breaking down compounds and using them as nutrients (Hellweger, van Sebille, & Fredrick, 2014; Van

Der Heijden et al., 2006). They can also be found in the sky, being carried by the winds or in clouds and may even influence weather patterns (DeLeon-Rodriguez et al., 2013). Many microbes can stick closely together to form biofilms (Donlan, 2002). In some situations, a microbe can be found in very low numbers, but suddenly explode in population size after the right trigger (Shade et al., 2014). This implies that even rare microbial species might be capable of producing dynamic effects under the correct conditions.

Many microbes are known as extremophiles, meaning they live in extreme conditions. They can be found on hydrothermal vents in the ocean, or even in the Marianas Trench (Emerson & Moyer, 2002; Takami, Inoue, Fuji, & Horikoshi, 1997). They can be found growing on rocks, in hot springs, and in tar lakes (Bott & Brock, 1969; Kim & Crowley, 2007; Walker & Pace, 2007). Some microbes can even survive on radioactive waste (Makarova et al., 2001). In these situations, unique genes allow microbes to survive and proliferate in these exotic environments.

In addition to environmental factors, community collaboration also play a role in where bacteria can survive. In some situations there exists see commensal relationships, where microbes live in an environment with the assistance of other organisms, without (normally) causing harm (Tlaskalová-Hogenová et al., 2004). In other situations symbiosis occurs, where one species of microbes is benefiting from, but also assisting another microbe or organism (Van Der Heijden et al., 2006). These relationships indicate that when studying microbial behavior, looking at one species at a time may not be sufficient. Instead, collections of microbes should be as a functional community and their capabilities should be assessed as a whole.

There have been hundreds of studies performed, characterizing the microbes in the environments listed. The functions of these microbes are closely linked to the location they are found in, and studying these microbes often provides new information about their habitat. However, one particular ecosystem that microbes have settled in is quickly rising to the forefront of human health research.

1.4 THE MICROBIOME

Microbes are found throughout the world in all different biomes, which are usually large areas classified based on geography and climate. Microbes also have the capability to make their homes in a much less stationary habitat; inside living organisms (NIH HMP Working Group et al., 2009). The collection of microbial cells living inside an organism is known as the organism's microbiome. The term microbiome may also be used to reference all the genomes present in the microbes living in a host. Some older estimates put the number of microbial cells an order of magnitude higher than the number of cells in their host, but now that ratio is predicted to be closer to 1:1 (Savage, 1977; Sender, Fuchs, & Milo, 2016). In addition to the vast number of microbial cells, the total number of genes found in the microbiome is extensive. This work is currently ongoing, but some large-scale studies have characterized over three million unique genes in the human microbiome (Qin et al., 2010), a stark contrast to the approximately 23 thousand genes within the entire human genome. Analogous to microbes living in large ecosystems like the ocean or in soils, the microbes living within organisms carry out many different functions that can have immense impacts on their hosts.

Based on the majority of previous research in infectious diseases, microbes residing in and on the human body have been given negative connotations. However, the microbes inhabiting living organisms are, for the most part, non-harmful (Martin J. Blaser, 2014). In fact, the microbiome is critical for the daily functioning of many species. In animals with rumen such as cows and sheep, microbes are able to break down large plant fibre such as cellulose, providing nutrients to the animal (Weimer, 1992). In a similar sense, termites are able to beneficially consume wood because microbes inside of them break down the cellulose and the termite utilizes the end products (Ikeda-Ohtsubo & Brune, 2009). In some insects such as aphids, particular microbes seem to have co-evolved with their host. These endosymbionts provide essential nutrients that the aphids have stopped consuming from their diets (Baumann, Moran, & Baumann, 1997).

The microbiome is also essential in humans. It has been termed the 'forgotten organ' due to its importance in our health and development (O'Hara & Shanahan, 2006).

The species of microbes that make up our microbiome is unique from person to person (M. J. Blaser, 2010). It was initially thought that microbes first colonized human bodies as babies exited their mother womb, but some evidence suggests that placental microbes may begin this process even earlier (Aagaard et al., 2014). Microbes then proceed to cover most of our bodies, including our skin, armpits, mouths, groins, lungs and most abundantly throughout our gastrointestinal tract. Comparable to microbes inhabiting different ecosystems, the microbes found in different parts of our bodies have some unique and distinct functions. (The Human Microbiome Project Consortium, 2012).

In early years, the microbiome is chaotic and changing, but during adulthood, it becomes relatively stable (Faith et al., 2013; Koenig et al., 2011). Over longer time periods, there is evidence of noticeable shifts in the microbiome (Yatsunenko et al., 2012). In addition to age, frailty may also be important in the composition of the microbiome (M. G. Langille et al., 2014). One major factor that alters the microbiome is diet (David et al., 2013). Use of antibiotics can wipe out many of the microbes living in our bodies, causing drastic shifts in microbial composition (Francino, 2016). Other drugs have been shown to alter the microbiome such as proton pump inhibitors, but this is more subtle shift in composition rather than an elimination of microbes that occurs with antibiotics (Zhernakova et al., 2016).

Human health relies on our microbiome. The microbes in the gut break down foods that are indigestible by our own systems (D'Argenio & Salvatore, 2015). Humans are unable to make certain essential vitamins on our own. These vitamins can be taken in through our diets, but the microbes in our gut are capable of synthesizing many of these vitamins for us (LeBlanc et al., 2013). This is particularly important in cases like vitamin K2, where we rely strongly on our microbiome to produce this compound to avoid deficiency (Conly & Stein, 1992). The microbiome may impact physiological factors such as blood pressure (Holmes et al., 2008), and the gut microbiome may play a role in influencing and regulating the brain and nervous system (Carabotti, Scirocco, Maselli, & Severi, 2015).

The immune system is closely linked to the human microbiome. The role of the immune system is to prohibit invading pathogens. During development the immune

system must learn to recognize and distinguish between the beneficial or non-harmful bacteria inhabiting the human body, and potentially dangerous pathogens trying to attack it. The gut microbiome does not just sit idly by. It actively mediates and assists our immune system in fighting off unwanted microbes (Round & Mazmanian, 2009). Dysregulation of this cooperation may cause severe problems due to the immune system not functioning properly.

The microbiome has been linked to several diseases. Diseases such as asthma, diabetes, and particular cardiovascular diseases have all been associated with the human microbiome (M. J. Blaser, Chen, & Reibman, 2008; Qin et al., 2010; Wang et al., 2011). Other conditions appear to have much stronger connections to 'dysbiosis' or irregularity in the microbiome.

Microbes in the gut have been strongly linked to inflammatory bowel disease (IBD). There are many different types of IBD with Crohn's disease and ulcerative colitis being the two principle diseases, each with slightly different symptoms (Gophna, Sommerfeld, Gophna, Doolittle, & Veldhuyzen van Zanten, 2006). In most cases of IBD there is dysbiosis in the microbiome, reduced diversity of microbes present, and an unhealthy response by the immune system to the gut microbes (Aleksandar D Kostic, Xavier, & Gevers, 2014). Since there appears to be an improper balance between the microbiome and the immune system, it has been speculated that the use of probiotics containing 'beneficial' microbes may be able to restore the microbiome to a state more similar to non-IBD patients (Jonkers, Penders, Masclee, & Pierik, 2012).

Obesity is a major health concern, especially in North America. Studies have shown that particular microbes are reduced in obese individuals (Ley, Turnbaugh, Klein, & Gordon, 2006). Additionally, particular functions of the gut microbiome appear to be responsible for translating the consumption of a 'Western' diet into increased levels of obesity (Bäckhed, Manchester, Semenkovich, & Gordon, 2007). Evidence from mouse models supports the connection between the gut microbiome and obesity, as researcher have demonstrated that transferring a microbiome associated with obesity to a mouse will result in increased fat gain compared to a non-obesity associated microbiome (Turnbaugh et al., 2006).

Many studies above highlight how the microbiome (or disruptions in it) can be linked to several disease conditions. Another side to this story is how the microbiome may play a role in recovery from unhealthy states. This is especially true when it comes to therapeutic drug treatment in patients.

Drug metabolism is influenced by the microbiome. This is not a new idea, as even in the 1970's researchers began noticing that the reactions performed by the microbes found in our gut could modify and break down drugs (Goldin, Peppercorn, & Goldman, 1973). For several decades this area was not explored in greater detail, presumably due to the complexity and variety of the microbiome. More recently, advancements in our sequencing technology and our computing power have allowed us to analyse the microbiome in a much more effective way than what was done 40 years ago. As a result, massive databases have been established such as the Human Microbiome Project (HMP) and the MetaHIT project, in order to try and catalogue all genes found in the human microbiome (Qin et al., 2010; The Human Microbiome Project Consortium, 2012). With this information, the field of pharmacomicrobiomics, or microbiome-drug interactions, is growing quickly and in the near future will likely play a large role in therapeutic treatment and personalized medicine (ElRakaiby et al., 2014).

1.5 PHARMACOLOGICAL IMPACT OF THE MICROBIOME

As previously outlined, one significant function of the human microbiome in regards to health is its ability to modify drugs. The microbes, mainly in our gut, may be modifying drugs so they are no longer as effective or potentially causing them to be toxic. Additionally, some drugs may actually rely on the microbiome for activity in humans. It is becoming increasingly clear that inclusion of the gut microbiome is a necessary component of studying drug metabolism in individuals (Nicholson, Holmes, & Wilson, 2005). This section outlines several microbiome-drug interactions.

1.5.1 DIRECT MODIFICATION BY THE MICROBIOME

This section provides evidence of the microbiome directly modifying drugs or particular compounds.

1.5.1.1 DIGOXIN

Digoxin is a cardiac glycoside that is extracted and purified from the plant *Digitalis lanata*. Digoxin works by inhibition of the Na/K-ATPase membrane pump, leading to increased intracellular levels of calcium and activation of contractile proteins (Haiser, Seim, Balskus, & Turnbaugh, 2014). Its use dates back to the 18th century, and although it is no longer a front-line drug, it may still be used to treat irregular heart beat as well as heart failure when these symptoms cannot be managed with other medication (Gheorghiade, 2006). One of the major issues concerning Digoxin is its narrow therapeutic window. Although Digoxin is dangerous when serum levels are above the therapeutic window, this case study arose from situations where Digoxin was actually below its therapeutic window.

Researchers studying Digoxin found that in about 10% of patients, the level of the drug in the body was greatly reduced (Lindenbaum, Rund, Butler, Tse-Eng, & Saha, 1981). Additionally, an inactive metabolite was found in this subset of patients that was not found in the others. When these researchers treated some of these patients experiencing reduced Digoxin levels with antibiotics, the inactive metabolite was greatly reduced or eliminated, and serum concentration of Digoxin rose, in some cases as high as double the Digoxin levels before antibiotics. Lindenbaum et al. concluded that it must be the gut microbes responsible for Digoxin metabolism.

Three decades later, researchers took another look into Digoxin in order to try and illustrate some of the mechanisms of action behind drug metabolism in the gut bacteria (Haiser et al., 2013). They were able to identify that one particular bacteria, *Eggerthella lenta*, was responsible for reducing the drug to its inactive metabolite. This bacteria was

successfully grown in culture, and it was found that in the presence of Digoxin, a newly discovered operon containing two genes was substantially upregulated. These two genes were named cardiac glycoside reductase (cgr) 1 and 2, and it was found that these genes respond to a broad range of cardiac glycosides, not only Digoxin. Finally, Haiser et al. also determined that an arginine-rich diet suppressed the cgr genes and therefore decreased the extent Digoxin was reduced, further highlighting the importance of diet on the composition and function of the gut microbiome.

1.5.1.2 L-DOPA

Parkinson's disease is a neurodegenerative disorder affecting over 50 million people worldwide and responsible for over 100 000 deaths annually (Global Burden of Disease Study 2013 Collaborators, 2015). Currently, the most effective and well tolerated drug for treating Parkinson's disease is L-DOPA (Vlaar, Hovestadt, van Laar, & Bloem, 2011). L-DOPA is orally administered, and it works by entering the body, crosses the blood brain barrier, and is converted to dopamine (Goole & Amighi, 2009). The increase of dopamine in the central nervous system relieves some of the symptoms of Parkinson's disease in these patients.

The crucial part of L-DOPA treatment is getting an acceptable level of dopamine in the brain, without increasing it to the point of causing schizophrenic symptoms (Swerdlow & Koob, 1987). As a result, doses of L-DOPA are kept relatively low to prevent these side effects. Unfortunately, an even smaller amount of L-DOPA might be crossing the blood brain barrier than initially expected. Researchers found that L-DOPA was metabolized when incubated with rat feces (Goldin et al., 1973). Therefore, it could be hypothesized that the gut microbiome could be capable of metabolizing L-DOPA before it can cross the blood-brain barrier, therefore reducing the amount of converted dopamine, and hindering the effectiveness of L-DOPA treatment.

1.5.1.3 AZO COMPOUNDS

Sulfasalazine is a prodrug used to treat arthritis and IBD (Sousa et al., 2008). The anti-inflammatory portion of the prodrug, 5-ASA, is linked to a sulfapyridine molecule by an azo bond. Researchers observed the two separated sulfasalazine metabolites in the feces of rats and when examining antibiotic-treated or germ-free rats, only the unchanged version of the drug remained (Peppercorn & Goldman, 1972). They concluded that azoreductases in the gut microbiome were responsible for reducing sulfasalazine. This finding was leveraged to ultimately create a prodrug called olsalazine, consisting of two 5-ASA molecules connected via azo bond (Wadworth & Fitton, 1991). This allowed the prodrug to be delivered to the gut, activated locally at the site of inflammation via microbial azoreductase, and avoided the side effects associated with the sulfapyridine portion of sulfasalazine.

Unfortunately, this azoreductase activity can also be harmful. Azo dyes are used for their vivid colours, but can enter the body by ingestion in inhalation. Researchers found that the gut microbiome was responsible for reducing these azo dyes into toxic aromatic amines (Xu, Heinze, Chen, Cerniglia, & Chen, 2007).

1.5.1.4 BERBERINE

Berberine (BBR) is a drug used to effectively treat high cholesterol by increasing low-density lipoprotein receptor expression levels (Kong et al., 2004). Initially, its mechanism of absorption was unknown as it was found to have poor solubility. Researchers found that incubation BBR in vitro with several different bacteria resulted in the production of an absorbable metabolite, dhBBR (Feng et al., 2015). From this finding, they suggested that the gut microbiota could be converting BBR to dhBBR in the intestine, allowing it to be absorbed by the body and then converted back to BBR in the blood and tissue. Further experiments using mice treated with BBR help support this claim. Mice given antibiotics showed a decrease in the conversion of BBR to dhBBR, as

well as a reduction of BBR in the blood stream compared to the control group not on antibiotics (Feng et al., 2015).

1.5.1.5 LEVAMISOLE

Levamisole is a now withdrawn drug that was previously used to treat parasitic worm infections as well as colon cancer (Dillman, 2011; Keiser & Utzinger, 2008). Several thiazole ring-opened metabolites were detected in the feces of patients on levamisole (Shu, Kingston, Van Tassell, & Wilkins, 1991). Through culturing experiments, Shu et al. demonstrated several bacteria found in the human gut were shown to have this ring-opening capability. Additionally, mixtures of bacterial strains were more effective at metabolising levamisole than any one strain in isolation.

1.5.1.6 SORIVUDINE

Sorivudine is an antiviral drug potent at treating infections by varicella-zoster virus and herpes simplex virus type 1. Drug-drug interactions are a concern when using sorivudine due to the metabolite of the drug, (E)-5-(2-bromovinyl)uracil (BVU). BVU has been shown to inhibit the breakdown of particular anticancer drugs such as 5-fluorouracil, resulting in accumulation and toxicity (Nakayama et al., 1997).

Nakayama et al. were interested in the conversion of sorivudine to BVU, as studies in rats showed very little enzymatic metabolism of this drug, even in the liver. They suspected the gut microbes were responsible, and measured phosphorolytic enzymes (the enzymes behind sorivudine metabolism) in several gut bacteria strains. It was discovered that many anaerobic strains had high levels of these enzymes. To strengthen their findings, they compared sorivudine metabolism in rats on general antibiotics versus rats on selective aerobic antibiotics. Rats given general antibiotics had reduced levels of BVU in their serum compared to untreated controls. On the other hand,

rats on aerobic antibiotics had an increase in BVU relative to the controls, highlighting that select anaerobic bacteria were metabolizing sorivudine into BVU (Nakayama et al., 1997).

1.5.1.7 CHOLINE

Choline is a nutrient that is essential for many functions in the human body (Zeisel & da Costa, 2009). Choline is ingested through normal diet, and it is also available as a supplement (Ueland, 2011). Unfortunately, high levels of choline is a risk factor for cardiovascular complications, especially atherosclerosis or the accumulation of plaques on the walls of arteries. Researchers discovered that gut microbes were capable of converting this dietary choline into trimethyl amine (TMA) via microbial TMA lyase (Wang et al., 2015). TMA is converted in the body to trimethylamine-N-oxide (TMAO), which is a compound strongly associated with cardiovascular events (Tang et al., 2013).

Researchers then wanted to decrease the risk of cardiovascular events caused by dietary choline without using antibiotics to wipe out the gut microbiota. Instead, their goal was non-lethal inhibition of these gut microbial enzymes. They discovered that a compound, 3,3-dimethyl-1-butanol (DMB), was able to inhibit TMA formation via microbial TMA lyase (Wang et al., 2015). This resulted in decreased TMA and TMAO production and was able to counter the increased risk of atherosclerosis in a choline-rich diet.

1.5.1.8 EQUOL

Another dietary compound of interest associated with the microbiome is equol. This compound is derived from the isoflavonoids in soy, and it has been linked to several health benefits such as prevention of prostate cancer and antioxidant activity (Akaza et al., 2004; Rüfer & Kulling, 2006). Studies comparing germ-free rats and rats containing

bacteria found in the human gastrointestinal tract have demonstrated that the gut bacteria are responsible for the conversion of isoflavonoids to equol (Matthies, Loh, Blaut, & Braune, 2012). Researchers were able to identify and isolate a novel gene in a particular strain of Lactococcus that is responsible for this metabolic production of equol (Shimada et al., 2010).

1.5.2 INDIRECT MODIFICATION BY THE MICROBIOME

The microbiome may influence drug metabolism without being exposed to the drug initially. This section discusses the actions of the microbiome involved with altering the course of a drug's movement through the body via enterohepatic circulation.

1.5.2.1 CPT-11

CPT-11 (also called irinotecan) is a drug used for the treatment of cancer, particularly colon cancer (Wallace et al., 2010). CPT-11 works by inhibiting topoisomerase 1, thereby preventing DNA from unwinding. CPT-11 is a prodrug given intravenously, which gets converted in the serum via carboxylesterase enzymes to its active form called SN-38. This active form eventually undergoes glucuronidation in the liver via UGT1A1, resulting in inactive SN-38G. Although this drug is listed on the WHO Model List of Essential Medicines, many patients experience debilitating side effects including severe nausea and diarrhea.

Wallace et al. were able to identify a major contributor to this toxicity. They discovered that microbial beta-glucuronidases in the gastrointestinal tract were targeting the inactive SN-38G, and reactivating it back into SN-38. The active form then re-entered circulation, causing increased serum levels and exacerbating the toxic side effects experienced by patients. (Wallace et al., 2010).

From this point, a simple suggestion might have been to co-administer antibiotics with CPT-11 to reduce toxicity. However, researches wanted to reduce the side effects of the drug without completely disrupting the bacteria in the gut and to demonstrate that they understood the precise mechanism of action. To do this, they performed high-throughput screening of potential bacterial beta-glucuronidase inhibitors, to identify ones that had no impact on the human form of the enzyme. They tested these inhibitors to show that they were effective at targeting bacterial beta-glucuronidase without killing other bacteria or human cells. Finally, they used a mouse model to show that oral administration of an inhibitor was able to prevent CPT-11 toxicity (A. B. Roberts, Wallace, Venkatesh, Mani, & Redinbo, 2013).

1.5.3 MICROBIOME-HOST INTERACTIONS

In addition to modification of drugs, the microbiome is also able to influence its host. These changes in the host may precipitate changes in the host's ability to metabolize drugs. This section explores two documented cases of microbiome-host interactions.

1.5.3.1 IMPACT ON HOST GENES

As discussed in Chapter 1.2, the human liver has a wide range of genes that play a role in xenobiotic metabolism. The abundance and expression of these genes may vary from person to person based on genetic factors such as polymorphisms, but it has also been documented that the gut microbiome may also influence the expression of these genes (Björkholm et al., 2009). In a study of hepatic gene expression comparing germ free and control mice, it was found that 112 genes related to xenobiotic metabolism were differentially expressed between the two groups. This finding implies that the gut microbiota may be able to influence drug metabolism indirectly by altering host gene expression.

1.5.3.2 IMPACT ON HOST METABOLISM PATHWAYS

Other researchers have investigated the impact the gut microbiome may have on particular human xenobiotic metabolism pathways. One study in particular looked at the metabolism of acetaminophen (Clayton, Baker, Lindon, Everett, & Nicholson, 2009). Acetaminophen undergoes phase II metabolism primarily via O-sulfonation and glucuronidation, resulting in acetaminophen sulfate and acetaminophen glucuronide, respectively. An interesting correlation was discovered between low levels of the sulfate metabolite, and high levels of a microbial metabolite called p-Cresol. Microbial p-Cresol is also metabolized through the hepatic O-sulfonation pathway (Gamage et al., 2006). Researchers proposed that p-Cresol was competing with acetaminophen for hepatic sulfotransferase activity, therefore reducing the amount of acetaminophen sulfate produced. This finding has implications not solely for acetaminophen metabolism, but any xenobiotic that O-sulfonation pathway, as well as promoting further investigation into potentially competitive microbial metabolites.

1.6 ANALYSING THE MICROBIOME

1.6.1 MOUSE MODELS

Mouse models can be used to analyse aspects of the microbiome on a much larger scale than in humans, with more flexibility and control of the variables in question. At first glance, one might assume the gastrointestinal structure of a mouse is vastly different from a human based off differences in size and obviously species. Although there are noted differences, generally speaking mouse physiology, anatomy, and even genetics are quite similar to humans (Nguyen, Vieira-Silva, Liston, & Raes, 2015) This, combined with the diversity of models available make mice an attractive candidate for studying the importance of the microbiome, although other animal models have been used successfully (A. D. Kostic, Howitt, & Garrett, 2013). These mouse models can be divided into three general groupings.

The first type of mouse model involves treating mice with drugs. Often times, researchers use antibiotics in order to eradicate the majority of bacteria in the gastrointestinal tract (Feng et al., 2015). From there, they can compare antibiotic-treated animals to the controls, providing strong evidence that the microbiome is responsible for differences observed between these two groups. In addition to antibiotic treatments, mice may be given inhibitors of specific bacterial enzyme (Wallace et al., 2010). These studies are beneficial as the majority of the microbiome's function remains intact, and researchers can analyse differences caused only by the inhibited enzyme.

Another group of mouse models involves using germ-free mice. These mice are grown in a sterile enclosure in order to prevent them from being exposed to any and all bacteria (Grover & Kashyap, 2014). This avoids the step of treating mice with antibiotics, and simply allows comparison between mice that have bacteria in their bodies, and mice that do not. This model is important as it avoids potential drug-drug interactions between antibiotics and the drug being studied. It also avoids the risk of antibiotic-resistant strains of bacteria resettling the gastrointestinal tract for antibiotic treated mice, making them no longer bacteria-free. However, there have been criticisms of using germ-free mice since there early life without microbes have been shown to alter development (Williams, 2014). These mice often have altered immune systems, organ function, and even reproductive capability and therefore differences observed between germ-free and control mice might not be due to treatment, but instead due to these underlying developmental differences (Williams, 2014).

The final group of mouse models used to study the microbiome is gnotobiotic models. These are mouse models in which only certain known microbial species are present (Williams, 2014). Technically speaking, germ-free mice are considered gnotobiotic mice as their microbial community is also 'known,' being void of species. This group also consist of mouse models in which known bacterial species from the human gut microbiome are introduced into the mice (Goodman et al., 2011). This allows the study of functional properties of specific microbes in a living system, while to some extent also controlling for the physiological complications present in germ-free mice.

1.6.2 SEQUENCING

Investigations into the microbiome began as early as the 1960's (Scheline, 1968). However, sequencing technology at this time was slow, expensive, and often focused on one cultured species at a time (Morgan & Huttenhower, 2012; Soleim & Scheline, 1972). There are major issues with this analysis, as it is estimated that upwards of 99% of bacteria cannot be grown in lab cultures (Pham & Kim, 2012). In the past decade, the development of 'next-generation' high-throughput sequencing techniques have revolutionized the field of microbial sequencing (Reis-Filho, 2009). Analogues to human genomic studies for personalized medicine, the microbiome can be sequenced in order to identify critical differences between individuals.

Previously, the most common technique used was Sanger sequencing, which used DNA chain-terminating inhibitors (Sanger, Nicklen, & Coulson, 1977). Although this was the most effective method at the time, it was clearly outclassed by next-generation sequencing methods. One of the earliest forms of next-generation sequencing was 454 pyrosequencing (Ronaghi, Uhlén, & Nyrén, 1998). From there, there was rapid progression in the development of these sequencing techniques, with pyrosequencing being overtaken by more accurate, less costly methods such as Illumina sequencing (Rodrigue et al., 2010). These advanced sequencing techniques enable analysis of the vast amounts of DNA in microbiome samples at an incredibly fast rate.

Sequencers such as the ones produced by Illumina use a process called sequencing by synthesis. This method starts with a flow cell containing adaptors (specific nucleotide sequences) bound to the cell. A single-stranded DNA fragment being sequenced is run through the flow cell, with adaptor regions of the fragment binding to complementing adaptors on the flow cell. A complementary strand of DNA is created from the DNA fragment and is therefore bound to the flow cell. This double-strand is then denatured so the original fragment can be washed away, leaving one single-strand of DNA bound to the flow cell. This strand is then amplified via 'bridge-amplification' where the strand folds over, and its end connects to a different adaptor on the flow cell. Then the complimentary strand is generated, and the bridge is denatured, resulting in two

single-stands of DNA attached to the cell. This process of folding over, is repeated many times creating a cluster of the DNA fragment. The reverse strands are removed leaving only forward strands, all identical to the initial DNA fragment. At this point, the 'sequencing by synthesis' begins. A primer is attached to the single strand, and fluorescently tagged nucleotides are used to create the complementary strand of DNA. After the correct nucleotide is added, the clusters are excited by light, and a fluorescent signal is produced, with each nucleotide having its own unique signal. Hence, the sequencer can read the light source coming from the cluster of identical strands to determine what nucleotide is added. This process continues until the desired length of the forward read (ie: the number of base pairs) is reached. The cluster then folds over, and the reverse strand is synthesised. The strands are separated, the forward strands are washed away, and the reverse strand is then sequenced by synthesis, in identical fashion to the forward strand sequencing. This provides both forward and reverse reads for the fragment at the desired read length (Illumina Inc., 2014).

Depending on the size of the DNA fragments and the read length, there may be overlap between the forward and reverse reads. These reads can be aligned so their paired-ends overlap resulting in a longer fragment that can provide more information for identifying matches in reference databases.

This process can be performed on vast numbers of DNA fragments simultaneously, allowing rapid sequencing of all DNA fragments obtained from a sample. Additionally, 'barcodes' or small sequences of nucleotides can be attached to DNA from a sample. With these barcodes, many samples can be pooled together, sequenced simultaneously, and sorted afterwards in a process called 'multiplexing'. This further increases the efficiency of DNA sequencing.

When sequencing microbiome samples, researchers are generally trying to determine either what microbes are there, or what those microbes are doing. For the former, sequencing for determining taxonomic composition is often done via 16S sequencing. The 16S gene encodes for a distinct part of the bacterial ribosome, 16S rRNA. This gene is useful for taxonomy as it contains both conserved and variable regions. As microbes become more distinct from each other, there will be more changes

in the variable regions. Based on the similarity of these variable regions, microbes can be grouped in to operational taxonomic units (OTUs). Hence, it can be determine how diverse the microbiome is based on the number and proportion of OTUs detected (Bartram, Lynch, Stearns, Moreno-Hagelsieb, & Neufeld, 2011).

To examine function, metagenomic sequencing can be performed. Unlike 16S sequencing which only examines one particular region of microbial DNA, metagenomics looks at all the DNA in a sample regardless of who or where it comes from. After sequencing, DNA reads can be matched to a reference database, in order to determine what genes are present in the microbiome sample. These genes then give us an idea of the functional capabilities of the microbes present.

1.6.3 REFERENCE DATABASES

References databases are an important tool in metagenomic analysis. From metagenomic sequencing millions of short reads of DNA can be obtained. However, this is essentially just a list of letters, and alone tells us nothing about what the DNA sequences actually do. References databases are essential in bridging the gap between raw sequences and biological analysis.

Many different reference databases exist, including eggNOG, COG, TIGRFAM, and KEGG (Haft, Selengut, & White, 2003; Huerta-Cepas et al., 2016; Kanehisa, Goto, Kawashima, Okuno, & Hattori, 2004; Tatusov, Galperin, Natale, & Koonin, 2000). Each can be used to functionally annotate metagenomic data, but each of these databases may have differences in the way they operate. To further explain how these databases work, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database will be used as an example.

Within KEGG, there are several different categories which metagenomic DNA can be related to, including information about pathways, reactions, and full genomes. One particularly useful section of KEGG is KEGG Orthology. Within this section, molecular-level functions are classified as KEGG Orthologs, or KOs. One, or several genes that are

thought to carry out a specific function may be categorized under a particular KO. That function is given a KO identifier, which is the letter K followed by five numbers. KEGG also provides information regarding what species a KO has been experimentally identified in.

Depending on the nature of the function, the genes, and the specific targets or pathways, multiple KO identifiers may be linked for one particular function. For example, some genes that have the general function 'alcohol dehydrogenase' are grouped together under the KO identifier of K00001. However, other genes with the 'alcohol dehydrogenase' function that act in different ways are grouped together in K00002, and others still can be grouped into K00055 and K00114. All of these KOs represent the function of alcohol dehydrogenase, but the gene(s) placed under each KO will depend on more specific information.

1.6.4 BIOINFORMATICS

With the advancement of high-throughput sequencing techniques came massive datasets, orders of magnitude larger than what biological researchers were analysing previously. In addition to these sequencing tools, bioinformatic tools have been developed to process these large amounts of data in an efficient and accurate manner.

Within metagenomic datasets such as the HMP dataset, there is an enormous amount of information and nearly limitless ways to manipulate and examine the data in order to identify biological information. For efficient examination, data must be available for multiple research groups to obtain and study. Servers such as MG-RAST are crucial in this process, as they contain thousands of publicly available metagenomic datasets with accompanying metadata files (F. Meyer et al., 2008).

Many pipelines have been developed to analyze bioinformatic data such as QIIME and Microbiome Helper (Caporaso et al., 2010; Comeau, Langille, & Douglas, 2016). These pipelines allow users to work through standard operating procedures (SOPs) in an operating system such as Linux. By running particular scripts, a user is able to

converts files of raw sequencing data into desired biological information by, for example, matching sequences to reference database in order to determine taxonomy or function.

Throughout these pipelines, specific bioinformatic programs are used to manipulate the dataset in a particular way. When trying to determine function from raw sequences, a tool known as DIAMOND can be used. Similar to BLAST, DIAMOND takes files of raw sequences (in particular, Illumina reads 100-150bp in length), and aligns them to a protein reference database (Buchfink, Xie, & Huson, 2014). The reason DIAMOND is used for metagenomic sequences instead of BLAST is because it is able to run up to 20 000 times faster while maintaining high accuracy of matching. Taking this one step further, a program called HUMAnN can be used in order to take the output from DIAMOND and match it to the KEGG Orthology database, thereby identifying the presence and abundance of functions found in the raw sequences from microbiome samples (Abubucker et al., 2012).

Another useful tool for functional annotation is PICRUSt. In many studies, 16S sequencing is performed to identify taxonomy in a sample, but this type of sequencing provides no direct information on function. PICRUSt is able to take the 16S information regarding what OTUs are present and, using a database of reference genomes, predict what functions are present based on the OTUs (M. G. I. Langille et al., 2013) without having to do metagenomic sequencing.

Once metagenomic data has been generated, visualization of the data allows for quick comparisons and analysis. Looking at the entire dataset as a whole, there are two commonly used methods to generate plots that show relationships within microbiome data. The first method is a principle component analysis (PCA) plot. The function of a PCA plot is to take complex data such as the abundance of many OTUs or KOs across many samples, and transform it into a new coordinate system. In this system, there are a series of principle components (PCs) which are used as axes in the coordinate system. Each PC is generated by manipulating the data down to a single point along an axis. These PCs are uncorrelated, and their function is to summarize the variation present in the dataset. The first PC is the one that explains the most amount of variation within the dataset, while the second PC explains the second most amount of variation, and so on.

With this transformation, the complex associations within the data can be plotted on a 2D axis showing (usually) PC1 vs PC2, or a 3D axis with the first three PCs. Thus, samples are plotted based on the variation between them. Samples clustered together indicate low levels of variation between the measured factor (OTU or KO abundance) in those samples. Samples far apart on the PCA plot indicate there is much more variation between those samples based on the measured factor.

To generate several visualizations including PCA plots, a program called STAMP can be used (Parks, Tyson, Hugenholtz, & Beiko, 2014). STAMP is able to accept a program file like the output from HUMAnN. This file is essentially a massive table, listing all samples sequenced as well as a particular factor of the dataset. When analyzing the microbiome, this factor could be the OTUs found in each sample, or the KOs found in each sample. This file is used by STAMP to generate a PCA plot showing the relationship and clustering of the samples based on the factor being analyzed. STAMP can then be provided with a metadata file containing information about samples such as what location they came from or if the sample came from a female or male. With this, STAMP is able to identify samples based on the categories provided in order to observe potential trends and associations. STAMP is also able to analyze each 'variable' individually. For example, STAMP can generate box plots showing the abundance of one particular KO across all samples, separating them based on what category each sample came from.

Statistical evaluation can also be performed in STAMP. If only two categories are present, a Welch's two-sided t-test can be performed to see if the mean value of the categories are equal or if they are statistically significantly different ($p<0.05$). For multiple categories, ANOVA is used to see if the mean value of multiple groups are equal. With such a large volume of data, there is a high chance that data will be found statistically significant just due to chance. To reduce the amount of these 'false discoveries', STAMP is able to calculate a modified p value using Benjamini-Hochberg false discovery rate (FDR) multiple test correction (MTC). This allows more confidence in significance values identified when using large datasets such as metagenomic sample sequences.

The second form of visualization is principle coordinate analysis (PCoA) plot. In general, a PCoA plot is very similar to a PCA plot. It takes complex data and tries to manipulate it so that large amounts variation between samples can be explained on a few axis. The major difference with PCoA is that axis are not calculated based on the raw data of OTU or KO abundance in each sample. Instead, a distance matrix is generated, and that is used to produce the PC axis. One way of generating this matrix is an ecological method called Bray-Curtis. This method was initially developed for ecological data, looking at the (dis)similarity of species between different sites based on the number of counts. However, it functions just the same for metagenomic data by calculating a distance between samples based on the relative abundance of OTUs or KOs. This Bray-Curtis matrix can then be used to generate PCs via PCoA. When samples are plotted along these PCs, points closer together are more similar in terms of OTU or KO abundance. Samples that are farther apart means there is more variation between those samples, just like the visualization of PCA plots.

For visualizing PCoA plots in 3D space, a program called Emperor can be used (Vázquez-Baeza, Pirrung, Gonzalez, & Knight, 2013). This program was initially designed to examine microbial communities in the environment to determine how samples varied by taxa. However, it functions the same way when the input is microbiome sample data based on the abundance of KOs. Emperor is especially useful when dealing with metadata containing multiple groups such as age. It is able to colour samples in a PCoA plot based on a gradient, so for example samples coloured from light to dark as they increase in age from young to old. This makes viewing the PCoA plot to identify clusters based on age much easier than if each individual age had its own unique, unrelated colour.

Statistical significance of the Bray-Curtis distance matrix can be calculated using a QIIME script called Adonis (Anderson, 2001). This method looks at the variation present in a distance matrix and determines how strong and significant a particular metadata variable is at determining said variation. Adonis is a non-parametric method, which means it does not assume the data (the abundance of OTUs or KOs for instance) fits a particular distribution. Additionally it can be used with both discrete and continuous

variables, making it useful when looking at metadata such as age across samples. Adonis will output an r^2 value indicating how well the variable correlates with the distribution of samples in the Bray-Curtis distance matrix, as well as a p value indicating if this correlation is statistically significant.

1.7 OBJECTIVES

This project's main objective was to explore the drug metabolism potential of the human microbiome. Although the case studies in Chapter 1.5 are informative, they only give specific targeted findings and do not provide information about the microbial drug metabolizing gene presence, prevalence, variability, and associations with host factors within the human microbiome. The first objective was to catalog functions that could be present in the human microbiome which are related to drug metabolism. The second objective was to determine the presence and abundance of these drug metabolism functions in the human microbiome including variance between body site and within the population, as well as identify how the abundance of particular functions may shift from person to person. The third objective was to determine if factors such as sex, age, frailty, and the use of certain medication could be correlated with trends in microbial drug metabolism functions.

This was an exploratory study, so we did not anticipate any of our results having immediate clinical significance. Instead, we hoped to identify many different microbial functions that display trends in abundance based on the factors we examined. This survey could then serve as a starting point for future studies by highlighting the most promising trends related to drug metabolism functions in the microbiome.

We hypothesized that there would be many functions identified in the human microbiome that could play a role in drug metabolism. We also hypothesized that the abundance of a subset of these functions would be correlated with the explored factors (body site, sex, age, frailty, and medication use), however due to this being an

exploratory study, we could not say for sure what particular functions would be affected, or in what direction their abundance would shift.

CHAPTER 2: MATERIALS AND METHODS

2.1 MANUAL SELECTION OF POSSIBLE DRUG METABOLIZING KEGG ORTHOLOGS

The investigation of the drug metabolism potential of the human microbiome was conducted with the use of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2004). The area of KEGG utilized was the KEGG Orthology section (accessed in October-November of 2014 at http://www.genome.jp/kegg/ko.html). Particular functions are organized into different KEGG Orthologs (KOs). Sequences of DNA can be matched to particular genes, which in turn, would be placed under a KO category.

The functions, or KOs, known to be present in microbes that potentially played a role in drug metabolism were identified. A list of general drug metabolism categories in humans was established (Table 1). These consisted of phase I and phase II enzymes, as well as additional phase III functions associated with modifying drugs.

Table 1 List of drug metabolism function categories.

Drug metabolism categories	
Cytochrome P450 monooxygenase	Flavin-containing monooxygenase
Alcohol dehydrogenase	Aldehyde dehydrogenase
Monoamine oxidase	Cytochrome P450 reductase
Esterase	Amidase
Epoxide hydrolase	Methyltransferase
Sulfotransferase	N-acetyltransferase
UDP-glucuronosyltransferase	Glutathione S-transferase
Acetyl CoA	Azoreductase
Beta-glucosidase	Beta-glucuronidase

One by one, each of the categories in Table 1 was searched for in the KEGG database. This generated a list of several KOs associated with that category. For every KO found in the category, the taxonomy this KO was biologically found in was examined. If there was evidence of that KO being found in bacteria, the KO was recorded. This process was repeated for each category. The end result was a list of 225 KOs known to be found in microbes which could potentially play a role in drug metabolism based on their human homologues. This collection of KOs will be referred to as microbial drug metabolizing genes (MDMGs).

2.2 DATASET AND SAMPLE INFORMATION

In this study three separate datasets were used for different areas of analysis. This section outlines information about the samples in each dataset.

2.2.1 HUMAN MICROBIOME PROJECT DATASET

The Human Microbiome Project (HMP) collected non-invasive microbiome samples from 242 individuals at one or two time points from up to 18 different body site locations (The Human Microbiome Project Consortium, 2012). These individuals ranged in age from 18-40 years old (mean age 27. They were clinically screened for absence of disease. In total, 749 samples were sequenced using 101bp paired-end Illumina shotgun metagenomic reads The HMP did extensive quality controls on these samples, resulting in a final set of 531 metagenomic samples. The 18 body sites were sorted into 5 general body sites: airways (62 samples), gastrointestinal tract (118 stool samples), oral (289 samples), skin (20 samples), and urogenital tract (42 samples). In this collection, 224 samples came from females, and 307 samples came from males.

The HMP set up inclusion criteria in order to be eligible to take part in the study. The individuals could be male or female subjects at least 18 years old, but not more than

40 years old at the time of enrollment. They must be able to provide informed consent. They must be healthy subjects, and must be willing and able to provide blood, oral cavity, skin, nasal cavity and stool specimens. Female subjects must be willing to provide vaginal specimens, and must also have a regular menstrual cycle (between 21-35 days) or for subjects on hormonal contraception, have a history of regular menstrual cycles before initiating contraception. Important exclusion criteria included antibiotic use within the last 6 months and females who were pregnant or on combination hormone vaginal ring contraception.

The metagenomic DNA sequences from the 531 HMP samples were made publically available and downloaded via the HMP website (http://hmpdacc.org/).

2.2.2 CHINESE DATASET

The second publicly available dataset used in this study was a collection of 368 stool samples from Chinese individuals, from here on out referred to as the Chinese dataset (Qin et al., 2012). These individuals ranged in age from 13 to 86 years old. For these samples, 157 came from female individuals, and 211 came from male individuals. Illumina GAIIx and HiSeq 2000 paired-end metagenomic sequencing was performed, with 75bp-90bp read length. These individuals had not received any antibiotic treatment within two months before sample collection. All individuals signed informed consent forms.

The metagenomic DNA sequences from the 368 stool samples were made publicly available and downloaded from MG-RAST.

2.2.3 NORTHWOOD DATASET

The third dataset used in this study came from researchers in collaboration with Dr. Langille who were studying the microbiome of individuals at Northwood, an assisted

care facility located in Halifax, Nova Scotia. This study looked at a total of 47 individuals (30 females, 17 males) and collected 216 stool samples to investigate the gut microbiome. Eleven samples were discarded after quality control, leaving 205 samples available for analyses.

These individuals ranged in age from 65 to 99 years old. A frailty index (FI) score was calculated for each of these individuals, based off clinical assessment of 70 physical, mental, and behavioral criteria (Rockwood, 2005). Scores were reported on a scale of 0 to 1, with 0 representing no signs of frailty, and 1 represent maximum frailty in every assessed aspect. The individuals sampled had FI scores ranging from 0.305 to 0.758.

The list of all medication and dosing information for each individual was recorded. In total, the 47 patients were on 291 different drug treatments. Each patient was on said treatment for at least one month in order to be included. No patients were on antibiotics in the month prior to the study. The clinician in the Northwood study noted 20 specific drug treatments that were thought to alter the microbiome (referred to as 'microbiome-altering drugs'). In total, 159 samples came from individuals on at least one microbiome-altering drug, and 46 samples came from individuals not on any microbiome-altering drugs.

Sequencing was done on the 205 samples using Illumina MiSeq paired end 300bp with 16S V6-V8 region primers through the Integrated Microbiome Resource (Comeau et al., 2016).

2.3 METAGENOMIC ANNOTATION

For the 531 Human Microbiome Project samples and the 368 Chinese samples, metagenomic functional annotation was performed in order to match DNA sequences to particular MDMGs. Within the Chinese data, the number of reads per sample were approximately 10 times the size of the HMP dataset average reads per sample. In order to process these sequences in a timely manner, the Chinese dataset was randomly subsampled to reduce it down to 10% its original size.

The metagenomic standard operating procedure (SOP) found on the Microbiome Helper GitHub (https://github.com/mlangill/microbiome_helper) was utilized for functional annotation. The datasets used paired-end Illumina sequencing, which means for a given fragment of DNA, both ends get sequenced. By combining these paired-ends together the quality of that sequenced fragment is improved, which makes it easier to align the fragment to a reference database. PEAR version 0.9.6 was used in order to stitch these paired-end reads together (Zhang, Kobert, Flouri, & Stamatakis, 2014). All samples are coming from human hosts, but only the sequences of microbial origin are of interest. As a result, all human DNA sequences must be screened out. A tool called Bowtie2 was used to align reads to reference sequence (Langmead, Trapnell, Pop, & Salzberg, 2009). In particular, Bowtie2 checks if any reads align to the human genome, and if so, those reads were discarded. The microbial sequences were then mapped to a protein reference database in order to identify what functions were present. For this a tool called DIAMOND was used, which is a high-throughput search tool that aligns short reads to a protein database (Buchfink et al., 2014). Finally, the output from DIAMOND was then run on HUMAnN version 0.99. HUMAnN takes the protein alignment information, and annotates them with the KEGG Orthology (KO) database (Abubucker et al., 2012). The end result is a file consisting of the relative abundance all annotated KOs in all samples. The relative abundance of our MDMGs could then be extracted from this file. The names and IDs of all annotated MDMGs can be found in Table 4 in the appendix.

2.4 16S ANNOTATION

Unlike the HMP and Chinese dataset, which contained metagenomic sequence data, the Northwood dataset initially performed 16S sequencing on the samples to generate an OTU table. From there, followed the 16S microbiome SOP found on the Microbiome Helper GitHub to identify OTUs in the samples.

This gave the taxonomy of the microbes present in the sampled gut microbiomes. However, this investigation was only interested in function, not taxonomy. In order to assess the functions present in these microbiome samples without having metagenomic

sequences available, a tool called PICRUSt is used. PICRUSt takes 16S information and accurately predict functional composition of a microbiome based on the taxonomy present (M. G. I. Langille et al., 2013). The PICRUSt workflow is also documented on Microbiome Helper. Following the workflow, the OTU table was converted into the proper format for PICRUSt. The table was then normalized by 16S copy number. PICRUSt was then run in order to predict the KOs found in the samples. The KOs were normalized for each sample and the file was converted to STAMP format, resulting in a table of (predicted) relative abundance for every annotated KO in all 205 samples. This table was copied and edited down to only consist of the abundance of the MDMGs in each sample. The names and IDs of all annotated MDMGs can be found in Table 4 in the appendix.

2.5 STATISTICAL METHODS

The annotated tables were used in order to compare trends in the relative abundance of MDMGs across several different variables in the dataset. Statistical evaluations and data visualizations for body site, sex, and medication use were done in STAMP version 2.0.9 (Parks et al., 2014). STAMP requires a profile file as well as a metadata file. The output from HUMAnN was reformatted so it could be used as the profile file, containing the relative abundance of all KOs or just our MDMGs in every sample. Metadata files for the HMP dataset indicated body site and sex of each sample, the Chinese dataset indicated the sex of each sample, and the Northwood dataset indicated medication use in each sample.

The first aspect of the HMP dataset examined was body site. The profile file containing the relative abundance of all annotated KOs from all 531 samples as well as the metadata file listing the body site where each sample came from was loaded into STAMP. A PCA plot was generated for the comparison between body sites based off the abundance of all KOs across the samples. The analysis was then performed using only our MDMGs. The same metadata file was used, but the profile file contained the relative abundance of only the MDMGs in all 531 samples. A PCA plot was generated for the

comparison between body sites based on the abundance of MDMGs across the samples. STAMP also generated box plots for the individual MDMGs, showing the relative abundance in the five body site locations. STAMP calculated the mean relative abundance of each MDMG in all five body sites, as well as the standard deviation. A Welch's two-sided t-test was performed with Benjamini-Hochberg FDR MTC to calculate if the mean relative abundance of each MDMG in the gut was significantly different ($p<0.05$) from the mean relative abundance of that MDMG in all four other body sites with a cut-off set at 0.05 to reduce false-positive correlations.

Next, the differences between the microbiomes of females and males was analyzed, in both the HMP dataset and the Chinese dataset. The profile file containing our MDMGs in all 531 HMP samples and the metadata file containing the sex of the samples were loaded into STAMP. This generated comparisons based on all 5 general body sites. Additionally, sex differences just in the gut microbiome was examined. For the HMP dataset, a profile file containing our MDMGs but only the 118 samples coming from the GI tract was loaded into STAMP. For the Chinese dataset, the profile file containing our MDMGs in all 368 gut samples and the metadata file containing the sex of the samples were loaded into STAMP. PCA plots were generated for the comparison between females and males separately for each dataset, as well as box plots for the relative abundance of individual MDMGs between females and males. A Welch's two-sided t-test was performed with Benjamini-Hochberg FDR MTC to calculate if the mean relative abundance of each MDMG was significantly different ($p<0.05$) between females and males.

The final variable examined using STAMP was medication use in the Northwood dataset. A metadata file was created which listed the samples into two groups. The 159 samples from patients on a microbiome-altering drug were placed in one group, and the 46 samples from individuals not taking any of the 20 microbiome-altering drugs were placed in the second group. The profile file containing the relative abundance of all the predicted KOs in all 205 samples, as well as the metadata file separating samples into the two drug groups were loaded into STAMP. Separately, the profile file containing the relative abundance of our MDMGs in all 205 samples, as well as the metadata file were

loaded into STAMP. PCA plots were generated for the comparison between 'taking' vs 'not taking' a microbiome altering drug. Box plots were also created in STAMP to compare relative abundance of individual MDMGs between these two groups. For each KO, a Welch's two-sided t-test was performed with Benjamini-Hochberg FDR MTC to calculate if the mean relative abundance was significantly different ($p<0.05$) between the two groups.

Next statistical evaluation and data visualization was performed for trends in age in the Chinese dataset, and trends in age and frailty in the Northwood dataset. A Bray-Curtis distance matrix was generated based on the relative abundance of MDMGs between samples in each dataset using the Microbiome Helper 16S SOP. This matrix was used to generate a PCoA plot, visualized in Emperor. Samples were coloured in Emperor as a gradient based on age or frailty index (FI) score. Adonis was implemented through QIIME to determine how strongly age or frailty correlated with the overall variation in the abundance of MDMGs across samples (as indicated by the calculated r^2 value), and to determine if this correlation was statistically significant ($p<0.05$).

For individual MDMGs, R was used to examine correlations between abundance and age or frailty (R Development Core Team, 2013). Using the R statistics package, corr.test was run. This generated scatterplots of relative abundance vs age or FI score for each MDMG, as well as calculated the strength of the correlation (r^2 values) and significance (Benjamini-Hochberg FDR p-values).

It is often difficult to identify the most appropriate statistical test for analyzing these complex datasets. Since this was an exploratory study, the focus was primarily on hypothesis generation, as opposed to explicit hypothesis testing. Therefore, the selection of statistical tests was based on what was built into the programs and pipelines, as long as it provided some indication on how particular variables were correlated with MDMG abundance trends.

CHAPTER 3: MICROBIAL DRUG METABOLIZING GENES WITHIN THE HUMAN MICROBIOME PROJECT SAMPLES

3.1 MANUAL SURVEY OF MICROBIAL DRUG METABOLIZING GENES

A total of 225 microbial drug metabolizing genes (MDMGs), were found in the manual surveying of the KEGG database. Although KEGG has over 19 000 KOs in the database, HUMAnN was only able to annotate 13 328 KOs. This means that out of our 225 total MDMGs, 174 were annotated. The acronym MDMG will henceforth be used to represent only the annotated microbial drug metabolizing KOs. The full list of annotated MDMGs can be found in the appendix (Table 4). The number of annotated MDMGs belonging to each metabolism category is listed below (Table 2).

Table 2 Number of microbial drug metabolizing genes (MDMGs) annotated by HUMAnN in each metabolism category.

Category	Number of MDMGs	Category	Number of MDMGs
Cytochrome P450 monooxygenase	2	Flavin-containing monooxygenase	0
Alcohol dehydrogenase	15	Aldehyde dehydrogenase	1
Monoamine oxidase	1	Cytochrome P450 reductase	1
Esterase	0	Amidase	19
Epoxide hydrolase	2	Methyltransferase	111
Sulfotransferase	0	N-acetyltransferase	11
UDP-glucuronosyltransferase	1	Glutathione S-transferase	2
Acetyl CoA	0	Azoreductase	2
Beta-glucosidase	5	Beta-glucuronidase	1

3.2 PRESENCE AND VARIABILITY OF MICROBIAL DRUG METABOLZING GENES ACROSS HUMAN MICROBIOME PROJECT SAMPLES

To determine if any of the MDMGs were present within the human gut microbiome, 531 metagenomic samples from the Human Microbiome Project were screened (The Human Microbiome Project Consortium, 2012). Of the 174 MDMGs that were annotatable, the presence of 168 MDMGs were identified within one or more samples (Figure 1). These ranged in average relative abundance from K05349: beta-glucosidase (9.47% of all annotated MDMG reads) to K13317: NDP-4-keto-2,6-dideoxyhexose 3-C-methyltransferase (6.14×10^{-6} %). The variability of each MDMG across samples ranged tremendously from K13317 (0.0000478 SD/0.0000614 mean = 7.79 CV) to K00783: 23S rRNA (pseudouridine1915-N3)-methyltransferase (0.197 SD/2.35 mean= 0.0838 CV)

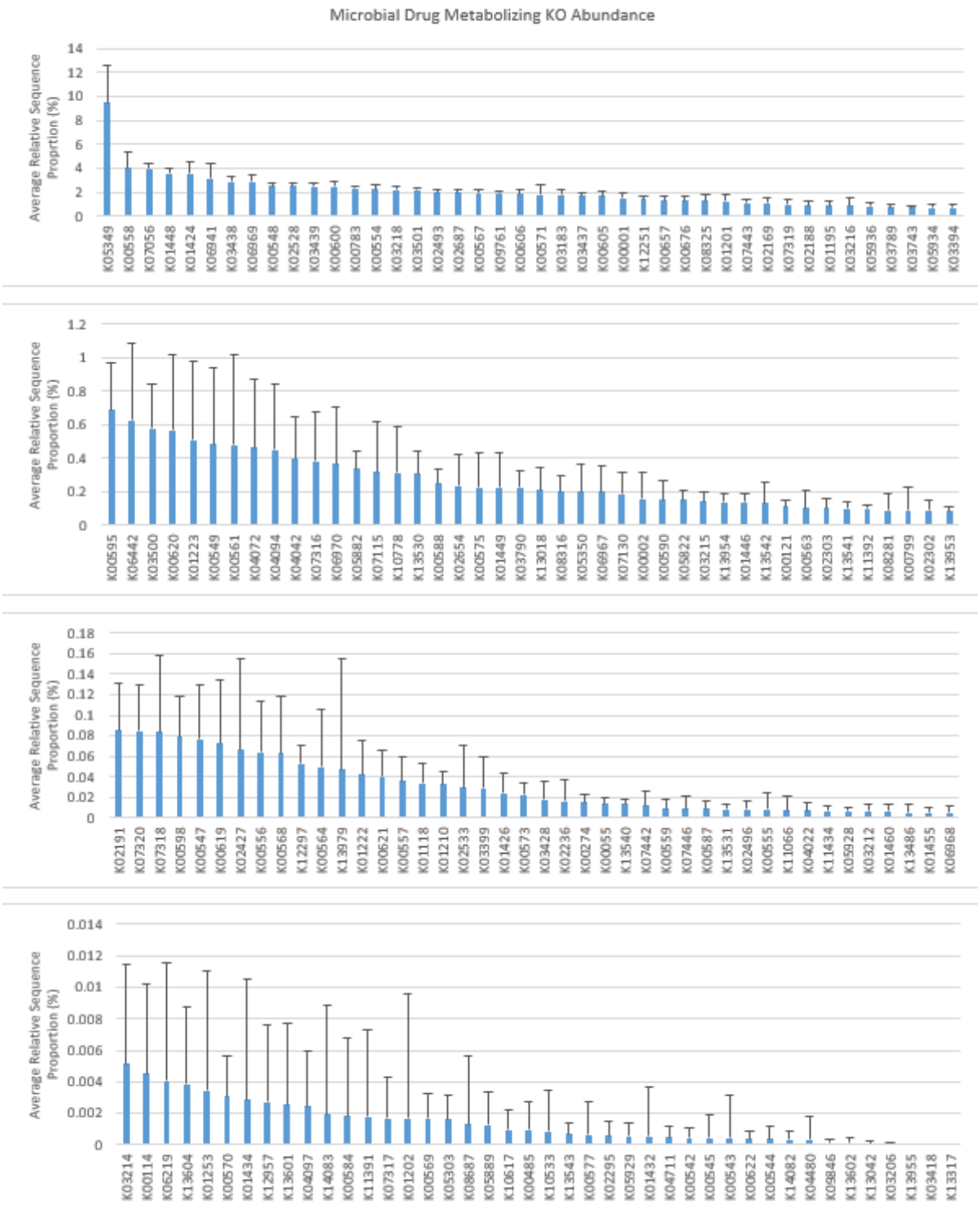

Figure 1 Average relative abundance of the 168 MDMGs found in the HMP microbiome gut samples. MDMGs are listed along the x-axis from highest average sequence proportion (K05349) to lowest average sequence proportion (K13317). Error bars represent one standard deviation.

3.3 VARIATION OF MICROBIAL DRUG METABOLIZING GENES ACROSS MULTIPLE BODY SITE MICROBIOMES

To examine if MDMGs existed within other human microbiomes besides the gut, MDMGs were screened for within other body sides sampled by the HMP. The HMP dataset contains samples from five distinct general body sites: airways (62 samples), GI tract (118 stool samples), oral (289 samples), skin (20 samples), and urogenital tract (42 samples, females only). As a baseline, body sites were compared using all 13 328 KOs. A principal component analysis showed general separation of all body sites with the exception of some overlap between skin and airways (Figure 2).

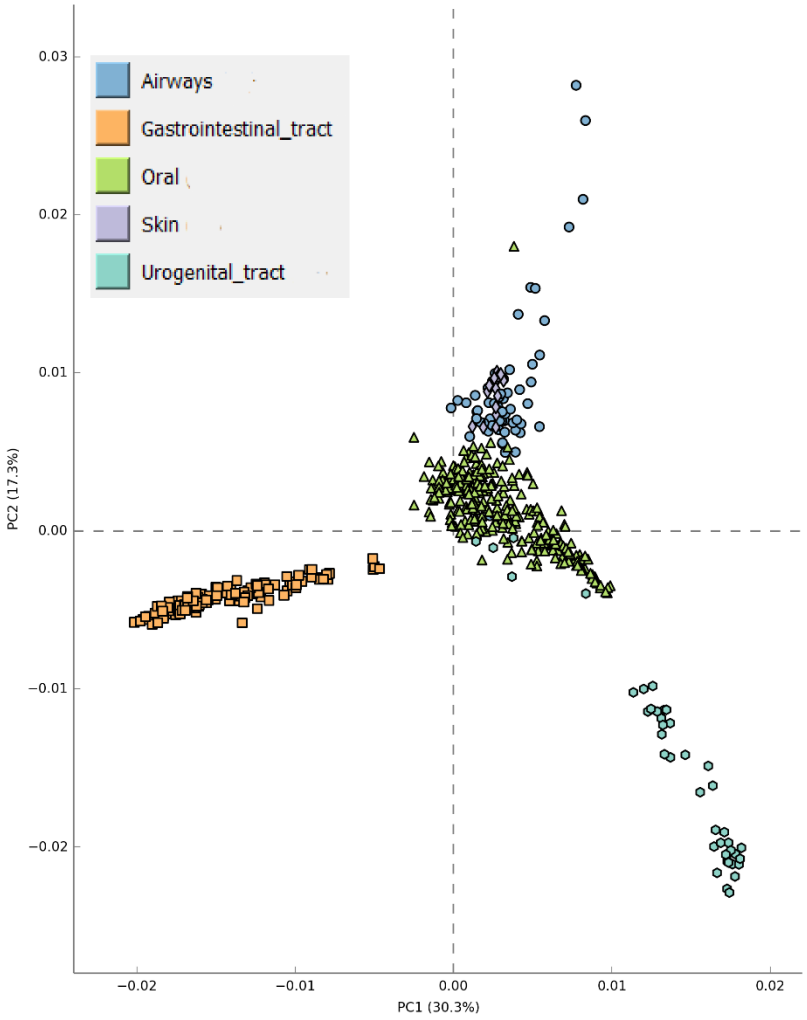

Figure 2 The gut microbiome is distinct from other body sites based on the abundance of all KOs. PCA plot based off the abundance of all 13 328 KOs showing all 531 HMP samples. Samples are coloured based on body site: airways (blue circles, 62 samples), GI tract (orange squares, 118 samples), oral (pale green triangles, 289 samples), skin (purple diamonds, 20 samples) and urogenital tract (teal circles, 42 samples).

Next body sites were compared using only the MDMGs. It was found that the GI tract was still very distinct from the other body sites being separated by PC1 representing 47.4% variation, while the other body sites were less distinct (Figure 3). The urogenital tract is the one exception, with samples from this body site being clearly distinguishable from the rest by PC2 representing 15.4% of the variation.

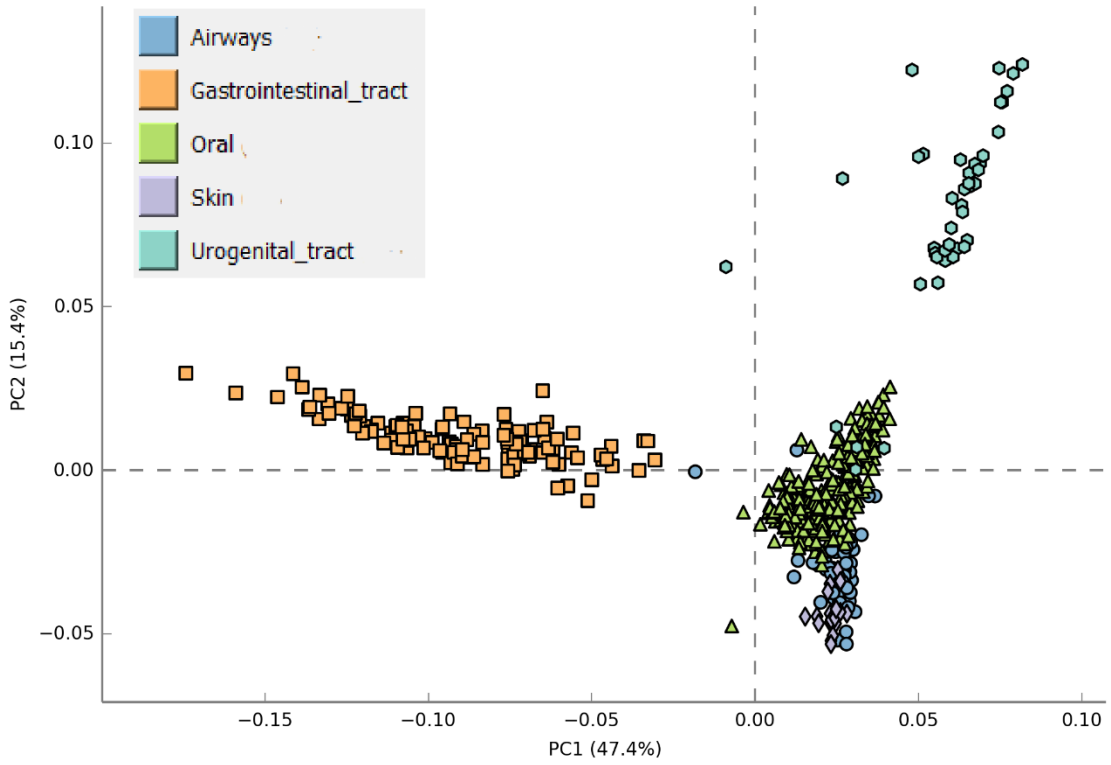

Figure 3 The gut microbiome is distinct from other body sites based on the abundance of MDMGs. PCA plot based off the abundance of MDMGs showing all 531 HMP samples. Samples are coloured based on body site: airways (blue circles, 62 samples), GI tract (orange squares, 118 samples), oral (pale green triangles, 289 samples), skin (purple diamonds, 20 samples) and urogenital tract (teal circles, 42 samples).

In addition to examining differences in body sites based off the collective set of MDMGs, the differences in individual MDMGs across body sites was investigated. Since the gastrointestinal tract was the primary focus, statistical comparisons were performed to see if the abundance of a MDMG in the GI tract microbiome was significantly different than the abundance in all four other body sites. Out of the 168 MDMGs identified in these samples, 148 showed significantly different ($p<0.05$) sequence proportion in the gut vs the other body sites after Benjamini-Hochberg FDR MTC. 54/148 of these MDMGs showed a higher average relative sequence proportion in the gut, and 94/148 had a lower average relative sequence proportion in the gut vs all other body sites (Table 4).

The most abundant and variable MDMG within the gut, K05349: beta-glucosidase, is also present in all other body sites but with significantly less relative abundance ($p < 1e\text{-}15$, ANOVA with Benjamini-Hochberg FDR) (Figure 4).

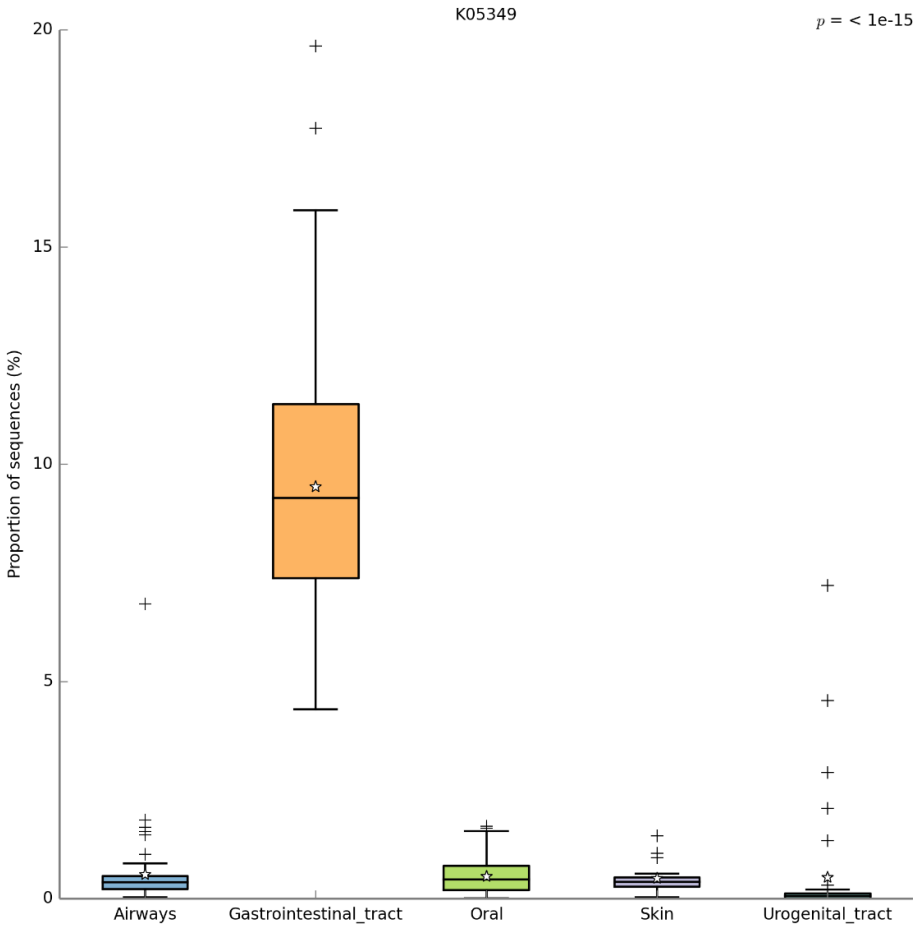

Figure 4 Average relative abundance of K05349: beta-glucosidase is significantly higher in the GI tract. Box plot showing relative abundance of K05349: beta-glucosidase amongst MDMGs (labeled as Proportion of sequences (%) on Y-axis). White stars in the box plots represent the average relative sequence proportion in each of the five body sites. Significance value of p < 1e-15 calculated using ANOVA with Benjamini-Hochberg FDR MTC.

K06442: 23S/16S rRNA-methyltransferase is an example of a MDMG that is found in lower relative abundance within the gut compared to other body sites ($p < 1e\text{-}15$, ANOVA with Benjamini-Hochberg FDR). Again this MDMG is present in all body sites ranging in relative abundance from approximately 0-4.5% (Figure 5).

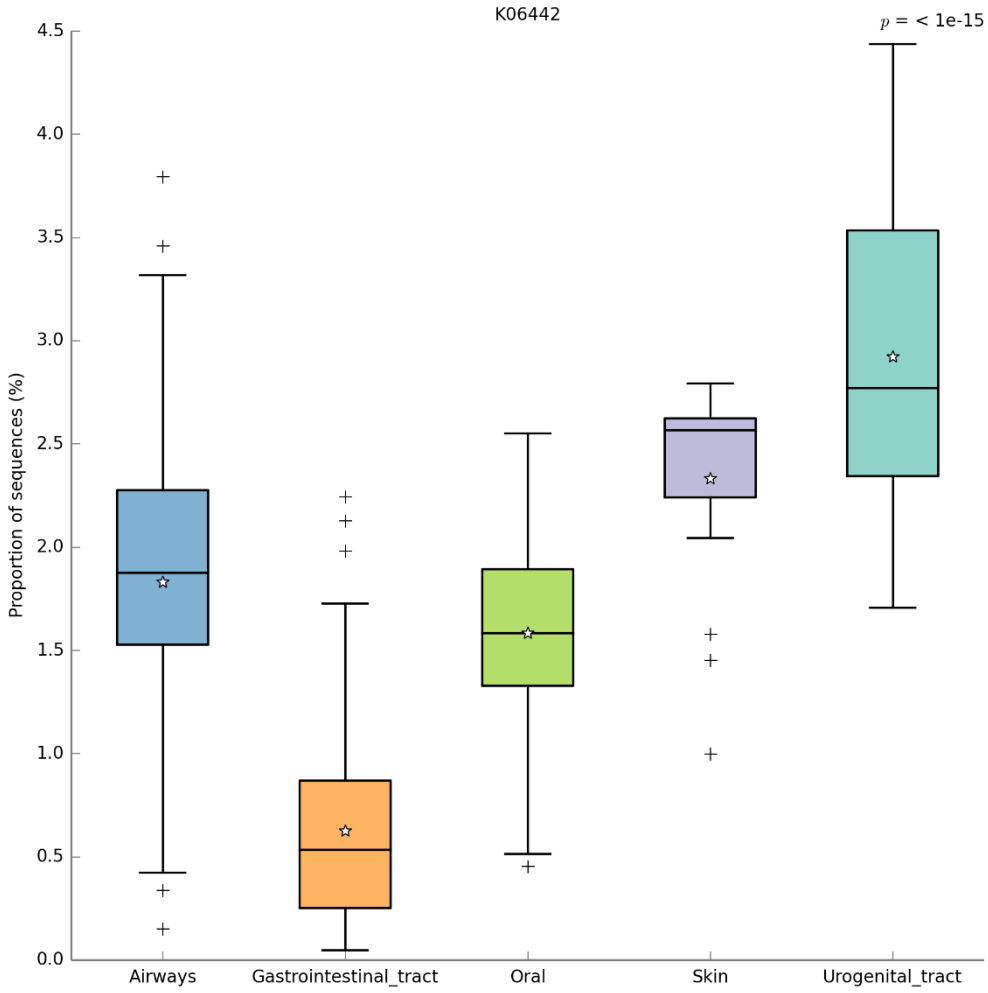

Figure 5 Average relative abundance of K06442: 23S/16S rRNA-methyltransferase is significantly lower in the GI tract. Box plot showing relative abundance of K06442: 23S/16S rRNA-methyltransferase amongst MDMGs (labeled as Proportion of sequences (%) on Y-axis). White stars in the box plots identify the average relative sequence proportion in each of the five body sites. Significance value of p < 1e-15 calculated using ANOVA with Benjamini-Hochberg FDR MTC.

Lastly, K01195: beta-glucuronidase represents the gene family that has been implicated in metabolism of CPT-11 (see Chapter 1.5.2.1). The mean relative abundance of K01195 varied significantly across body sites ((p < 1e-15, ANOVA with Benjamini-Hochberg FDR) with 0.3-1.8% in the gut and even higher levels on the skin, 0.1-2.7%.

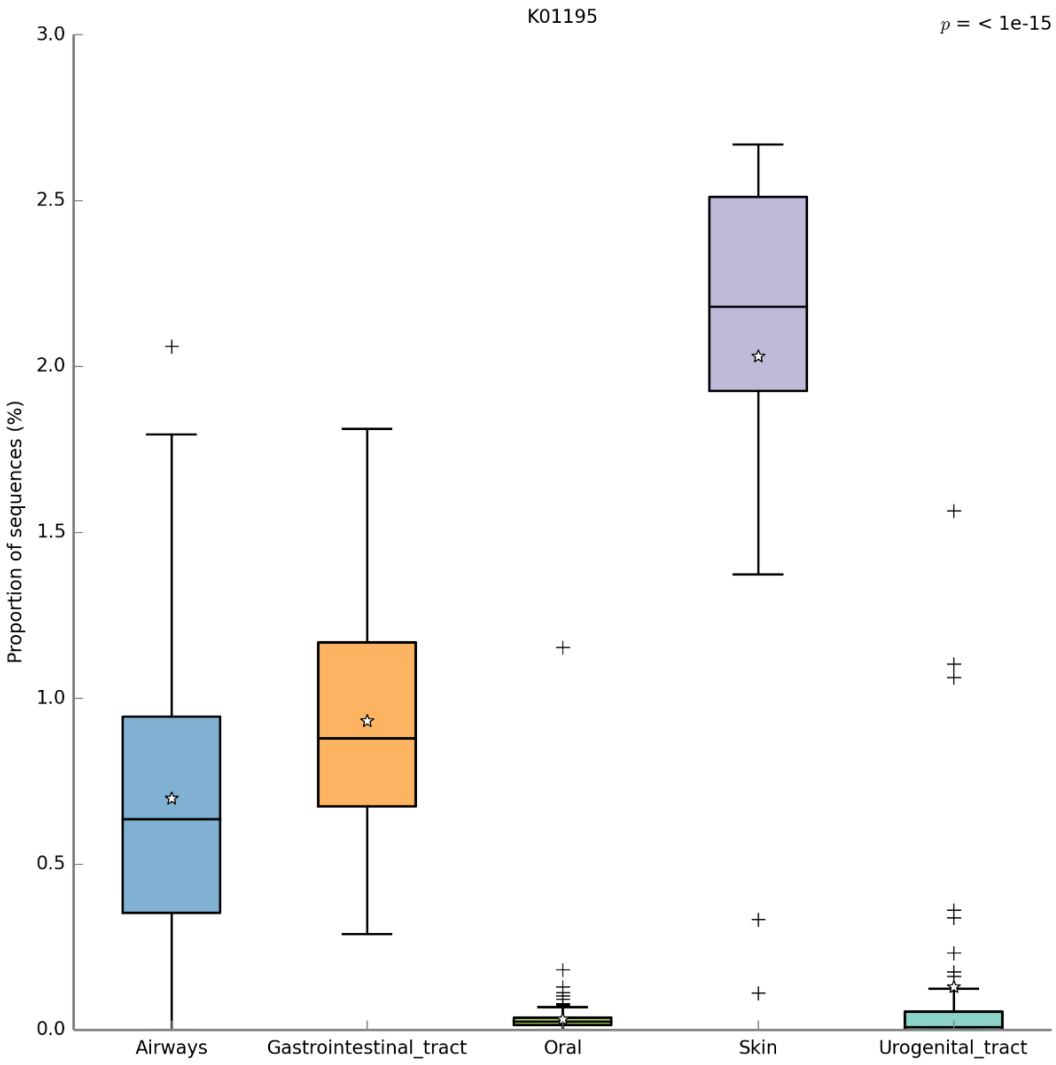

Figure 6 Average relative abundance of K01195: beta-glucuronidase is moderate in the GI tract. Box plot showing relative abundance of K01195: beta-glucuronidase amongst MDMGs (labeled as Proportion of sequences (%) on Y-axis). White stars in the box plots identify the average relative sequence proportion in each of the five body sites. Significance value of p < 1e-15 calculated using ANOVA with Benjamini-Hochberg FDR MTC.

3.4 MICROBIAL DRUG METABOLIZING GENES DO NOT VARY BY SEX IN THE HMP DATASET

The MDMGs within all body sites except those from urogenital (female only) were compared between males and females from which the samples originated. There

were no significant differences ($p<0.05$) between females and males in terms of MDMG abundance.

Additionally, sex differences within the 118 GI tract samples were tested on their own (Figure 7) which did not identify any distinct clustering between female and male samples. Looking at individual MDMGs, none showed a significant difference ($p<0.05$) between the two groups (using Welch's two-sided t-test, with Benjamini-Hochberg FDR MTC).

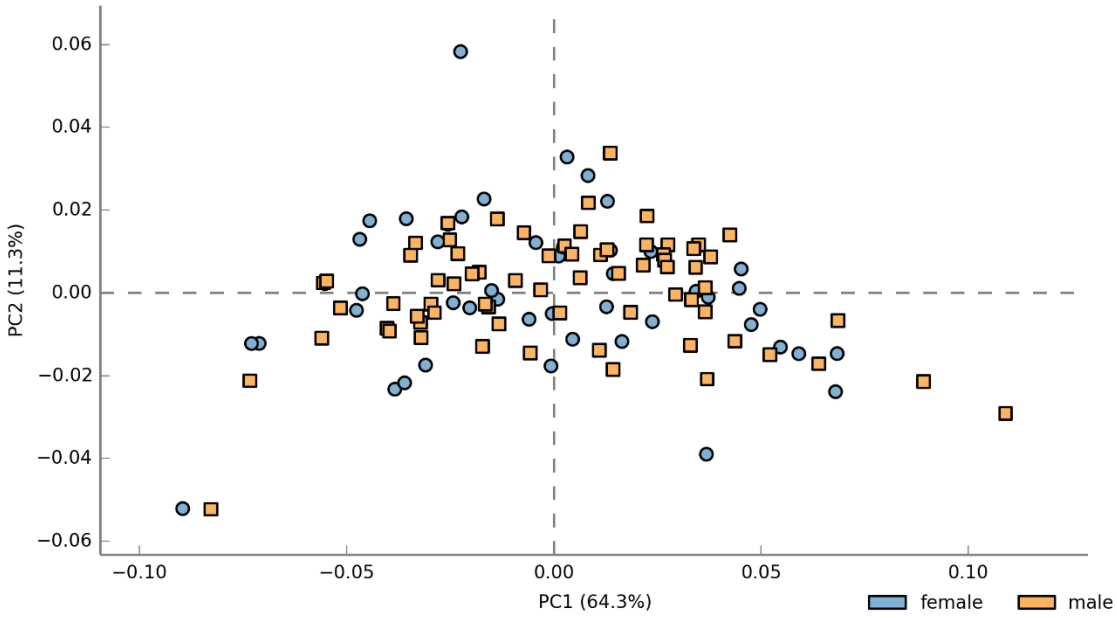

Figure 7 No differences in MDMGs between females and males in the HMP gut samples. PCA plot based on the abundance of 174 drug metabolism MDMGs. Only the 118 HMP samples coming from the GI tract ae included. Female samples (48) are shown as blue circles, male samples (70) are shown as orange squares.

3.5 SUMMARY

From the Human Microbiome Project dataset, we confirmed that nearly all of our microbial drug metabolizing genes (168/174) were found in the human microbiome. We also discovered there is a wide spread in terms of relative sequence abundance between MDMGs, and large differences from person to person for the relative abundance of an individual MDMG. The gastrointestinal tract was shown to be distinct not only across all KOs, but across the MDMGs as well. For individual MDMGs, the microbiome of the GI

tract was also shown to have higher average relative abundance for 54 MDMGs compared to the other four body sites, and a lower average relative abundance for 94. When comparing females and males, we found that no MDMGs were significantly different between these two groups in the gut microbiome.

CHAPTER 4: MICROBIAL DRUG METABOLIZING GENES WITHIN THE GUT MICROBIOME OF CHINESE INDIVIDUALS

4.1 PRESENCE AND VARIABILITY OF MICROBIAL DRUG METABOLZING GENES ACROSS CHINESE SAMPLES

Out of the 174 MDMGs annotated by HUMAnN, 163 of them were found in the microbiome stool samples from the Chinese dataset. The 11 not found were: K01432: Arylformamidase, K03183: ubiquinone/menaquinone biosynthesis methyltransferase, K05882: aryl-alcohol dehydrogenase, K13042: dimethylglycine N-methyltransferase, K13317: NDP-4-keto-2,6-dideoxyhexose 3-C-methyltransferase, K13602: bacteriochlorophyll C12 methyltransferase, K13604: bacteriochlorophyll C20 methyltransferase, K13955: zinc-binding alcohol dehydrogenase, K14369: erythromycin 3"-O-methyltransferase, K14372: cytochrome P450 hydroxylase, and K14374: avermectin B 5-O-methyltransferase. Interestingly, last 3 were also missing from the HMP samples. The remaining 3 MDMGs missing from the HMP samples (K14080, K14338, and K14568) were found in the Chinese samples.

The mean relative abundance and standard deviation was calculated for each of these 163 MDMGs (Figure 8). They ranged in average relative abundance from K05349: beta-glucosidase (7.13% of all annotated MDMG reads) to K05929: phosphoethanolamine N-methyltransferase (1.38×10^{-6} %). The variability of each MDMG across samples ranged tremendously from K09846: hydroxyneurosporene methyltransferase (0.0000562 SD/0.00000293 mean = 19.2 CV) to K03501: glucose inhibited division protein B (0.356 SD/2.56 mean= 0.139 CV).

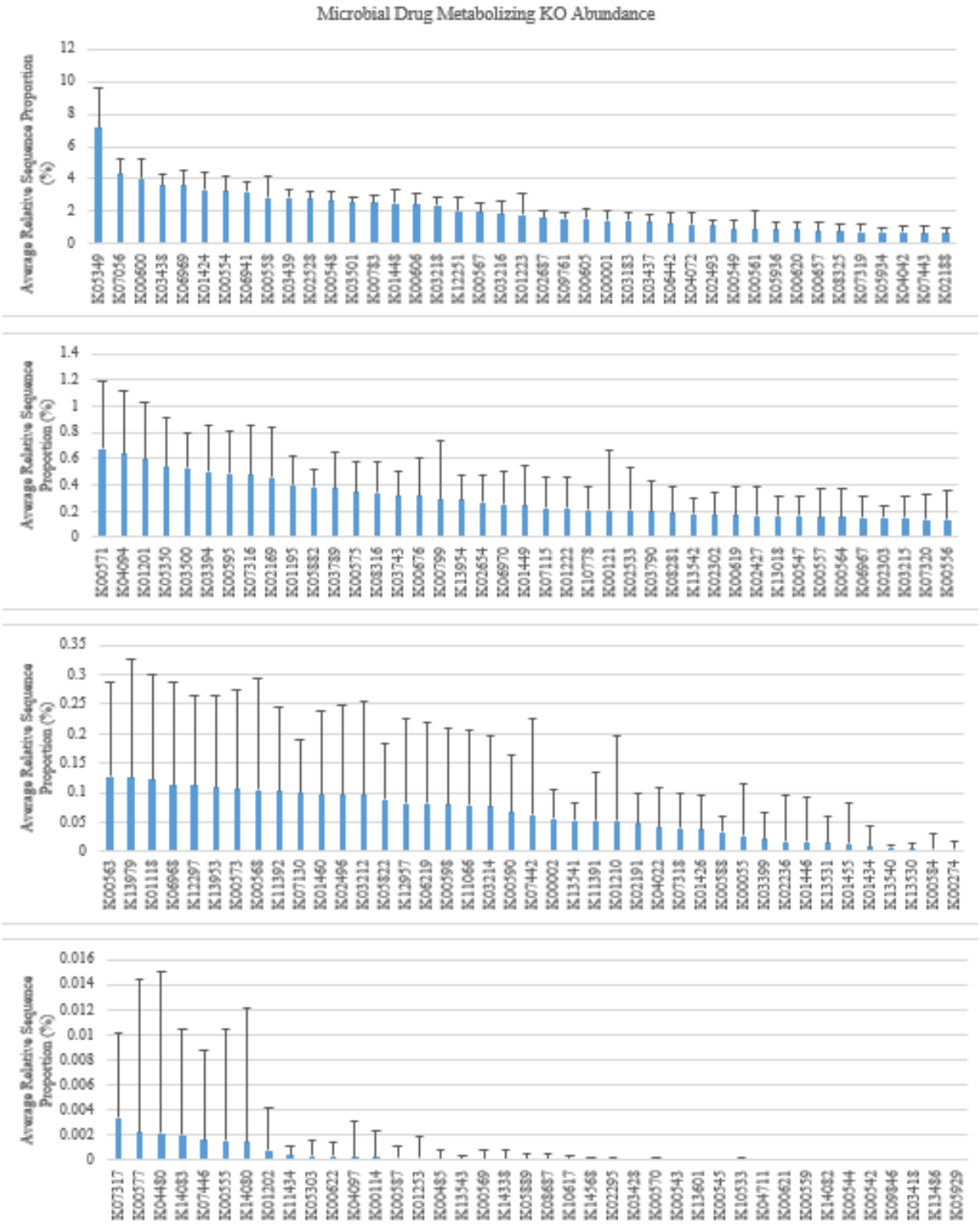

Figure 8 Average relative abundance of 163 MDMGs found from the Chinese microbiome gut samples. MDMGs are listed along the x-axis from highest average sequence proportion (K05349) to lowest average sequence proportion (K05929). Error bars represent one standard deviation.

4.2 MICROBIAL DRUG METABOLIZING GENES DO NOT VARY BY SEX IN THE CHINESE DATASET

This dataset collected 157 gut microbiome samples from females and 211 gut microbiome samples from males. Differences in MDMGs between the sexes were examined. There was no distinct clustering between female and male samples (Figure 9). Looking at individual MDMGs, none were found to have a significant difference in relative abundance ($p<0.05$) between the two groups (using Welch's two-sided t-test, with Benjamini-Hochberg FDR MTC).

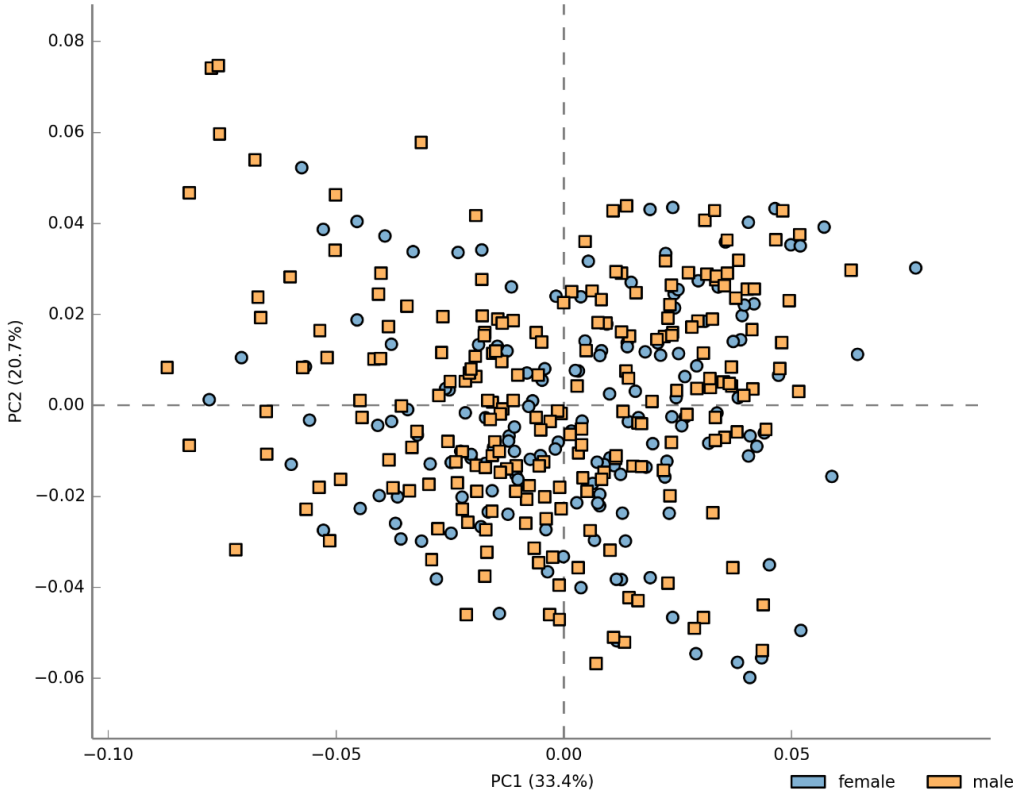

Figure 9 No differences in MDMGs between females and males in the Chinese samples. PCA plot showing 368 Chinese GI tract samples, based off MDMG abundance. Female samples (157) are shown as blue circles, male samples (211) are shown as orange squares.

4.3 AGE INFLUENCES THE ABUNDANCE OF MICROBIAL DRUG METABOLZING GENES IN THE CHINESE DATASET

The metadata for the Chinese dataset also included the age of each individual sampled and contained a large comprehensive range in age from 13 to 86 years old (Figure 10). Trends in MDMGs found in the microbiome related to the age of the individual were examined.

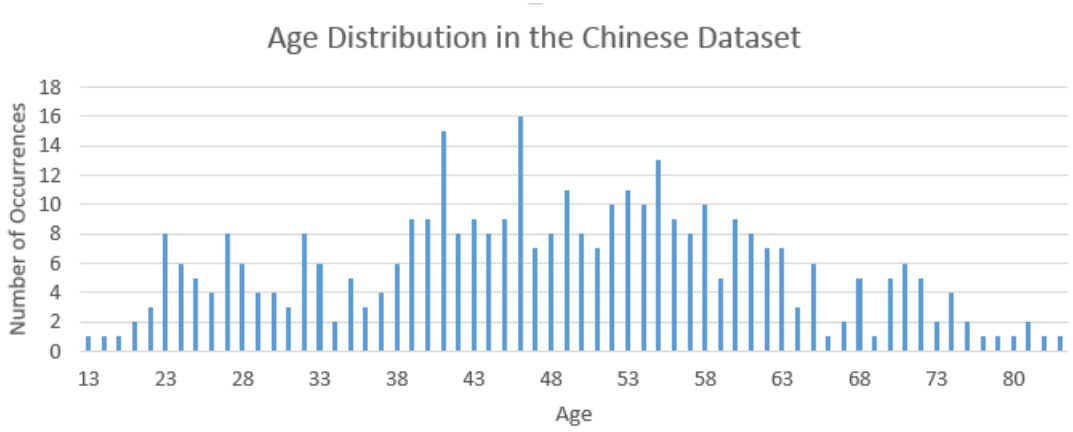

Figure 10 Age distribution of individuals sampled in the Chinese dataset. Individuals ranged from 13 to 86 years old. The number of individuals of each specific age are counted along the Y-axis.

First the relationship between the MDMGs and age was examined as a whole (Figure 11). Using the statistical method Adonis, the trend between age and MDMG abundance was calculated to have an r^2=0.018 and a p-value of p=0.001. This tells us that there is a statistically significant correlation between age and the abundance of MDMGs, but it is weak in terms of explaining all the variation within MDMGs across samples.

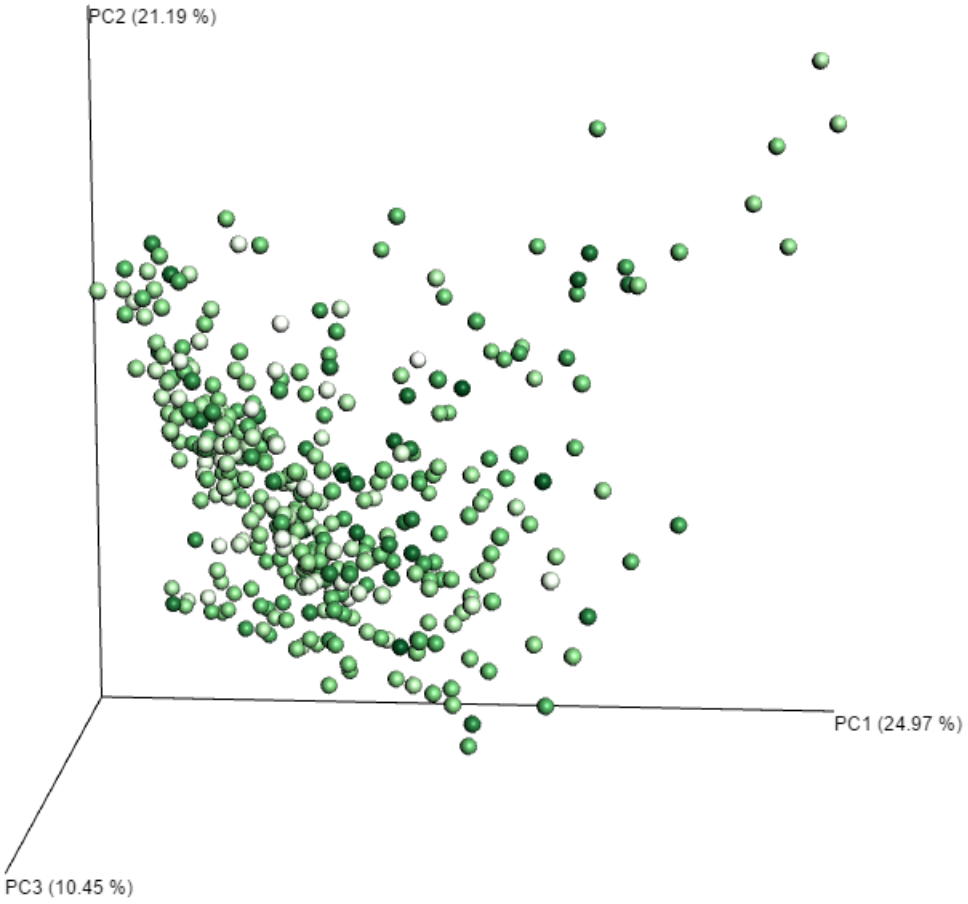

Figure 11 Age correlates weakly but significantly with variation of MDMGs in the Chinese samples. PCoA plot showing 368 Chinese GI tract samples, based off similarities in MDMGs. Samples are coloured from the youngest individual (13 y/o) as white, to the oldest individual (86 y/o) as the darkest green. Adonis calculations determine r^2=0.018 and p=0.001

The impact of age is more pronounced when examining MDMGs individually. From this, 36 MDMGs show statistically significant correlation in abundance with age. Specifically, 19 show a positive trend in relative abundance as age increases, and 17 show a downward trend. Table 4 in the appendix details all significant MDMGs. Figures 12 through 16 highlight these trends in some particular MDMGs of interest.

K03743: nicotinamide-nucleotide amidase had the strongest correlation with age (r^2=0.0611 and FDR p=0.0001) out of all MDMGs. Interestingly, many samples have similar relative abundances for this MDMG (just below $1x10^{-4}$). However, once

individuals hit around 40 years old, quite a few samples show dramatic increase in abundance (Figure 12).

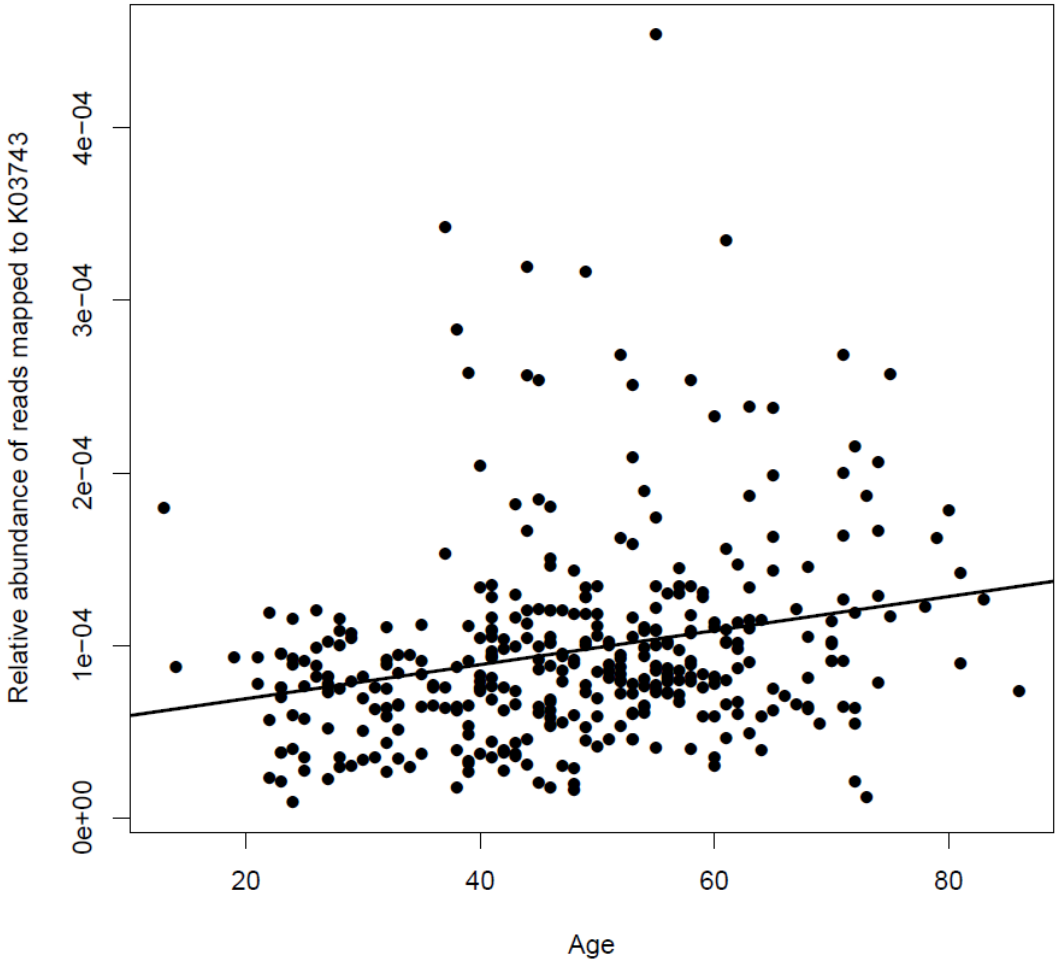

Figure 12 Relative abundance of K03743: nicotinamide-nucleotide amidase is positively correlated with age in the Chinese samples. Scatterplot showing the relative abundance of K03743 in all Chinese dataset samples. R^2=0.0611 and FDR p=0.001 calculated in R. Plot generated in R.

Following this, K01448: N-acetylmuramoyl-L-alanine amidase was examined (Figure 13). There is a great deal of variation for this MDMG from person to person (r^2=0.0475), but there is a statistically significant decrease in relative abundance as individuals get older (FDR p=0.0006).

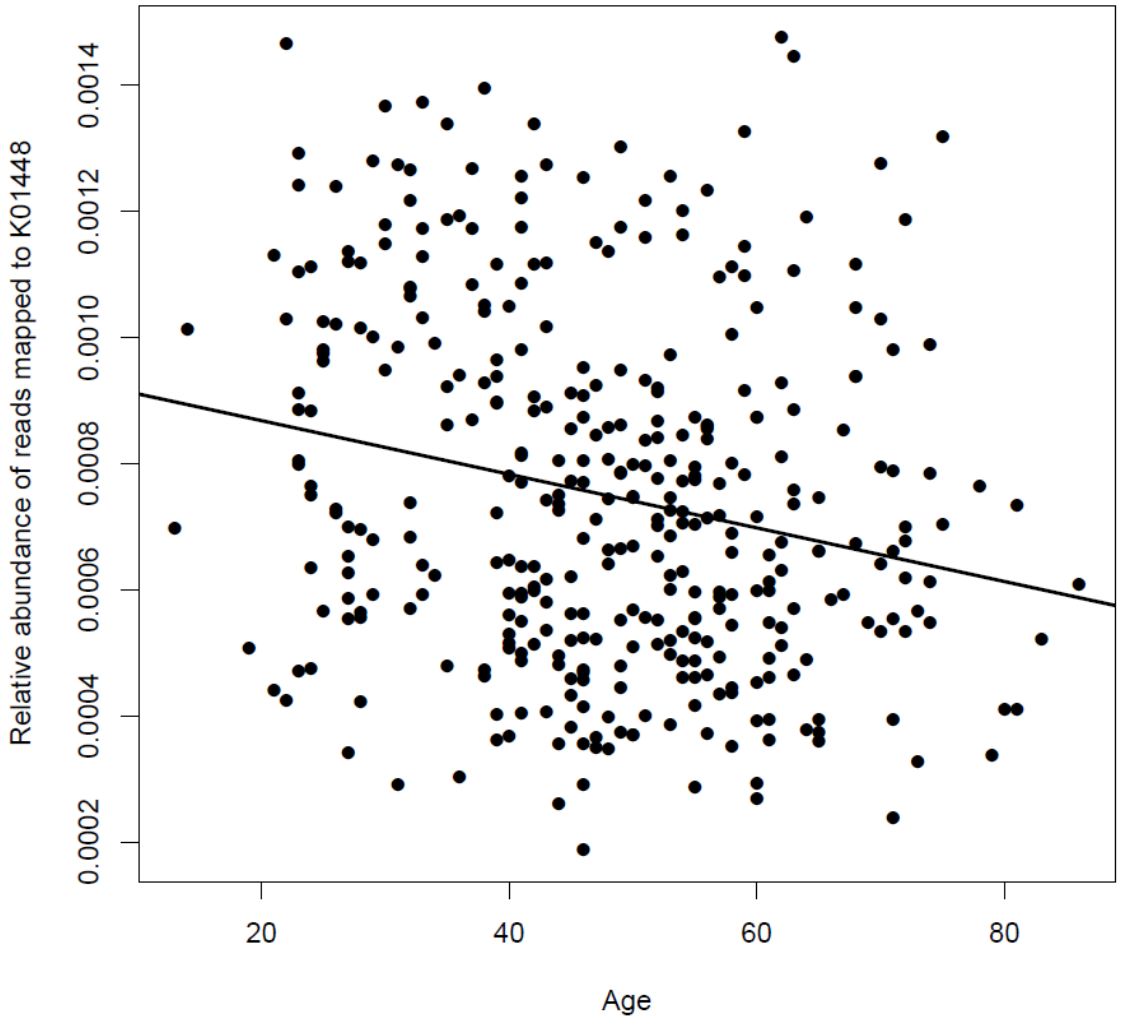

Figure 13 Relative abundance of K01448: N-acetylmuramoyl-L-alanine amidase is negatively correlated with age in the Chinese samples. Scatterplot showing the relative abundance of K01448 in all Chinese dataset samples. R^2=0.0475 and FDR p=0.0006 calculated in R. Plot generated in R.

K00001: alcohol dehydrogenase had the second strongest correlation out of all MDMGs examined (r^2=0.0602) (Figure 14), with K03743 being the only MDMG with a higher r^2 value. There is a large amount of variation in relative abundance between samples, but the trend with age is statistically significant (FDR p=0.0001).

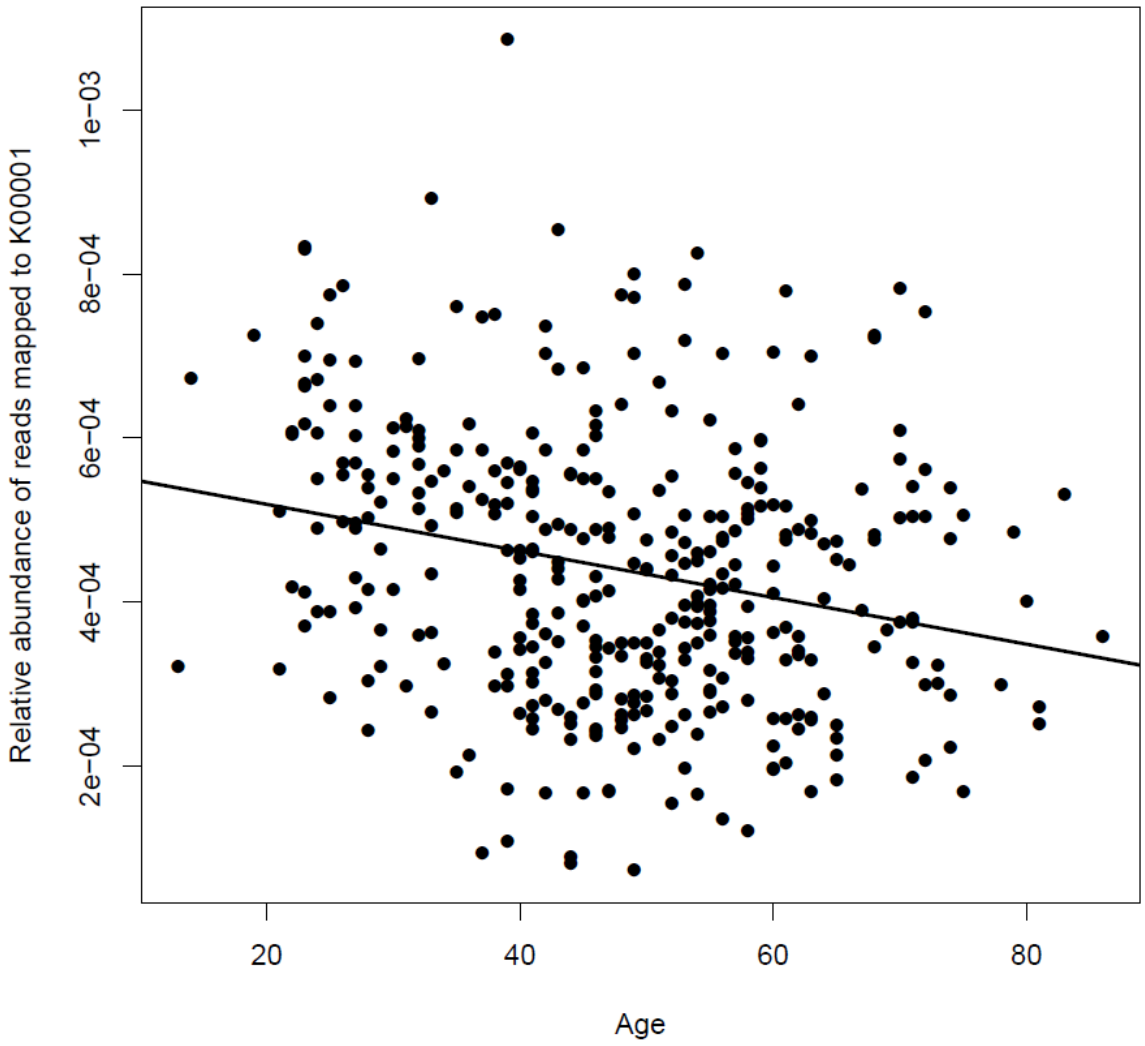

Figure 14 Relative abundance of K00001: alcohol dehydrogenase is negatively correlated with age in the Chinese samples. Scatterplot showing the relative abundance of K00001 in all Chinese dataset samples. R^2=0.0602 and FDR p=0.0001 calculated in R. Plot generated in R.

K00600: serine hydroxymethyltransferase (SHMT) was the 3rd most abundant MDMG in the Chinese dataset (4.02% mean relative abundance), and the most abundant MDMG shown to be statistically significant (FDR p=0.036). The abundance of SHMT from person to person is quite variable (r^2=0.0197), but overall there is a trend in increased abundance as age increases (Figure 15).

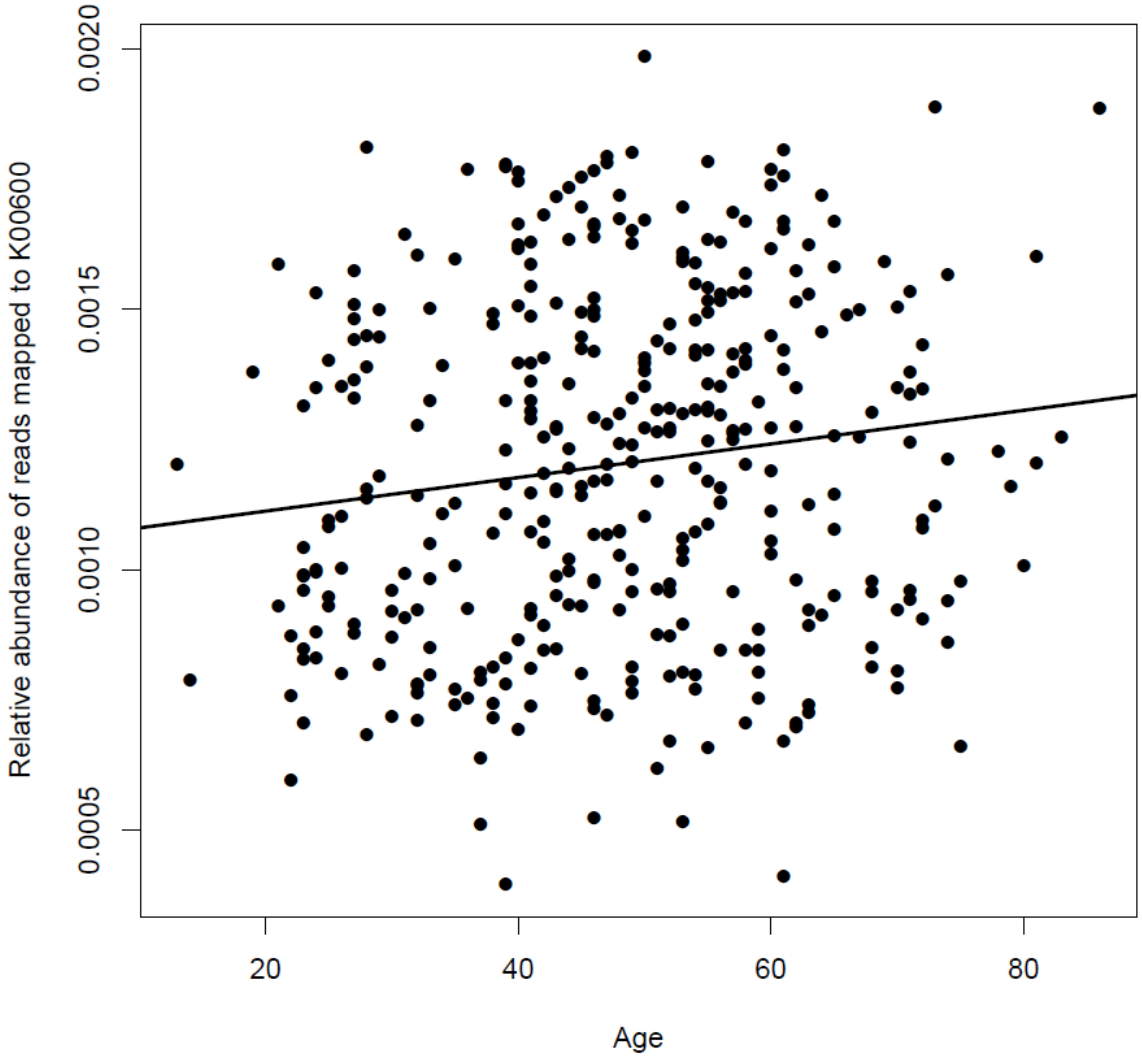

Figure 15 Relative abundance of K00600: SHMT is positively correlated with age in the Chinese samples. . Scatterplot showing the relative abundance of K00600 in all Chinese dataset samples. R^2=0.0197 and FDR p=0.036 calculated in R. Plot generated in R.

The final MDMG examined in this chapter is K04042: UDP-N-acetylglucosamine pyrophosphorylase / glucosamine-1-phosphate N-acetyltransferase. The abundance of this MDMG is positively correlated with age (r^2=0.0481, p=0.0008) (Figure 16).

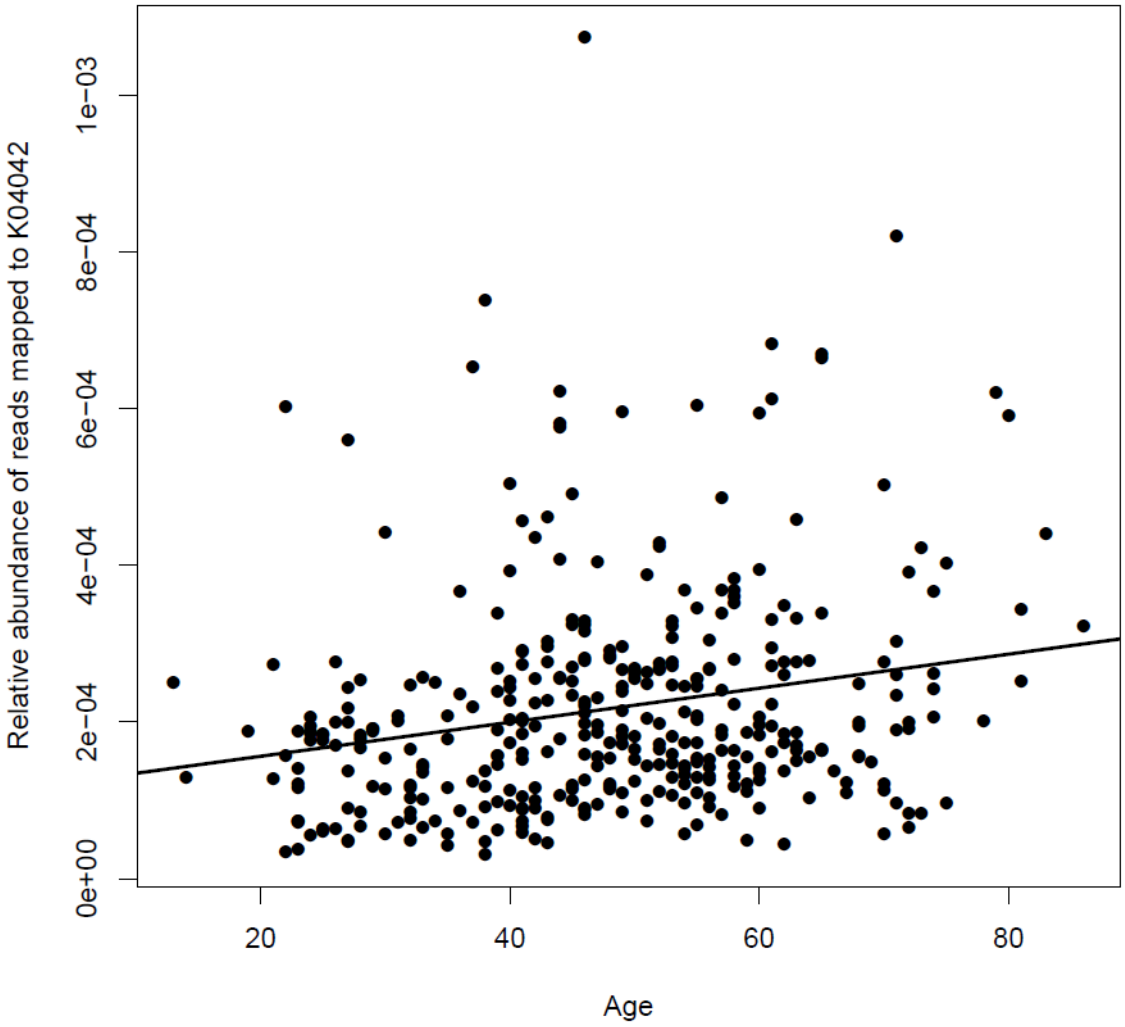

Figure 16 Relative abundance of K04042: UDP-N-acetylglucosamine pyrophosphorylase / glucosamine-1-phosphate N-acetyltransferase is positively correlated with age in the Chinese samples. Scatterplot showing the relative abundance of K04042 in all Chinese dataset samples. R^2=0.0481 and FDR p=0.0008 calculated in R. Plot generated in R.

4.4 SUMMARY

From the Chinese dataset, we found that 163/174 of our MDMGs were actually present in the gut microbiome in these samples. Three of these were also absent in the Human Microbiome Project dataset, but also the Chinese dataset had three MDMGs that were not found in the HMP samples. There were large amounts of variation within MDMGs, especially the ones found in lower abundances. Overall, age had a small but

statistically significant impact on the abundance of MDMGs. The abundance of 36 MDMGs were significantly correlated with age, 19 displaying positive trends, and 17 with negative trends.

CHAPTER 5: MICROBAL DRUG METABOLIZING GENES WITHIN THE GUT MICROBIOME OF NORTHWOOD RESIDENTS

5.1 PRESENCE AND VARIABILITY OF MICROBIAL DRUG METABOLZING GENES ACROSS NORTHWOOD SAMPLES

The Northwood dataset contained 205 stool samples coming from 47 individuals. 16S sequencing was performed on these samples to identify OTUs. PICRUSt was then used to predict what functions were present in the samples. From all the KOs in KEGG, PICRUSt was able to annotate 6909 KOs. A total of 4912 KOs were predicted to be in the Northwood samples. Out of our initial list of 225 potential MDMGs, PICRUSt was able to successfully annotate 174 KOs. This list is not exactly the same as the one that HUMAnN was able to annotate. PICRUSt was unable to annotate three MDMGs that HUMAnN annotated. They were K01202: Galactosylceramidase, K02295: gentamicin 3'-N-acetyltransferase, or K14568: rRNA small subunit pseudouridine methyltransferase. On the other hand, PICRUSt was able to annotate K14681: argininosuccinate lyase, K14682: mino-acid N-acetyltransferase, and K15023 5- iron sulfur protein methyltransferase, all of which were unannotated by HUMAnN.

From the 16S sequencing data of the Northwood samples, PICRUSt predictions found 139/174 MDMGs present amongst these samples. Figure 17 below shows the average relative abundance of each detected MDMG as well as its standard deviation. K05349: beta-glucosidase was the most prevalent (6.54% average relative abundance) and K13955: zinc-binding alcohol dehydrogenase/oxidoreductase was the least abundant (0.000364%). When looking at variability, K00144: alcohol dehydrogenase is the most variable (0.0161 SD/0.00125 mean = 12.9 CV) and K00600: SHMT is the least variable (0.186 SD/2.01 mean = 0.0928 CV).

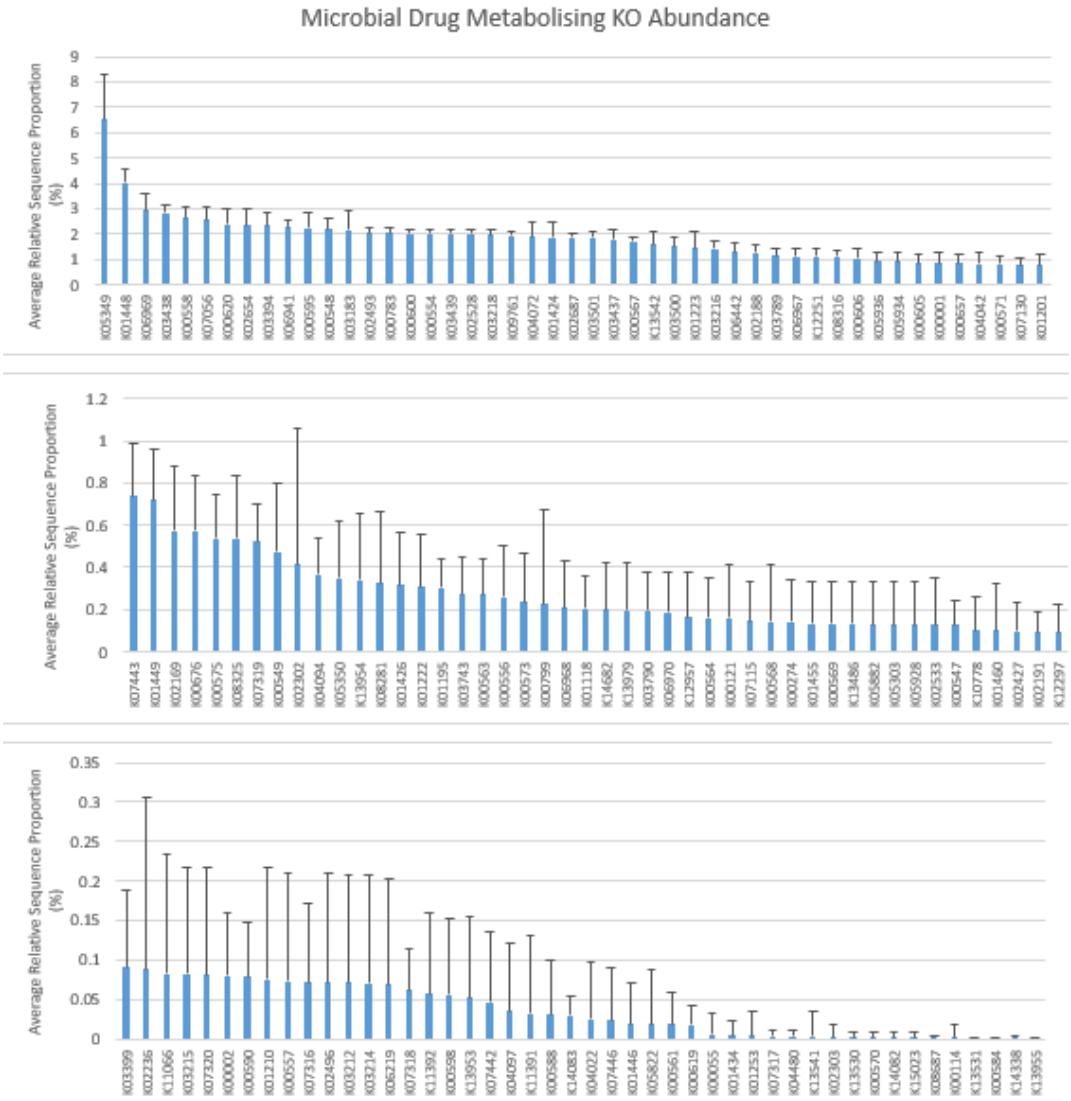

Figure 17 Average relative abundance of 139 MDMGs found from the Northwood microbiome gut samples after PICRUSt annotation. MDMGs are listed along the X-axis from highest average sequence proportion (K05349) to lowest average sequence proportion (K13955). Error bars represent one standard deviation**.**

5.2 AGE INFLUENCES THE ABUNDANCE OF MICROBIAL DRUG METABOLZING GENES IN THE NORTHWOOD DATASET

The age of each person sampled was also recorded in Northwood dataset. These individuals ranged in age from 65 to 99 years old (Figure 18).

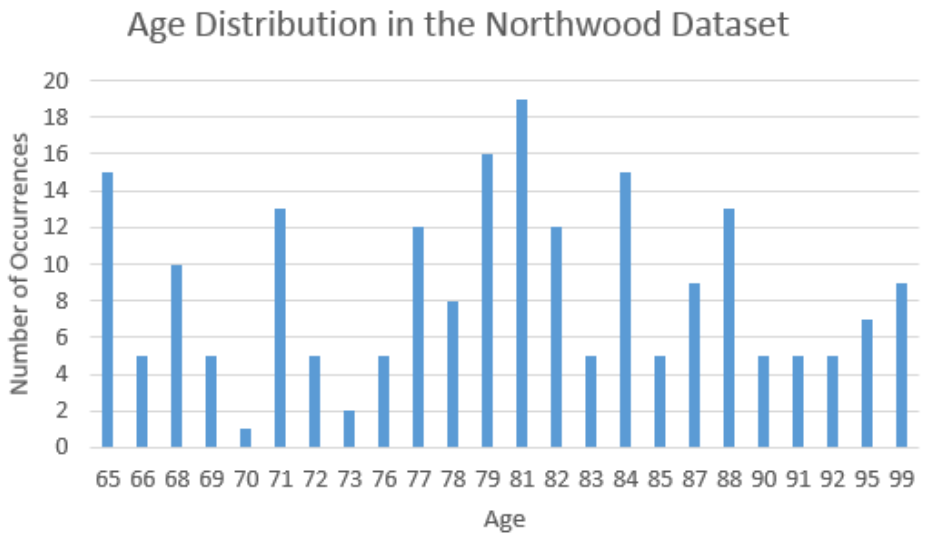

Figure 18 Age distribution of samples in the Northwood dataset.

First, the relationship between age and the overall abundance of MDMGs in the Northwood samples was investigated (Figure 19). Adonis calculations result in r^2=0.0228 and p=0.005. This tells us that age does explain a statistically significant proportion of the variation in the abundance of MDMGs, but the correlation is weak as there is a lot of variation not attributed to age.

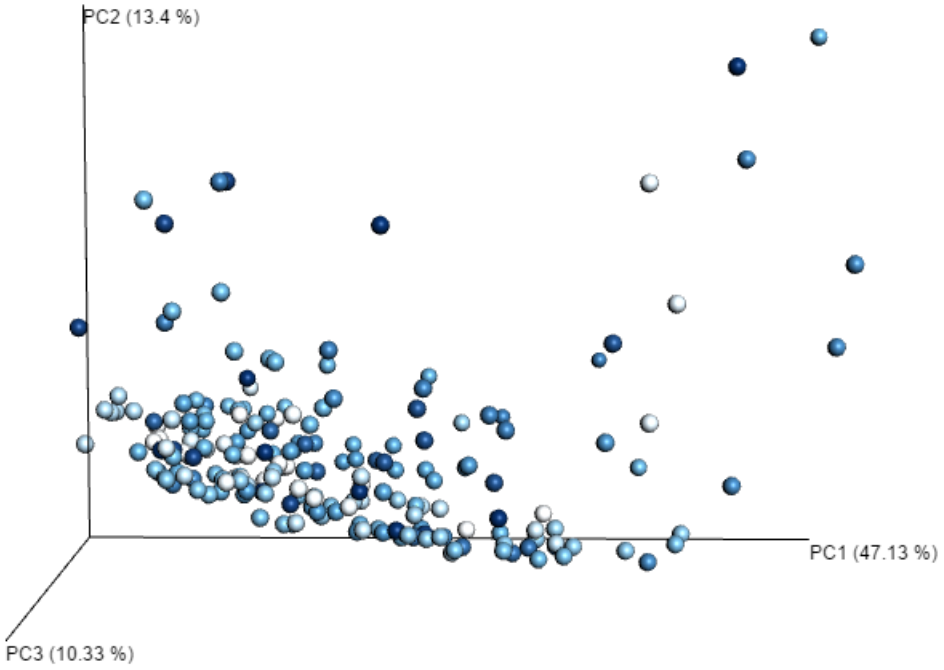

Figure 19 Age correlates weakly but significantly with variation of MDMGs in the Northwood samples. PCoA plot showing 205 Northwood samples based off similarities in MDMGs from PICRUSt predictions. Samples are coloured from the youngest individual (65 y/o) as white, to the oldest individual (99 y/o) as the darkest blue. Adonis calculations determine r^2=0.0228 and p=0.005

Examining individual MDMGs, the abundance of 39 are significantly correlated with age in this dataset (FDR p<0.05). Out of these, 24 MDMGs show a positive correlation of their abundance vs age, and 15 MDMGs show a negative correlation with age. Interestingly, three MDMGs that showed positive correlations were also found to be positively correlated with age in the Chinese dataset, and seven of the negatively correlated MDMGs in this dataset were also found to be negatively correlated with age in the Chinese dataset. Furthermore, the abundance of K00554 and K07442 (both listed as tRNA methyltransferases) were found to be positively correlated with age in the Northwood samples, but negatively correlated with age in the Chinese samples. Table 4 in the appendix details all significant MDMGs. Figures 20 through 23 highlight some MDMGs of interest.

The first MDMG examined in this dataset was K01195: beta-glucuronidase (Figure 20). This MDMG does not have the strongest correlation with r^2=0.0401, as there is significant amounts of variation across ages. The slope of this correlation is not

drastically steep, however, it was still found to be a statistically significant correlation (FDR p=0.018).

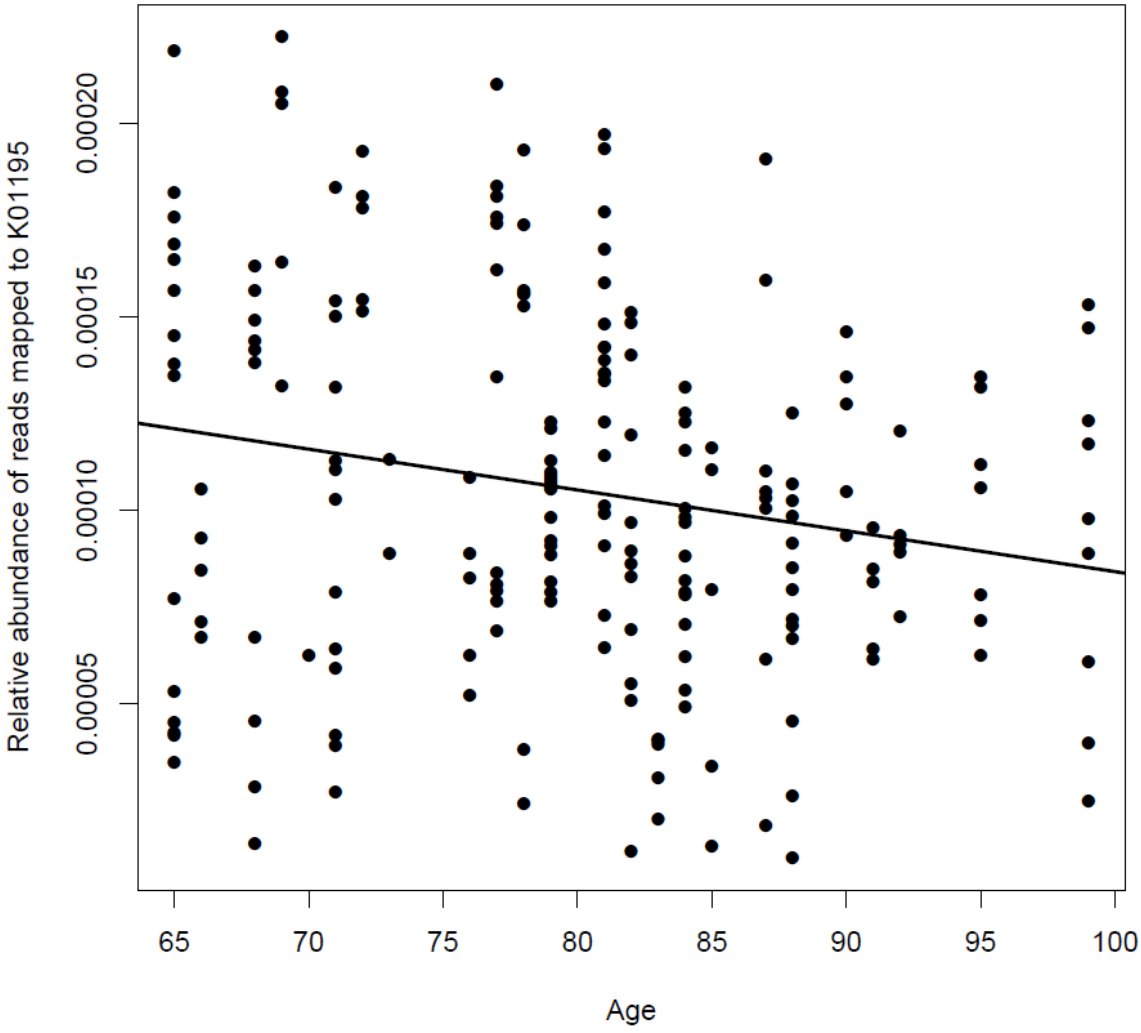

Figure 20 Relative abundance of K01195: beta-glucuronidase is negatively correlated with age in the Northwood samples. Scatterplot showing the relative abundance of K01195 in all Northwood dataset samples R^2=0.0401 and FDR p=0.018 calculated in R. Plot generated in R.

K01118: FMN-dependent NADH-azoreductase was interesting as it was one of the only functionally annotated azoreductase (Figure 21). This MDMG had a moderate correlation with age (r^2=0.0707, FDR p=0.002). This is another situation where many samples have similar abundances (close to 0.00005%) but particular individuals,

especially around 80 years and older, have a substantial increase in the abundance of this MDMG, driving the positive trend.

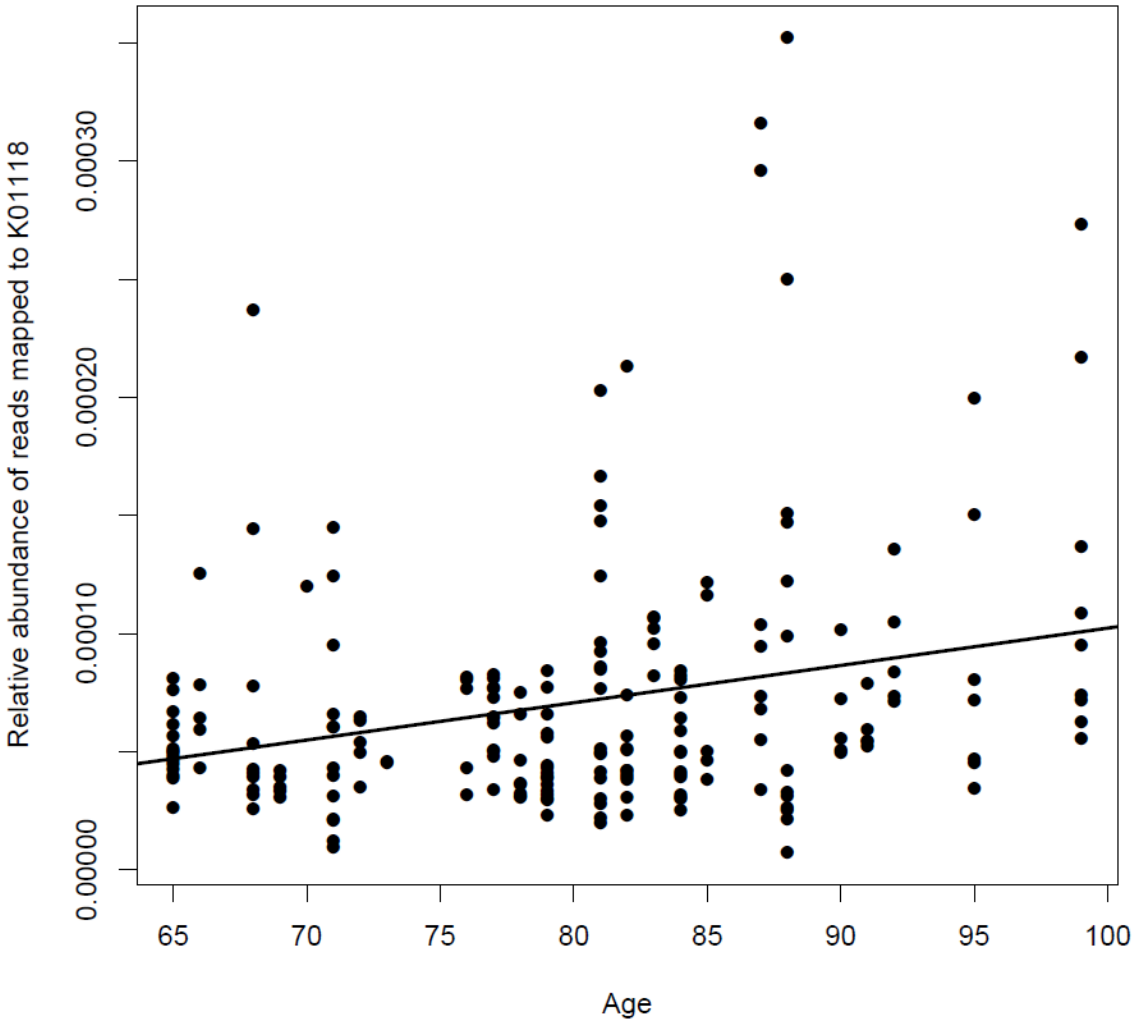

Figure 21 Relative abundance of K01118: FMN-dependent NADH-azoreductase is positively correlated with age in the Northwood samples. Scatterplot showing the relative abundance of K01118 in all Northwood dataset samples R^2=0.0707 and FDR p=0.002 calculated in R. Plot generated in R.

Our last analysis in this section looks at a few MDMGs. K03218: RNA methyltransferase (TrmH family) is examined first (Figure 22). This MDMG has the strongest negative correlation out of all our Northwood age plots, with $r^2 = 0.1352$ and an FDR p value of p=0.000008.

Another MDMG found to be significantly correlated with age was K03214: RNA methyltransferase (TrmH family), (FDR p= 0.002) (Figure 23). This MDMG is less

statistically significant than K03218, but still had a moderate correlation with age (r^2=0.0639). Unlike K03218 which displayed a negative correlation, K03214 actually shows a positive correlation in abundance vs age.

In fact, there is even a third significant MDMG, K03437, also classified as RNA methyltransferase (TrmH family). This one, which showed a downward trend in abundance vs age, had a weaker correlation than the others (r^2 = 0.0378) but was still statistically significant (FDR p=0.02).

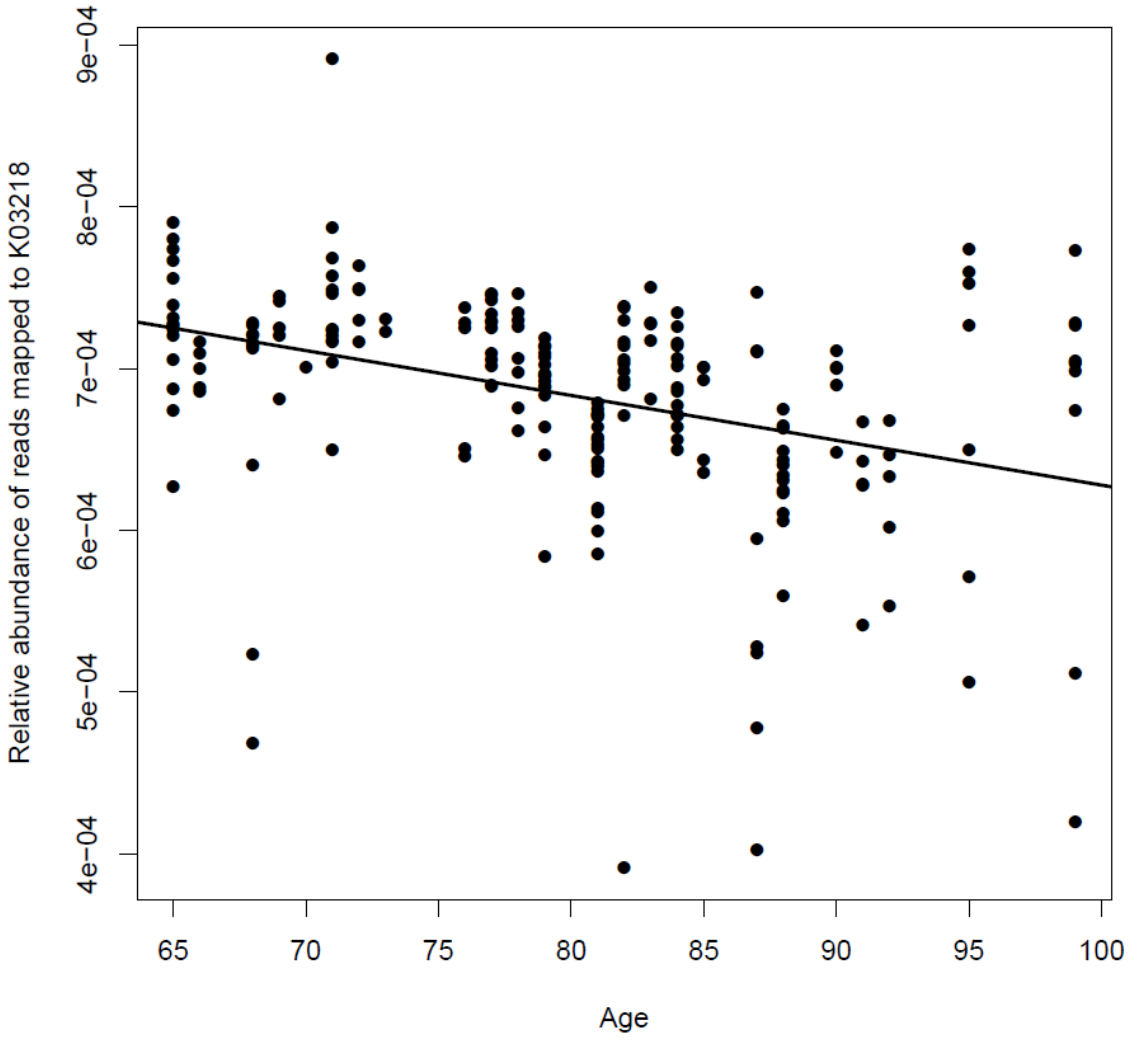

Figure 22 Relative abundance of K03218: RNA methyltransferase, TrmH family is negatively correlated with age in the Northwood dataset. Scatterplot showing the relative abundance of K03218 in all Northwood dataset samples. R^2 = 0.1352 and FDR p=0.000008 calculated in R. Plot generated by R.

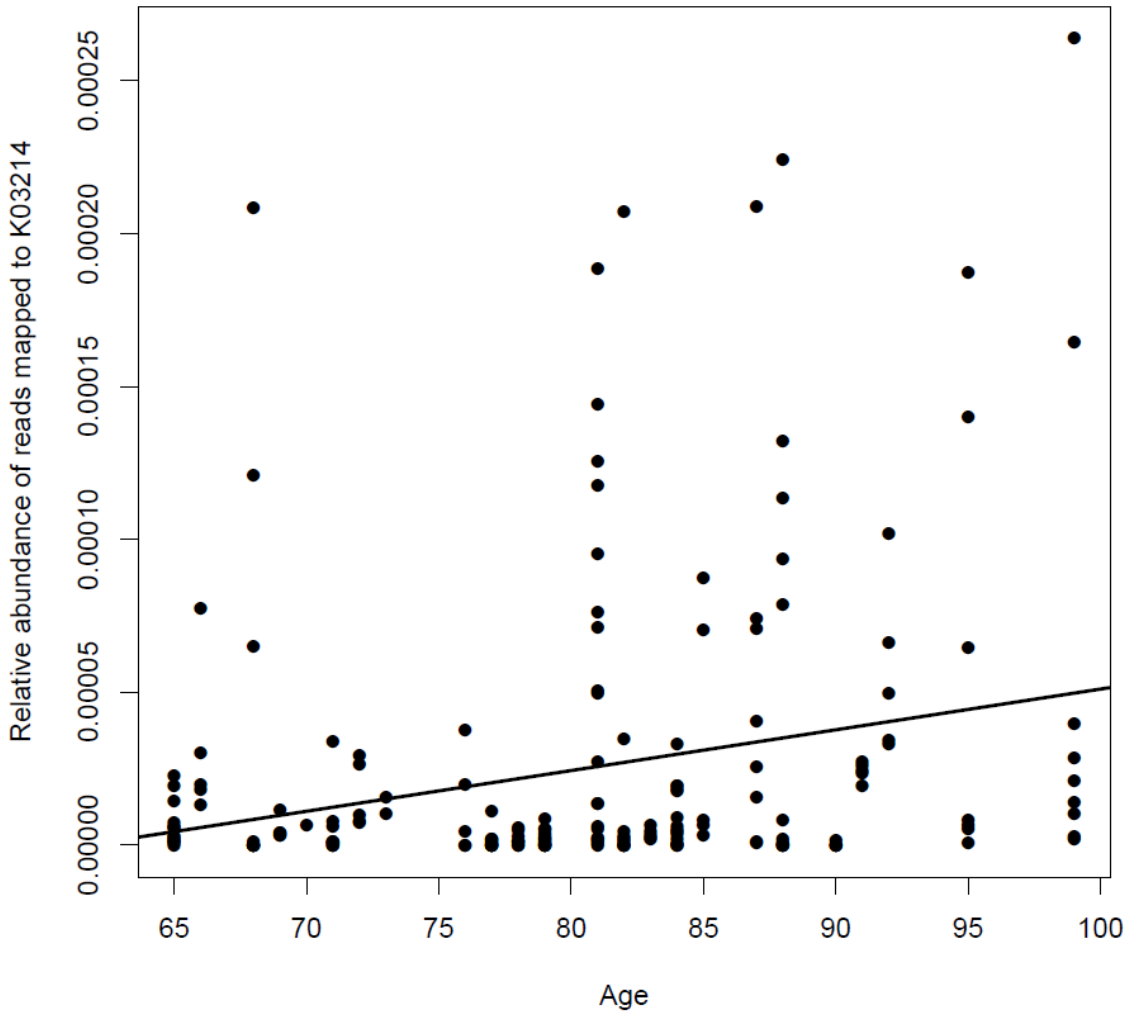

Figure 23 Relative abundance of K03214: RNA methyltransferase, TrmH family is positively correlated with age in the Northwood dataset. Scatterplot showing the relative abundance of K03214 in all Northwood dataset samples. R^2=0.0639 and FDR p=0.002 calculated in R. Plot generated by R**.**

5.3 FRAILTY INFLUENCES THE ABUNDANCE OF MICROBIAL DRUG METABOLZING GENES IN THE NORTHWOOD DATASET

In the Northwood dataset, frailty of individuals was also assessed using a frailty index (FI) scoring system (Rockwood, 2005). In this system, individuals are evaluated on several physical, mental, and behavioral factors. Scores are represented on a scale from 0 to 1, with 0 being the least frail, and 1 being the frailest in all categories. The Northwood

individuals had FI scores ranging from 0.305 to 0.758 (Figure 24). Additionally, the age of an individual and their frailty are not strongly correlated, with an r^2=0.004 (Figure 25).

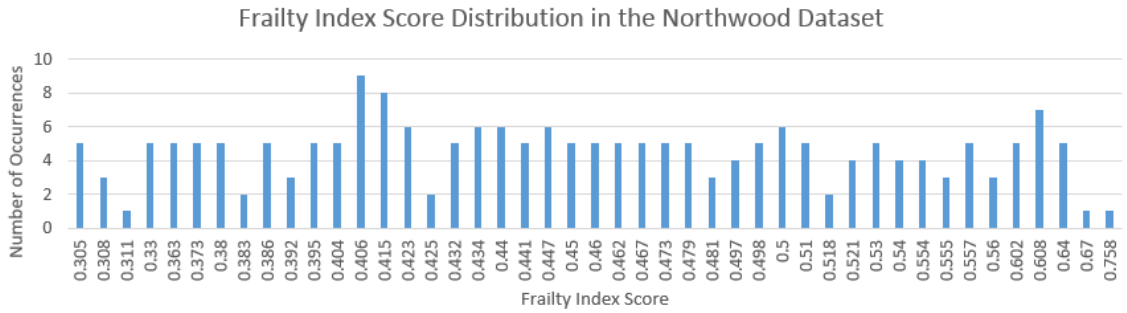

Figure 24 Frailty Index score distribution of samples in the Northwood dataset.

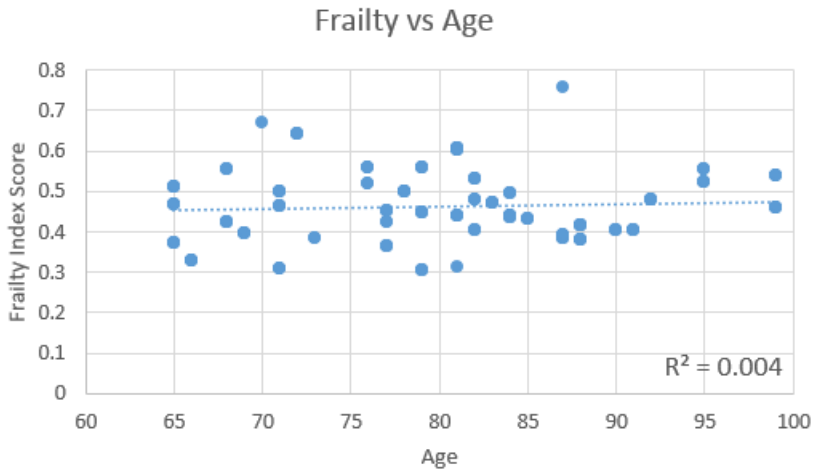

Figure 25 No correlation between age and FI score in the Northwood dataset. Scatterplot representing the age and FI score of the 47 individuals in the Northwood dataset. R^2=0.004, p>0.05.

Using the frailty index score calculated for each individual, changes in MDMG abundance can be explored to see if the correlated with frailty (Figure 26). The Adonis calculation for this dataset based on frailty gives us r^2=0.01379 and p=0.033. This suggests that frailty explains a small, yet statistically significant proportion of the variation in MDMGs between samples. However, this correlation had a lower r^2 value and a higher p value than the Adonis calculation with age on this dataset (r^2=0.0228 and p=0.005 based off age).

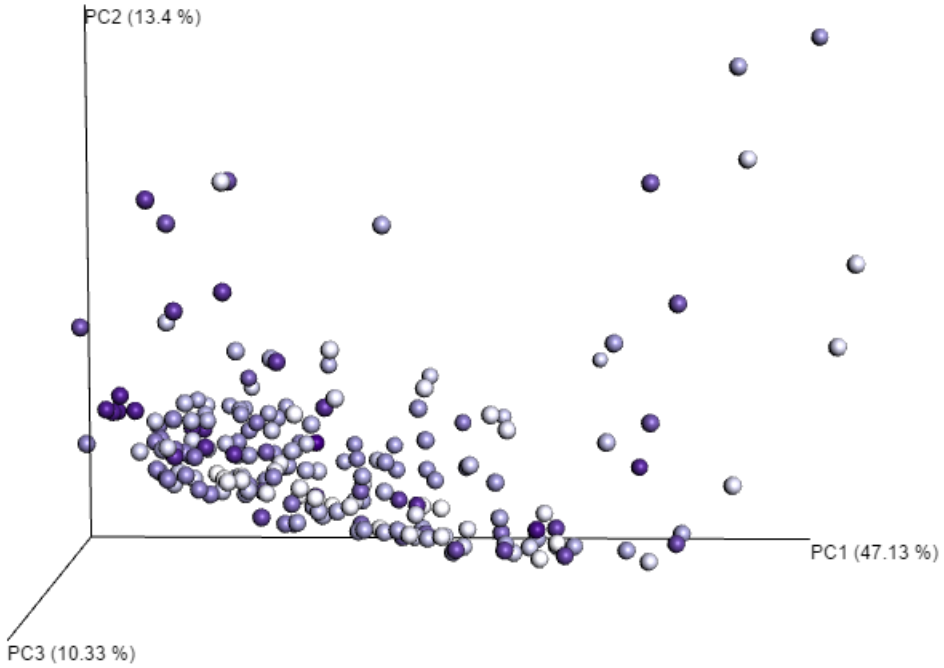

Figure 26 Frailty correlates weakly but significantly with variation of MDMGs in the Northwood samples. PCoA plot showing 205 Northwood samples based off similarities in MDMGs from PICRUSt predictions. Samples are coloured from the individual with the lowest frailty index score (0.305) as white, to the individual with the highest frailty index score (0.758) as the darkest purple. Adonis calculations determine r^2=0.01379 and p=0.033.

Examining individual MDMGs, 24 were found to be significantly correlated with FI score (FDR $p<0.05$). Out of these significant MDMGs, 11 show a positive correlation in their abundance vs frailty, and 13 show a negative correlation with frailty. Interestingly, only one MDMG that was significantly correlated with frailty was also significantly correlated with age. Table 4 in the appendix details all significant MDMGs. Figures 27 through 29 highlight some MDMGs of interest.

The first MDMG examined was K01424: L-asparaginase. The correlation between this MDMG and FI score had an r^2=0.0559, and FDR p=0.0095 (Figure 27). There is a large amount of variation in abundance of this MDMG for all FI scores, but overall there is a positive trend in abundance as FI score increases.

K01424 was the only MDMG that was significant in both age and frailty. Interestingly, the abundance decreased as individuals got older. With age, the correlation

was weaker than the frailty correlation (r^2=0.0317 in the age correlation), and it was only just barely classified as statistically significant based on age (FDR p=0.039).

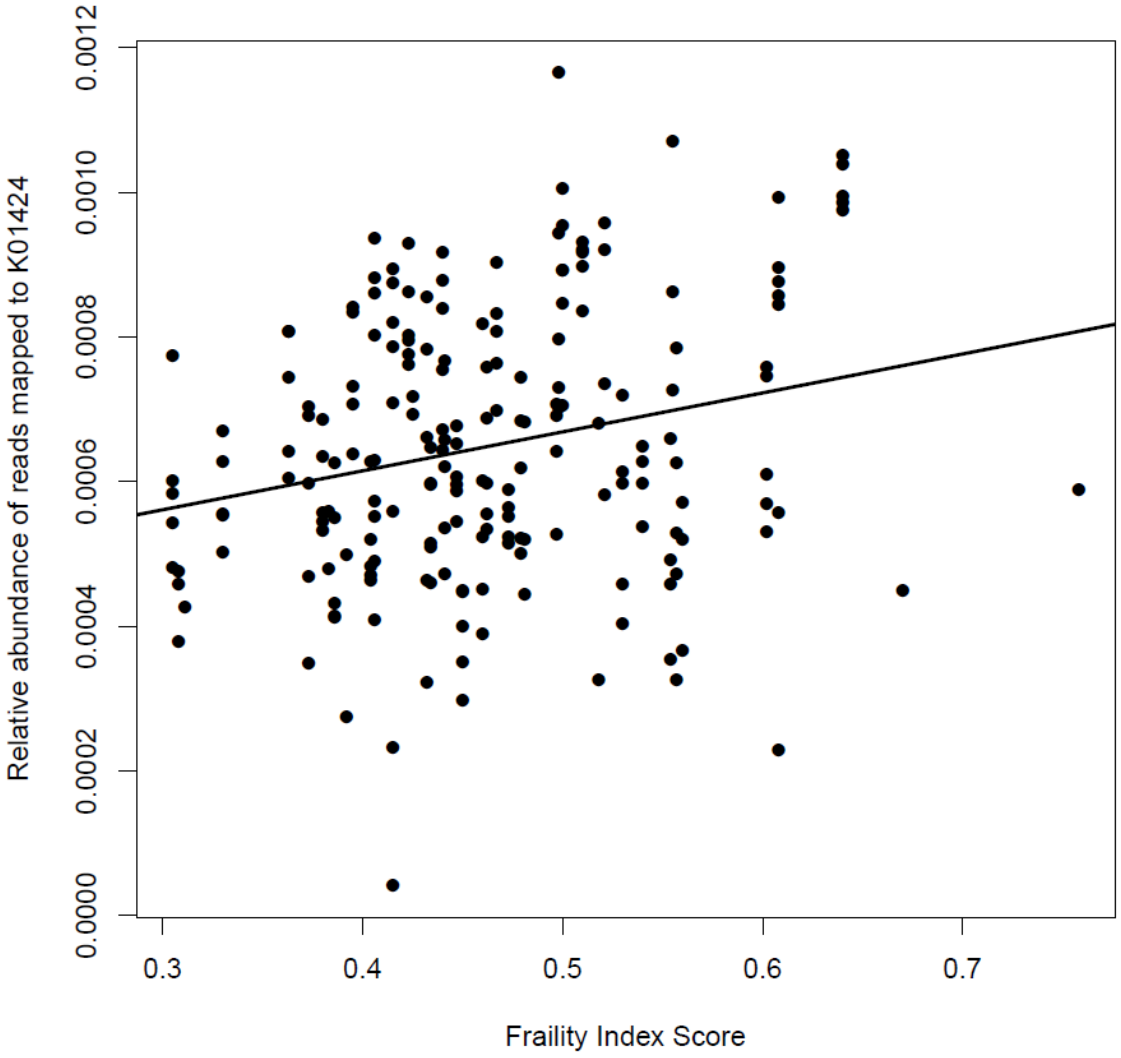

Figure 27 Relative abundance of K01424: L-asparaginase is positively correlated with FI score in the Northwood samples. Scatterplot showing the relative abundance of K01424 in all Northwood dataset samples. R^2=0.0559 and FDR p=0.0095 calculated by R. Plot generated in R.

K02654: prepilin peptidase had an r^2=0.0827, making it the strongest correlation based on frailty, as well as an FDR p=0.004 (Figure 28). Once again there is lots of variation in abundance, particularly in the mid-range of frailty. However, there is a clear downward trend in abundance as FI score increases.

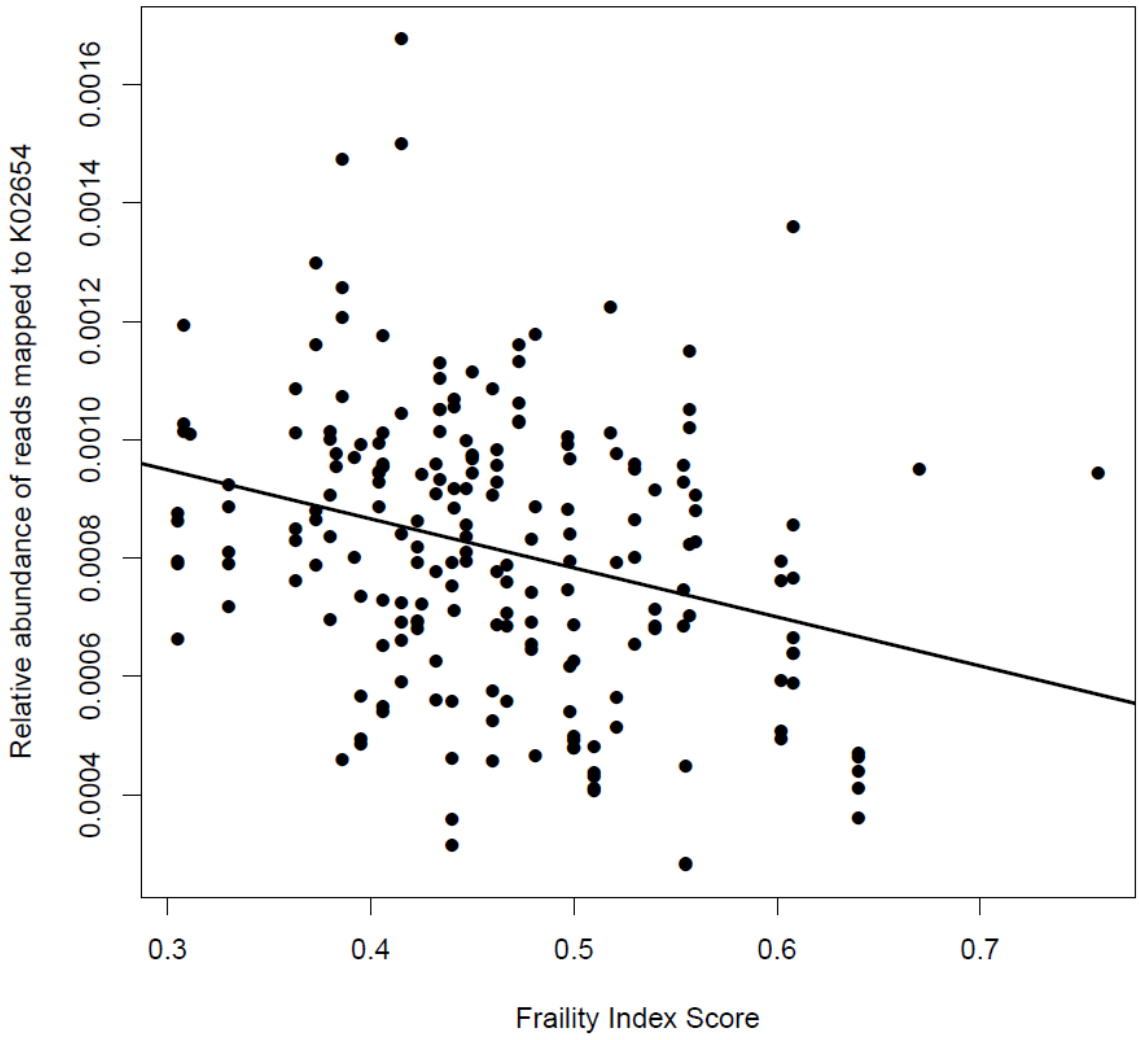

Figure 28 Relative abundance of K02654: prepilin peptidase is negatively correlated with FI score in the Northwood samples. Scatterplot showing the relative abundance of K02654 in all Northwood dataset samples. R^2=0.0827 and FDR p=0.004 calculated by R. Plot generated in R.

K08325: NADP-dependent alcohol dehydrogenase had the strongest positive correlation between abundance and frailty with r^2=0.0746, and FDR p=0.005 (Figure 29). There is a lot of variation in abundance of this MDMG, but there is a notable clustering of points at high abundance for individuals with FI scores greater than 0.5. The relative abundance for K08325 is quite low overall compared to the others analysed in this section, but it was still found to some extent across all Northwood samples.

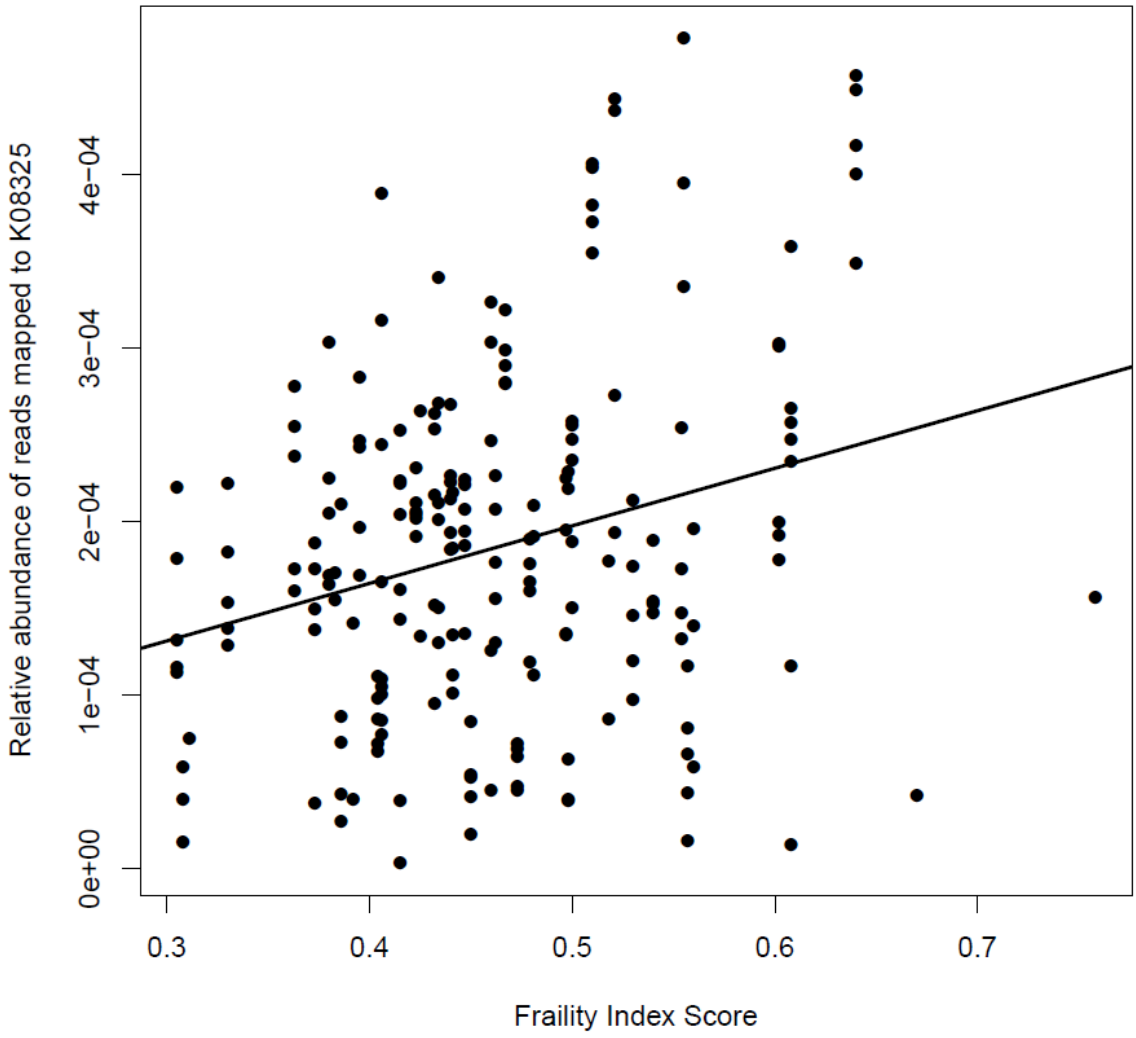

Figure 29 Relative abundance of K08325: NADP-dependent alcohol dehydrogenase is positively correlated with FI score in the Northwood samples. Scatterplot showing the relative abundance of K08325 in all Northwood dataset samples. R^2=0.0746 and FDR p=0.005 calculated by R. Plot generated in R.

5.4 MEDICATION INFLUENCES THE ABUNDANCE OF MICROBIAL DRUG METABOLZING GENES IN THE NORTHWOOD DATASET

The Northwood study also collected information on what drugs each patient was taking with a total of 291 different drugs being taken across one or more patients. Drugs that would likely affect the gut microbiome through GI related treatments were identified (Table 3).

Table 3 List of drugs thought to alter the microbiome. Daily dose of each drug is listed. A * after the dose indicates a generic version of the drug was used. The probiotic Bacid and the laxative Lax-A Day were also included as drugs thought to alter the microbiome.

Drug name	Dose
Ferrous Gluconate	300mg, 300mg*
Ferrous Sulfate	300mg, 300mg*
Omeprazole	20mg*
Pantoprazole	40mg*
Prednisone	5mg, 5mg*
Bacid	N/A (probiotic)
Ranitidine	150mg*, 300mg*
Feramax	150mg
Gaviscon	200mg
Lax-A Day	N/A (laxative/powder)
Magnesium Oxide	420mg
Tecta	40mg
Rabeprazole EC	20mg*

In total, 159 samples came from patients on at least one of the 'microbiome-altering drugs' and 46 samples came from patients not on any of these drugs. The samples were split into two groups to examine if these microbiome-altering drugs change the abundance of KOs in general and for our MDMGs.

First all KOs predicted by PICRUSt in these samples were examined. PICRUSt was able to annotate a total of 6909 KOs, with 4912 KOs being predicted in the actual Northwood samples (Figure 30). Overall the two groups do not show distinct separation. Testing for individual KO differences between the drug groups resulted in the identification of 24.6% (1206/4912) significant KOs ($p < 0.05$, Welch's two-sided t-test, with Benjamini-Hochberg FDR MTC).

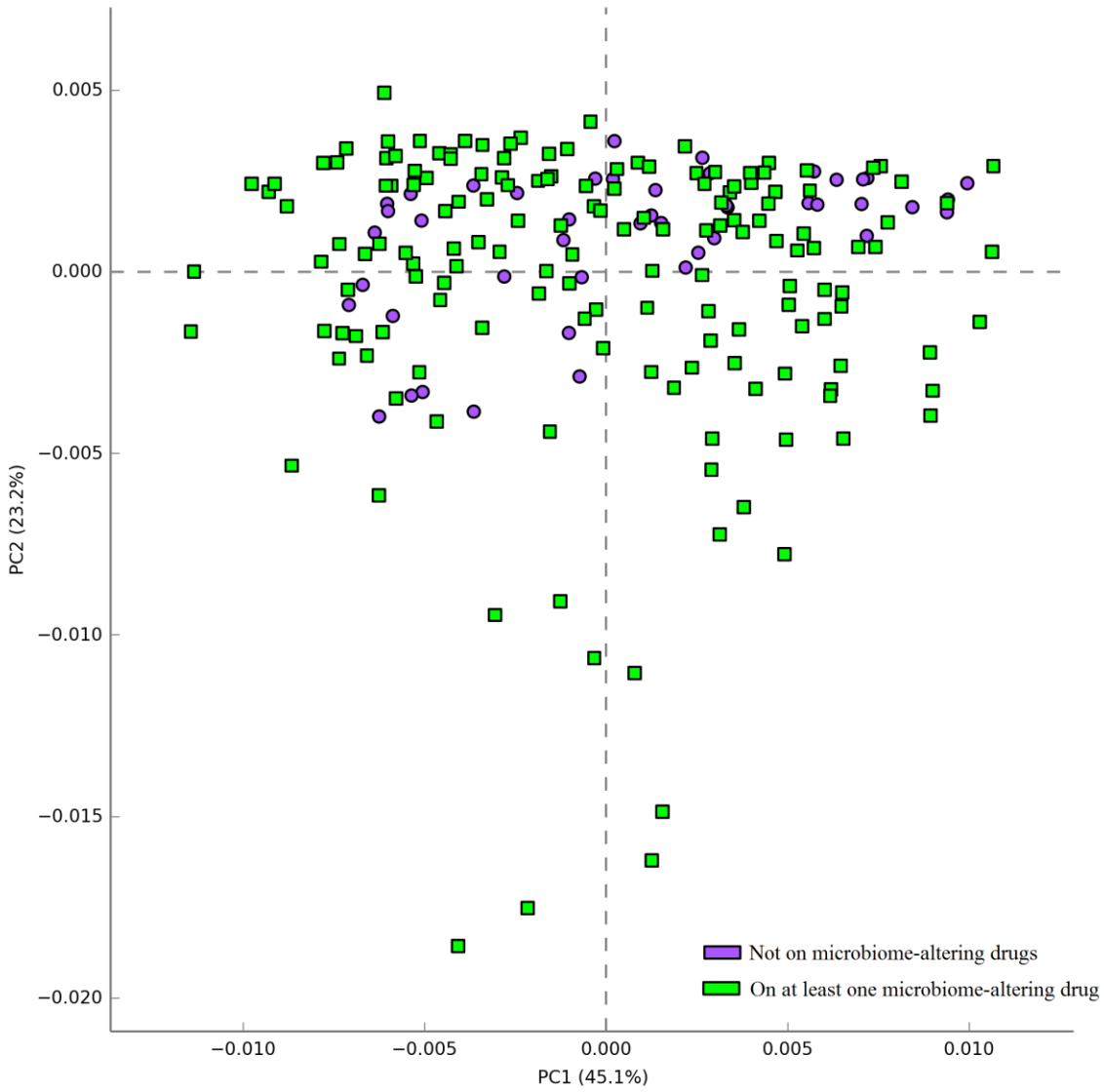

Figure 30 Microbiome differences between individuals on vs off microbiome-altering drugs based off all KOs annotated by PICRUSt from Northwood samples. PCA plot showing 205 Northwood samples based off similarities in all 4912 KOs from PICRUSt predictions. It compares samples from individuals not on any identified microbiome-altering drug (purple circles, 46 samples) vs samples from individuals on at least one identified microbiome-altering drug (green squares, 159 samples).

Next, only the 139 MDMGs in the Northwood samples were examined (Figure 31). Once again, most of the samples both groups are intertwined and there is no distinct clustering.

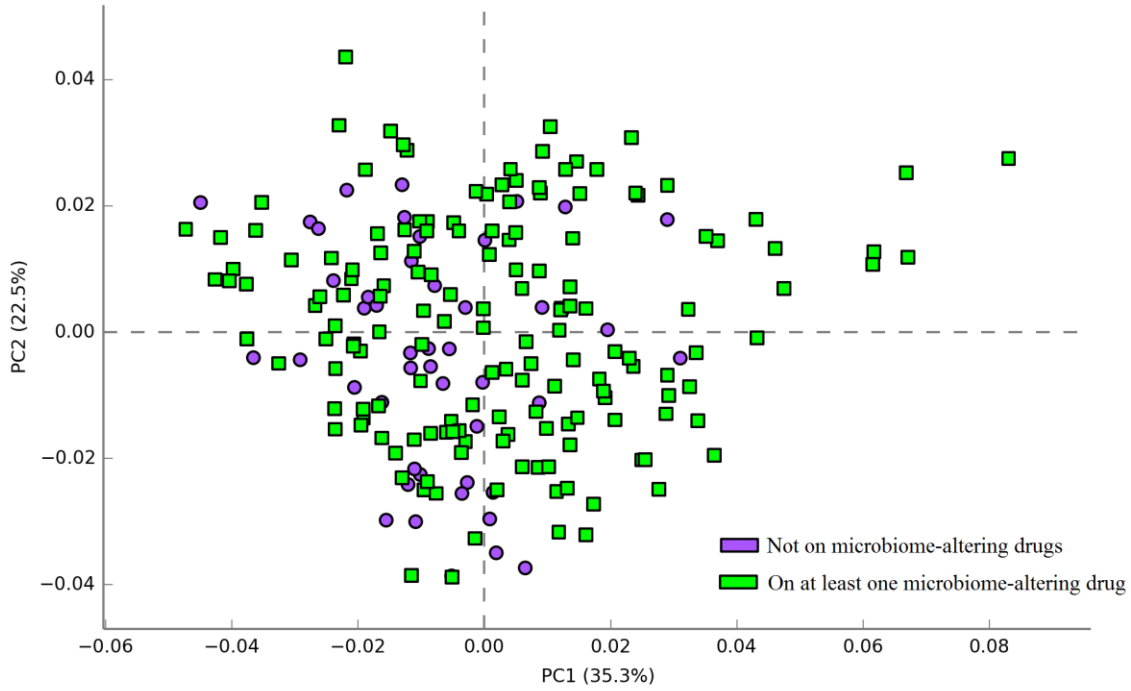

Figure 31 Microbiome differences between individuals on vs off microbiome-altering drugs based off MDMGs annotated by PICRUSt from Northwood samples. PCA plot showing 205 Northwood samples based off similarities in MDMG abundance from PICRUSt predictions. It compares samples from individuals not on any identified microbiome-altering drug (purple circles, 46 samples) vs samples from individuals on at least one identified microbiome-altering drug (green squares, 159 samples).

Individually, 40 MDMGs out of the 139 are statistically different ($p<0.05$) between the two groups after MTC, representing 28.8% of the MDMGs predicted in these samples. Although this is a higher percent of KOs affected when compared to the full KO list, a chi-squared test tells us that this increase is not statistically significant.

Out of the 40 MDMGs that are statistically different between the two groups, only five have a higher average relative abundance in the group not on microbiome-altering drugs: K12251: N-carbamoylputrescine amidase, K05349: beta-glucosidase, K03438: S-adenosyl-methyltransferase, K03394: precorrin-2 C20-methyltransferase, and K00620: glutamate N-acetyltransferase. The remaining 35 MDMGs show a significant increase in average relative abundance in the group on microbiome-altering drugs.

Comparing significant MDMGs in the Northwood samples, 15 of the MDMGs that are higher abundance in the drug group are also positively correlated with age. Additionally, four MDMGs with increased abundance in the drug group are also positively correlated with frailty, and one of the MDMGs with decreased relative abundance in the drug group was also negatively correlated with frailty. Table 4 in the appendix details all significant MDMGs. The box plots in Figure 32 and Figure 33 compare the abundance of MDMGs between the two groups.

K12251: N-carbamoylputrescine amidase was one of the five to show a decreased relative abundance in the group on microbiome-altering drugs (Figure 32). Even though the mean relative abundance for this amidase is significantly lower in the group on microbiome altering drugs, there is an extensive amount of overlap across the two groups. Most of the samples in the microbiome-altering drug group fall within the box of the non-drug group. Only portion samples appear to be 'dragging down' the average abundance. This pattern of extensive overlap is also seen in the other four MDMGs where the mean relative abundance is lower in the group on microbiome-altering drugs.

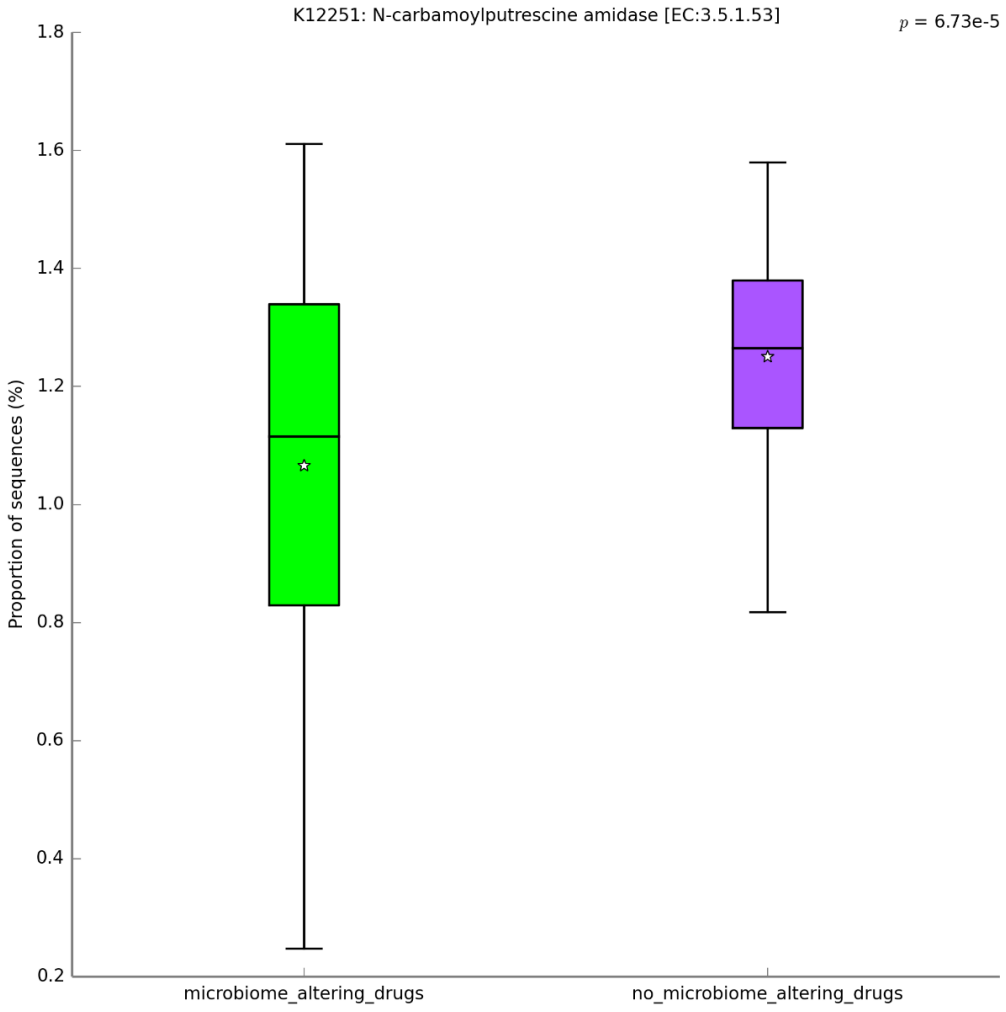

Figure 32 K12251: N-carbamoylputrescine amidase has decreased average relative abundance in individuals on microbiome altering drugs. The green box represents individuals taking at least 1 microbiome altering drug, and the purple box represents individuals not on any microbiome-altering drugs. White stars in the boxes indicate mean relative abundance for the group. p=0.00007 calculated using Welch's two-sided t-test with Benjamini-Hochberg FDR MTC.

Examining the 35 MDMGs that were increased in the microbiome-altering drug group there is a similar pattern. Using K00799: glutathione S-transferase as an example, the drug group does have a higher average relative abundance than those not on microbiome altering drugs (Figure 33). However, a large portion of these two boxes overlap, and it is clear that around 20 samples with much higher abundances are the driving force behind the increased average abundance in the drug group.

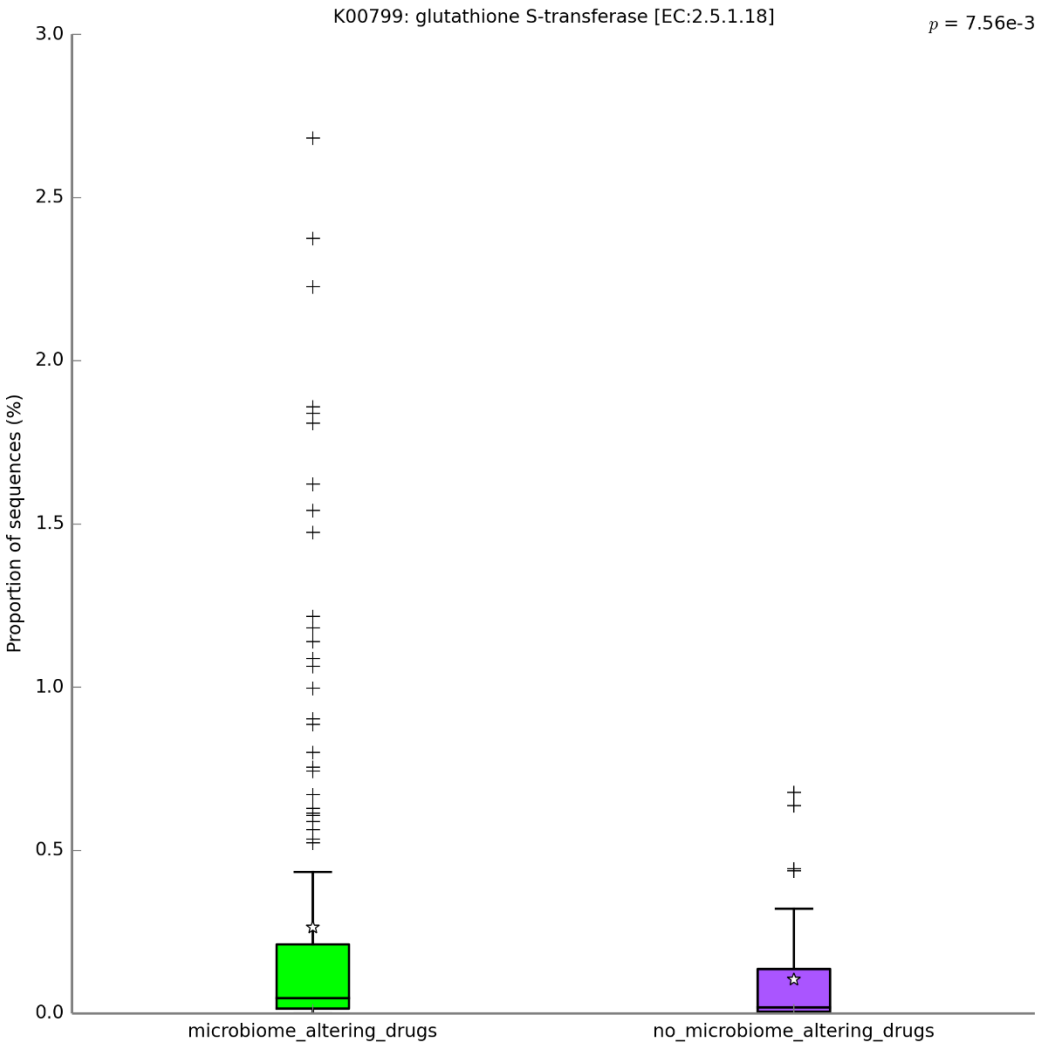

Figure 33 Figure 30 K00799: glutathione S-transferase has increased average relative abundance in individuals on microbiome altering drugs. The green box represents individuals taking at least 1 microbiome altering drug, and the purple box represents individuals not on any microbiome-altering drugs. White stars in the boxes indicate mean relative abundance for the group. p=0.008 calculated using Welch's two-sided t-test with Benjamini-Hochberg FDR MTC.

5.5 SUMMARY

Out of the 174 MDMGs annotatable by PICRUSt, 139 were predicted to be present within the Northwood samples. The individuals in this study ranged in age from 65-99 years old, and we found the abundance of 39 MDMGs correlated with age. Frailty

index scores were calculated for these individuals, ranging from 0.305-0.758, and 24 MDMGs were found to be significantly correlated with FI score. We also divided the samples into two groups: those on drugs thought to alter the microbiome and those not on drugs known to alter the microbiome. We find overall the abundance of 1206 KOs out of the 4912 predicted in the dataset are significantly different between these two groups. Examining only MDMGs, we find that 40 out of 139 have significantly different abundances between the two groups. However, it appears only a subset of the microbiome-altering drug group is responsible for driving this difference in abundance.

CHAPTER 6: DISCUSSION

6.1 VARIABILITY OF MICROBIAL DRUG METABOLIZING GENES THROUGHOUT THE HUMAN MICROBIOME

One key finding in all three datasets that may have been overshadowed throughout the results section is the number and types of MDMGs found in the microbiome. It is important to remember that the MDMGs are all functions related to drug metabolism in humans. The fact that we were able to identify nearly all investigated functions in the microbiome highlights its versatility in relation to metabolizing xenobiotic compounds. Reviews on microbiome-drug interactions such as the one by Haiser and Turnbaugh is informative as it directly links the microbiome to the metabolism of particular compounds (Haiser & Turnbaugh, 2013), yet these reviews contain a relatively small amount of functions carried out by the microbiome. Our research here indicates that there are far more microbiome-drug interactions than previously described through other microbial pathways that have yet to be investigated.

Another thing that is important about our findings is the amount of variation from person to person. Each person has a unique microbiome with altered abundances of MDMGs. Take K00543: acetylserotonin N-methyltransferase in the HMP dataset as an example. Most samples have a relative abundance close to the mean for this MDMG (which was 0.0004%), but relative abundance in one sample was approximately 0.03%, nearly 100x the mean. Biologically, this variability could mean loss of efficacy (as demonstrated with Digoxin and cgr genes (Haiser et al., 2013)) or increased activity pushing levels over the therapeutic window into toxicity (like with CPT-11 and beta-glucosidase (Wallace et al., 2010)). Although we may be able to find trends in particular functions, individual assessment will still be necessary in many situations if clinicians want to incorporate these trends into treatment modifications.

The abundance graphs at the beginning of each chapter (Figures 1, 8, and 17) also show that there is a gradual change in relative abundance from the most to least abundant MDMG. This may seem like an obvious result, but identifying that all MDMGs are not

found in similar proportions is important. The most reasonable approach for future research could be to focus on the most abundant MDMGs; however, even MDMGs with small relative abundances could be pharmacologically relevant.

Take K01195: beta-glucuronidase as an example. In the gut microbiome, K01195 represents less than 1% of sequences amongst all MDMGs (Figure 1 and 6). Yet this MDMG has very notable clinical implications in its ability to re-activate glucuronidated compounds in enterohepatic circulation (Chapter 1.5.2.1). Therefore, we can speculate that any MDMG in the gut microbiome with a sequence proportion similar to or above K01195 should be abundant enough to cause a noticeable biological effect if a drug or compound vulnerable to that MDMG was encountered by the gut microbiome. This is assuming that all else remains equivalent, which of course in not the case. There may be differences in how readily certain MDMGs are transcribed, or perhaps the beta-glucuronidases produced by the gut microbes are more efficient at metabolizing their targets than other MDMGs. This MDMG at least gives us a point of reference for the abundance required to cause some clinical impact. Functional studies are required to determine the level of raw/transcribed sequences required in order to produce a response on vulnerable drug to an extent that is biologically significant.

6.2 MICROBIAL DRUG METABOLIZING GENES ACROSS BODY SITES

As expected, the gut microbiome was unique from the other four body sites examined in relation to the relative abundance of MDMGs. Nearly all MDMGs were identified in the gut. Samples from the gastrointestinal tract are clustered away from the samples coming from the other four body sites (Figure 3). This reveals that the drug metabolism functions present in the gut microbiome are distinct from other areas in the body. Also, the GI samples are not tightly clustered themselves, but quite spaced out. This emphasizes the interpersonal differences in the abundances of drug metabolism functions across the population.

Our initial rational for examining drug metabolism functions across different body sites was that there might be some type of selection occurring. In theory, gut microbes with particular genes would respond better when they encountered particular compounds, for example, by being able to break them down and use them for their own cellular purpose, or by degrading toxic drug metabolites. As a result, those bacteria as well as those genes would increase in relative abundance. This explanation seems to fit the pattern for K05349: beta-glucosidase. The relative abundance is highest compared to all other MDMGs in all three of our datasets, and it is also far more abundant in the gut than the other body sites in the HMP dataset. It appears the gut bacteria are using this function to break down carbohydrates, possibly large carbohydrates that are indigestible by humans such as cellulose, for their own benefit.

6.3 MICROBIAL DRUG METABOLIZING GENES DO NOT VARY BY SEX

There have been documented differences in drug metabolism between females and males (O'Malley, Crooks, Duke, & Stevenson, 1971). Additionally, there has been some studies that suggest a difference in the gut microbiome between these two sexes, but other studies have found no significant differences between them (Markle et al., 2013; The Human Microbiome Project Consortium, 2012). However, within our research we found no significant differences in the abundance of MDMGs between females and males.

It could be hypothesized that a hormonal difference between the two groups might cause a shift in the bacteria present and also would influence the abundance of particular functions. The microbiome is known to shift during pregnancy (Koren et al., 2012). However, the datasets examined excluded females on many types of hormonal birth controls as well as those who were pregnant or recently gave birth. While this is useful for eliminating variables in an initial study, it does not represent the situation in the general population, as many women would fall in these categories. Further investigation with less exclusion criteria might demonstrate that in a less controlled subset of the population, there are noticeable differences in MDMGs between females and males.

6.4 AGE AND FRAILTY CORRELATE WITH MICROBIAL DRUG METABOLIZING GENE ABUNDANCE

It has been documented for several decades that aging may alter drug metabolism (O'Malley et al., 1971). Furthermore, it has been found that the microbiome shifts with age and frailty (M. G. Langille et al., 2014). This suggested that particular microbial genes may be altered in old or frail individuals, thereby impacting drug metabolism.

In the Chinese dataset and Northwood dataset, 36 and 39 MDMGs respectively were identified that had statistically significant correlations between their relative abundance and the age of the sampled individual. Only 12 of these overlap between the datasets (two of which showed opposite trends). In the Northwood dataset, 24 MDMGs were correlated with frailty index score.

K03743: nicotinamide-nucleotide amidase, and K01448: N-acetylmuramoyl-L-alanine amidase were both correlated with age. Amidase enzymes catalyse the hydrolysis of an amide and although the specific classifications of these MDMGs are different, they are both still amidases and therefore may both be able to target drugs with vulnerable amide groups. Although K03743 has a stronger positive correlation with age, the relative abundance of K01448 seems higher for the vast majority of samples. As a result, if a drug susceptible to metabolism via amidase was introduced to a person's gut microbiome, we would predict that an older individuals' microbiome would metabolise less of this drug compared to a younger individual. Whether this difference would have biologically significant impacts is not obvious, as many other factors would be involved in said drug's metabolism. These first MDMGs also show that the correlations between age and abundance are very weak. As a result, there is still a vast amount of unexplained variation that could be explored in more detail regarding the abundance trends of MDMGs.

Alcohol dehydrogenase (K00001) catalyses the oxidation of alcohol functional groups to aldehydes or ketones and was found to decrease in the gut microbiome with age. It was shown to be negatively correlated with age. As previously discussed in Chapter 1.5.1.7 the gut microbiome may increase risk of atherosclerosis and other cardiovascular conditions by metabolising dietary choline into TMAO and that the

compound DMB could be used to block microbial TMA lyase (Wang et al., 2015). However, Wang et al. also reported DMB to be rapidly metabolized by alcohol dehydrogenase. Based on our findings, DMB could possibly be metabolized by the gut microbiome resulting in differential efficacy of DMB to inhibit TMAO production. This inhibition could be less for older individuals as microbes with alcohol dehydrogenase decline, but large individual variability would require metagenomic profiling.

There are two other statistically significant alcohol dehydrogenase MDMGs correlated with age in the Chinese dataset (K04072 and K13953). Unlike K00001, these other two show positive correlations in abundance vs age. However, these two MDMGs were found to be some of the few positively associated with T2D (Qin et al., 2012). As a result, the upward trend observed in these other alcohol dehydrogenases might just be resulting from the number of T2D patients in this collection of samples. In order to make any definitive statements about these MDMGs in the general population, it would be best to analyse samples from disease-free individuals, or have the proportion of individuals sampled with T2D be representative of the incidence of T2D in the population.

K08325: NADP-dependent alcohol dehydrogenase was also a significant MDMG, except it was correlated with FI score. In contrast to K00001 which had a negative correlation with age, K08325 had a positive correlation with frailty. This example is another point of caution around the complexity of the microbiome. Using DMB treatment as an example, although the age studies suggest older individuals would need a lower dose of DMB since they have less alcohol dehydrogenase breaking it down, the frailty data suggests more frail individuals would need a higher dose.

K00600: SHMT had a relatively weak but significant positive correlation with age and was one of the most abundant MDMGs across datasets. The known function of SHMT is to transfer a hydroxymethyl group to tetrahydrofolate in order to form 5, 10-methylene-tetrahydrofolate (Renwick, Snell, & Baumann, 1998). This is an important biological function as this ‘one-carbon unit’ is a major source of carbon for certain cells. Moreover, increase in SHMT activity has been linked to cells rapidly proliferating, namely in the case of tumors (Locasale, 2013). Reducing or inhibiting the activity of SHMT has thus become an attractive option for treating cancer patients. The results

presented in this study suggest that the gut microbiome has the potential via microbial SHMT activity to interact with the growth rate of tumor cells, especially in nearby areas such as the gastrointestinal tract and the colon. No research has been done on the impact this microbial function has compared to human SHMT, but incorporating the microbiome into anti-cancer treatments could result in increased efficacy. Additionally, if selective targeting to microbial genes could be done and demonstrate efficacy, the side-effects that stem from targeting human cells in cancer therapy could be greatly reduced.

K01195: beta-glucuronidase was another significant MDMG that decreased in relative abundance as individuals got older. This function has documented biological relevance in the gut microbiome as it plays a role in enterohepatic circulation, and inhibitors have been developed to reduce their activity (Wallace et al., 2010). Consequently, middle-aged individuals may benefit more from these inhibitors as they, on average, had higher relative abundance of this function. However, there is significant variation throughout all ages, and so use of microbial beta-glucuronidase inhibitors might be utilized even in the most elderly patients depending on an individual's unique microbiome.

K01118: FMN-dependent NADH-azoreductase, had increased abundance with age in the Northwood dataset. This was also the only azoreductase function identified in both the Chinese and Northwood samples (although in the HMP samples, K03206: azobenzene reductase was detected). Azoreductase enzymes catalyse the breakdown of nitrogen double-bonds, also called azo bonds. Chapter 1.5.1.3 discussed the functional implications of azoreductase activity, in a beneficial sense for prodrugs like sulfasalazine, as well as in a negative context through its role in creating toxic metabolites from azo dyes. While situations like inhaling dyes should be avoided regardless of a person's microbial composition, the abundance of this function in the microbiome could affect dose and response in elderly individuals. If testing for azo bond prodrugs was done on younger individuals, there might be a marked increase of the active form of the drug in older patients.

One very interesting MDMG found to be positively associated with frailty was K01424: L-asparaginase. This MDMG was negatively correlated with age. It is intriguing

to see an example of age suggesting functionality in one direction, while frailty suggests functionality in the opposite direction. K01424 has very important biological significance, as it is used to break down circulating asparagine into aspartic acid. As a result, it can be used as a drug so certain cancer cells like acute lymphoblastic leukemia cells that cannot make their own asparagine, are depleted of this amino acid which leads to cell death (Broome, 1981). This gene in the microbiome may therefore have some protective effect when it comes to more frail individuals.

K02654: prepilin peptidase had the strongest correlation based on FI score. This MDMG was interesting for several reasons. First, this enzyme is found on the surface of bacteria with its primary function is to cleave specific regions of exported proteins. This function in itself may play a role in how the gut microbiome can impact drugs, depending on what proteins are released and how cleaving these proteins affects their functions (Dalbey, 1991). In addition to this, it also has documented function as an N-methyltransferase. Since its action is nonspecific, this MDMG might, for example, work as an amine N-methyltransferase. As a result, an amine drug could be methylated by this enzyme, subsequently affecting the function of that drug.

In Northwood, the individuals were aged 65 to 99 whereas in the Chinese samples, individuals ranged in age from 13 to 86 years old. This may indicate that significant functions in the Chinese dataset are ones that shift throughout life, whereas MDMGs in the Northwood dataset shift in the later stages of life. Additionally, Northwood individuals all lived at the same location. People living in assisted care facilities could be under environmental factors that change the microbiome besides age. In regards to frailty, although there was a wide range in FI scores, even the lowest scoring individual was still reasonably frail. Further exploration into populations with a larger separation in frailty might bring forth additional MDMGs not identified in the Northwood dataset.

There are many possible explanations as to why there are shifts in the abundance of MDMGs as people age or get frailer. It has been documented that the microbiome overall shifts in composition over longer time periods. With that, changes in the functions present are likely to occur. This may or may not be related to lifestyle shifts as people get

older. A person's body as a whole slows down in terms of functional capabilities. Since the microbiome is closely linked to our own health, this may drive some of the changes in functions present. Furthermore, elderly people may be on more drugs or treatments in order to help maintain good health. It is possible that these medical practices are altering the microbiome and influencing the abundance of particular functions. Old and frail individuals may have changes in diet, exercise, or living situations, which would alter the environment in the gut and therefore change the microbes present.

6.5 MEDICATION USE IMPACTS MICROBIAL DRUG METABOLIZING GENE ABUNDANCE

Many studies have examined how drugs (mostly antibiotics) alter the general composition of the microbiome (Francino, 2016). Other drugs related to the gastrointestinal tract have been shown to affect the genes present in the microbiome (Zhernakova et al., 2016). We investigated the impact particular drugs related to the GI tract had on the abundance of drug metabolizing genes in the microbiome. The abundance of several MDMGs were found to be significantly associated with the use of these drugs.

K12251: N-carbamoylputrescine amidase, was one of the five MDMGs that showed a lower average relative abundance in the microbiome-altering drug group. We have previously discussed how this function may be able to metabolize drugs with amide groups. The other MDMG examined was K00799: glutathione S-transferase, which was one of the 35 MDMGs with a higher average relative abundance in the microbiome-altering drug group. The mammalian version of this gene functions by detoxifying electrophilic xenobiotics, and this reduction of this function is linked to susceptibility to carcinogens and disease states (Hayes, Flanagan, & Jowsey, 2005). If the microbial version of this function remains the same, microbes in the gut would be able to reduce toxic impact of particular compounds, thus benefiting the human host.

It is important to remember that all the microbiome-altering drugs were grouped together in this study. This was done to provide statistical power to determine if certain drugs were correlated to changes in drug metabolism functions, given the limited number of patients compared to the vast amount of different drugs. If a large enough group could be analysed with more people on one particular treatment and less overlap of drugs, functional changes in the microbiome might be more accurately assessed.

6.6 LIMITATIONS

6.6.1 SELECTION OF MICROBIAL DRUG METABOLIZING GENES

In this study, our assessment of the functional capabilities of the microbiome is determined by our selection of KOs (that we label as MDMGs). While this approach is useful for getting a general understanding of drug metabolism potential, there are also obvious setbacks.

The first issue appears when looking at the categories of functions these KOs fall into. Throughout the results section, the number of MDMGs showing significant differences or trends based on the variable being assessed was stated. These numbers should not be taken to directly represent how important a particular variable is in terms of drug metabolism functional abundance. Each particular function may not be equally important when it comes to drug metabolism. For example, there are only a handful of beta-glucuronidase MDMGs but over 100 methyltransferase MDMGs. There has been direct documentation of microbial beta-glucuronidase interfering with drug metabolism (see Chapter 1.5.2.1), but as of yet there are no case studies directly tying microbial methyltransferase to drug modifications. The actual number of abundant MDMGs correlated with individual variables should not be ignored, but these numbers alone cannot be used to determine what variables are ‘most important’ in terms of microbiome-drug interactions.

Along that same line, there is a bit of caution surrounding the methyltransferase KOs. To be inclusive, every methyltransferase KO that had evidence of being found in bacteria was included in our drug metabolism list. Some of these KOs were given general names, but there are also many ones like K00590: site-specific DNA-methyltransferase. The name of KOs such as this indicate they are known to act on specific targets. Since it is possible that the genes responsible for this function may also target other xenobiotic compounds, we chose to leave them in the dataset. However, based on classification alone, some KOs may have more broad applications and as such are more important in terms of significant drug metabolism functions.

There is an additional issue with using KOs to represent functional capacity of the microbiome. We only looked at functions that are known to play a role in drug metabolism in humans and that were also found in bacteria. However, the microbes in our gut have 100 times the genes of humans. Some, perhaps many of these genes play a role in xenobiotic metabolism. These genes may have human homologs in the KEGG database, but if there was no documentation of these genes found in bacteria they were excluded from our study. Therefore, our current findings are just giving a general overview of the drug metabolism potential. With increased classification of microbial genes we will be able to update this overview into a more complete picture.

6.6.2 ANNOTATION

This study is also relies on particular sequences being annotated to functions in the KEGG database, which creates some limitations. First of all, HUMAnN cannot annotate a sequence as a particular KO if that function does not have an associated KO. This seems like an obvious statement but an important one to keep in mind. Rare or unclassified functions without a KO would be invisible in our current approach. Genes like the cgr genes found in *E. lenta* are not represented in the KEGG and would therefore remain unexamined in this analysis. Only the functions present in a reference database can be studied, but this represents just a portion of the many functions present in the microbiome.

An example of this problem is found when discussing CYP enzymes. As discussed in Chapter 1.2, these enzymes play a critical role in human drug metabolism in the liver. Yet, these enzymes were not well represented when searching the KEGG database for metabolism functions found in microbes. Does this mean that the microbiome lacks CYP450s? Not necessarily. Microbial CYP enzymes have been researched and identified, just not to the extent of their human counterparts (McLean et al., 2006). In the millions of genes in the microbiome, it is very possible that some of these genes carry out functions similar to human CYP450s. However, with these functions largely absent in the KEGG database, there is no way of annotating sequences to these functions, even if those sequences do encode for microbial CYP-like enzymes.

This limitation also holds true for PICRUSt predictions. PICRUSt can only predict functions if they are actually found in the KEGG database. Additionally, PICRUSt relies on an additional database to relate taxa with functions. This leaves the predictions vulnerable to missing rare and unclassified functions as previously stated, but also opens up slightly more room for error. If a function is newly acquired by a particular taxa, say via LGT, the reference database would not be able to associate that function with that taxa. Hence in our analysis this function would be missing. The idea of studying metagenomic samples without using a reference genome has been investigated, but it has not advanced to the forefront of the field (Nielsen et al., 2014).

There is also an issue with the annotation process in general. Many different genes man be grouped together under one function. These genes may all carry out the same general function, but there may be detailed factors that distinguish these genes from one-another such as the specific structure of compounds they are able to act on. As a result we cannot claim with certainty that all microbiomes containing a particular MDMG will be able to modify a drug in the same way.

6.6.3 BIOLOGICAL IMPACT OF FUNCTIONS

The objective of this study was to survey drug metabolism functions across the microbiome. This is useful for getting an overview of what is happening in the microbiome, but is lacking in terms of clinical significance. We have identified that the abundance of certain MDMGs change based on several variables. It is important to remember that we are just looking at DNA sequences. This does not provide us with any information about how quickly or in what situations these gene would be transcribed and translated. The level of mRNA and protein production are not necessarily correlated (Greenbaum, Colangelo, Williams, & Gerstein, 2003). Even if they were all translated into proteins at the same rate, we cannot say for sure that the amount translated would cause clinically significant effects. Further biological work must be done in order to determine if these differences and correlations actually have an impact on bioavailability of drugs in patients.

6.6.4. RELATIVE ABUNDANCE

Another issue related to interpretation of the results is relative abundance. Relative abundance is a useful way of looking at data to identify particular trends. The problem with using relative abundance is that it may hide some important information. Particular studies have commented on this issue and proposed new statistical methods to examine microbial samples, but these methods are still in development (Olesen et al., 2016).

As an example, consider metagenomic samples from different body sites. The gut microbiome contains a MDMG with a very high relative proportion across all samples (K05349). As a result, every other MDMG in the gut will make up a smaller proportion of the drug metabolism sequences. Due to this, other MDMGs might have smaller relative abundances in the gut compared to their relative abundance in the other body sites which do not have a MDMG making up such a large proportion of the sequences.

We can speculate that for some (possibly many) of these functions, it is not the case that they are being downregulated in the gut microbiome, but their proportion is being overshadowed by the highly abundant K05349.

Furthermore, we are looking at relative abundance of these functions in a sample, not the raw number of sequences found for each MDMG. Therefore, we cannot say for sure if the MDMGs found with lower relative abundances in the gut actually have lower number of sequences compared to the other body sites, or if some MDMGs are more abundant in the gut, but appear small due to MDMGs with an even higher amount of sequences present. This limits our ability to speculate biological consequences between groups based on the relative abundance of MDMGs.

This is likely less of an issue comparing similar environments like the gut across multiple people. However, the issue is still present. Take K00600: SHMT as an example, whose abundance was positively correlated with age. It can be said that overall, SHMT makes up a larger proportion of drug metabolism functions in these individuals as they get older. However, the actual number of genes in the microbiome may decrease with age. As a result even though older individuals may have a higher relative proportion of these genes, the raw number of sequences could be lower.

Relative abundance is a simple and effective way to compare certain aspects of data. However, to translate these findings into therapeutics, further analysis would require quantitative levels of these functions in the microbiome.

6.6.5 CROSS-COMPARABILITY

Each of the datasets utilized in this study provided important information regarding the individuals sampled. Unfortunately, one major limitation is the inability to directly compare findings across different datasets. This is not just an issue with this study but with metagenomics in general. Ideally, all methods of analysis would yield the same results but rarely does this occur. Slight differences in sequencing procedures may create particular biases. Most of the times if you were to plot two different datasets on a

PCA plot, they would cluster primarily based on datasets regardless of any other variable being assessed.

This problem is made even more difficult as individuals between studies grow more separate. Different environments, diets, and genetic backgrounds will introduce variation between datasets that overpowers any other factor. Additionally, different annotation methods (such as HUMAnN vs PICRUSt) make comparisons across datasets even less appropriate. HUMAnN is based off metagenomic data whereas PICRUSt, although accurate, are still predictions based off 16S sequences. Therefore, trends within our datasets can be examined, but directly comparing MDMG abundance between the datasets cannot be justified.

6.6.6 GLOBAL EXTRAPOLATION

Our findings in this study apply to the specific population examined. The question then becomes: how wide do these trends extend? Take the Northwood samples as an example. Would the same trends exist if a population outside of Northwood was examined? What about throughout all of Nova Scotia, or across all Canada? Geographical location plays a major role in shaping the microbiome, with changes in diet being likely responsible for a large portion of the changes (David et al., 2013). For any of these trends to translate into clinical significance, identifying what population(s) they are consistently found in is critical. At the point where these trends disappear, determining what specific factors caused the changes in the microbiome may also provide valuable insight.

6.6.7 DRUG METABOLISM DATABASE

In this study, many different functions present in the microbiome that could potentially play a role in drug metabolism were identified. Unfortunately, even with

databases containing thousands of functions, there is no well-established database surrounding drug metabolism. Currently, one of the only ways to determine how a drug is metabolized is to look up the individual drug's profile and look to see if the metabolism pathway is listed in its pharmacokinetics. For some drugs, the exact method of their elimination is unknown. With the remaining drugs, we had hoped that there would be some collection put together detailing, at least in a broad sense, how drugs our primarily metabolized. With this information we could look at the MDMGs identified in our samples, and quickly establish a shortlist of drugs that may be vulnerable to metabolism by the microbiome. Since such a system does not currently exists, we are left with an extremely inefficient process of identifying drugs that interact with the microbiome.

6.7 FUTURE WORK

6.7.1 DRUG IDENTIFICATION

With several metabolism functions identified in the microbiome, the next step would be to identify drugs whose metabolism are affected by the gut microbes. As mentioned in the previous chapter, there is no well-established database classifying drugs based on how they are metabolized. As a result, individual drugs could be researched on known pathways that cause their metabolism. If that function is one documented in the microbiome, it would be a candidate for testing. This process could be tedious.

A more efficient way of performing the task would be large scale screening. Since we have only got a glimpse of the drug metabolism potential of the microbiome, it is likely that many other functions currently unknown in the microbiome can modify drugs. Performing large-scale screening experiments would illuminate what functions appear to be most relevant in modifying drugs.

There are many different approaches to performing these large scale screenings. One could start by introducing as many drugs as possible through two mouse groups, a control normal mouse and a germ free group. The blood, urine, and/or feces could be

screened for metabolites in order to assess the impact of the gut microbiome in modifying drugs based on the differences between the two groups.

This would be a relatively fast and efficient initial comparison, but there are a few problems with this approach. First, the metabolites of the drug being studied must be known or there must be some way of detecting it. If the microbes are breaking down the drug in a new way creating different metabolites, they may be missed.

One way around this issue would be looking at the actual physiological impact of drugs. This would only be measurable for certain drugs. If testing a drug that for example lowers blood pressure, you could compare the reduction in blood pressure between the two groups. If there are significant changes between the groups, microbial metabolism is likely the cause. Unfortunately, germ free mice are developmentally different than normal mice and that difference may alter the internal drug metabolism potential of the mice. To resolve this issue, the two groups could be germ free mice, and initially germ free mice colonized by human gut microbes. This process would be a little more laborious, but would reduce the amount of noise from physiological differences between the two groups while also giving a closer representation of what is the human microbiome is capable of.

6.7.2 FUNCTIONAL ANALYSIS

Identifying MDMGs, and finding changes in the abundance of these functions is important in understanding the microbiome. However, simply pointing out differences is not enough. From a clinical perspective, we must be able to determine if changes in these MDMGs produce biologically significant impacts.

In regards to functional significance, there are two general situations to consider. The easier situation is a comparison between individuals that have a particular metabolism function in their microbiome versus people without that function. This is similar to the case of Digoxin, where individuals who had a particular microbial function breaking down had serum digoxin levels reduced to as low as half (Haiser et al., 2013). If

we are able to identify microbial functions such as this, that break down a particular drug but is not found in all individuals, then the clinical relevance of this function should be easily assessed.

The second situation is much more complicated. With many of the MDMGs identified, there are differences or trends in the abundance of these functions. The question we then need to investigate is if these differences in abundance cause biologically noticeable changes.

One step that could be performed would be using metatranscriptomics or metaproteomics to get a better functional representation of what is happening. These processes are more complicated than metagenomics, and also have issues associated with them. With metatranscriptomics, the turnover rate of mRNA might mean that the functions assessed are very time sensitive and may not represent the microbiome as a whole. Metaproteomics is also troublesome as identifying and classifying proteins is more complicated that looking at DNA sequences. However, with improved mass spectrometry methods, this method has recently been used to identify microbial proteins (Grassl et al., 2016).

Another form of functional analysis is simply testing drugs on microbial communities with known amounts of particular MDMGs. Using gnotobiotic mice, you can set up several 'concentrations' of functions in their microbiomes. It would then be possible see if the differences in functional abundances cause significant difference in physiological response or the metabolites produced.

6.7.3 KEY MICROBES

Throughout or study we focused on functional capabilities of the microbiome rather than studying particular taxa or OTUs. Many different research groups have found that even with changes in diversity and composition in the microbiome, the abundance of general functional categories remain quite stable (Zhernakova et al., 2016). For an initial

survey we decided that looking at the abundance of functions as a whole would give us more insight than looking at OTU composition.

This does not mean individual taxa are not important in shaping the drug metabolism potential of the microbiome. As with Digoxin in Chapter 1.5.1.1, one species, *E. lenta,* was responsible for carrying the genes that metabolised the drug. This may be a very rare and unique case, but we cannot rule out the possibility that other species of microbes have unique genes that have not been shared via LGT. If such microbes are identified, it will make clinical assessment of drug metabolism far easier, as we would simply need to identify particular taxa to predict a drug will be altered by the microbiome.

There is also a case to be made for conditionally rare taxa (CRT) (Shade et al., 2014). As the name suggests, these taxa are normally in very low abundance compared to the number of other microbes. However, occasionally CRT will bloom to an abundance orders of magnitude larger than their baseline. If these CRT carry unique, or even common genes related to drug metabolism, their peaks in abundance could alter how the microbiome processes drugs as a whole. These taxa would become even more relevant if the conditions that dictate their blooming phase were related to particular disease conditions or treatments.

If key microbes were suspected of playing a role in drug metabolism, drug testing could be done on them. If culturable, these microbes could be grown, introduced to a drug, and analysed for the extent of drug metabolism occurring. If taxa were not culturable but the specific gene(s) thought to be related to drug metabolism were known, these genes cloned and tested into a model organism such as *E. coli.* This would not be a perfect representation of what's occurring in the microbiome, as interactions with other microbes and human cells would likely impact the taxa being studied. However, it might provide some early estimates regarding how effective that taxa is in modifying a particular drug.

6.7.4 DRUG USE VS TIME PROFILE

Our studies exploring the impacts drugs had on the microbiome yielded some interesting results. However, many areas of this study could be explored in greater detail. Aside from having a larger patient group, and more patients on the same drug treatment, a profile of drug use over time would be a critical investigation. This type of analysis could be done for drugs already on the market, as well as drugs in ongoing clinical trials. The Northwood study looked at medication use retrospectively, and therefore had no information on patients before starting particular drug treatments. The problem with this is that an individual might have had a very distinct microbiome before treatment began. If an individual was examined only after drug treatment and compared them to the rest of the population, it cannot be determined whether the changes are a result of the drugs, or just baseline differences in the individual. Since the Northwood study was only exploratory, this type of information was not a primary concern. In the future, a study analyzing how microbial functions shift throughout a drug treatment would provide a more precise insight on how the microbiome adapts to medication use.

6.7.5 ADDITIONAL FACTORS INFLUENCING FUNCTIONAL ABUNDANCE

In this study, we explored many factors and their connection to changes in the abundance of drug metabolism functions. As the results from this investigation showed, many of these factors were significant but only explained a small proportion of the variation observed. Further studies done in the same fashion could help tease out additional variation.

It would be interesting to examine many different disease states, to determine if the functional capabilities of the microbiome is altered compared to healthy individuals. It would also be interesting to see how the microbiome shifts as the disease progresses. Many diseases such as autoimmune disease, cardiovascular disease, neurological diseases, and cancer have many contributing factors associated with them, the

microbiome being one. Changes in the drug metabolism functions in the microbiome as a result of these diseases may influence the effectiveness of treatment.

The pediatric population is another area of interest. There has been much debated surrounding the microbiome of babies born naturally versus caesarean section (Neu & Rushing, 2011). Some researcher have even tried swabbing C-section babies with their mother's vaginal microbes in order to establish a more 'natural' microbial community (Dominguez-Bello et al., 2016). The long-term differences in these two birthing methods are still unclear. In the short-term, it would be intriguing to see if these methods cause differences in the genes, particularly drug metabolism functions, we observe in these children.

Antibiotics are another major category. In the datasets we examined, all patients had been free of antibiotics for several months as they cause chaos in the microbial environment of the gut. For initial investigations, avoiding this chaos is helpful as we are able to investigate drug metabolism functions in a more stable environment. However, antibiotic use in the real world is not something that is easily avoided. It will be important to investigate how antibiotic use alters the functional capabilities of the microbiome as well.

These studies should not be restricted to the gut microbiome either. Although this environment is the one most often studied, other areas of the human body contain microbes that may play a role in drug metabolism. Take the lung microbiome for example. Respiratory conditions such as asthma are often treated by drugs in inhalers. The microbes in the lungs may contain functions that can modify these drugs before they reach their target. Classifying the impact of these functions would allow for better dosing of patients using inhalers.

6.7.6 PROBIOTICS

Probiotics have been used for many years with claims of improved health and a 'better' gut bacteria. As our knowledge of the microbiome grows, the role probiotics can

play in human health will be better understood. A potential avenue for probiotics in the future is their use to complement a drug treatment. These probiotics may contain microbes enriched with particular functions depending on the treatment. These functions could help metabolize slowly-absorbed drugs so toxic accumulation does not occur in the GI tract. They could also be used in combination with prodrugs, to improve the rate of conversion to the active compound thus increasing bioavailability. This type of implementation into medicine is probably many years away, but it is still important to consider actively manipulating the microbiome as a treatment strategy.

6.7.7 DRUG-MICROBIOME INTERACTION DATABASE

The results presented throughout this study show a first-pass exploration of the drug metabolism potential of the human microbiome. With continued research in this field comes a better understanding of how the microbes in our bodies interact with drug treatments. This research could contribute to the establishment of a database of drugs vulnerable to metabolism by the microbiome, and particular microbial genes responsible for said metabolism. Ideally, this information would be gathered together in a comprehensive drug-microbiome interaction database. In a similar fashion to how human genes are screened to find genetic predispositions that may alter treatment, a person's microbiome may also be sequenced. These sequences could be screened through the microbiome database to identify any potential treatments that may be influenced by the microbiome. From there, the treatment could be altered by dosing adjustments or use of different drugs to ensure maximum therapeutic benefit.

This database would also be useful in the process of drug development. If early pharmacokinetic studies identify that a candidate drug is metabolized primarily by a particular function, that function can be researched in the database to determine how abundant it is in the microbiome, both in general and specific populations. During clinical trials, the microbiome of participants can be sequenced and screened through the database. Based off drugs of similar structure or specific metabolic pathways, it could be

possible to determine if microbial metabolism may be linked increased or decreased bioavailability of the drug.

This could also extend beyond in-vivo analysis. Programs like Simcyp use previously generated experimental data to simulate the processes of absorption, distribution, metabolism and excretion for a compound in a given population (Jamei et al., 2009). By incorporating a drug-microbiome database into Simcyp, this program would be able to identify particular microbial functions that would impact a certain drug. As a result, it would be able to more accurately simulate how a drug is likely to move through the body in a given population.

Although this database would be useful in drug treatment modifications, the incorporation of the microbiome into personalized medicine shares similar challenges as human genetic information. Some of these problems include confidentiality of microbial makeup, discrimination by physicians and insurance companies, reduced access to particular treatment based on microbial gene abundance, and the rate at which research can be translated into the clinic (Joly, Saulnier, Osien, & Knoppers, 2014). There is no easy solution to these problems. It is therefore a necessity that regulatory agencies such as the FDA become more involved in microbiome-drug interactions to insure that this information is able to benefit patients in an effective and ethical manner.

6.8 CONCLUSIONS

The human microbiome is a fascinating and complex system. The functional composition of the microbiome varies from person to person, with many factors contributing to its uniqueness. In this study, we have explored several of these factors and their connection to drug metabolism functions in the microbiome. This process has illuminated many potential pathways the microbiome has at its disposal to alter drugs. Importantly, this is only an initial exploration. There is still much work to do in order to translate the functions present in the microbiome into useful clinical applications. With increased exploration of the capabilities of the microbiome, we move closer and closer to its incorporation into personalized medical treatment.

REFERENCES

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science Translational Medicine*, *6*(237), 237ra65. doi:10.1126/scitranslmed.3008599

Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., … Huttenhower, C. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology*, *8*(6), e1002358. doi:10.1371/journal.pcbi.1002358

Akaza, H., Miyanaga, N., Takashima, N., Naito, S., Hirao, Y., Tsukamoto, T., … Pantuck, A. J. (2004). Comparisons of percent equol producers between prostate cancer patients and controls: case-controlled studies of isoflavones in Japanese, Korean and American residents. *Japanese Journal of Clinical Oncology*, *34*(2), 86–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15067102

Amon, R. M. W., & Benner, R. (1996). Bacterial utilization of different size classes of dissolved organic matter. *Limnology and Oceanography*, *41*(1), 41–51. doi:10.4319/lo.1996.41.1.0041

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*, 32–46. doi:10.1111/j.1442-9993.2001.tb00081.x

Angert, E. R. (2005). Alternatives to binary fission in bacteria. *Nature Reviews Microbiology*, *3*(3), 214–224. doi:10.1038/nrmicro1096

Bäckhed, F., Manchester, J. K., Semenkovich, C. F., & Gordon, J. I. (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(3), 979–84. doi:10.1073/pnas.0605374104

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G., & Neufeld, J. D. (2011). Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Applied and Environmental Microbiology*, *77*(11), 3846–3852. doi:10.1128/AEM.02772-10

Bates, S. (2010). Progress towards personalized medicine. *Drug Discovery Today*, *15*(3-4), 115–120. doi:10.1016/j.drudis.2009.11.001

Baumann, P., Moran, N. A., & Baumann, L. (1997). The Evolution and Genetics of Aphid Endosymbionts. *BioScience*, *47*(1), 12–20. doi:10.2307/1313002

Baumgart, D. C., & Sandborn, W. J. (2012). Crohn's disease. *Lancet (London, England)*, *380*(9853), 1590–605. doi:10.1016/S0140-6736(12)60026-9

Björkholm, B., Bok, C. M., Lundin, A., Rafter, J., Hibberd, M. L., & Pettersson, S. (2009). Intestinal Microbiota Regulate Xenobiotic Metabolism in the Liver. *PLoS ONE*, *4*(9), e6958. doi:10.1371/journal.pone.0006958

Blaser, M. J. (2010). Harnessing the power of the human microbiome. *Proceedings of the National Academy of Sciences*, *107*(14), 6125–6126. doi:10.1073/pnas.1002112107

Blaser, M. J. (2014). Missing Microbes: How the Overuse of Antibiotics Is Fueling Our Modern Plagues. *Henry Holt and Company LLC, New York*, *35*(9), 261. doi:10.3201/eid2011.141052

Blaser, M. J., Chen, Y., & Reibman, J. (2008). Does Helicobacter pylori protect against asthma and allergy? *Gut*, *57*(5), 561–567. doi:10.1136/gut.2007.133462

Blumenthal, T. (2004). Operons in eukaryotes. *Briefings in Functional Genomics & Proteomics*, *3*(3), 199–211. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15642184

Bott, T. L., & Brock, T. D. (1969). Bacterial Growth Rates above 90 C in Yellowstone Hot Springs. *Science*, *164*(3886), 1411–1412. doi:10.1126/science.164.3886.1411

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. doi:10.1038/nmeth.3176

Burke, W., & Psaty, B. M. (2007). Personalized Medicine in the Era of Genomics. *JAMA*, *298*(14), 1682. doi:10.1001/jama.298.14.1682

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–6. doi:10.1038/nmeth.f.303

Carabotti, M., Scirocco, A., Maselli, M. A., & Severi, C. (2015). The gut-brain axis: Interactions between enteric microbiota, central and enteric nervous systems. *Annals of Gastroenterology*, *28*(2), 203–209.

Cardno, A. G., & Owen, M. J. (2014). Genetic Relationships Between Schizophrenia, Bipolar Disorder, and Schizoaffective Disorder. *Schizophrenia Bulletin*, *40*(3), 504–515. doi:10.1093/schbul/sbu016

Chen, G.-B., Lee, S. H., Brion, M.-J. A., Montgomery, G. W., Wray, N. R., Radford-Smith, G. L., & Visscher, P. M. (2014). Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Human Molecular Genetics*, *23*(17), 4710–4720. doi:10.1093/hmg/ddu174

Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R., & Nicholson, J. K. (2009). Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(34), 14728–33. doi:10.1073/pnas.0904489106

Comeau, A., Langille, M., & Douglas, G. (2016). Microbiome Helper. Retrieved from https://github.com/mlangill/microbiome_helper/wiki

Conly, J. M., & Stein, K. (1992). The production of menaquinones (vitamin K2) by intestinal bacteria and their role in maintaining coagulation homeostasis. *Progress in Food & Nutrition Science*, *16*(4), 307–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1492156

D'Argenio, V., & Salvatore, F. (2015). The role of the gut microbiome in the healthy adult status. *Clinica Chimica Acta*, *451*, 97–102. doi:10.1016/j.cca.2015.01.003

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., … Turnbaugh, P. J. (2013). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. doi:10.1038/nature12820

DeLeon-Rodriguez, N., Lathem, T. L., Rodriguez-R, L. M., Barazesh, J. M., Anderson, B. E., Beyersdorf, A. J., … Konstantinidis, K. T. (2013). Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Sciences*, *110*(7), 2575–2580. doi:10.1073/pnas.1212089110

den Hoed, M., Strawbridge, R. J., Almgren, P., Gustafsson, S., Axelsson, T., Engström, G., … Lind, L. (2015). GWAS-identified loci for coronary heart disease are associated with intima-media thickness and plaque presence at the carotid artery bulb. *Atherosclerosis*, *239*(2), 304–310. doi:10.1016/j.atherosclerosis.2015.01.032

Díaz, E. (2004). Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *International Microbiology : The Official Journal of the Spanish Society for Microbiology*, *7*(3), 173–80. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15492931

Dillman, R. O. (2011). Cancer immunotherapy. *Cancer Biotherapy & Radiopharmaceuticals*, *26*(1), 1–64. doi:10.1089/cbr.2010.0902

Dixon, R., & Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology*, *2*(8), 621–631. doi:10.1038/nrmicro954

Dominguez-Bello, M. G., De Jesus-Laboy, K. M., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., … Clemente, J. C. (2016). Partial Restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature Medicine*, *22*(3), 250–253. doi:10.1038/nm.4039

Donlan, R. M. (2002). Biofilms: Microbial Life on Surfaces. *Emerging Infectious Diseases*, *8*(9), 881–890. doi:10.3201/eid0809.020063

ElRakaiby, M., Dutilh, B. E., Rizkallah, M. R., Boleij, A., Cole, J. N., & Aziz, R. K. (2014). Pharmacomicrobiomics: the impact of human microbiome variations on systems pharmacology and personalized therapeutics. *Omics : A Journal of Integrative Biology*, *18*(7), 402–14. doi:10.1089/omi.2014.0018

Emerson, D., & Moyer, C. L. (2002). Neutrophilic Fe-Oxidizing Bacteria Are Abundant at the Loihi Seamount Hydrothermal Vents and Play a Major Role in Fe Oxide Deposition. *Applied and Environmental Microbiology*, *68*(6), 3085–3093. doi:10.1128/AEM.68.6.3085-3093.2002

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., … Gordon, J. I. (2013). The Long-Term Stability of the Human Gut Microbiota. *Science*, *341*(6141), 1237439–1237439. doi:10.1126/science.1237439

Feng, R., Shou, J.-W., Zhao, Z.-X., He, C.-Y., Ma, C., Huang, M., … Jiang, J.-D. (2015). Transforming berberine into its intestine-absorbable form by the gut microbiota. *Scientific Reports*, *5*, 12155. doi:10.1038/srep12155

Fernald, G. H., Capriotti, E., Daneshjou, R., Karczewski, K. J., & Altman, R. B. (2011). Bioinformatics challenges for personalized medicine. *Bioinformatics*. doi:10.1093/bioinformatics/btr295

Flotte, T. R., & Carter, B. J. (1995). Adeno-associated virus vectors for gene therapy. *Gene Therapy*, *2*(6), 357–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7584109

Francino, M. P. (2016). Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances. *Frontiers in Microbiology*. doi:10.3389/fmicb.2015.01543

Gamage, N., Barnett, A., Hempel, N., Duggleby, R. G., Windmill, K. F., Martin, J. L., & McManus, M. E. (2006). Human sulfotransferases and their role in chemical metabolism. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, *90*(1), 5–22. doi:10.1093/toxsci/kfj061

Gheorghiade, M. (2006). Contemporary Use of Digoxin in the Management of Cardiovascular Disorders. *Circulation*, *113*(21), 2556–2564. doi:10.1161/CIRCULATIONAHA.105.560110

Gibney, M. J., & Walsh, M. C. (2013). The future direction of personalised nutrition: my diet, my phenotype, my genes. *The Proceedings of the Nutrition Society*, *72*(2), 219–25. doi:10.1017/S0029665112003436

Global Burden of Disease Study 2013 Collaborators. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)*, *386*(9995), 743–800. doi:10.1016/S0140-6736(15)60692-4

Goldin, B. R., Peppercorn, M. A., & Goldman, P. (1973). Contributions of host and intestinal microflora in the metabolism of L-dopa by the rat. *The Journal of Pharmacology and Experimental Therapeutics*, *186*(1), 160–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4723308

Goldman, P. (1978). Biochemical pharmacology of the intestinal flora. *Annual Review of Pharmacology and Toxicology*, *18*(5), 523–39. doi:10.1146/annurev.pa.18.040178.002515

Goodman, A. L., Kallstrom, G., Faith, J. J., Reyes, A., Moore, A., Dantas, G., & Gordon, J. I. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proceedings of the National Academy of Sciences*, *108*(15), 6252–6257. doi:10.1073/pnas.1102938108

Goole, J., & Amighi, K. (2009). Levodopa delivery systems for the treatment of Parkinson's disease: an overview. *International Journal of Pharmaceutics*, *380*(1-2), 1–15. doi:10.1016/j.ijpharm.2009.07.026

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F., & Veldhuyzen van Zanten, S. J. O. (2006). Differences between Tissue-Associated Intestinal Microfloras of Patients with Crohn's Disease and Ulcerative Colitis. *Journal of Clinical Microbiology*, *44*(11), 4136–4141. doi:10.1128/JCM.01004-06

Grassl, N., Kulak, N. A., Pichler, G., Geyer, P. E., Jung, J., Schubert, S., … Mann, M. (2016). Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Medicine*, *8*(1), 44. doi:10.1186/s13073-016-0293-0

Greenbaum, D., Colangelo, C., Williams, K., & Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, *4*(9), 117. doi:10.1186/gb-2003-4-9-117

Grover, M., & Kashyap, P. C. (2014). Germ-free mice as a model to study effect of gut microbiota on host physiology. *Neurogastroenterology & Motility*, *26*(6), 745–748. doi:10.1111/nmo.12366

Haffen, E., Paintaud, G., Berard, M., Masuyer, C., Bechtel, Y., & Bechtel, P. R. (2000). On the assessment of drug metabolism by assays of codeine and its main metabolites. *Therapeutic Drug Monitoring*, *22*(3), 258–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10850391

Haft, D. H., Selengut, J. D., & White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Research*. doi:10.1093/nar/gkg128

Haiser, H. J., Gootenberg, D. B., Chatman, K., Sirasani, G., Balskus, E. P., & Turnbaugh, P. J. (2013). Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. *Science (New York, N.Y.)*, *341*(6143), 295–8. doi:10.1126/science.1235872

Haiser, H. J., Seim, K. L., Balskus, E. P., & Turnbaugh, P. J. (2014). Mechanistic insight into digoxin inactivation by Eggerthella lenta augments our understanding of its pharmacokinetics. *Gut Microbes*, *5*(2), 233–238. doi:10.4161/gmic.27915

Haiser, H. J., & Turnbaugh, P. J. (2013). Developing a metagenomic view of xenobiotic metabolism. *Pharmacological Research : The Official Journal of the Italian Pharmacological Society*, *69*(1), 21–31. doi:10.1016/j.phrs.2012.07.009

Hanley, M. J., Abernethy, D. R., & Greenblatt, D. J. (2010). Effect of obesity on the pharmacokinetics of drugs in humans. *Clinical Pharmacokinetics*, *49*(2), 71–87. doi:10.2165/11318100-000000000-00000

Hayes, J. D., Flanagan, J. U., & Jowsey, I. R. (2005). Glutathione transferases. *Annual Review of Pharmacology and Toxicology*, *45*, 51–88. doi:10.1146/annurev.pharmtox.45.120403.095857

He, K., Iyer, K. R., Hayes, R. N., Sinz, M. W., Woolf, T. F., & Hollenberg, P. F. (1998). Inactivation of cytochrome P450 3A4 by bergamottin, a component of grapefruit juice. *Chemical Research in Toxicology*, *11*(4), 252–9. doi:10.1021/tx970192k

Hellweger, F. L., van Sebille, E., & Fredrick, N. D. (2014). Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science*, *345*(6202), 1346–1349. doi:10.1126/science.1254421

Hogan, C. (2011). Bacteria. In *Encyclopedia of Earth*.

Holmes, E., Loo, R. L., Stamler, J., Bictash, M., Yap, I. K. S., Chan, Q., … Elliott, P. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, *453*(7193), 396–400. doi:10.1038/nature06882

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., … Bork, P. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, *44*(D1), D286–D293. doi:10.1093/nar/gkv1248

Hutkins, R. (2006). *Microbiology and Technology of Fermented Foods*. John Wiley & Sons.

Ikeda-Ohtsubo, W., & Brune, A. (2009). Cospeciation of termite gut flagellates and their bacterial endosymbionts: Trichonympha species and "Candidatus Endomicrobium trichonymphae". *Molecular Ecology*, *18*(2), 332–42. doi:10.1111/j.1365-294X.2008.04029.x

Illumina Inc. (2014). Intro to Sequencing by Synthesis: Industry-leading Data Quality. Retrieved from https://www.youtube.com/watch?v=HMyCqWhwB8E

Ingram, L. O., Aldrich, H. C., Borges, A. C. C., Causey, T. B., Martinez, A., Morales, F., … Zhou, S. (1999). Enteric Bacterial Catalysts for Fuel Ethanol Production. *Biotechnology Progress*, *15*(5), 855–866. doi:10.1021/bp9901062

Jamei, M., Marciniak, S., Feng, K., Barnett, A., Tucker, G., & Rostami-Hodjegan, A. (2009). The Simcyp ® Population-based ADME Simulator. *Expert Opinion on Drug Metabolism & Toxicology*, *5*(2), 211–223. doi:10.1517/17425250802691074

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*. doi:10.1128/JCM.01228-07

Joly, Y., Saulnier, K. M., Osien, G., & Knoppers, B. M. (2014). The ethical framing of personalized medicine. *Current Opinion in Allergy and Clinical Immunology*, *14*(5), 404–408. doi:10.1097/ACI.0000000000000091

Jonkers, D., Penders, J., Masclee, A., & Pierik, M. (2012). Probiotics in the Management of Inflammatory Bowel Disease. *Drugs*, *72*(6), 803–823. doi:10.2165/11632710-000000000-00000

Kahindi, J. H. P., Woomer, P., George, T., de Souza Moreira, F. M., Karanja, N. K., & Giller, K. E. (1997). Agricultural intensification, soil biodiversity and ecosystem function in the tropics: the role of nitrogen-fixing bacteria. *Applied Soil Ecology*, *6*(1), 55–76. doi:10.1016/S0929-1393(96)00151-5

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Research*, *32*(Database issue), D277–80. doi:10.1093/nar/gkh063

Keiser, J., & Utzinger, J. (2008). Efficacy of current drugs against soil-transmitted helminth infections: systematic review and meta-analysis. *JAMA*, *299*(16), 1937–48. doi:10.1001/jama.299.16.1937

Kim, & Crowley. (2007). Microbial Diversity in Natural Asphalts of the Rancho La Brea Tar Pits. *Applied and Environmental Microbiology*, *73*(14), 4579–4591. doi:10.1128/AEM.01372-06

Kim, D. H., Yeo, S. H., Park, J.-M., Choi, J. Y., Lee, T.-H., Park, S. Y., … Cha, H.-J. (2014). Genetic markers for diagnosis and pathogenesis of Alzheimer's disease. *Gene*, *545*(2), 185–193. doi:10.1016/j.gene.2014.05.031

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., … Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, *108*(Supplement_1), 4578–4585. doi:10.1073/pnas.1000081107

Kong, W., Wei, J., Abidi, P., Lin, M., Inaba, S., Li, C., … Jiang, J.-D. (2004). Berberine is a novel cholesterol-lowering drug working through a unique mechanism distinct from statins. *Nature Medicine*, *10*(12), 1344–1351. doi:10.1038/nm1135

Koren, O., Goodrich, J. K., Cullender, T. C., Spor, A., Laitinen, K., Kling Bäckhed, H., … Ley, R. E. (2012). Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy. *Cell*, *150*(3), 470–480. doi:10.1016/j.cell.2012.07.008

Kostic, A. D., Howitt, M. R., & Garrett, W. S. (2013). Exploring host-microbiota interactions in animal models and humans. *Genes & Development*, *27*(7), 701–718. doi:10.1101/gad.212522.112

Kostic, A. D., Xavier, R. J., & Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology*, *146*(6), 1489–99. doi:10.1053/j.gastro.2014.02.009

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., … Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821. doi:10.1038/nbt.2676

Langille, M. G., Meehan, C. J., Koenig, J. E., Dhanani, A. S., Rose, R. A., Howlett, S. E., & Beiko, R. G. (2014). Microbial shifts in the aging mouse gut. *Microbiome*, *2*(1), 50. doi:10.1186/s40168-014-0050-9

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Leahy, J. G., & Colwell, R. R. (1990). Microbial degradation of hydrocarbons in the environment. *Microbiological Reviews*, *54*(3), 305–315.

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D., & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current Opinion in Biotechnology*, *24*(2), 160–168. doi:10.1016/j.copbio.2012.08.005

Lesko, L. J. (2007). Personalized Medicine: Elusive Dream or Imminent Reality? *Clinical Pharmacology & Therapeutics*, *81*(6), 807–816. doi:10.1038/sj.clpt.6100204

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: Human gut microbes associated with obesity. *Nature*, *444*(7122), 1022–1023. doi:10.1038/4441022a

Lindenbaum, J., Rund, D. G., Butler, V. P., Tse-Eng, D., & Saha, J. R. (1981). Inactivation of digoxin by the gut flora: reversal by antibiotic therapy. *The New England Journal of Medicine*, *305*(14), 789–94. doi:10.1056/NEJM198110013051403

Liston, H. L., Markowitz, J. S., & DeVane, C. L. (2001). Drug glucuronidation in clinical psychopharmacology. *Journal of Clinical Psychopharmacology*, *21*(5), 500–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11593076

Locasale, J. W. (2013). Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nature Reviews Cancer*, *13*(8), 572–583. doi:10.1038/nrc3557

Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V, & Daly, M. J. (2001). Genome of the extremely radiation-resistant bacterium Deinococcus radiodurans viewed from the perspective of comparative genomics. *Microbiology and Molecular Biology Reviews : MMBR*. doi:10.1128/MMBR.65.1.44-79.2001

Markle, J. G. M., Frank, D. N., Mortin-Toth, S., Robertson, C. E., Feazel, L. M., Rolle-Kampczyk, U., … Danska, J. S. (2013). Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science (New York, N.Y.)*, *339*(6123), 1084–8. doi:10.1126/science.1233521

Matthies, A., Loh, G., Blaut, M., & Braune, A. (2012). Daidzein and genistein are converted to equol and 5-hydroxy-equol by human intestinal Slackia isoflavoniconvertens in gnotobiotic rats. *The Journal of Nutrition*, *142*(1), 40–6. doi:10.3945/jn.111.148247

McLean, K. J., Clift, D., Lewis, D. G., Sabri, M., Balding, P. R., Sutcliffe, M. J., … Munro, A. W. (2006). The preponderance of P450s in the Mycobacterium tuberculosis genome. *Trends in Microbiology*, *14*(5), 220–8. doi:10.1016/j.tim.2006.03.002

Mergaert, J., Webb, A., Anderson, C., Wouters, A., & Swings, J. (1993). Microbial degradation of poly(3-hydroxybutyrate) and poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in soils. *Applied and Environmental Microbiology*, *59*(10), 3233–8. doi:10.1007/BF02067443

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., … Edwards, R. A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, *9*, 386. doi:10.1186/1471-2105-9-386

Meyer, J. M., & Ginsburg, G. S. (2002). The path to personalized medicine. *Current Opinion in Chemical Biology*. doi:10.1016/S1367-5931(02)00340-X

Morgan, X. C., & Huttenhower, C. (2012). Chapter 12: Human Microbiome Analysis. *PLoS Computational Biology*, *8*(12), e1002808. doi:10.1371/journal.pcbi.1002808

Nakayama, H., Kinouchi, T., Kataoka, K., Akimoto, S., Matsuda, Y., & Ohnishi, Y. (1997). Intestinal anaerobic bacteria hydrolyse sorivudine, producing the high blood concentration of 5-(E)-(2-bromovinyl)uracil that increases the level and toxicity of 5-fluorouracil. *Pharmacogenetics*, *7*(1), 35–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9110360

Neu, J., & Rushing, J. (2011). Cesarean Versus Vaginal Delivery: Long-term Infant Outcomes and the Hygiene Hypothesis. *Clinics in Perinatology*. doi:10.1016/j.clp.2011.03.008

Neurath, M. F. (2015). Advances in imaging to allow personalized medicine in Crohn's disease. *Current Opinion in Pharmacology*, *23*, 6–10. doi:10.1016/j.coph.2015.05.001

Nguyen, T. L. A., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Disease Models & Mechanisms*, *8*(1), 1–16. doi:10.1242/dmm.017400

Nicholson, J. K., Holmes, E., & Wilson, I. D. (2005). Opinion: Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews Microbiology*, *3*(5), 431–438. doi:10.1038/nrmicro1152

Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., … Ehrlich, S. D. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology*, *32*(8), 822–828. doi:10.1038/nbt.2939

NIH HMP Working Group, Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., … Guyer, M. (2009). The NIH Human Microbiome Project. *Genome Research*, *19*(12), 2317–23. doi:10.1101/gr.096651.109

O'Hara, A. M., & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, *7*(7), 688–693. doi:10.1038/sj.embor.7400731

O'Malley, K., Crooks, J., Duke, E., & Stevenson, I. H. (1971). Effect of Age and Sex on Human Drug Metabolism. *BMJ*, *3*(5775), 607–609. doi:10.1136/bmj.3.5775.607

Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, *405*(6784), 299–304. doi:10.1038/35012500

Olesen, S. W., Vora, S., Techtmann, S. M., Fortney, J. L., Bastidas-Oyanedel, J. R., Rodríguez, J., … Alm, E. J. (2016). A Novel Analysis Method for Paired-Sample Microbial Ecology Experiments. *PloS One*, *11*(5), e0154804. doi:10.1371/journal.pone.0154804

Olopade, O. I., Grushko, T. A., Nanda, R., & Huo, D. (2008). Advances in Breast Cancer: Pathways to Personalized Medicine. *Clinical Cancer Research*, *14*(24), 7988–7999. doi:10.1158/1078-0432.CCR-08-1211

Paine, M. F., Widmer, W. W., Hart, H. L., Pusek, S. N., Beavers, K. L., Criss, A. B., … Watkins, P. B. (2006). A furanocoumarin-free grapefruit juice establishes furanocoumarins as the mediators of the grapefruit juice-felodipine interaction. *The American Journal of Clinical Nutrition*, *83*(5), 1097–105. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16685052

Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics (Oxford, England)*, *30*(21), 3123–4. doi:10.1093/bioinformatics/btu494

Peppercorn, M. A., & Goldman, P. (1972). The role of intestinal bacteria in the metabolism of salicylazosulfapyridine. *The Journal of Pharmacology and Experimental Therapeutics*, *181*(3), 555–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4402374

Pham, V. H. T., & Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends in Biotechnology*, *30*(9), 475–84. doi:10.1016/j.tibtech.2012.05.007

Pond, S. M., & Tozer, T. N. (1984). First-pass elimination. Basic concepts and clinical consequences. *Clinical Pharmacokinetics*, *9*(1), 1–25. doi:10.2165/00003088-198409010-00001

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., … Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. doi:10.1038/nature08821

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., … Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, *490*(7418), 55–60. doi:10.1038/nature11450

R Development Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. *R Foundation for Statistical Computing, Vienna, Austria.*

Reis-Filho, J. S. (2009). Next-generation sequencing. *Breast Cancer Research*, *11*(Suppl 3), S12. doi:10.1186/bcr2431

Remmer, H. (1970). The role of the liver in drug metabolism. *The American Journal of Medicine*, *49*(5), 617–629. doi:10.1016/S0002-9343(70)80129-2

Renwick, S. B., Snell, K., & Baumann, U. (1998). The crystal structure of human cytosolic serine hydroxymethyltransferase: a target for cancer chemotherapy. *Structure (London, England : 1993)*, *6*(9), 1105–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9753690

Roberts, A. B., Wallace, B. D., Venkatesh, M. K., Mani, S., & Redinbo, M. R. (2013). Molecular insights into microbial β-glucuronidase inhibition to abrogate CPT-11 toxicity. *Molecular Pharmacology*, *84*(2), 208–17. doi:10.1124/mol.113.085852

Roberts, M. S., Magnusson, B. M., Burczynski, F. J., & Weiss, M. (2002). Enterohepatic Circulation. *Clinical Pharmacokinetics*, *41*(10), 751–790. doi:10.2165/00003088-200241100-00005

Rockwood, K. (2005). A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal*, *173*(5), 489–495. doi:10.1503/cmaj.050051

Rodrigue, S., Materna, A. C., Timberlake, S. C., Blackburn, M. C., Malmstrom, R. R., Alm, E. J., & Chisholm, S. W. (2010). Unlocking short read sequencing for metagenomics. *PloS One*, *5*(7), e11840. doi:10.1371/journal.pone.0011840

Ronaghi, M., Uhlén, M., & Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science (New York, N.Y.)*, *281*(5375), 363, 365. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9705713

Round, J. L., & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews Immunology*, *9*(5), 313–323. doi:10.1038/nri2515

Rowland, M. (1972). Influence of Route of Administration on Drug Availability. *Journal of Pharmaceutical Sciences*, *61*(1), 70–74. doi:10.1002/jps.2600610111

Rüfer, C. E., & Kulling, S. E. (2006). Antioxidant activity of isoflavones and their major metabolites using different in vitro assays. *Journal of Agricultural and Food Chemistry*, *54*(8), 2926–31. doi:10.1021/jf053112o

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/271968

Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annual Review of Microbiology*, *31*, 107–33. doi:10.1146/annurev.mi.31.100177.000543

Scheline, R. R. (1968). The Metabolism of Drugs and Other Organic Compounds by the Intestinal Microflora Studies on the metabolic fate of drugs and foreign organic compounds in the tissues . These reactions include oxidation , reduction , hydrolysis Investigations carried out in, 332–342.

Sender, R., Fuchs, S., & Milo, R. (2016). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*, *164*(3), 337–40. doi:10.1016/j.cell.2016.01.013

Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., & Gilbert, J. A. (2014). Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. *mBio*, *5*(4), e01371–14–e01371–14. doi:10.1128/mBio.01371-14

Sheweita, S. A. (2000). Drug-metabolizing enzymes: mechanisms and functions. *Current Drug Metabolism*, *1*(2), 107–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11465078

Shimada, Y., Yasuda, S., Takahashi, M., Hayashi, T., Miyazawa, N., Sato, I., … Hishigaki, H. (2010). Cloning and expression of a novel NADP(H)-dependent daidzein reductase, an enzyme involved in the metabolism of daidzein, from equol-producing Lactococcus strain 20-92. *Applied and Environmental Microbiology*, *76*(17), 5892–901. doi:10.1128/AEM.01101-10

Shu, Y. Z., Kingston, D. G., Van Tassell, R. L., & Wilkins, T. D. (1991). Metabolism of levamisole, an anti-colon cancer drug, by human intestinal bacteria. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems*, *21*(6), 737–50. doi:10.3109/00498259109039513

Smith, H. S. (2009). Opioid Metabolism. *Mayo Clinic Proceedings*, *84*(7), 613–624. doi:10.1016/S0025-6196(11)60750-7

Soleim, H. A., & Scheline, R. R. (1972). Metabolism of xenobiotics by strains of intestinal bacteria. *Acta Pharmacologica et Toxicologica*, *31*(5), 471–80. doi:10.1111/j.1600-0773.1972.tb03610.x

Sousa, T., Paterson, R., Moore, V., Carlsson, A., Abrahamsson, B., & Basit, A. W. (2008). The gastrointestinal microbiota as a site for the biotransformation of drugs. *International Journal of Pharmaceutics*, *363*(1-2), 1–25. doi:10.1016/j.ijpharm.2008.07.009

Stokes, H. W., & Gillings, M. R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiology Reviews*, *35*(5), 790–819. doi:10.1111/j.1574-6976.2011.00273.x

Swerdlow, N. R., & Koob, G. F. (1987). Dopamine, schizophrenia, mania, and depression: Toward a unified hypothesis of cortico-striatopallido-thalamic function. *Behavioral and Brain Sciences*, *10*(02), 197. doi:10.1017/S0140525X00047488

Takami, H., Inoue, A., Fuji, F., & Horikoshi, K. (1997). Microbial flora in the deepest sea mud of the Mariana trench. *FEMS Microbiology Letters*, *152*(2), 279–285. doi:10.1016/S0378-1097(97)00211-5

Tang, W. H. W., Wang, Z., Levison, B. S., Koeth, R. A., Britt, E. B., Fu, X., … Hazen, S. L. (2013). Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *The New England Journal of Medicine*, *368*(17), 1575–84. doi:10.1056/NEJMoa1109400

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33–36. doi:10.1093/nar/28.1.33

Teh, L. K., & Bertilsson, L. (2012). Pharmacogenomics of CYP2D6: molecular genetics, interethnic differences and clinical importance. *Drug Metabolism and Pharmacokinetics*, *27*(1), 55–67. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22185816

The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–14. doi:10.1038/nature11234

Thomson, P. D. (1973). Lidocaine Pharmacokinetics in Advanced Heart Failure, Liver Disease, and Renal Failure in Humans. *Annals of Internal Medicine*, *78*(4), 499. doi:10.7326/0003-4819-78-4-499

Tlaskalová-Hogenová, H., Štěpánková, R., Hudcovic, T., Tučková, L., Cukrowska, B., Lodinová-Žádníková, R., … Kokešová, A. (2004). Commensal bacteria (normal microflora), mucosal immunity and chronic inflammatory and autoimmune diseases. *Immunology Letters*, *93*(2-3), 97–108. doi:10.1016/j.imlet.2004.02.005

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027–131. doi:10.1038/nature05414

Turnheim, K. (2004). Drug therapy in the elderly. *Experimental Gerontology*, *39*(11-12), 1731–8. doi:10.1016/j.exger.2004.05.011

Ueland, P. M. (2011). Choline and betaine in health and disease. *Journal of Inherited Metabolic Disease*, *34*(1), 3–15. doi:10.1007/s10545-010-9088-4

Van Der Heijden, M. G. A., Bakker, R., Verwaal, J., Scheublin, T. R., Rutten, M., Van Logtestijn, R., & Staehelin, C. (2006). Symbiotic bacteria as a determinant of plant community structure and plant productivity in dune grassland. In *FEMS Microbiology Ecology* (Vol. 56, pp. 178–187). doi:10.1111/j.1574-6941.2006.00086.x

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., & Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience*, *2*(1), 16. doi:10.1186/2047-217X-2-16

Venter, J. C. (2001). The Sequence of the Human Genome. *Science*, *291*(5507), 1304–1351. doi:10.1126/science.1058040

Visscher, P. T., Reid, R. P., Decho, A. W., Stolz, J. F., Bebout, B. M., Dupraz, C., … DesMarais, D. J. (2000). No Title. *Nature*, *406*(6799), 989–992. doi:10.1038/35023158

Vlaar, A., Hovestadt, A., van Laar, T., & Bloem, B. R. (2011). The treatment of early Parkinson's disease: levodopa rehabilitated. *Practical Neurology*, *11*(3), 145–52. doi:10.1136/practneurol-2011-000011

Wadworth, A. N., & Fitton, A. (1991). Olsalazine. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in inflammatory bowel disease. *Drugs*, *41*(4), 647–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1711964

Wagner, M., Rath, G., Koops, H., Flood, J., & Amann, R. (1996). Analysis of nitrifying bacteria in sewage treatment plants. *Water Science and Technology*, *34*(1-2), 237–244. doi:10.1016/0273-1223(96)00514-8

Walker, J. J., & Pace, N. R. (2007). Phylogenetic composition of rocky mountain endolithic microbial ecosystems. *Applied and Environmental Microbiology*, *73*(11), 3497–3504. doi:10.1128/AEM.02656-06

Wallace, B. D., Wang, H., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., … Redinbo, M. R. (2010). Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science (New York, N.Y.)*, *330*(6005), 831–5. doi:10.1126/science.1191175

Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., DuGar, B., … Hazen, S. L. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, *472*(7341), 57–63. doi:10.1038/nature09922

Wang, Z., Roberts, A. B., Buffa, J. A., Levison, B. S., Zhu, W., Org, E., … Hazen, S. L. (2015). Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. *Cell*, *163*(7), 1585–95. doi:10.1016/j.cell.2015.11.055

Weimer, P. J. (1992). Cellulose Degradation by Ruminal Microorganisms. *Critical Reviews in Biotechnology*, *12*(3), 189–223. doi:10.3109/07388559209069192

Williams, S. C. P. (2014). Gnotobiotics: *Proceedings of the National Academy of Sciences*, *111*(5), 1661–1661. doi:10.1073/pnas.1324049111

Xu, H., Heinze, T. M., Chen, S., Cerniglia, C. E., & Chen, H. (2007). Anaerobic metabolism of 1-amino-2-naphthol-based azo dyes (Sudan dyes) by human intestinal microflora. *Applied and Environmental Microbiology*, *73*(23), 7759–62. doi:10.1128/AEM.01410-07

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., … Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222–7. doi:10.1038/nature11053

Zeisel, S. H., & da Costa, K.-A. (2009). Choline: an essential nutrient for public health. *Nutrition Reviews*, *67*(11), 615–23. doi:10.1111/j.1753-4887.2009.00246.x

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. doi:10.1093/bioinformatics/btt593

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., … Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, *352*(6285), 565–569. doi:10.1126/science.aad3369

Zhou, S.-F., Liu, J.-P., & Chowbay, B. (2009). Polymorphism of human cytochrome P450 enzymes and its clinical impact. *Drug Metabolism Reviews*, *41*(2), 89–295. doi:10.1080/03602530902843483

Zhou, S.-F., Zhou, Z.-W., Yang, L.-P., & Cai, J.-P. (2009). Substrates, inducers, inhibitors and structure-activity relationships of human Cytochrome P450 2C9 and implications in drug development. *Current Medicinal Chemistry*, *16*(27), 3480–675. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19515014

APPENDIX

Table 4 Microbial drug metabolizing gene master list. The first column contains all annotated MDMG IDs and names between HUMAnN and PICRUSt, along with Enzyme Commission (EC) numbers if applicable. The 'Gut vs All' column indicates if the mean relative abundance of a particular MDMG was increased (up) or decreased (down) in the gut compared to the four other body sites in the HMP dataset. The 'China Age' column indicates if the relative abundance of a MDMG was positively correlated (up) or negatively correlated (down) with age in the Chinese dataset. The 'NW Age' column indicates if the relative abundance of a MDMG was positively correlated (up) or negatively correlated (down) with age in the Northwood dataset. The 'NW FI' column indicates if the relative abundance of a MDMG was positively correlated (up) or negatively correlated (down) with FI score in the Northwood dataset. The 'NW Med' column indicates if the mean relative abundance of a MDMG was increased (up) or decreased (down) in the samples of individuals on microbiome-altering drugs compared to the samples of individuals not on microbiome-altering drugs in the Northwood Samples. In all columns, 'nosig' indicates that the difference or correlation for a KO's relative abundance in the given analysis was not statistically significant. Blank spaces indicate that the MDMG was not found at all in the dataset, or the MDMG was not annotated.

KO ID and Name	Gut vs All	China Age	NW Age	NW FI	NW Med
K00001: alcohol dehydrogenase [EC:1.1.1.1]	down	down	nosig	nosig	nosig
K00002: alcohol dehydrogenase (NADP+) [EC:1.1.1.2]	down	nosig	nosig	nosig	nosig
K00055: aryl-alcohol dehydrogenase [EC:1.1.1.90]	down	nosig	nosig	nosig	nosig
K00114: alcohol dehydrogenase (acceptor) [EC:1.1.99.8] or alcohol dehydrogenase (cytochrome c) [EC:1.1.2.8]	up	nosig	nosig	nosig	nosig
K00121: S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase [EC:1.1.1.284 1.1.1.1]	down	nosig	up	nosig	up
K00274: monoamine oxidase [EC:1.4.3.4]	down	nosig	nosig	nosig	nosig
K00485: dimethylaniline monooxygenase (N-oxide forming) [EC:1.14.13.8]	down	nosig			
K00542: guanidinoacetate N-methyltransferase [EC:2.1.1.2]	nosig	nosig			
K00543: acetylserotonin N-methyltransferase [EC:2.1.1.4]	down	nosig			
K00544: betaine-homocysteine S-methyltransferase [EC:2.1.1.5]	down	nosig			
K00545: catechol O-methyltransferase [EC:2.1.1.6]	down	nosig			
K00547: homocysteine S-methyltransferase [EC:2.1.1.10]	down	nosig	nosig	nosig	nosig
K00548: 5-methyltetrahydrofolate--homocysteine methyltransferase [EC:2.1.1.13]	up	nosig	nosig	nosig	nosig
K00549: 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase [EC:2.1.1.14]	down	nosig	nosig	nosig	nosig
K00554: tRNA (guanine-N1-)-methyltransferase [EC:2.1.1.31]	down	up	down	nosig	nosig
K00555: tRNA (guanine-N2-)-methyltransferase [EC:2.1.1.32]	down	nosig			
K00556: tRNA (guanosine-2'-O-)-methyltransferase [EC:2.1.1.34]	down	nosig	nosig	up	up
K00557: tRNA (uracil-5-)-methyltransferase [EC:2.1.1.35]	down	nosig	up	nosig	up
K00558: DNA (cytosine-5-)-methyltransferase [EC:2.1.1.37]	up	nosig	nosig	nosig	nosig
K00559: sterol 24-C-methyltransferase [EC:2.1.1.41]	down	nosig			
K00561: 23S rRNA (adenine2085-N6)-dimethyltransferase [EC:2.1.1.184] or rRNA (adenine-N6-)-methyltransferase [EC:2.1.1.48]	up	down	nosig	nosig	nosig

K00563: 23S rRNA (guanine745-N1)-methyltransferase [EC:2.1.1.187] or rRNA (guanine-N1-)-methyltransferase [EC:2.1.1.51]	down	nosig	nosig	nosig	up
K00564: ribosomal RNA small subunit methyltransferase C [EC:2.1.1.172] or ribosomal RNA small subunit methyltransferase C [EC:2.1.1.52]	down	nosig	nosig	nosig	up
K00567: methylated-DNA-[protein]-cysteine S-methyltransferase [EC:2.1.1.63]	nosig	nosig	nosig	down	nosig
K00568: 3-demethylubiquinone-9 3-methyltransferase [EC:2.1.1.- 2.1.1.64]	down	nosig	up	nosig	up
K00569: thiopurine S-methyltransferase [EC:2.1.1.67]	down	nosig	nosig	nosig	nosig
K00570: phosphatidyl-N-methylethanolamine N-methyltransferase [EC:2.1.1.71]	up	nosig	nosig	nosig	up
K00571: site-specific DNA-methyltransferase (adenine-specific) [EC:2.1.1.72]	up	nosig	nosig	nosig	nosig
K00573: protein-L-isoaspartate(D-aspartate) O-methyltransferase [EC:2.1.1.77]	down	nosig	nosig	nosig	nosig
K00575: chemotaxis protein methyltransferase CheR [EC:2.1.1.80]	up	nosig	up	nosig	nosig
K00577: tetrahydromethanopterin S-methyltransferase subunit A [EC:2.1.1.86]	nosig	nosig			
K00584: tetrahydromethanopterin S-methyltransferase subunit H [EC:2.1.1.86]	up	nosig	nosig	nosig	up
K00587: protein-S-isoprenylcysteine O-methyltransferase [EC:2.1.1.100]	down	nosig			
K00588: caffeoyl-CoA O-methyltransferase [EC:2.1.1.104]	up	nosig	nosig	nosig	nosig
K00590: site-specific DNA-methyltransferase (cytosine-N4-specific) [EC:2.1.1.113]	up	nosig	nosig	nosig	nosig
K00595: precorrin-6Y C5,15-methyltransferase / precorrin-8W decarboxylase [EC:2.1.1.132 1.-.-.-]	up	down	nosig	nosig	nosig
K00598: trans-aconitate 2-methyltransferase [EC:2.1.1.144]	up	up	up	nosig	nosig
K00600: glycine hydroxymethyltransferase [EC:2.1.2.1]	down	up	nosig	nosig	nosig
K00605: aminomethyltransferase [EC:2.1.2.10]	up	up	nosig	up	nosig
K00606: 3-methyl-2-oxobutanoate hydroxymethyltransferase [EC:2.1.2.11]	up	nosig	nosig	nosig	nosig
K00619: amino-acid N-acetyltransferase [EC:2.3.1.1]	down	up	nosig	up	nosig
K00620: glutamate N-acetyltransferase / amino-acid N-acetyltransferase [EC:2.3.1.35 2.3.1.1]	down	nosig	nosig	down	down
K00621: glucosamine-phosphate N-acetyltransferase [EC:2.3.1.4]	up	nosig			
K00622: arylamine N-acetyltransferase [EC:2.3.1.5]	down	nosig			
K00657: diamine N-acetyltransferase [EC:2.3.1.57]	up	nosig	nosig	up	nosig
K00676: ribosomal-protein-alanine N-acetyltransferase [EC:2.3.1.128]	up	nosig	nosig	nosig	nosig

K00783: 23S rRNA (pseudouridine1915-N3)-methyltransferase [EC:2.1.1.177] rlmH	nosig	nosig	down	nosig	nosig
K00799: glutathione S-transferase [EC:2.5.1.18]	down	nosig	up	nosig	up
K01118: FMN-dependent NADH-azoreductase [EC:1.7.-.-]	down	nosig	up	nosig	up
K01195: beta-glucuronidase [EC:3.2.1.31]	up	nosig	down	nosig	nosig
K01201: glucosylceramidase [EC:3.2.1.45]	Up	nosig	down	nosig	nosig
K01202: Galactosylceramidase (GALC) [EC:3.2.1.46]	nosig	nosig			
K01210: glucan 1,3-beta-glucosidase [EC:3.2.1.58]	nosig	nosig	nosig	nosig	nosig
K01222: 6-phospho-beta-glucosidase [EC:3.2.1.86]	down	nosig	up	nosig	nosig
K01223: 6-phospho-beta-glucosidase [EC:3.2.1.86]	down	up	nosig	nosig	nosig
K01253: microsomal epoxide hydrolase [EC:3.3.2.9]	down	nosig	nosig	nosig	nosig
K01424: L-asparaginase [EC:3.5.1.1]	up	nosig	down	up	nosig
K01426: amidase [EC:3.5.1.4]	down	nosig	nosig	nosig	nosig
K01432: arylformamidase [EC:3.5.1.9]	down				
K01434: penicillin amidase [EC:3.5.1.11]	down	nosig	nosig	nosig	nosig
K01446: N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28]	down	nosig	nosig	nosig	up
K01448: N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28]	up	down	nosig	nosig	nosig
K01449: N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28]	up	nosig	nosig	down	nosig
K01455: formamidase [EC:3.5.1.49]	down	nosig	nosig	nosig	nosig
K01460: glutathionylspermidine amidase/synthetase [EC:3.5.1.78 6.3.1.8]	down	nosig	up	nosig	nosig
K02169: biotin synthesis protein BioC or malonyl-CoA O-methyltransferase [EC:2.1.1.197]	up	nosig	nosig	up	nosig
K02188: cobalamin biosynthesis protein CbiD or cobalt-precorrin-5B (C1)-methyltransferase [EC:2.1.1.195]	up	down	nosig	nosig	nosig
K02191: precorrin-8W decarboxylase [EC:1.-.-.-] or cobalt-precorrin-7 (C15)-methyltransferase [EC:2.1.1.196]	down	nosig	nosig	nosig	nosig
K02236: leader peptidase (prepilin peptidase) / N-methyltransferase [EC:3.4.23.43 2.1.1.-]	down	nosig	nosig	nosig	up
K02295: gentamicin 3'-N-acetyltransferase [EC:2.3.1.60]	down	nosig			
K02302: uroporphyrin-III C-methyltransferase / precorrin-2 dehydrogenase / sirohydrochlorin ferrochelatase [EC:2.1.1.107 1.3.1.76 4.99.1.4]	down	nosig	up	nosig	up
K02303: uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	down	up	nosig	nosig	nosig
K02427: ribosomal RNA large subunit methyltransferase E [EC:2.1.1.-]	down	nosig	up	nosig	nosig
K02493: methyltransferase [EC:2.1.1.-]	down	down	down	nosig	nosig
K02496: uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	down	nosig	up	nosig	up
K02528: 16S rRNA (adenine1518-N6/adenine1519-N6)-dimethyltransferase [EC:2.1.1.182] or dimethyladenosine transferase [EC:2.1.1.-]	down	down	down	nosig	nosig
K02533: tRNA/rRNA methyltransferase [EC:2.1.1.-]	down	nosig	up	nosig	nosig
K02654: leader peptidase (prepilin peptidase) / N-methyltransferase [EC:3.4.23.43 2.1.1.-]	down	nosig	nosig	down	nosig
K02687: ribosomal protein L11 methyltransferase [EC:2.1.1.-]	up	down	nosig	nosig	nosig

K03183: ubiquinone/menaquinone biosynthesis methyltransferase [EC:2.1.1.163 2.1.1.-] or ubiquinone/menaquinone biosynthesis methyltransferase [EC:2.1.1.-]	up	down	nosig	up	nosig
K03206: azobenzene reductase [EC:1.7.1.6]	down				
K03212: RNA methyltransferase, TrmA family [EC:2.1.1.-]	down	nosig	up	nosig	up
K03214: RNA methyltransferase, TrmH family [EC:2.1.1.-]	down	nosig	up	nosig	up
K03215: RNA methyltransferase, TrmA family [EC:2.1.1.-]	down	nosig	up	nosig	up
K03216: RNA methyltransferase, TrmH family, group 2 [EC:2.1.1.-] or tRNA (cytidine/uridine-2'-O-)-methyltransferase [EC:2.1.1.207]	down	nosig	nosig	down	nosig
K03218: RNA methyltransferase, TrmH family [EC:2.1.1.-]	down	nosig	down	nosig	nosig
K03394: precorrin-2 C20-methyltransferase / cobalt-factor-2 C20-methyltransferase [EC:2.1.1.130 2.1.1.151] or precorrin-2/cobalt-factor-2 C20-methyltransferase [EC:2.1.1.130 2.1.1.151]	up	down	nosig	nosig	down
K03399: precorrin-6Y C5,15-methyltransferase [EC:2.1.1.132]	down	nosig	nosig	nosig	nosig
K03418: N,N-dimethylformamidase [EC:3.5.1.56]	nosig	nosig			
K03428: magnesium-protoporphyrin O-methyltransferase [EC:2.1.1.11]	nosig	nosig			
K03437: RNA methyltransferase, TrmH family	down	down	down	nosig	nosig
K03438: S-adenosyl-methyltransferase [EC:2.1.1.-] or 16S rRNA (cytosine1402-N4)-methyltransferase [EC:2.1.1.199]	nosig	nosig	nosig	nosig	down
K03439: tRNA (guanine-N7-)-methyltransferase [EC:2.1.1.33]	down	down	down	nosig	nosig
K03500: ribosomal RNA small subunit methyltransferase B [EC:2.1.1.-]	down	nosig	nosig	down	nosig
K03501: glucose inhibited division protein B [EC:2.1.-.-] or ribosomal RNA small subunit methyltransferase G [EC:2.1.1.170]	down	nosig	nosig	nosig	nosig
K03743: None	up	up	nosig	nosig	up
K03789: ribosomal-protein-alanine N-acetyltransferase [EC:2.3.1.128]	down	nosig	nosig	down	nosig
K03790: ribosomal-protein-alanine N-acetyltransferase [EC:2.3.1.128]	down	nosig	nosig	nosig	up
K04022: alcohol dehydrogenase	up	nosig	up	nosig	nosig
K04042: bifunctional UDP-N-acetylglucosamine pyrophosphorylase / Glucosamine-1-phosphate N-acetyltransferase [EC:2.7.7.23 2.3.1.157] or bifunctional protein GlmU [EC:2.7.7.23 2.3.1.157]	down	up	nosig	nosig	up
K04072: acetaldehyde dehydrogenase / alcohol dehydrogenase [EC:1.2.1.10 1.1.1.1]	down	up	nosig	nosig	nosig
K04094: methylenetetrahydrofolate--tRNA-(uracil-5-)-methyltransferase [EC:2.1.1.74] or glucose inhibited division protein Gid	down	nosig	nosig	nosig	nosig
K04097: glutathione S-transferase [EC:2.5.1.18]	down	nosig	nosig	nosig	up

K04480: methanol---5-hydroxybenzimidazolylcobamide Co-methyltransferase [EC:2.1.1.90]	up	nosig	nosig	nosig	nosig
K04711: dihydroceramidase [EC:3.5.1.-]	down	nosig			
K05303: macrocin O-methyltransferase [EC:2.1.1.101]	up	nosig	nosig	nosig	nosig
K05349: beta-glucosidase [EC:3.2.1.21]	up	nosig	nosig	nosig	down
K05350: beta-glucosidase [EC:3.2.1.21]	down	nosig	nosig	nosig	nosig
K05822: tetrahydrodipicolinate N-acetyltransferase [EC:2.3.1.89]	nosig	nosig	nosig	nosig	up
K05882: aryl-alcohol dehydrogenase (NADP+) [EC:1.1.1.91]	up	down	nosig	nosig	nosig
K05889: polyvinyl alcohol dehydrogenase (cytochrome) [EC:1.1.2.6] or polyvinyl-alcohol dehydrogenase (acceptor) [EC:1.1.99.23]	nosig	nosig			
K05928: tocopherol O-methyltransferase [EC:2.1.1.95]	up		nosig	nosig	nosig
K05929: phosphoethanolamine N-methyltransferase [EC:2.1.1.103]	nosig	nosig			
K05934: precorrin-3B C17-methyltransferase [EC:2.1.1.131]	up	nosig	nosig	down	nosig
K05936: precorrin-4 C11-methyltransferase [EC:2.1.1.133]	up	nosig	nosig	down	nosig
K06219: S-adenosylmethionine-dependent methyltransferase	down	nosig	up	nosig	up
K06442: putative hemolysin	down	nosig	nosig	down	nosig
K06941: ribosomal RNA large subunit methyltransferase N [EC:2.1.1.-]	up	down	down	nosig	nosig
K06967: or tRNA (adenine-N(1)-)-methyltransferase [EC:2.1.1.36]	down	nosig	nosig	down	nosig
K06968: ribosomal RNA large subunit methyltransferase M [EC:2.1.1.-]	down	nosig	nosig	nosig	nosig
K06969: putative SAM-dependent methyltransferase or ribosomal RNA large subunit methyltransferase I [EC:2.1.1.-]	up	down	nosig	nosig	nosig
K06970: ribosomal RNA large subunit methyltransferase F [EC:2.1.1.181] or ribosomal RNA large subunit methyltransferase F [EC:2.1.1.48]	up	nosig	nosig	up	up
K07056: or 16S rRNA (cytidine1402-2'-O)-methyltransferase [EC:2.1.1.198]	up	nosig	down	nosig	nosig
K07115: 23S rRNA (adenine2030-N6)-methyltransferase [EC:2.1.1.266] rlmJ	down	nosig	nosig	up	up
K07130: arylformamidase [EC:3.5.1.9] (kynB)	down	nosig	nosig	nosig	nosig
K07316: adenine-specific DNA-methyltransferase [EC:2.1.1.72]	down	nosig	nosig	nosig	up
K07317: adenine-specific DNA-methyltransferase [EC:2.1.1.72]	down	down	down	nosig	up
K07318: adenine-specific DNA-methyltransferase [EC:2.1.1.72]	nosig	nosig	nosig	nosig	nosig
K07319: putative adenine-specific DNA-methyltransferase [EC:2.1.1.72]	up	nosig	nosig	nosig	nosig
K07320: putative adenine-specific DNA-methyltransferase [EC:2.1.1.72]	down	nosig	up	nosig	up

K07442: tRNA (adenine-N1-)-methyltransferase [EC:2.1.1.36] or tRNA (adenine-N1-)-methyltransferase catalytic subunit [EC:2.1.1.36]	down	up	down	nosig	nosig
K07443: methylated-DNA-protein-cysteine methyltransferase related protein	up	nosig	nosig	nosig	nosig
K07446: N(2),N(2)-dimethylguanosine tRNA methyltransferase [EC:2.1.1.32] or putative methyltransferase	up	nosig	nosig	nosig	nosig
K08281: nicotinamidase/pyrazinamidase [EC:3.5.1.19 3.5.1.-]	down	up	nosig	nosig	nosig
K08316: ribosomal RNA small subunit methyltransferase D [EC:2.1.1.171] or ribosomal RNA small subunit methyltransferase D [EC:2.1.1.52]	down	nosig	nosig	down	nosig
K08325: NADP-dependent alcohol dehydrogenase [EC:1.1.-.-]	up	nosig	nosig	up	up
K08687: N-carbamoylsarcosine amidase [EC:3.5.1.59]	up	nosig	nosig	up	nosig
K09761: ribosomal RNA small subunit methyltransferase E [EC:2.1.1.-]	nosig	down	down	nosig	nosig
K09846: hydroxyneurosporene methyltransferase [EC:2.1.1.-]	down	nosig			
K10533: limonene-1,2-epoxide hydrolase [EC:3.3.2.8]	nosig	nosig			
K10617: p-cumic alcohol dehydrogenase	nosig	nosig			
K10778: AraC family transcriptional regulator, regulatory protein of adaptative response / methylated-DNA-[protein]-cysteine methyltransferase [EC:2.1.1.63]	up	up	up	nosig	up
K11066: N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28]	down	nosig	up	nosig	up
K11391: ribosomal RNA large subunit methyltransferase G [EC:2.1.1.174] or ribosomal RNA large subunit methyltransferase G [EC:2.1.1.52]	nosig	nosig	up	nosig	nosig
K11392: ribosomal RNA small subunit methyltransferase F [EC:2.1.1.-]	up	nosig	up	nosig	nosig
K11434: protein arginine N-methyltransferase 1 [EC:2.1.1.-]	down	nosig			
K12251: N-carbamoylputrescine amidase [EC:3.5.1.53]	up	nosig	nosig	nosig	down
K12297: ribosomal RNA large subunit methyltransferase L [EC:2.1.1.52] or ribosomal RNA large subunit methyltransferase L [EC:2.1.1.173]	down	up	up	nosig	up
K12957: uncharacterized zinc-type alcohol dehydrogenase-like protein [EC:1.-.-.-]	down	nosig	nosig	nosig	nosig
K13018: UDP-D-GlcNAc3NA acetyltransferase [EC:2.3.1.-]	up	up			
K13042: dimethylglycine N-methyltransferase [EC:2.1.1.161]	nosig				
K13317: NDP-4-keto-2,6-dideoxyhexose 3-C-methyltransferase	nosig				
K13486: chemotaxis protein methyltransferase WspC	nosig	nosig	nosig	nosig	nosig
K13530: AraC family transcriptional regulator, regulatory protein of adaptative response / methylphosphotriester-DNA alkyltransferase methyltransferase [EC:2.1.1.-]	up	up	nosig	nosig	nosig
K13531: methylated-DNA-[protein]-cysteine S-methyltransferase [EC:2.1.1.63]	down	nosig	nosig	nosig	nosig

K13540: precorrin-2 C20-methyltransferase / precorrin-3B C17-methyltransferase [EC:2.1.1.130 2.1.1.131]	down	nosig			
K13541: cobalamin biosynthesis protein CbiG / precorrin-3B C17-methyltransferase [EC:2.1.1.131] or cobalt-precorrin 5A hydrolase / precorrin-3B C17-methyltransferase [EC:3.7.1.12 2.1.1.131]	up	up	nosig	nosig	nosig
K13542: uroporphyrinogen III methyltransferase / synthase [EC:2.1.1.107 4.2.1.75]	down	nosig	nosig	down	nosig
K13543: uroporphyrinogen III methyltransferase / synthase [EC:2.1.1.107 4.2.1.75]	down	nosig			
K13601: bacteriochlorophyll C8 methyltransferase [EC:2.1.1.-]	down	nosig			
K13602: bacteriochlorophyll C12 methyltransferase [EC:2.1.1.-]	down				
K13604: bacteriochlorophyll C20 methyltransferase [EC:2.1.1.-]	up				
K13953: alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]	down	up	nosig	nosig	up
K13954: alcohol dehydrogenase [EC:1.1.1.1]	up	up	nosig	nosig	nosig
K13955: zinc-binding alcohol dehydrogenase/oxidoreductase	down		nosig	nosig	nosig
K13979: uncharacterized zinc-type alcohol dehydrogenase-like protein [EC:1.-.-.-]	down	nosig	nosig	nosig	nosig
K14080: methylcobalamin:coenzyme M methyltransferase [EC:2.1.1.-]		nosig			
K14082: methylcobamide:CoM methyltransferase [EC:2.1.1.-]	down	nosig	nosig	nosig	up
K14083: trimethylamine methyltransferase [EC:2.1.1.-]	up	nosig	nosig	nosig	nosig
K14338: cytochrome P450 / NADPH-cytochrome P450 reductase [EC:1.14.14.1 1.6.2.4]		nosig	nosig	nosig	nosig
K14369: O-methyltransferase [EC:2.1.1.-]					
K14372: cytochrome P450 hydroxylase [EC:1.14.-.-]					
K14374: avermectin B 5-O-methyltransferase [EC:2.1.1.-]					
K14568:		nosig			
K14681: argininosuccinate lyase / amino-acid N-acetyltransferase [EC:4.3.2.1 2.3.1.1]					
K14682: amino-acid N-acetyltransferase [EC:2.3.1.1]			nosig	nosig	nosig
K15023: 5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase			nosig	nosig	up