Interaction between host genetics and drug therapy in influencing gut bacterial diversity

BY

Mulalo Magadze

Dissertation presented for the degree
MSc (Med) Human Genetics
Department of Pathology
Faculty of Health Sciences
University of Cape Town
December 2022

Supervisor: Professor Collet Dandara (UCT)

Plagiarism declaration

DECLARATION

I, *Mulalo Magadze*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 29/11/22

Abstract

Background: Response to drug treatment shows inter-individual variability attributed mainly to environmental and genetic factors. It is now accepted that human genome variation and microbiome profiles play significant roles in patients' response to medication. This study aimed to investigate host genetic variation that affects an individual's microbiome profile and, ultimately, its effects on the therapeutic drug response of efavirenz (EFV) and warfarin. Efavirenz and warfarin are widely prescribed drugs for treatment of HIV and clotting disorders, respectively. Genetic polymorphisms in fucosyltransferase 2 (FUT2), Neural Cell Adhesion Molecule 1 (NCAM1), Shroom Family Member 3 (SHROOM3), Vitamin D receptor (VDR) and Lactase (LCT) were characterised and evaluated on their role on microbiome profiles and how this affected drug responses.

Methods: Six hundred and forty-seven (n=647) African adults were recruited from Zimbabwe and blood for DNA extraction, stool samples for microbiome analysis and plasma for drug analyses were obtained. Genetic characterisation of SNPs, FUT2 (rs601338), NCAM1(rs17115310), SHROOM3 (rs11724031), VDR (rs1544410), VDR (rs7975232), VDR (rs731236), LCT (rs2164210) and LCT (rs7579771) was performed. Correlations were then made among genotypes and microbiome profiles on participants with different drug exposures (efavirenz and warfarin).

Results: The five genes exhibited genetic variation. Baseline allele frequencies for each of the SNPs was as follows; FUT2 rs601338A (0.433), NCAM1 rs17115310G (0.233), SHROOM3 rs11724031A (0.192), VDR rs1544410T (0.255), VDR rs7975232C (0.270), VDR rs731236G (0.242), LCT rs2164210C (0.188) and LCT rs7579771T (0.292). The frequencies of the variant alleles showed differences when compared to other populations. FUT2 rs601338 was shown to have an influence on warfarin C_{max}. Genera *Actinomycetospora* and *Brevibacterium*; and species *Corynebacterium kroppenstedtii, Macrococcus caseolyticus* and *Kocuria kristinae* were significantly abundant in the gut bacterial composition of FUT2 rs601338 AA genotype than the GG+GA genotypes. Bacterial order *Bifidobacteriales* were relatively abundant in the warfarin group

than in the untreated control group whereas *Symbiobacteriaceae* was significantly plentiful in the efavirenz group than the untreated control group.

Discussion: Genes that affect body colonising microbiomes, although not directly important in pharmacogenomics, influence treatment outcomes. There were some differences and similarities in the distribution of minor allele for the 8 variants studied when compared to other global populations, which could point to possible differences in microbiome profiles among individuals. Different microbiome profiles identified associated with the polymorphisms are likely to affect drug treatment outcomes, through host genetics, microbiome and drug interactions.

Conclusion: These results show that African populations carry different polymorphisms which are likely to affect microbiome profiles and drug responses. Thus, when considering pharmacogenomics, it is important to take into account, microbiome profiles as these affect and are affected by therapeutic drugs.

Conflict of interest and disclosure: None

Acknowledgements

- My supervisor, Professor Collet Dandara for giving an opportunity to dive deep into the exciting branch genetics infused with gut bacteriome.
- The amazing Pharmacogenetics group for the support, now identified as “SAMRC Platform for Pharmacogenomics Research and Translation (PREMED) Unit”
- The siblings that academia gave me, Arinao Ndadza, Kagiso Montjane and Zimasa Mdlulwa for all fun and support.
- My parents, family and everyone at home for their love.
- Dr Harris Onywera for comprehensive lessons in microbiome analysis.

Table of Contents

Contents

List of Abbreviations

°C	degrees Celsius
µg	microgram
µL	microlitres
3TC	lamivudine
ADME	absorption, distribution, metabolism and excretion
ART	antiretroviral therapy
ASV	amplicon sequence variant
BIOM	biological observation matrix
bp	base pair
cART	combination antiretroviral therapy
CD36	cluster of differentiation 36
CD4	cluster of differentiation 4
C_{max}	maximum plasma concentration
CNS	central nervous system
CYP	cytochrome P450
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EFV	efavirenz
EM	extensive metaboliser
FDA	Food and Drug Administration
FTC	emtricitabine
FUT2	fucosyltransferase 2
GALT	gut-associated lymphoid tissue
GIT	gastrointestinal tract
GNA12	Guanine Nucleotide-Binding Protein Subunit Alpha-12
HIV	human immunodeficiency virus

HMP	Human Microbiome Project
IBD	inflammatory bowel disease
IBS	irritable bowel syndrome
INR	international normalised ratio
IQR	interquartile range
ITSN1	intersectin 1
kb	kilobase pair
kg	kilogram
LCT	Lactase
LDA	Linear discriminant analysis
LEfSe	Linear discriminant analysis Effect Size
LINGO2	Leucine Rich Repeat And Ig Domain Containing
MAF	minor allele frequency
mGWAS	microbiome-wide association studies
mL	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
MT	microbial translocation
NCAM1	Neural Cell Adhesion Molecule 1
NM	normal metaboliser
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
OTU	operational taxonomic unit
PCR	polymerase chain reaction
PD	pharmacodynamics
PDZ	postsynaptic density-95 (PSD-95)/Discs Large (Dlg)/Zonula occludens-1
PERMANOVA	Permutational multivariate analysis of variance
PI	protease inhibitor
PK	pharmacokinetics

PLEKHA5	Pleckstrin homology domain-containing family A member 5
PM	poor metaboliser
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	quantitative polymerase chain reaction
RCF	relative centrifugal force
RFLP	restriction fragment length polymorphism
rRNA	ribosomal ribonucleic acid
SHROOM3	Shroom Family Member 3
SNP	single nucleotide polymorphism
T_a	annealing temperatures
TBE	Tris/Borate/EDTA
TDF	tenofovir disoproxil fumarate
T_m	melting temperature
UM	ultra-rapid metaboliser
UNAIDS	Joint United Nations Programme on HIV and AIDS
UTR	untranslated region
V	volts
VANGL1	Van Gogh-Like Protein 1
VDR	Vitamin D receptor
VKORC1	vitamin K epoxide reductase complex 1
WHO	World Health Organization

List of Figures

List of Tables

CHAPTER ONE: INTRODUCTION AND BACKGROUND

1.1 Introduction

Response to drug treatment shows inter-individual variability attributed mainly to environmental and genetic factors (Ingelman-Sundberg, 2001). Genetic factors are thought to account for nearly 60% of the variability observed among patients when administered therapeutic drugs (Belle and Singh, 2008). The intersection or interaction between genomic variation and therapeutic drug response is what is commonly referred to as pharmacogenomics. Pharmacogenomics is a branch of genetics that deals with how genomic profiles affect the response to drugs (Adams, 2008). Genetic polymorphisms that are important in pharmacogenomics alter the pharmacokinetics (PK) and pharmacodynamics (PD) of drugs. Pharmacokinetics is the measure of absorption, distribution, metabolism and excretion (ADME) of a drug, which deals with its kinetics within the body. PK particularly deals with how the body handles drugs it is exposed to. On the other hand, PD refers to the biological effect of a drug on the body by interacting with the receptors, ion channels, and enzymes to bring about the action (Ahmed et al., 2016). In contrast to PK, PD generally refers to what the drug does to the body. Based on an individual's drug metabolism activity, they can be categorised into poor metabolizers (PMs), normal metabolizers (NMs), extensive metabolizers (EMs), and ultra-rapid metabolizers (UMs) (Bank et al., 2018). However, it has become apparent that pharmacogenomics alone is not adequate to elucidate the observed variation for each drug response phenotype (Doestzada et al., 2018). Recently, it has been shown that gut microbiota plays an important role in drug responses (Dikeocha et al., 2022). Thus, the profile of one's microbiota is likely to affect their responses to therapeutic drug treatments.

1.2 Microbiomes in health and disease

Communities of microorganisms colonize particular locations in and on human bodies, such as skin, nose, ears and the gastrointestinal tract (GIT), the gut. The human microbiota is made up of trillions of cells in communities of bacteria, viruses, archaea, fungi, and protozoans. These microorganisms contain their own

genes, collectively known as the gut microbiome (Ursell et al., 2012) and is thought to weigh nearly 2 kilograms (kg). With over 20 million genes, microbiome is considered as a human's second genome (Zhu et al., 2010). The microbiome is gaining momentum for its fundamental role in human health, thanks to information that is coming out of the Human Microbiome Project (HMP) and associated projects (Human Microbiome Project Consortium, 2012a, 2012b, "The Integrative Human Microbiome Project," 2014), which have improved knowledge on these microorganisms. However, from all the anatomical regions of the human body, the gut has received considerable attention, thus, the bacterial microbiome is the most studied. Several factors have been reported to alter the composition of the gut microbiota and this includes diet (Wan et al., 2019), mode of birth-delivery (Wampach et al., 2018), one's age (Nagpal et al., 2018), alcohol consumption (Bjørkhaug et al., 2019), use of antibiotics (Cully, 2019) and geographical location (Fontana et al., 2019). These factors are believed to explain more than 20 percent of inter-person variability in gut microbiome profiles.

These organisms have been implicated in nutrient metabolism, an example being the ability to aid in plant cellulose digestion (Zhang et al., 2018). Gut microbiome produce several vitamins B and K_2, these vitamins are beneficial to the host as they are involved in certain human host pathways (Rowland et al., 2018). Apart from gut microbiota's ability to root out energy from otherwise indigestible diet, the human gut microbiome serves as an immune organ. It has an influence on maintaining host health through immune sensitisation (Takiishi et al., 2017); inhibiting pathogen colonization, and keeping the intestinal mucosa healthy (Alessandri et al., 2019). The gut microbiome is unique for every individual and the definition of a healthy gut is still being debated (McBurney et al., 2019). However *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, and *Proteobacteria* phyla are the most dominant in the healthy human gut (Rinninella et al., 2019). Abnormalities in the gut microbiome have been considered to contribute to the large spectrum of diseases and disorders ranging from inflammatory bowel diseases (IBD), irritable bowel syndrome (IBS), diabetes, allergic disease, cancer to neurodevelopmental illnesses (Dąbrowska and Witkiewicz, 2016). There is still a lot of debate on what shapes an individual's microbiota profile, and

the focus of this study is on the role of the host genomic variation and exposure to selected therapeutic drug exposures on the gut bacterial diversity.

1.3 Human genetic variation shaping the microbiome

Previously, host genetic profiles with variations in microbiome composition were mostly done in mice models (Jin et al., 2015). Progressively, studies on humans have established a component of gut microbiome heritability. This is supported by observations from a UK study by Goodrich et al. (2014) which showed that microbiome profiles were more similar among monozygotic than dizygotic twins, an indication of the role of the underlying human genome on microbiome profiles. The study became a point of reference and gave rise to microbiome-wide association studies (mGWAS) (Blekhman et al., 2015; Bonder et al., 2016; Davenport et al., 2015; Goodrich et al., 2016; Turpin et al., 2016; Kolde et al., 2018) using candidate-gene approaches to evaluate the interactions between microbiomes and host genetic variability. However, Scepanovic et al. (2019) did a similar study but none of the loci reached genome-wide significance threshold. Although Rothschild et al. (2018) did a similar study to Goodrich et al. (2014), heritability index indicated that the environment had a huge effect in shaping the gut microbiome profiles compared to genetics. Several host genes have been shown to influence human microbiota profiles and examples include Van Gogh-Like Protein 1 (VANGL1), Leucine Rich Repeat And Ig Domain Containing (LINGO2), Neural Cell Adhesion Molecule 1 (NCAM1), Pleckstrin homology domain-containing family A member 5 (PLEKHA5) and Intersectin 1 (ITSN1) (Bonder et al., 2016); Lactase (LCT), cluster of differentiation 36 (CD36) and Guanine Nucleotide-Binding Protein Subunit Alpha-12 (GNA12) (Goodrich et al., 2016); Vitamin D receptor (VDR), Shroom Family Member 3 (SHROOM3) (Wang et al., 2016); Fucosyltransferase 2 (FUT2) (Kolde et al., 2018) and many more genes. Below we give more details on host genes that form a major part of the focus of this project.

1.3.1 Fucosyltransferase 2 (*FUT2)* gene

The fucosyltransferase 2 gene (FUT2

) is located on chromosome 19 and encodes α(1,2) fucosyltransferase protein (Kaur et al., 2022; McGovern et al., 2010). The gene has nine exons in total and more than 2000 single nucleotide polymorphisms (SNPs). The protein changes type 1 N-acetyllactosamine glycan chains to H antigen required in the soluble A and B antigen synthesis pathway (Sugiyama et al., 2016). Non-secretor individuals do not express active enzyme due to a non-sense mutation in the *FUT2* gene and therefore they are not able to express the ABH antigens in their mucus and other secretions (Kaur et al., 2022). The mucosal ABH blood group antigens are known to serve as energy source and adhesion receptors for many microbes in the gut (Josenhans et al., 2020). This gene is a known risk factor for Crohn's disease. SNP rs601338 is located in exon 2 (Velkova et al., 2017) and the rs601338A variant results in a premature stop codon, rendering A/A genotype carriers as non-secretors (Wacklin et al., 2011). The allele frequencies of rs601338 are scarcely reported on African populations, however a study by Oussalah et al. (2012) conducted on western Africans of Togo and Benin reported 43% of the A "non-secretor" allele. Non-secretors have been found to have lower abundance and diversity of *Bifidobacteria longum* as compared to the secretors (Kolde et al., 2018). *Bifidobacteria longum* is known to utilise mucus-derived fucose as a carbon source for energy. *Bifidobacteria longum* has been associated with a number of benefits, such as to enhance the intestinal barrier function and regulate the gut microbiota, to name a few (Yao et al., 2021).

1.3.2 Neural cell adhesion molecule 1 (*NCAM1*)

The neural cell adhesion molecule 1 (*NCAM1*) gene encodes a cell adhesion protein. This gene is located on chromosome 11 and contains nearly 70000 SNPs. The encoded protein is a member of the immunoglobulin superfamily and play a role in cell-to-cell/matrix interactions during embryonic development and differentiation (Deak et al., 2005). NCAM1 also plays a role in the expansion of T cells and dendritic cells, which in turn play a role in immune surveillance (Ziegler et al., 2017). Reduced expression of NCAM1 on the epithelia could alter the anti-bacterial immune responses, therefore shaping the gut microbiome profiles. Alternative splicing results into nearly 30 different isoforms of NCAM1 protein of varying lengths, such as NCAM-120, NCAM-140, and NCAM-180, to name a few.

The rs17115310 SNP is located in the intronic region, its allele frequencies have not yet been published anywhere else except in The 1000 Genomes Project Consortium (2015), indicating that the G allele frequency of this SNP was 28%. Only Bonder et al. (2016) was able to associate this SNP with family *Acidaminococcaceae* of the gut microbiome.

1.3.3 Shroom Family Member 3 (*SHROOM3*) gene

SHROOM3 gene is located on chromosome 4 with more than 70000 SNPs. This gene encodes a postsynaptic density-95 (PSD-95)/Discs Large (Dlg)/Zonula occludens-1 (PDZ) domain-containing protein. SHROOM3 has emerged as a gene that is essential in the regulation of cell shape changes during the gut morphogenesis in humans (Chung et al., 2010). Intestinal epithelial cells link and form tight junctions which act as a physical barrier between the immune system and the gut microbiome (Lee et al., 2018). Changes in the structure and function of these epithelial junctions affects the permeability of the gut barrier. However, literature has no information on how the encoded protein shapes the gut microbiome.

The rs11724031 SNP is located in an intron of this gene. According to “The 1000 Genomes Project Consortium (2015)”, the frequency of rs11724031A allele is about 17% in African populations, 10% in Europeans and 7% in Asians. Wang et al. (2016) reported an association between rs11724031 and *Marinilabiliaceae* family of the gut bacteria. However, like rs17115310, there are very few studies that have reported on the rs11724031 SNP.

1.3.4 Vitamin D receptor (*VDR*) gene

The VDR is a well-known receptor for secondary bile acids and the gene is located on the long arm of chromosome 12 and frequently studied for vitamin D activities. The *VDR* gene comprises of 30 exons with more than 2200 SNPs reported. Variation in this gene may alter the expression or the function of the VDR protein and the most studied polymorphisms are denoted BsmI TaqI, and ApaI, respectively.

The BsmI polymorphism is located in the eighth intron at the 3′-untranslated region (UTR) of this gene and the polymorphism does not change the amino acid sequence of the protein, resulting in a silent polymorphism. However, this polymorphism may influence *VDR* messenger ribonucleic acid (mRNA) expression and alter its stability (Vasilovici et al., 2019). ApaI polymorphism lies adjacent to BsmI, also in intron 8 at the 3′UTR of VDR (Vasilovici et al., 2019). TaqI polymorphism is located in exon 9, but like BsmI, this polymorphism results in a synonymous codon change (Vasilovici et al., 2019).

It is also known that VDR promotes homeostasis within the gut microbiota with its antimicrobial properties by regulating the expression of antimicrobial peptides and defensins (Jin et al., 2015). In *Vdr* knockout mice, there was reduced abundance in Lactobacillus but the opposite in Clostridium and Bacteroides (Jin et al., 2015). Parabacteroides showed to be the most significant taxon whose abundance correlated with the polymorphism within this gene in humans.

Olesen et al. (2007) reported minor allele frequencies (MAFs) to be 51%, 33% and 31% for ApaI, BsmI and TaqI, respectively in Guinea-Bissau. Another study in Ethiopia (Ahmed et al., 2019) reported that the minor allele frequencies of TaqI(G) and ApaI(G) to be 38% and 39% respectively. In a German cohort, an intron variant of the rs7974353 SNP was found to be associated with beta-diversity and abundance of *Parabacteroides* (Wang et al., 2016).

1.3.5 Lactase (*LCT*) gene

The gene coding for the lactase enzyme, LCT, is located on chromosome 2. It encodes for an enzyme responsible for the catabolism of lactose, a sugar found dairy products, to glucose and galactose via its β-galactosidase site which are ready for absorption in the GIT. Lactase, the enzyme, is mainly expressed in cells lining the small intestine. Its pattern of expression decreases with age from neonatal period of weaning and minimally retained in adulthood. However, some individuals retain high levels of expression of this enzyme into adulthood, due to their genetic makeup. Variation in the expression of the *LCT* gene in adulthood has been associated with abundance of *Bifidobacteria*, which is one of the main commensal microbial taxa in the gut. Species in this order possess β-

galactosidase activity which allows them to breakdown lactose and use it as a primary carbon source (Forsgård, 2019). A study by Liu et al. (2015), showed a positive relationship between carbon source and abundance of *Bifidobacterium*, therefore, in high expression of LTC one can speculate that there would be low levels of *Bifidobacterium* since a small portion of ingested lactose reaches the colon for bacterial use (Goodrich et al., 2016).

The TwinsUK study by Goodrich et al. (2016) reported that a regulatory region located in the minichromosome maintenance complex component 6 (*MCM6*), a nearby gene, controls the expression of *LCT*. Few variations in this regulatory region have been identified, including C/T-13910 (rs4988235) which has been associated with lactose persistence trait in European populations (Bonder et al., 2016; Goodrich et al., 2016). However, this single nucleotide polymorphism (SNP) does not explain this trait in African populations as the MAF is almost zero (Torniainen et al., 2009). Moreover, Goodrich et al. (2016) found rs2164210 to be associated with *Bifidobacterium*. rs7579771 is just a few base pairs away from rs2164210. Both SNPs remain understudied probably because they are found in the non-coding regions of the gene. Their allele frequencies in African populations are high, as per The 1000 Genomes Project Consortium (2015) C=12% and T=27%, respectively and could point to the variability in microbiome profiles observed among different people. In addition to host genetics, this study also examined the effects of exposure to antiretroviral therapy containing efavirenz and anti-coagulation therapy with warfarin.

1.4 HIV management and treatment

The Joint United Nations Programme on HIV and AIDS (UNAIDS, 2019) statistics have shown that there were 37.9 million people on average living with HIV in 2018 worldwide with about 23.3 million on antiretroviral therapy (ART). As of November 2019, there were 35 Food and Drug Administration (FDA) registered ART drugs. Common Combination ARV therapy (cART) regiments include; efavirenz (EFV) + emtricitabine (FTC)/lamivudine (3TC) + tenofovir disoproxil fumarate (TDF).

1.4.1 Use of Efavirenz in HIV treatment

Efavirenz is one of the frequently used oral non–nucleoside reverse transcriptase inhibitors (NNRTIs) used as a first line treatment against human immunodeficiency virus type 1 (HIV-1) infections. It is preferred for its lower cost in both developed and low-income countries (Naidoo et al., 2014). To maximise its efficacy, it is usually co-administered with either a protease inhibitor (PI) and/or with nucleoside reverse transcriptase inhibitors.

Central nervous system (CNS) toxicity and psychiatric occurrences have been reported in more than 50% of patients who initiated efavirenz therapy (Mukonzo et al., 2013). Indicators of CNS toxicity include insomnia, depression, and hallucinations (Naidoo et al., 2014). The psychiatric adverse events are more frequent during the first days of treatment immediately after consuming the drug (Robarge et al., 2017). It is these side effects and high toxicity profile that lead to the discontinuation of the therapy. On the other hand, lower concentrations of efavirenz has been associated with failure of antiviral action of the drug.

Efavirenz hinders the formation of viral double-stranded DNA from the single-stranded viral RNA genome via binding to the hydrophobic region of the HIV-1 reverse transcriptase enzyme which results in the enzyme conformation changes and prevents access to the substrates (Robarge et al., 2017). This action interferes with the synthesis of new virions and results in the reduction of viral RNA load and increase in cluster of differentiation 4 $(CD4)^{+}$ cell count.

Efavirenz is mainly cleared by the liver and the process is catalysed by cytochrome P450 (CYP) enzymes, mainly the isoform *CYP450 2B6* (CYP2B6); and CYP1A2, CYP2A6, CYP3A4 and CYP3A5 to a lesser extent (Swart et al., 2016; Wang et al., 2019). Through hydroxylation, this drug is bio-converted to mainly 8-hydroxyefavirenz. Some drugs taken in tandem with efavirenz affect its action by acting on these enzymes. Polymorphisms in the genes that encode for these enzymes are being extensively studied across different populations. Moreover, there are no studies showing how these drug metabolising enzymes of efavirenz lead to variability in the gut microbiome profiles.

Despite the improved efficacy of ART and its effectiveness against HIV, responses from patients have remained non-uniform. Some individuals experience gastrointestinal (GI) discomfort, yet others appear responding well (Gupta et al., 2012).

1.5 Thromboembolic disorders management and treatment

Thromboembolic diseases were reported to be accountable for a quarter of deaths in 2010 globally (Wendelboe and Raskob, 2016). These diseases encompass conditions such as venous thrombosis, pulmonary embolisms, stroke, and ischemic attacks. They occur when blood clots develop within blood vessels, subsequently moving through the bloodstream and obstructing the blood flow. Ischemic heart disease and stroke are the leading causes of death globally in 2019, as reported in the World Health Organization (WHO). (2020) The top 10 causes of death (accessed 10/06/2023). However, in the African region, ischemic heart disease and stroke ranked in fourth and seventh places respectively. Anticoagulants are used to antagonise blood clotting cascades. The statistics on the use of these anticoagulants in Africa is not well published.

1.5.1 Warfarin treatment

Warfarin, known as Coumadin, is a derivative of coumarin that is commonly employed as a direct oral anticoagulant, as shown in Figure 1.1, to counteract blood clotting. It exerts its anticoagulant effects by effectively antagonizing the vitamin K-dependent clotting pathway. However, other oral anticoagulants such as dabigatran, rivaroxaban, edoxaban and apixaban have been introduced and their use is increasing worldwide but warfarin remains a drug of choice for more than five decades in most of Africa due to several factors that include affordability (Dalby et al., 2013; Pirmohamed, 2018; Mekaj et al., 2015; Laäs and Naidoo, 2018). Although warfarin remains the most used, its use is constantly declining when compared to other anticoagulants.

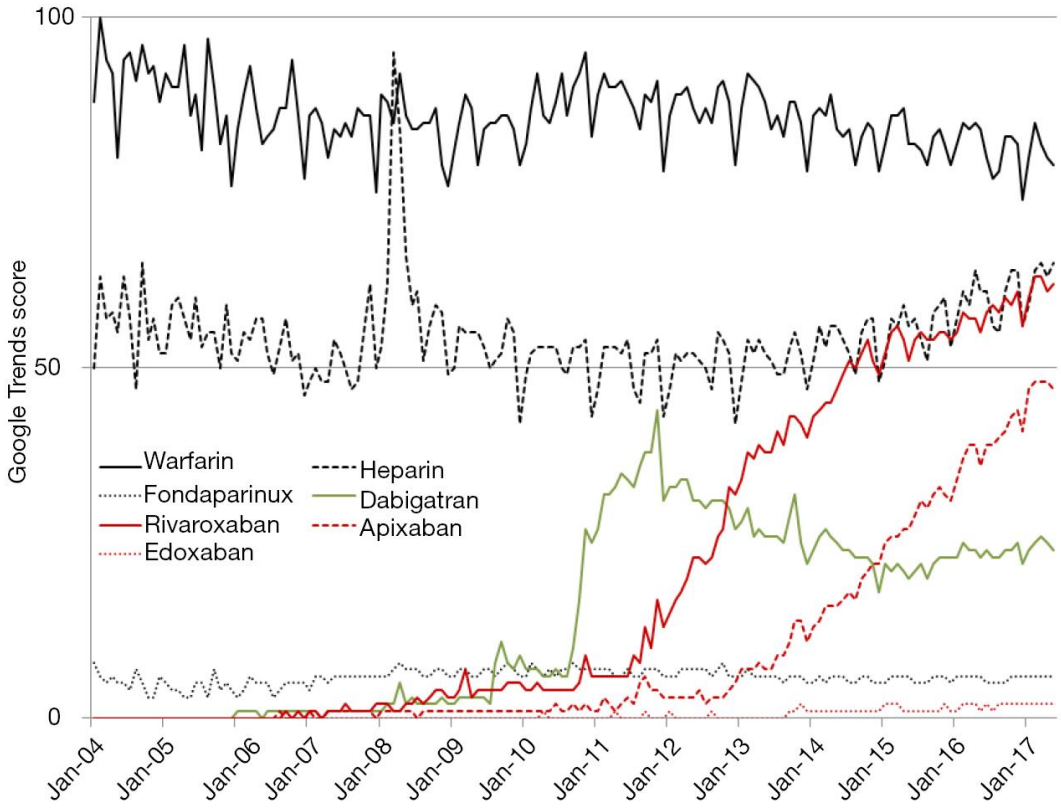


Figure 1. 1 Comparison of the extent of use of different anticoagulants across the world *(Lippi et al., 2017)*

Warfarin is composed of R and S enantiomers in racemic proportions, however the S-enantiomer is more potent that than the R-enantiomer (Wittkowsky, 2003). Warfarin exerts its anticoagulative action by competitively inhibiting the action of vitamin K epoxide reductase complex 1 (VKORC1), an essential enzyme involved in activating the vitamin K available (Stafford, 2005). This mechanism reduces the availability of active vitamin K. This vitamin is a cofactor for the synthesis of all of these vitamin K-dependent clotting factors (II, VII, IX, and X), therefore its depletion will reduce synthesis of these clotting factors (Patel et al., 2019).

Lately, literature on the influence of genetics on the metabolism of warfarin in African populations is expanding (Perera et al., 2013; Bader and Elewa, 2016; Kudzi et al., 2016; Ndadza et al., 2019). The S-enantiomer is predominantly metabolised by CYP2C9 and on the other hand, the R-enantiomer by CYP3A4. CYP1A1, CYP1A2, CYP2C8, CYP2C9, CYP2C18 and CYP2C19 are other enzymes that play minor roles. Variation in these genes involved in warfarin

metabolism continues to be studied in both African and non-African contexts (Stack and Maurice, 2016; Ndadza et al., 2019).

Significant haemorrhage has been reported with the use of this anticoagulant. In addition, vomiting, abdominal pain, and change in sense of taste are some of the observed adverse effects (Patel et al., 2019). With a narrow therapeutic index; lengthy onset of action; volatile response demanding a recurrent monitoring of the International Normalised ratio (INR) and dose adjustment, the use of warfarin proved to be very complex (Dalby et al., 2013).

1.6 Gut microbiome and drug metabolism

Variation in human gut microbiome has also shown to play a role in drug metabolism (Das et al., 2016). This was well summarised in ElRakaiby et al. (2014); Wilson and Nicholson (2017); Das et al. (2016) that gut microbiome co-metabolises drugs and xenobiotics. The microbiome can have a direct effect on drug metabolism by encoding enzymes that bio-transform drug molecules that are usually poorly metabolised or conjugated that reach the intestines. In addition, the microbiome may compete with the drug molecules over the metabolising enzymes (Zimmermann et al., 2019). Microbiomes are said to have the ability to synthesize molecules which can affect the expression levels of important drug metabolising enzymes, such as CYP and N-acetyltransferases (Das et al., 2016; Clarke et al., 2019). All these can alter the efficacy and toxicity of the drug. Studies show more than 40 drugs that are metabolised by gut microbiome and most of this drug-microbiome induction is in an online platform "PharmacoMicrobiomics: The Drug-Microbiome Portal; http://pharmacomicrobiomics.com/" (Aziz et al., 2011). However, there has been a very little progress on this platform and very little is known about the gut microbiome interaction with efavirenz and warfarin.

Recently, Zimmermann-Kogadeeva et al. (2019) built a pharmacokinetic model of drug metabolism in the mouse that takes into consideration both host and microbiota contributions to metabolism of three drugs. This model successfully predicted serum levels of metabolites of three different drugs, showing the

potential inclusion of gut microbiome in predictive pharmacokinetic models of drug metabolism.

Vitamin K is a cofactor necessary for a number coagulation factors (Scarpellini et al., 2009). Apart from the dietary vitamin K, several members such as *Bacteroides* and *Enterobacter* of the normal flora can synthesize vitamin K_2 or menaquinone. Alteration of the microbiota which synthesize vitamin K, that antagonises warfarin action may therefore increase the risk of arterial thromboembolic disease (Fischer et al., 2010).

1.7 Drug influence on gut microbiome

HIV infection on its own reciprocally depletes the CD4+ T cells, mostly the Th17 cells. The Th17 cells are mostly found in the gut-associated lymphoid tissue (GALT). When the gastrointestinal barrier integrity is destroyed, it results in impaired immunity and microbial translocation (MT) which leads to chronic immune activation and/or inflammation leading disease progression. These immune activation/inflammatory changes are reduced but persist in the presence of ART (Li et al., 2016). Although initiation of ART does not completely restore the gut microbiome, studies have shown that the use of ART also exert changes to the gut microbiome. In an in vitro study aiming to show that ART has an antibacterial effect on selected strains found in gut microbiota, EFV was among the drugs that showed antibacterial activity against *Bacillus subtilis* (Shilaih et al., 2018). Pinto-Cardoso et al. (2017) compared the gut microbiome of both HIV negative and positive subjects on ART based on either EFV or ritonavir. Several taxa of family *Ruminococcaceae* and specifically *Faecalibacterium prausnitzii* were depleted in ART treated individuals as compared to the HIV negative ones. A longitudinal study by Ji et al. (2018) where EFV was part of the regimen, the abundance of *Proteobacteria* and *Fusobacteria* phyla was relatively higher after the initiation of ART. This evidence supported the findings in (Pinto-Cardoso et al., 2017). In a recent study (Sortino et al., 2019) on age and sex-matched Thai population who had EFV as part of their regimen, higher proportions of *Proteobacteria* and *Fusobacteria*; and reduced abundance of *Bacteroidetes* members were observed post-ART compared with pre-ART. These three studies

show a common decrease in abundance of *Fusobacteria* members following exposure to ART. Literature on gut microbiome changes in humans due to warfarin exposure are very limited. A study in mice reported a significant increase in a warfarin dose dependent *Actinobacteria* phylum pre and post-warfarin exposure (Quarta et al., 2015).

1.8 Rationale

African populations have the largest genetic diversity; however, these populations have been underrepresented in scientific research as compared to other populations (Campbell and Tishkoff, 2008; Choudhury et al., 2018). There is extensive information on the host genetic influence on gut microbiome profiles, mainly on the European populations (Goodrich et al., 2014; Blekhman et al., 2015; Bonder et al., 2016; Goodrich et al., 2016; Turpin et al., 2016; Kolde et al., 2018; Wang et al., 2016). To date, comprehensive studies on the role of host genetics in shaping the gut microbiome profiles in African populations remain strikingly low. With so much genetic diversity in African populations, conclusions cannot be made in an African context based on the existing information. Large number of genetic loci have been identified and associated with gut microbiome profiles but few of them have been replicated across the existing studies, thus it is important to investigate host genetics-microbiome interaction.

Significant work has been done on the pharmacogenetics of both efavirenz (Swart et al., 2012a, 2012b; Pinillos et al., 2016; Swart et al., 2016; Reay et al., 2017) and warfarin (Schapkaitz and Sithole, 2017; Jacobs et al., 2017; Ndadza et al., 2019) in an African context. Also, the influence of drug exposure (mostly antibiotics) on the gut microbiome profiles has been explored but the information relating to efavirenz and warfarin is limited. Moreover, the attempt to address these gaps on possible influence of both host genetics and individual drug (efavirenz and warfarin) in shaping the gut microbiota remain poorly understood, it would therefore be important to look at this complex interaction.

1.9 Aims and Objectives

1.9.1 Aim

The overall aim of this study was to investigate the effects of host genetic variation and drug exposure (efavirenz and warfarin) on gut microbiome profiles in African patients.

1.9.2 Objectives

In order achieve the abovementioned aim, the following objectives were outlined:

1) To identify participants within the Pharmacogenomics group with information on warfarin or efavirenz use as well as availability of quality DNA
2) To use literature to select candidate host genes that have been reported to affect host microbiomes
3) To characterise SNPs in the candidate genes, FUT2, NCAM1, SHROOM3, VDR & LCT
4) To compare distribution of genetic variants of this study with African American, Luhya, Mende, Yoruba, European and East Asian populations
5) To determine the plasma drug levels in selected patients
6) To characterise the gut microbiome profiles
7) To correlate host genome, drug treatments (efavirenz and warfarin) and microbiome profile

CHAPTER TWO: MATERIALS AND METHODS

2.1 Study participants' recruitment, ethics and sample collection

Six hundred and forty-seven participants of Zimbabwean origin were included in this project from the Pharmacogenomics biorepository, based on the availability of DNA. These participants included some participants who were on treatment with either warfarin (i.e. warfarin group), antiretroviral therapy (ART) containing efavirenz (EFV-group) or were not on any treatment (control group). All the participants were 18 years of age and above. Of the participants on warfarin, 21 had plasma warfarin maximum concentration levels while of those on ART, 38 had plasma efavirenz concentrations. In addition to DNA and/or plasma drug concentration levels, both groups had stool samples available, including 35 of the control group. Demographic and clinical information of the participants were also available. The study was approved by the University of Cape Town, Faculty of Health Sciences Human Research Ethics Committee, HREC Ref: 707/2018.

2.2 Genotyping for single nucleotide polymorphisms

2.2.1 Primer designing

In order to successfully amplify the targeted genic regions, primers were designed to amplify DNA regions that included the SNPs or gene fragments of interest. The primers were designed using Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and IDT PrimerQuest Tool (https://eu.idtdna.com/primerquest/home/index) online platforms. Preferred pairs of primers were generally 18-22 bases long, had guanine-cytidine (GC) content between 40 and 60%, melting temperature (T_m) ranging from 50˚C to 64˚C and self-3' complementarity less than 4. The values of Gibbs free energy were also considered to assess the formation of secondary structures. An online tool, OligoAnalyzer from IDT, was used to determine the possibilities of formation of secondary structures, by mainly focusing on the formation of hairpins, self-dimers and hetero-dimers within and between the primer pairs. Primers that required more than 5 dG (kcal/mol) to break these secondary structures were ruled out. Primer-BLAST tool was used to confirm if the selected pair of primers was amplifying the desired gene and expected amplicon size without the possibility of

amplifying undesired regions. The selected primers were synthesized by Inqaba Biotec™ (Inqaba Biotechnical Industries (Pty) Ltd, Sub-Saharan Africa).

2.2.2 Polymerase Chain Reaction (PCR)

To attain the maximum yield during amplification, PCRs were optimised for each primer pair. Each reaction was carried out in a total volume of twenty-five microliters, comprising 2 µL of at least 50 ng DNA template, 0.5 U GoTaq Polymerase (Promega Cooperation, Madison, USA), 0.4 µM each of primers (Inqaba Biotechnical Industries (Pty) Ltd, Sub-Saharan Africa), 1X Green GoTaq® Reaction Buffer (Promega Cooperation, Madison, USA) and 0.4 mM deoxyribonucleotide triphosphates (dNTPs) (Bioline, Memphis, USA). Using a T100™ Thermal Cycler (Biorad Laboratories, California, USA), temperature gradient ranging from 48 °C to 65 °C was evaluated to determine the optimal annealing temperatures (T_a) for each reaction. The best annealing temperatures which resulted in clean and clear DNA bands on agarose gels were used for PCR. The number of cycles for PCR ranged from 25 to 35 times. Depending on the expected length of the product, primer extension times differed from 40 s to 45 s. Once the optimal conditions were obtained for each primer set, the DNA fragments were then amplified using T100™ Thermal Cycler (Biorad Laboratories, California, USA). The optimal PCR conditions used in this study are listed in Table 2.1. Successful PCR was confirmed by electrophoresing 5 µL of the product on 1.5% (or 1% for VDR SNPs) agarose gel stained with either SYBR® Safe DNA Gel Stain or GelRed® Nucleic Acid Gel Stain (Biotium, CA, USA) at 90 Volts (V) for an hour. 100 (base pair) bp plus Gene Ruler (ThermoScientific, Wilmington, USA) molecular weight marker was included during electrophoresis to track the migration of the fragments. Since DNA is negatively charged, it therefore migrated towards the anode under the imposed electrical current from the cathode in 1X Tris/Borate/EDTA (TBE) buffer. Agarose gel images were captured using UVITEC Cambridge Gel Documentation System.

2.2.3 Restriction Fragment Length Polymorphism (RFLP)

In order to distinguish different allelic variants and therefore different genotypes, RFLP is still one of the preferred techniques to date. The technique detects the polymorphisms by use of restriction enzymes (Williams, 1989). These enzymes bind on their recognition sites of the amplicons and cleave the DNA resulting in unique banding patterns on the gel. An online tool, NEBcutter V2.0 (New England Bio systems®) (http://nc2.neb.com/NEBcutter2/) was used to select respective enzymes whose restriction sites or sequences were affected by the polymorphisms. NEBcutter V2.0 shows all the possible enzymes that recognise the input DNA sequence. Furthermore, the position of restriction and number of sites is presented for each enzyme. An enzyme was selected if it recognised the polymorphism of interest and based on the number of recognition sites it had on the amplicon, enzymes with too many binding sites was excluded. Additionally, NEBcutter V2.0 also shows the expected band sizes for the selected enzyme. Enzymes whose resulting bands had less than 40 bp were ruled out, since it may be difficult to separate using agarose gel electrophoresis. Polymorphisms either create or abolish recognition sites for the enzymes. After the incubation, the digested DNA forms a unique banding pattern observed on appropriate percentage gels (ranging from 1.5% up to 3%) stained with GelRed® Nucleic Acid Gel Stain and ran at 120V for an hour, followed by 30 minutes at 80V. In this study, FUT2 rs601338G/A, NCAM1 rs17115310C/G, SHROOM3 rs11724031G/A and VDR gene SNPs were genotyped using PCR-RFLP. However, for VDR 1.5% agarose gel was used and ran for 2 hour 30 minutes at 75 V, adapted from Uitterlinden et al. (1996).

2.2.3.1 Genotyping Fucosyltransferase 2 (FUT2) rs601338G/A

This SNP was genotyped using PCR-RFLP method. The PCR reaction was set up as described above (in 2.2.2) and the cycling conditions were described on Table 2.1 to yield a 195 bp amplicon. The relative amplicon size was visualised on 1.5% agarose gel that was run at 90 V for one hour, stained with SYBR® Safe DNA Gel Stain. For RFLP, 7 µL of the PCR product was subsequently digested using 10 units of BfaI enzyme by incubating at 37 °C for 16 hours in a CutSmart®

Buffer. The enzyme has one restriction site on this amplicon in the presence of A allele, resulting in the generation of two fragments; 138 bp and 57 bp. The presence of G allele results in the loss of restriction site, leaving the amplicon as it is, 195 bp. These banding patterns were visualised on a 2.5% agarose gel stained with GelRed® Nucleic Acid Gel Stain and ran at 120 V for an hour, followed by 30 minutes at 80 V against 100 bp DNA ladder. Figure 2.1 below is a schematic diagram of where the rs601338G/A is located in the *FUT2* gene.

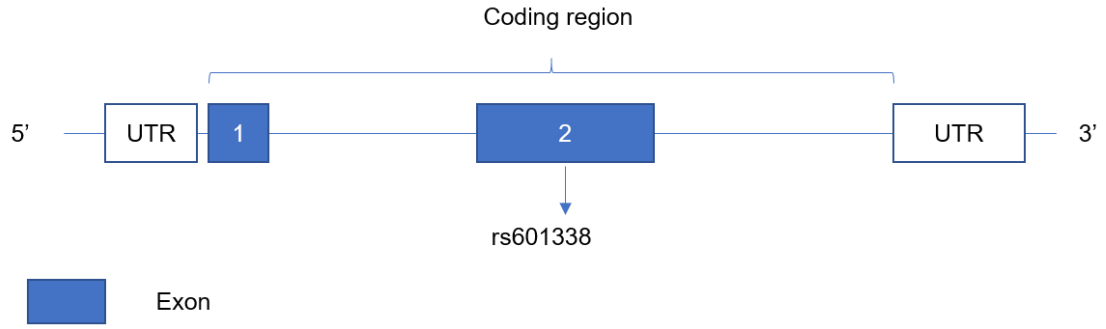


Figure 2. 1: Schematic representation of FUT2 gene. Adapted from "AceView: Gene:FUT2, a comprehensive annotation of human, mouse and worm genes with mRNAs or ESTsAceView.," (accessed 16/07/2020)

2.2.3.2 Genetic characterisation of Neural cell adhesion molecule 1 (NCAM1) rs17115310C/G

This SNP was also genotyped using PCR-RFLP method. The PCR reactions was set up to 25 µL as described above. The cycling conditions were described on Table 2.1, yielding a 351 bp amplicon. The relative amplicon size was visualised on 1.5% agarose gel that was run at 90 V for one hour, stained with SYBR® Safe DNA Gel Stain. In the step for RFLP, 10 µL of the PCR product was digested using 10 units of PvuII enzyme by incubating at 37 °C for 16 hours in a NEBuffer™ 3.1. In the presence of G allele, this enzyme has one restriction site on this amplicon, resulting into two fragments; 210 bp and 141 bp. The presence of C allele results in the loss of restriction site, resulting in no cleavage and leaving the amplicon as it is, 352 bp. To visualise these bands, 2.5% agarose gel stained with GelRed® Nucleic Acid Gel Stain was ran at 120 V for an hour, followed by 30 minutes at 80 V against 100 bp DNA ladder. Figure 2.2 below is a schematic representation of rs17115310 SNP location.

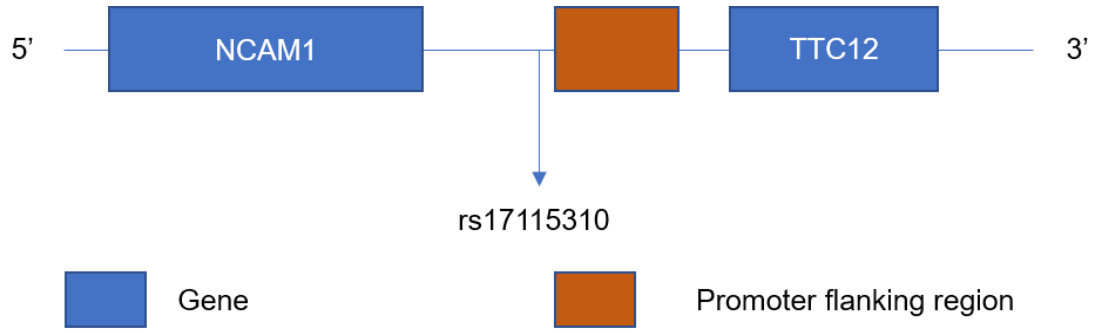


Figure 2. 2: Schematic representation of rs17115310 SNP location

2.2.3.3 Evaluating the presence of SHROOM3 rs11724031G/A polymorphism

This SNP too was genotyped using PCR-RFLP method. After the PCR described on Table 1, the resulting amplicon was 369 bp in length. To do the RFLP, 10 µL of the PCR product was digested using 10 units of HinP1I enzyme by incubating at 37°C for 16 hours in a CutSmart® Buffer. The presence of G allele brings about one restriction site for this enzyme on this amplicon, resulting in the generation of two fragments; 244 bp and 125 bp. The presence of A allele destroys the restriction site, leaving the amplicon as it is, 369 bp. This banding patterns were visualised on a 2,5% agarose gel stained with GelRed® Nucleic Acid Gel Stain and ran at 120V for an hour, followed by 30 minutes at 80V against 100bp DNA ladder

2.2.3.4 Genotyping for mutations in the vitamin D receptor (VDR) gene

Three SNPs in this gene, namely *Bsm*I, *Apa*I and *Taq*I were investigated. These SNPs, named after the enzymes that digest each SNP are relatively close to one another, making it simpler to cover them in one amplicon as shown in Figure 2.3.

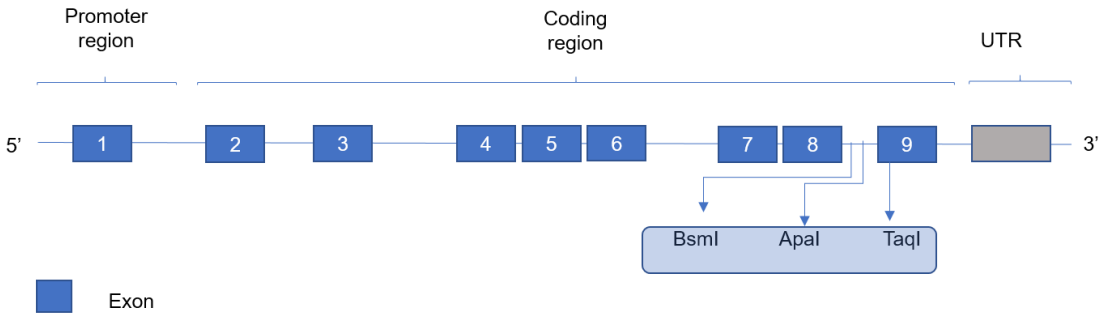


Figure 2. 3: Schematic representation of the SNPs studied in the VDR gene *(Colombini et al., 2016)*.

PCR was set up as described in Table 2.1, adapted from Uitterlinden et al. (1996) yielding an amplicon that was 2229 bp long. The PCR gel was 1% agarose, ran at 80 V for 90 minutes, stained with GelRed® Nucleic Acid Gel Stain using O'GeneRuler 1 kb DNA Ladder (ThermoScientific, Waltham, USA) as reference of migration. RFLP was run simultaneously, 5 units of each of the three enzymes in one reaction with CutSmart® Buffer. Incubation was done at 37 °C for 2 hours and at 65 °C for 2 hours.

2.2.4 DNA preparation for sequencing

In some cases, after PCR, some samples were prepared for sequencing, either to confirm the results of RFLP or using Sanger sequencing as the genotyping method. The preparation method is an enzymatic procedure that removes excess reagents unincorporated during the PCR and involves Exonuclease I (*ExoI*) and FastAP™ Thermosensitive alkaline phosphatase (ThermoScientific, Waltham, USA). *Exo*I degrades the single stranded DNA, therefore targeting unincorporated primers whereas FastAP gets rid of excess dNTPs through dephosphorylation. In each reaction, 5 µL of the PCR product, 1U of FastAP, and 4U of Exo1 were added, and the reaction volume was adjusted to 20 µL using nuclease-free water. The clean-up reaction was performed at 37 °C for an hour followed by inactivation of both enzymes at 75 °C for 15 minutes in a BioRad MyCycler Thermal cycler (BioRad Laboratories, California, USA). The success this procedure was confirmed by electrophoresing 5 µL of the product on 1.5% agarose gel stained with either SYBR® Safe DNA Gel Stain or GelRed® Nucleic Acid Gel Stain (Biotium, CA, USA) at 90 V for an hour against 100 bp plus Gene Ruler (ThermoScientific, Wilmington, USA). A clear gel picture without any nonspecific PCR products, excess primers and dNTPs was expected before commencing with downstream applications such as sequencing.

2.2.4.1 Sanger sequencing

Sanger sequencing, also known as chain termination method, is a genotyping method that determines the linear arrangement of nucleotides on a DNA

fragment. Proceeding from the post PCR clean-up step, the PCR was subjected to sequencing. Cleaned PCR products were sequenced using Big-Dye Terminator V3.1 cycle sequencing kit (Life Technologies, CA, USA). Ten microliters reaction consisted of 1.5 µL of PCR product, 2 µL of 2.5 X Big-Dye Terminator mix (Life Technologies, CA, USA), 2 µL of 5X Big-Dye Sequencing buffer (Life Technologies, CA, USA), 3.5 µL of nuclease-free water (Adcock Ingram, Johannesburg, South Africa) and 1 µL 10 mM (final concentration of 0.4 µM) primer of choice. Sequencing was carried out in an Applied Biosystems SimpliAmp Thermal Cycler (Applied Biosystems, CA, USA) with the following conditions: an initial denaturation at 98 °C for 5 minutes, followed by 40 cycles of denaturation at 96 °C for 30 seconds, annealing at 50 °C for 15 seconds, and finally extension at 60 °C for 4 minutes. Post sequencing clean-up was done with ethanol/EDTA precipitation. Capillary electrophoresis of sequenced products was done on the ABI 3730xl DNA Analyzer (Applied Biosystems, CA, USA). SNPs in the LCT (Minichromosome Maintenance Complex Component 6, MCM6) gene were genotyped using this method. It is regarded as gold standard and was also used to validate the genotypes of the SNPs characterised through PCR-RFLP method. Figure 2.4 below is a schematic representation of the SNPs upstream of the LCT gene where rs2164210 and rs7579771 are located.

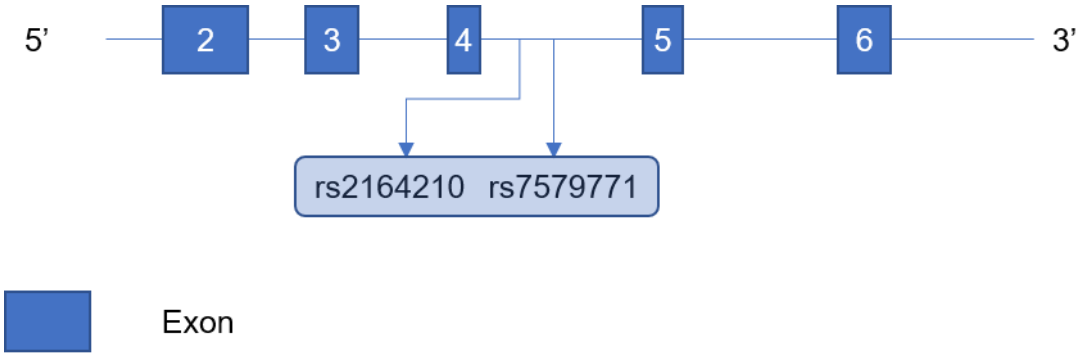


Figure 2. 4: This is a schematic representation of the SNPs upstream of the LCT gene the primers of this study amplify a fragment

Table 2. 1: PCR and RFLP conditions for all SNPs genotyped using the method

SNP	Primers (5'-3')	Cycling conditions	Amplicon size (bp)	Restriction enzyme	Recognition site	Digested fragments
FUT2 rs601338 G/A	F: GAGGAATACCGCCACATCCCGGGGGAGTAC R: ATGGACCCCTACAAAGGTGCCCGGCCGGCT	94 °C for 3 min; 94 °C for 30 s, 72 °C for 40 s, 30 cycles; 72 °C for 10 min	195	BfaI	5'...CTAG...3' 3'...GATC...5'	A: 138 bp, 57 bp G: 195 bp
NCAM1 rs17115310 C/G	F: CTGGGTTGAGAGACTGAGAAAG R: CAGCATCAGGACATGGACTAC	94 °C for 3 min; 94 °C 30 s, 57 °C for 30 s, 72 °C for 45 s, 30 cycles; 72 °C for 5 min	351	PvuII	5'...CAGCTG...3' 3'...GTCGAC...5'	G: 210 bp, 141 bp C: 351 bp
SHROOM3 rs11724031 G/A	F: CTCTTTGAAGCTGAGTAGAGT R: AACAGGATTGACAGACAAGA	94 °C for 3 min; 94 °C 30 s, 54 °C for 30 s, 72 °C for 45 s, 30 cycles; 72 °C for 5 min	369	HinP1I	5'...GCGC...3' 3'...CGCG... 5'	G: 244 bp, 125 bp A: 369 bp
VDR rs1544410 G/A	F: CAACCAAGACTACAAGTACCGCGTCAGTGA R: GCAACTCCTCATGGCTGAGGTCTC	95 °C for 3 min; 94 °C for 1 min, 60 °C for 1 min, 72 °C for 1 min, 35 cycles; 72 °C for 10 min	2229	BsmI	5'…GAATGCN…'3 5'…CTTACGN…'3	G: A:
rs7975232 C/A				ApaI	5'…GGGCCC…'3 5'…CCCGGG…'3	C: A:
rs731236 T/C				TaqI	5'…TCGA…'3 5'…AGCT…'3	T: C:
LCT rs2164210 T/C	F: GAAGTGCCTGCAAACAGATAAA R: GTCTCAAGTAGGGCTCAAGAAT	94°C for 3 min; 94°C 30 s, 53°C for 30 s, 72°C for 40 s, 30 cycles; 72°C for 10 min	574	Not applicable		
rs7579771 T/A						

2.2.5 Gut microbial characterisation by sequencing 16S ribosomal RNA (rRNA)

16S ribosomal Ribonucleic Acid (16S rRNA) gene

The 16S ribosomal RNA (rRNA) is a 1.5 kilobase pairs (kbp) gene that forms part of 30S ribosomal subunit. This gene essential in mRNA translation and is widely present in all bacteria. It contains both conserved and variable sequence regions. Culture-based methods of microbial identification are slowly becoming a thing of the past. In an era where DNA sequencing is improving and decreasing in cost, 16S rRNA sequencing technology has enabled wide spectrum identification of microbial taxa and it has now become the standard for prokaryotic taxonomy (Osman et al., 2018). It is among the most well-studied and characterised genes whose information is well documented on databases. The gene has both highly conserved regions and hypervariable regions, V1 to V9 (Osman et al., 2018). The V4 region is the most studied and MetaHIT consortium (Lozupone et al., 2013) has recommended it as a gold standard to characterise the human gut microbiome composition. Conversely, whole genome or shotgun metagenomics sequencing can be used as an alternative to 16S rRNA sequencing, however, this technique is costly compared to the 16S rRNA sequencing.

2.2.5.1 Gut bacterial DNA extraction

Analysis of the bacterial composition within the gut begins with the extraction of the DNA. In this study, the gut bacterial DNA was extracted from adults' stool samples using ZymoBIOMICS™ DNA Miniprep Kit as per manufacturer's instructions. Fifty milligrams of stool were carefully weighed in a bead-containing tube. This was proceeded by addition of 750 µL ZymoBIOMICS™ Lysis Solution and then fixed in Qiagen TissueLyser LT (Qiagen, Hilden, Germany) for 5 minutes at maximum speed. The small beads are highly effective to disrupt the bacterial cell walls and therefore maximise the yield. The tubes were centrifuged at 10000 relative centrifugal force (rcf) to separate DNA from the unwanted cell debris. The resultant supernatant with DNA was filtered through Zymo-Spin™ III-F Filter to a Collection Tube and centrifuged at 8000 rcf for 60 seconds. The use of binding buffer enabled the DNA to bind onto the column. The wash buffers removed the traces of salts and proteins before elution with DNase/RNase Free Water. Further

filtration was done in ZymoBIOMICS™ HRC Prep Solution and stored in -20 °C until further analysis. DNA was quantified using NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

In preparation of 16S rRNA library, quantitative PCR (qPCR) was performed using CFX Real Time Quantitative PCR Detection System (Biorad Laboratories, California, USA). A set of primers: 16S-F1 5′-CGAAAGCGTGGGGAGCAAA-3′; 16S-R1 5′-GTTCGTACTCCCCAGGCGG-3′ and 16S-P1 6-FAM FAM-ATTAGATACCCTGGTAGTCCA –MGB and the conditions were: 50 °C for 2 min; 95 °C for 10 min; and 45 cycles of 15 seconds denaturation at 95 °C and 1minute annealing at 60 °C.

2.2.5.2 Library Preparation

Both long and short 16S PCRs were used successively to amplify the V4 region of the 16S rRNA gene using the universal PCR primers 515F (GTGCCAGCHGCYGCGGT) and 806R (GGACTACNNGGGTNTCTAAT). To assess validity of the experiments, between run repeat, within run repeat, water used during PCR, water spiked with known bacterial DNA, bacterial mock community DNA, sample storage medium or elution buffer, sample storage medium or elution buffer spiked, and extraction controls were included throughout. The PCR conditions were as follows: initial denaturation at 95 °C for 3 minutes, cycle denaturation 95 °C for 30 seconds, annealing 50 °C for 30 seconds, extension 72 °C for 1 second then final denaturation 72 °C for 5 minutes. The difference was in number of cycles, 10 for short and 30 for long PCR reactions. The amplicons were purified using Agencourt AMPure XP System (Beckmann Coulter, Germany) on Agencourt SPRIPlate 96 super Magnet Plate. To confirm the size of the purified amplicons, they were electrophoresed in 1.5% agarose gel for 90 minutes at 110 V against Trackit 1kb Plus DNA ladder, followed by quantification using QuantiFluor™ dsDNA System. Purified amplicons were pooled, further cleaned-up and quantified using Qubit dsDNA Broad-Range (BR) Assay Kit (Thermo Fisher Scientific, USA). Library size was determined in 1.6% agarose at approximately 400bp before sending them to

Centre for Proteomic and Genomic Research (CPGR) for Illumina sequencing. The Illumina MiSeq platform generated 16S rRNA gene sequences in the form of fastq files which are compatible with Quantitative Insights Into Microbial Ecology (QIIME) pipeline.

2.3. Data analysis

2.3.1 Genotype analysis

Baseline genotype and allele frequencies were calculated for all participants using SHEsis (Shi and He, 2005). Hardy Weinberg Equilibrium was tested, also using SHEsis (Shi and He, 2005) to test for genetic variation within the participants. Linkage disequilibrium (LD) analyses to assess the non-random association of alleles at different loci of the same gene in this population was carried out using SHEsis (Li et al. 2009). Other statistical tests were done using Stata 15.1 (StataCorp LP, Texas USA).

2.3.2 16S rRNA Data processing

Raw sequence data from Illumina was analysed using QIIME 2 2019.4 release (Bolyen et al., 2019). The visual computations were carried out through facilities provided by the University of Cape Town's ICTS High Performance Computing team (UCT eResearch HPC, accessed 02 November 2019). Stages of microbiome data analysis are shown in Figure 2.5. The sequences were already demultiplexed at CPGR. In this stage, multiple sequences that were pooled together during the library preparation are traced back to the individual samples using the DNA barcodes which are unique to each sample. This stage of demultiplexing also includes the removal of these barcodes together with the primers.

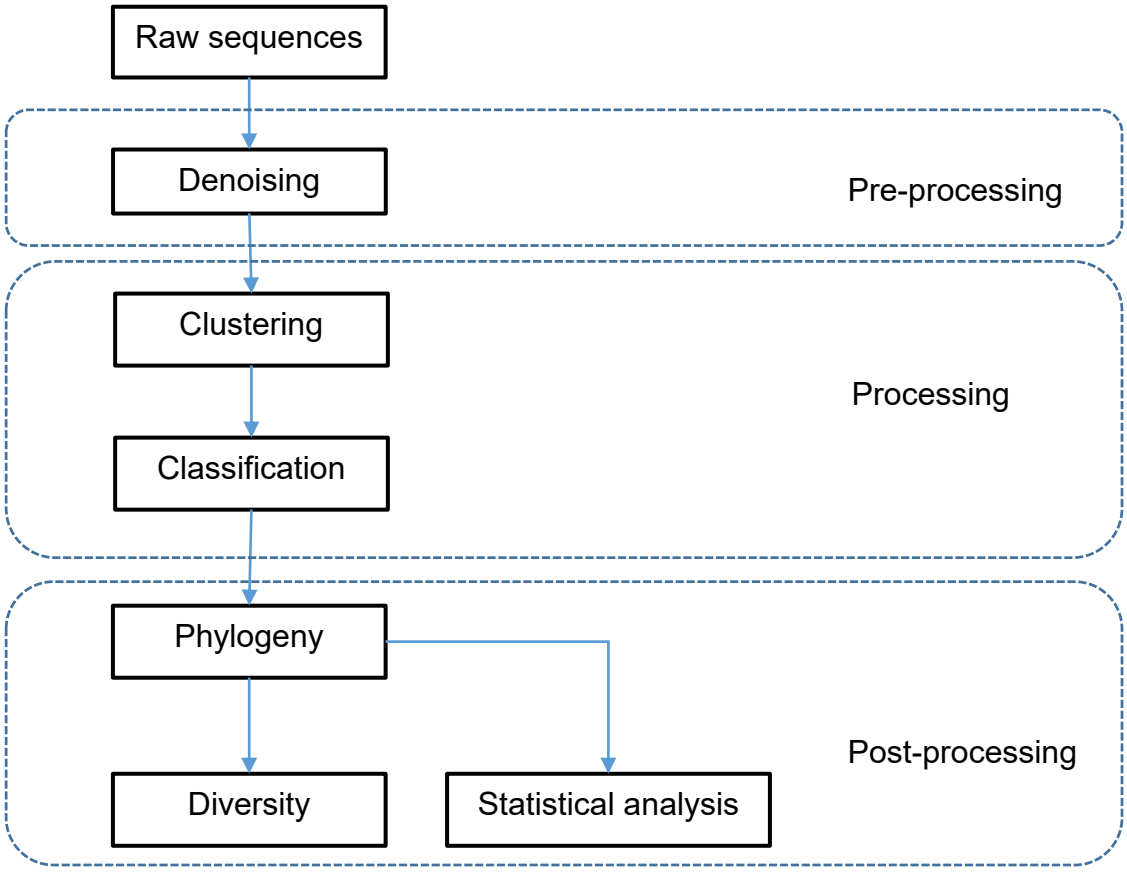


Figure 2. 5: Microbiome data analysis workflow. Adapted from Kachambwa (2017)

After demultiplexing, comes the stage of denoising using DADA2 for sequence quality control. This stage involves removal of noisy sequences. During PCR, two or more sequences of different parental origins may hybridize and become chimeric which do not exist in nature (de la Cuesta-Zuluaga and Escobar, 2016). DADA2 package automatically removes these chimeric sequences. Furthermore, isolated sequences that are only observed once, called singletons, are removed as they are considered erroneous by creating many false taxa, thereby overestimating diversity (Auer et al., 2017). Denoised sequence reads (forward and reverse) are joined in a paired-end manner, and then dereplicated to reduce

computational time and file volume requirements in subsequent analysis producing amplicon sequence variants (ASVs).

The stage of dereplication is also part of clustering. However, in the stage of clustering, there are three strategies of clustering sequences: de novo, closed- and open-reference clustering. In this study, de novo clustering was performed at 99% similarity to create 99% operational taxonomic units (OTUs).

After clustering, the last stage of processing is classification, where organisms present in the samples are identified. This is done by aligning our ASVs with reference database of sequences with known taxonomy. In this study, GreenGenes dataset was used to for classification. Phylogeny was created and later used for diversity analysis. Biological Observation Matrix (BIOM) format was generated for further statistical analysis.

Subsampling was performed at the lowest sequence depth (Horner-Devine et al., 2004; Cárcer et al., 2011) to generate rarefaction curves for the estimation of gut bacterial richness. Alpha diversity was assessed using Shannon index and further supported using observed OTUs, Faith's PD and Pielou's evenness. Principal Coordinates Analysis (PCoA) (weighted UniFrac distance) was used to measure beta diversity in the gut bacteria composition. Both alpha and beta diversities were tested using permutation multivariate analysis of variance (PERMANOVA).

Potential bacterial taxa in the study drug groups and their genetics were identified using Linear discriminant analysis Effect Size (LEfSe) tool (Segata et al., 2011). The alpha value for the factorial Kruskal-Wallis test among classes was set at 0.05. The threshold of the logarithmic Linear discriminant analysis (LDA) score for discriminative features was set at the default 2.0. To account for multiple comparisons, p-values were adjusted based on the false discovery rate (FDR) method. The resulting q-values less than 0.2 were considered significant as in most studies (Human Microbiome Project Consortium, 2012; Aagaard et al., 2013; Onywera et al., 2019).

CHAPTER THREE: RESULTS

3.1. Demographic and clinical characteristics of participants

The study cohort comprised a total of 647 participants of which 265 were on antiretroviral therapy (ART), 154 on warfarin and 228 were not on any therapy. Of the 467 participants, only 93 had stool samples from which microbiome samples could be analysed. Table 3.1 shows demographic features only for the participants who were further characterised for gut microbiome. All participants were above 18 years of age and the average age was 33 years. The cohort had an average BMI of 27 kg/m^2 of which only one participant was underweight. The normal BMI ranges from 18.5 to <25 kg/m^2 and our cohort was overweight. There were 92% women and the remaining 8% of men were from the warfarin group. All the participants were from generally the same geographical area and consumed mainly maize meal porridge.

Table 3. 1 Demographic parameters for the participants include in this study

	Combined n=93	**Untreated control n=34**	**Efavirenz n=38**	**Warfarin n=21**
Age Median (IQR)	31 (26-38)	27 (23-33)	30 (26-34)	46 (41-52)
BMI Median (IQR)	26.15 (24.05-29.1)	26.5 (24.7-31.2)	25.7 (24-28)	27.94 (22.86-33.06)
Normal	22.7 (21.75-24.2)	22.75 (22.2-24.5)	23.6 (22.2-24.3)	20.21 (19.98-22.86)
Overweight	27.5 (25.9-28.3)	26.6 (25.6-28.3)	27.7 (25.9-28.4)	27.94 (26.84-28.09)
Obese	34.55 (32.33-37.05)	33.6 (31.2-36)	35.6 (34.2-39.1)	35.35 (33.06-39.79)
Sex	n=93	n=34	n=38	n=21
Female n (%)	86 (0.92)	34 (1.00)	38 (1.00)	14 (0.67)

Key: Only one person was under-weight with a BMI of 17.3. IQR= interquartile range, BMI= body mass index.

3.2 Quality, quantity and success of biological material preparation

Genomic DNA was retrieved from the Pharmacogenomics biorepository. DNA purity and concentration were measured using a NanoDrop machine for all the available samples (647), 154 of warfarin group, 265 efavirenz group and 228 of the unexposed group. Furthermore, DNA integrity check was done in all 647 samples using gel electrophoresis and the DNA was still intact in all samples. The DNA was used for characterisation of variation of the selected polymorphism associated with gut microbiome profiles.

3.3 Genotyping for single nucleotide polymorphisms

Primers were designed as described in section 2.2.1 to amplify the DNA regions in/near FUT2, NCAM1, SHROOM3 and VDR genes. All the PCRs were successfully optimised, however the PCRs for the SNPs in the VDR and LCT genes required an additional 0.5mM of magnesium (Mg^{+}) to maximise the yield. Following the successful PCRs, RFLP and Sanger sequencing genotyping techniques were used as described in section 2.3 and 2.5 respectively. Below we show examples of successful PCR (Figure 3.1), RFLP (Figure 3.2) and Sanger sequencing techniques (Figure 3.3), respectively.

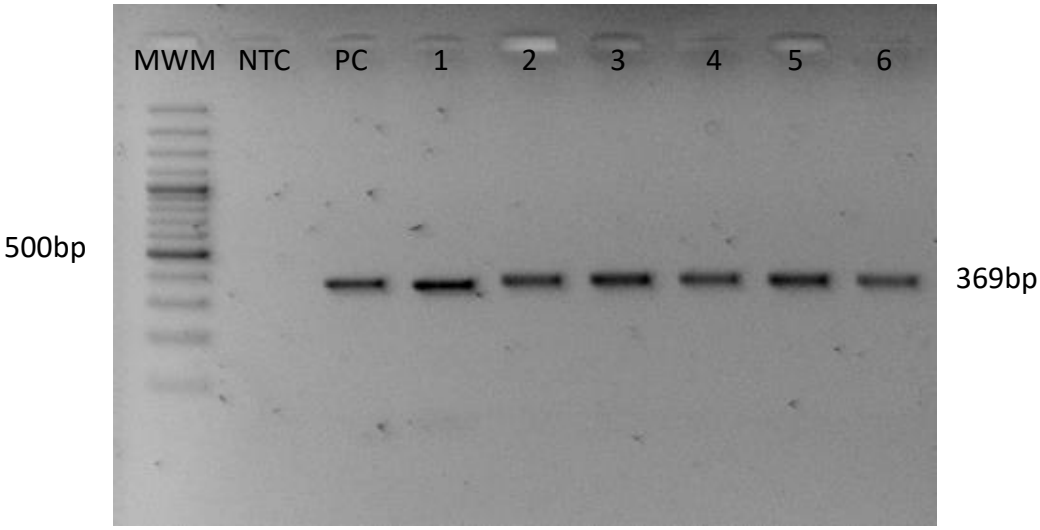


Figure 3. 1: A representative picture of a gel showing PCR fragments for SHROOM3 rs11724031G/A polymorphism. 1.5% agarose gel stained with SYBR® Safe DNA Gel Stain electrophoresed for 1 hour at 90V, resulting in a 351bp fragment. The wells are in the order of molecular weight maker (MWM) of 100bp, non-template control (NTC), positive control (PC) and samples (1-6).

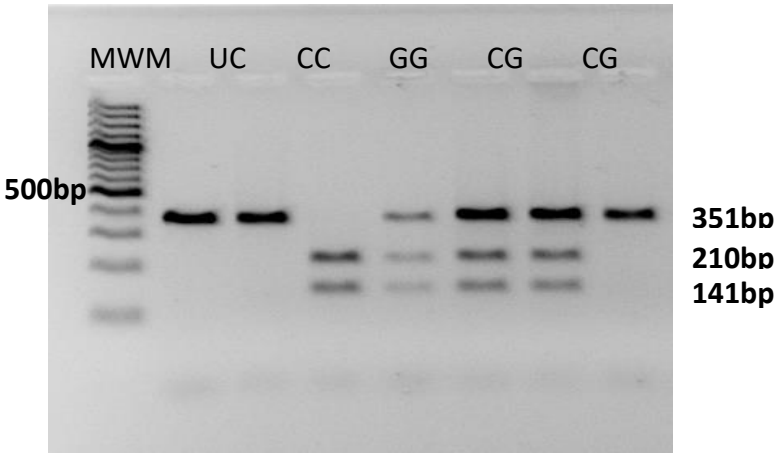


Figure 3. 2: A representative gel showing a RFLPon a 2.5% GelRed stained agarose for NCAM1 rs17115310C>G SNP after digestion with PvuII restriction enzyme. The G allele digests into two fragments while the C allele doesn't have the restriction site for this enzyme. The gel was ran at 120 V for an hour, followed by 30 minutes at 80 V.

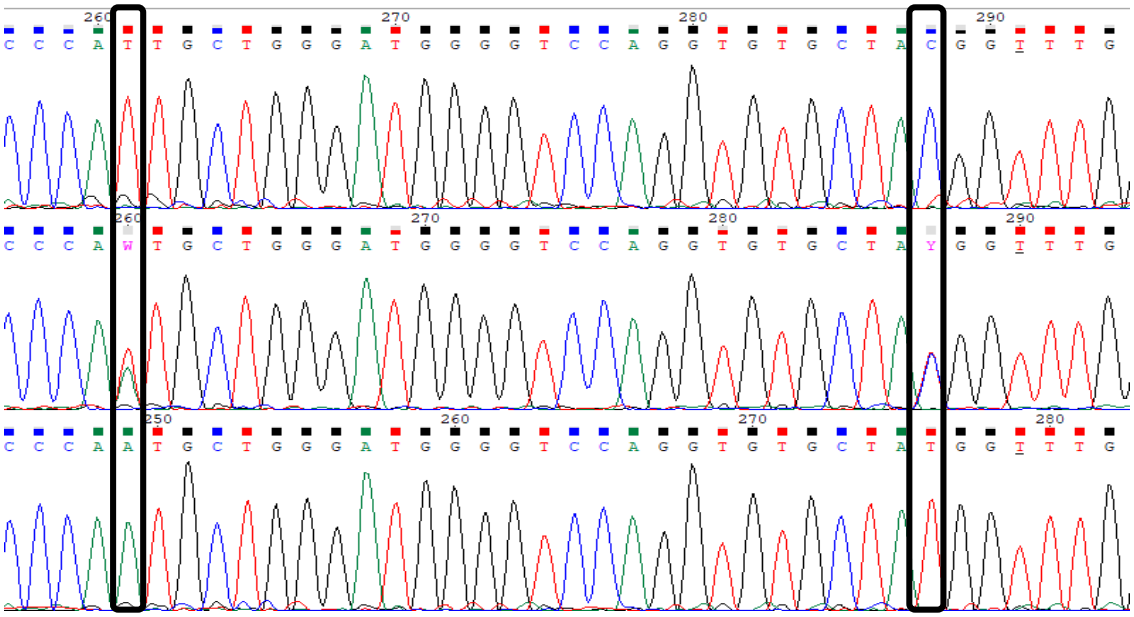


Figure 3. 3: A representative electropherogram DNA sequence analysis for the SNPs rs2164210C and rs7579771T of LCT. Each peak and a colour represent a nucleotide. The highlighted regions of the electropherogram indicate variation in these SNPs, respectively.

3.4 Distribution of genotypes for SNPs associated with microbiome profiles

After genotyping candidate SNPs associated with gut microbiome profiles, genotype calls were made from the RFLP gels (Figure 3.2) and electropherogram (Figure 3.3) and shown in Table 3.2. Most of the SNPs were in Hardy-Weinberg equilibrium (HWE) except rs7975232 of VDR and rs7579771 of LCT.

Table 3. 2 Genotype distribution for variants associated with microbiome profiles as observed in the Zimbabwean population cohort.

SNP Genotype or allele	Combined, N (freq)	Untreated, N (freq)	Efavirenz treated, N (freq)	Warfarin treated, N (freq)
FUT2 rs601338				
G/G	199 (0.308)	69 (0.303)	78 (0.294)	52 (0.338)
G/A	336 (0.519)	122 (0.535)	142 (0.536)	72 (0.468)
A/A	112 (0.173)	37 (0.162)	45 (0.170)	30 (0.195)
A	0.433	0.430	0.438	0.420
NCAM1 rs17115310				
C/C	380 (0.587)	128 (0.561)	159 (0.600)	93 (0.604)
C/G	232 (0.359)	88 (0.386)	91 (0.343)	53 (0.344)
G/G	35 (0.054)	12 (0.053)	15 (0.057)	8 (0.052)
G	0.233	0.246	0.228	0.224
SHROOM3 rs11724031				
G/G	424 (0.655)	158 (0.693)	169 (0.638)	97 (0.630)
A/G	198 (0.306)	64 (0.281)	81 (0.306)	53 (0.344)
A/A	25 (0.039)	6 (0.026)	15 (0.057)	4 (0.026)
A	0.192	0.167	0.209	0.198
VDR rs1544410				
G/G	332 (0.554)	126 (0.633)	127 (0.502)	80 (0.544)
A/G	228 (0.381)	61 (0.307)	108 (0.427)	58 (0.395)
A/A	39 (0.065)	12 (0.060)	18 (0.071)	9 (0.061)
A	0.255	0.214	0.285	0.259
VDR rs7975232				
A/A	292 (0.487)	106 (0.533)	123 (0.486)	
A/C	291 (0.486)	91 (0.457)	116 (0.458)	
C/C	16 (0.027)	2 (0.010)	14 (0.055)	
C	0.270	0.239	0.285	
VDR rs731236				
T/T	344 (0.574)	123 (0.618)	128 (0.506)	93 (0.633)
T/C	220 (0.367)	62 (0.312)	112 (0.443)	46 (0.313)
C/C	35 (0.058)	14 (0.070)	13 (0.051)	8 (0.054)
C	0.242	0.226	0.272	0.211
LCT rs2164210				
T/T	49 (0.681)	15 (0.750)	26 (0.743)	13 (0.619)
T/C	19 (0.264)	4 (0.200)	7 (0.200)	6 (0.286)
C/C	4 (0.056)	1 (0.050)	2 (0.057)	2 (0.095)
C	0.188	0.150	0.157	0.238
LCT rs7579771				
A/A	42 (0.583)	13 (0.650)	23 (0.657)	10 (0.476)
A/T	18 (0.250)	4 (0.200)	9 (0.257)	4 (0.190)
T/T	12 (0.167)	3 (0.150)	3 (0.086)	7 (0.333)
T	0.292	0.250	0.214	0.429

Key: FUT2: fucosyltransferase 2; NCAM1: neural cell adhesion molecule 1; SHROOM3: Shroom Family Member 3; VDR: vitamin D receptor; LCT: Lactase

3.5 Baseline allele frequencies and their comparison to other populations

All eight SNPs in/near the five genes known to shape the gut microbiome were characterised for. The allele frequencies of this study were calculated and then compared to six other populations as per The 1000 Genomes Project Consortium (2015) accessed via Ensembl (Cunningham et al., 2019) as shown in Table 3.3. The populations include four African, the African American (African Ancestry in Southwest US), Luhya in Webuye, Kenya, Mende of Sierra Leone and Yoruba of Ibadan Nigeria; and two non-African, East Asian and Europe in general.

FUT2 rs601338A had the highest allele frequency (43%) as compared to other SNPs in our Zimbabwean Cohort. This was followed by LCT rs7579771T (29%), VDR rs7975232C (27%), VDR rs1544410T (26%), VDR rs731236G (24), NCAM1 rs17115310G (23%); lastly SHROOM3 rs11724031A and LCT rs2164210C (both at 19%). The MAF of this study did not show any difference from that of the African American population in all the SNPs except rs7975232C variant of the VDR gene. The Luhya population allele frequencies only diverted from the Zimbabwean population on SNP rs2164210C of the LCT and is expected because they are both of African origin. When comparing the Zimbabwean population with that of Mende, it was only rs17115310G of NCAM1, rs7975232C of VDR and rs2164210C of LCT that reached the statistical significance. Unexpectedly, the Yoruba population showed a drift in allele frequencies of many SNPs (rs601338A, rs17115310G, rs7975232C, rs731236G, rs2164210C and rs7579771T) as compared to the Zimbabwean population of this study although they are both African populations.

Table 3. 3: Allele frequency distribution and comparison between the current study and four other populations as published on The 1000 Genomes Project Consortium (2015) online database.

Variant	Current study	African American	Luhya	Mende	Yoruba	European	East Asian
rs601338A	0.433	0.508	0.439	0.388	**0.542 ****	0.441	**0.004 ****
rs17115310G	0.233	0.213	0.222	**0.312***	**0.315 ****	0.201	**0.013 ****
rs11724031A	0.192	0.139	0.182	0.135	0.222	**0.098 *****	**0.139 ****
rs1544410T[#]	0.255	0.270	0.253	0.235	0.296	**0.404 *****	**0.064 ****
rs7975232C[#]	0.270	**0.369***	0.283	**0.371 *****	**0.398 ****	**0.445 *****	**0.709 ****
rs731236G[#]	0.242	0.246	0.268	0.247	**0.310***	**0.400 ****	**0.067 ****
rs2164210C[##]	0.188	0.246	**0.101***	**0.088 ****	**0.046 *****	**0.595 ****	**0.400 ****
rs7579771T[##]	0.292	0.402	0.278	0.206	**0.157 ****	**0.595 *****	**0.400 ***

[#] indicates n=72 were characterised, [##] indicates n=599 were characterised, others n=647 was characterised for the others. *p<0.05, **p<0.01, ***p<0.001

It was also observed that the frequencies of alleles rs601338A and rs17115310G in this study (43% and 23%, respectively) did not differ from the existing ones of the European population (44% and 20%, respectively) and the rest showed a significant difference. In addition to the comparison of the allele frequencies between the current study and the European populations, all the variants of all the SNPs showed statistically significant different allele frequencies compared to those of the East Asian population. This can be expected from the two populations since they are not African.

3.6 Linkage disequilibrium (LD) analysis

Haplotypes were generated for SNPs that were on the same genes and calculated using SHEsis (Shi and He, 2005; Li Z et al., 2009). Figure 3.4 shows the LD plots for LCT gene. The two SNPs had high D' (D'=0.94) but with average R^2 (R^2=0.50) due to difference in allele frequency. The haplotypes for the two SNPs (rs17115310T>C - rs7579771A>T) showed that the T-A was the most common (0.70) followed by C-T (0.18), T-T (0.11) and the least common being C-A (0.01).

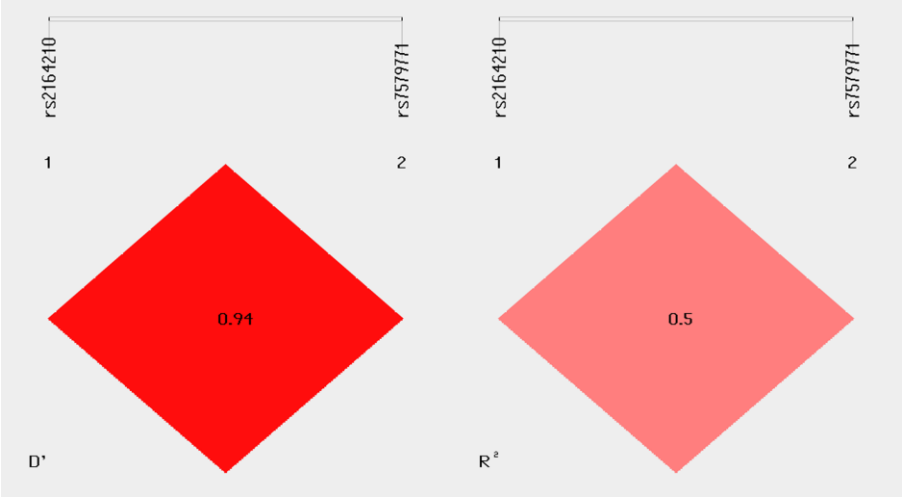


Figure 3. 4: Linkage disequilibrium plots for LCT variants

Figure 3.5 shows the LD plots for SNPs in the VDR gene. The D' is very strong between the *Bsm*I and *Apa*I (D'=0.95) polymorphisms; and between *Apa*I and TaqI (D'=0.91) polymorphisms. However, these findings are conflicted with the R^2 which shows that the SNPs are in weak LD. The GAT haplotype was found to be the most common at 39%.

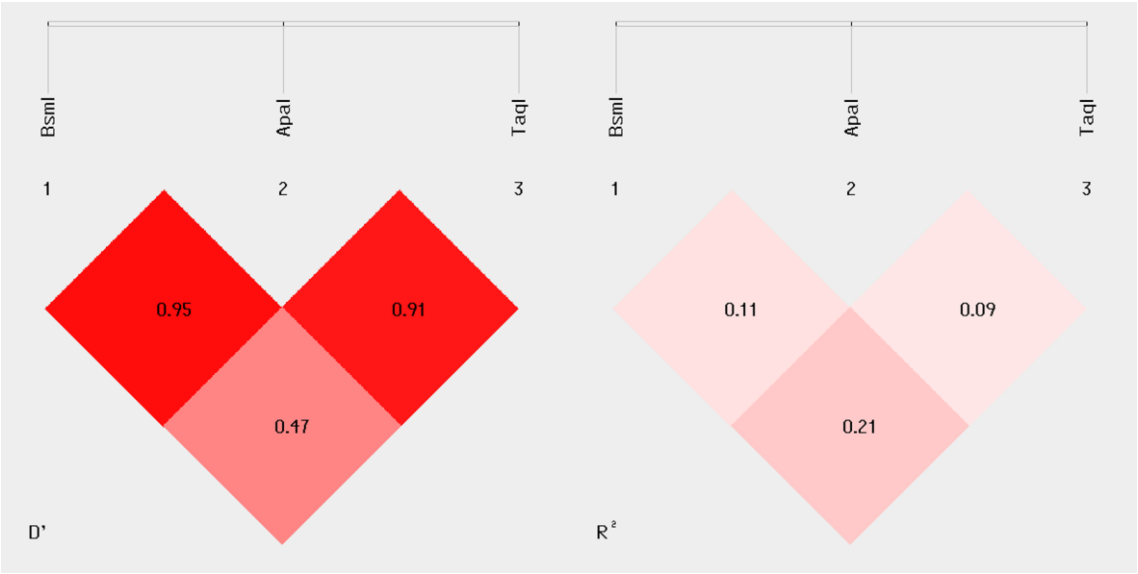


Figure 3. 5: Linkage disequilibrium plots for SNPs in the VDR gene

Haplotype frequencies from the three SNPs in the VDR gene (rs1544410-rs7975232-731236) were as follows, A-A-C (0.14), A-A-T (0.11), G-C-T (0.26), G-A-T (0.39), and G-A-C (0.091), respectively.

3.7 Variation on efavirenz and warfarin plasma levels

From the 93 patients that were characterised for gut microbiome, plasma drug levels were available for 54 participants (i.e., 21 on warfarin and 33 on efavirenz treatment). With this data available, this study went further to evaluate the genetic models on the distribution of both efavirenz and warfarin although none of these SNPs had been reported to have any pharmacogenetic significance.

3.7.1 Distribution of efavirenz plasma concentration

The mean efavirenz levels was 2.78 µg/ml. The normal therapeutic range is from 1 to 4 µg/mL. All the 33 patients were females. Most (85%) of the efavirenz levels in this study were below the acceptable maximum level of 4 µg/mL. Only 15% (n=5/33) of the participants had efavirenz plasma level above 4 µg/mL, 58% (n=19/33) were within the normal range (1 to 4 µg/mL) whereas the remaining 27% (n=9/33) of the participants had efavirenz plasma levels below 1 µg/mL. Figure 3.6 shows the distribution of the log of the efavirenz plasma levels.

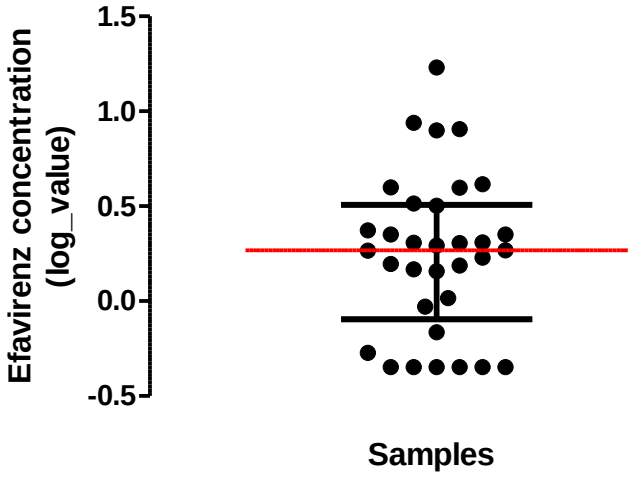
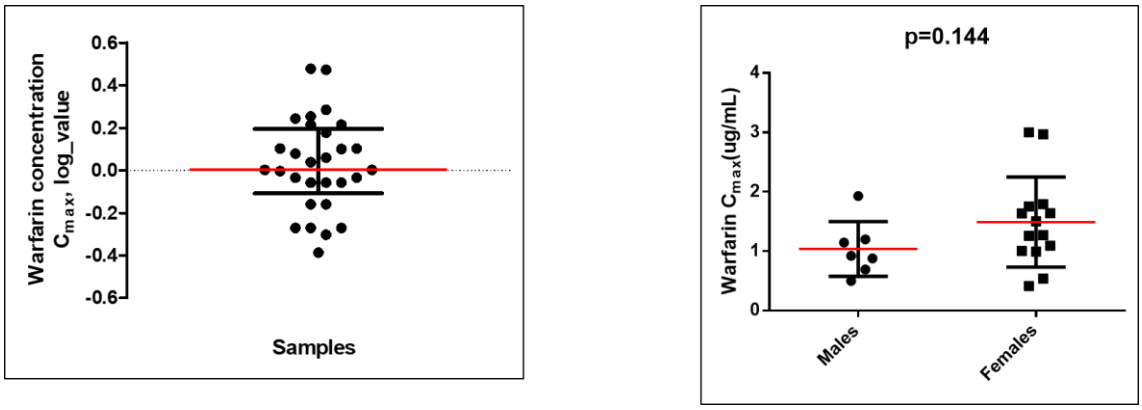


Figure 3. 6: Distribution of efavirenz plasma concentration among the participants

3.7.2 Distribution for warfarin maximum plasma concentration (C_{max})

Warfarin C_{max} levels was available for 21 participants who were on warfarin. The mean C_{max} was 1.34 µg/mL, ranging from 0.411 to 3.001 µg/mL; this was shown in Figure 3.7A. Warfarin C_{max} levels were further divided based on the sex, shown in Figure 3.7B. The mean of the male group (n=7/21) was 1.04 ug/mL whereas the female group (n=14) had a mean of 1.49 ug/mL. The Mann Whitney test was done on the concentrations for the two groups and there was no statistically significant difference between the two (p=0.144).



A B

Figure 3. 7: Distribution of the warfarin plasma concentration

Genetic models were tested on the plasma concentrations of both efavirenz and warfarin groups as shown on Table 3.4. It is observed that genetic variation in host genes NCAM1, SHROOM3, VDR and LCT did not affect plasma drug levels for both efavirenz and warfarin, however, there was a significant effect of genetic variation in FUT2 on warfarin levels (p=0.036).

Table 3. 4: Effects of host genetic variants that affect microbiome profiles, on the plasma levels of efavirenz and warfarin

	EFV Median (IQR)	P-value	Warfarin Median (IQR)	P-value
FUT2 rs601338				
GG*	1,850 (0,533-2,356)		1,006 (0,615-1,351)	
GA	1,959 (0,934-3,694)	0,844	1,259 (0,958-1,772)	0,162
AA	1.036 (1.036-1.036)	-	1.927 (1.267-3.001)	0.036
NCAM1 rs17115310				
CC*	1,570 (0,492-2,809)		1.092 (0.730-1.840)	
CG	2,034 (1,644-3,969)	0,212	1.229 (0.944-1.604)	0,914
GG	2.099 (1,644-13,340)	0,2488	1,502 (1,502-1,502)	-
SHROOM3 rs11724031				
GG*	1.656 (0.4708-2.188)		1.146 (0.765-2.117)	
GA	3.219 (1.514-4,995)	0,161	1.199 (0.877-1.752)	0,939
AA	1.570 (0.692- 1.969)	0,521	1.502 (1.502-1.502)	-
VDR rs1544410				
GG*	1,770 (0,872-3,761)		1,049 (0,831-1,354)	
GA	2.024 (0.450-2.188)	0,541	1.634 (1.035-2.447)	0.113
AA	6.015 (3.975-8.055)	-	0.411 (0.411-0.411)	-
VDR rs7975232				
AA*	1,844 (0,450- 2,239)		1,070 (0,476-1,826)	
AC	2,101 (1,547- 3,896)	0,100	1,229 (0,912-1,667)	0,536
CC	1,468 (1,468- 1,468)	-	-	-
VDR rs731236				
TT*	1.555 (0.627-3.761)		1,146 (0,935-1,636)	
TC	1.989 (0,824-2,190)	0.7343	1.095 (0.623-2.976)	0,965
CC	6.091 (4.126-8.055)	0.1287	1.792 (1.792-1.792)	_
LCT rs2164210				
TT*	1,617 (3,586- 0,509)		1.259 (0.9580-1.783)	
TC	1,997 (0,813- 2,240)	1,000	1.103 (0.831-1.664)	0,630
CC	4,126 (4,126- 4,126)	-	1.102 (0.4110-1.792)	-

LCT rs7579771				
AA*	1,695 (0,450-3,262)		1.263 (0.879-1.970)	
AT	1,959 (0,934-2,242)	0,931	1.379 (0.771-1.883)	0.945
TT	4,126 (4,126-4,126)	-	1.146 (0.8770-1.634)	0.475

* represents a reference genotype for that SNP

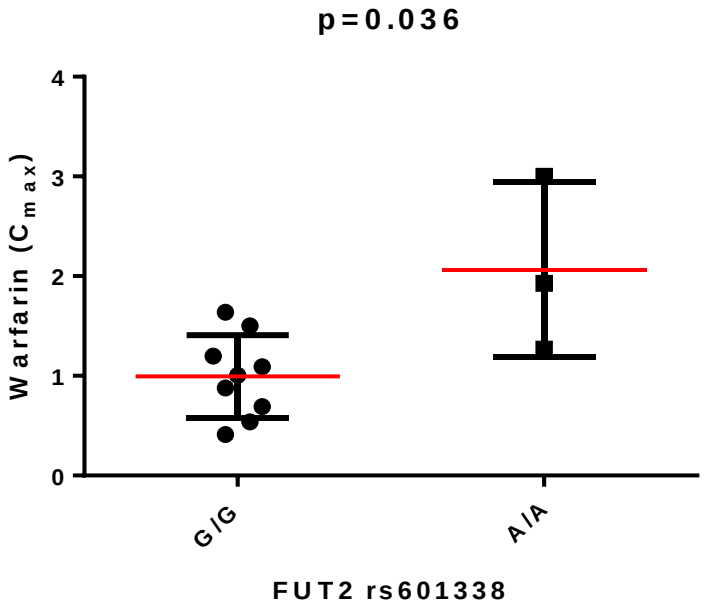


Figure 3. 8: Comparison of warfarin maximum concentration between GG and AA genotypes of rs601338 SNP.

3.8 Gut bacterial profiles

Bioinformatics pipeline, QIIME 2 2019.4 release, was used to further analyse the demultiplexed sequences from Illumina. High quality reads (2979 reads per sample) cut-off was selected for rarefaction curves for a satisfactory coverage of bacterial diversity within the samples. Contaminants from the kit were identified and removed.

3.8.1 Sample quality control

In pre-processing stage, noisy sequences were removed successfully for sequence quality control. This is shown in Figure 3.9 as a downtrend from input samples to filtered, denoised, and merged until the removal of chimeras. It can be observed that the non-chimeric are almost similar.

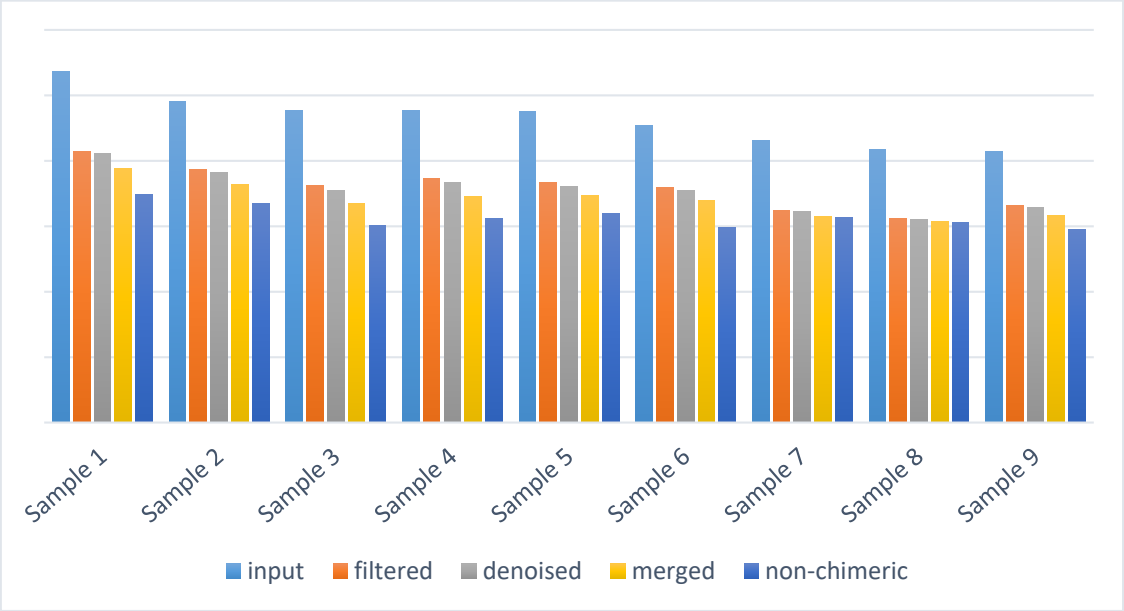


Figure 3. 9: Samples demonstrating the success of denoising of gut microbiome pre-processing stage, the y-axis is an estimate quantity of gut microbiome.

3.8.2 Alpha diversity

Alpha diversity is the measure of species variance within a sample, in this case, the diversity within a participant’s gut microbiome. Alpha diversity is measured in different indices, such as Shannon, Simpson, Dominance and Faith's Phylogenetic diversity (PD). Alpha diversity rarefaction curves that reached a considerable plateau and this includes Pielou’s evenness, Simpson, Shannon, dominance, observed OTUs, Faith’s PD, chao1 and Good’s coverage.

Assessment of alpha diversity within the three groups

The gut microbiome was firstly clustered according to the drug groups under study, that is untreated control, efavirenz and warfarin groups. In this study, QIIME2 was used to compute the alpha diversity difference within our groups via Shannon, observed OTUs, Faith's PD and Pielou's evenness. The microbiome composition of the untreated control and efavirenz groups was significantly more diverse than the warfarin group (observed OTUs and Faith's PD). However, diversity and richness of the gut microbiome in the untreated control group was not significantly different to that of the efavirenz (p&q = 0.910) and warfarin (p = 0.795, q = 0.910), as per the Shannon's index. The number of observed OTUs were significantly lower in the gut microbiome of warfarin group as compared to that of untreated control group (p = 0.014, q = 0.021), however, there was no difference in OTUs observed when comparing the gut microbiome of the untreated control and efavirenz groups (p&q = 0.964). Phylogenetic diversity shows the richness of gut microbiome in the warfarin group to be less diverse when compared to the untreated control group (p = 0.033, q = 0.050), but there was no significant difference when comparing the untreated control and efavirenz groups (p&q = 0.955). The richness in the gut microbiome of the warfarin group was significantly even (Pielou's evenness) than in the untreated control (p = 0.010, q = 0.015). Nevertheless, Pielou's evenness was not significantly different between the untreated control and efavirenz groups (p&q = 0.883).

Alpha diversity of efavirenz plasma concentration

Breaking down the efavirenz group further, participants were further divided into three groups according to efavirenz plasma concentrations, those with normal plasma concentrations (1-4 µg/mL), EFV < 1 µg/mL and those with efavirenz plasma concentration above the therapeutic limit, EFV > 4 µg/mL. There was no indication of difference in gut microbiome composition between those within the normal therapeutic range and the sub-therapeutic range with all the alpha diversity indices, Shannon (p&q = 0.814); observed OTUs (p = 0.289, q = 0.525); Faith's PD (p = 0.258, q = 0.579) and Pielou's evenness (p&q = 0.480). The alpha diversity between those within the therapeutic range and those above showed Shannon (p = 0.377, q = 0.695); observed OTUs (p&q = 0.587); Faith's PD (p&q = 0.786) and Pielou's evenness (p = 0.103, q = 0.309).

Alpha diversity due to warfarin C_{max}

Participants in the warfarin group were further categorised on whether they had C_{max} ≤ 1.34 ug/mL or C_{max} > 1.34 ug/mL, of which 1.34 ug/mL was the mean concentration. None of the alpha diversity indices showed difference in the gut microbiome composition, Shannon (p&q= 0.436); observed OTUs (p&q= 0.586); Faith's PD (p&q= 0.815) and Pielou's evenness (p&q= 0.350).

Comparison of alpha diversity relative to the host genetics

Three genetic models, that is, additive, dominant and recessive, in four groups, that is the combined, untreated control, efavirenz and warfarin; were employed in QIIME2 platform to test for association using the permutational multivariate anova (PERMANOVA) statistical tests. Haplotypes of SNPs found in one gene were added into the analysis to assess the effect of the combined SNPs on the gut microbiome. Difference in gut microbiome composition in this study were assesses by comparing the alpha diversity using Faith's PD, observed OTUs and Pielou's evenness.

In the combined group, the Faith's PD and observed OTUs indices did not reach statistical difference in all the genetic models as shown in Figure 3.5. However, the abundance was evenly distributed between the GG and GA genotypes of VDR rs1544410 ($p = 0.005$, $q = 0.014$). Pielou's evenness indicated the difference in gut microbiome composition in the VDR rs1544410 and LCT rs7579771 recessive models (p&q = 0.004; 0.019, respectively). The alpha diversity in untreated control group (Table 3.6) significantly varied with the recessive model of NCAM1 rs17115310, p&q = 0.041 and p&q = 0.049 in both Faith's PD and observed OTUs, respectively. Nonetheless, there was no significant difference as per Pielou's evenness index (p&q = 0.940). Comparison of the Faith's PD of participants with AA and TT genotypes of the LCT rs7579771 SNP was at a cut-off of statistical significance, $p = 0.051$, however it drifted away from significance after adjustment, $q = 0.077$. Again, Faith's PD significantly varied in the dominant model of LCT rs7579771 SNP (p&q = 0.028). In the efavirenz group (Table 3.7), none of the genetic models' Faith's PD and observed OTUs of the alpha diversity

differed statistically. The evenness between the TT and TC genotypes and recessive model of VDR rs731236 was significantly different (p = 0.014, q = 0.043 and p&q = 0.009 respectively). However, in the warfarin group (Table 3.8), FUT2 rs601388's GG vs AA (p = 0.021, q = 0.062) and dominant (p&q = 0.021) models indicated difference in Pielou's evenness index. LCT rs7579771's recessive model showed variation in the phylogenetic diversity, p&q = 0.028 and accompanied by difference in observed OTUs, p&q = 0.009. For the same SNP, comparison of the AA and TA genotypes indicated that more OTUs were observed in TA than in the AA genotype (p = 0.034). Furthermore, more OTUs were observed in the TT genotype than in AA (p = 0.040), however, the significance was lost following adjustment in both comparisons (q = 0.061). The gut microbiome for participants with GG genotype of VDR rs1544410 was significantly more even when compared to the GA genotype (p = 0.014, q = 0.043). Also, there was significant difference in the evenness of the recessive model for this very same SNP (p&q = 0.010). When it comes to VDR rs731236 SNP, the pairwise Kruskal-Wallis showed that the gut microbiome taxa were significantly less even in the participants with TT genotype than those with the TC genotype (p = 0.054), however, the significance was lost with the adjustment for false discovery (q = 0.161). Moreover, significant difference was observed in the evenness of the recessive model for rs731236 SNP (p&q = 0.043).

Table 3. 5: Alpha diversity by candidate SNPs in the general population.

	P values								
Genotypes/ haplotypes	Shannon		Faith's PD			Observed OTUs		Evenness	
FUT2 rs601388									
GG*					0.545				
GA	0.692	0.692	0.341	0.671		0.500	0.830	0.939	0.939
AA	0.327	0.692	0.982	0.982		0.553	0.830	0.305	0.604
GG+GA/AA	0.408	0.408	0.617	0.617	0.617	0.737	0.737	0.343	0.343
GG/GA+AA	0.562	0.562	0.393	0.393	0.393	0.461	0.461	0.760	0.760
NCAM1 rs17115310									
CC*					0.186				
CG	0.483	0.483	0.385	0.385		0.960	0.960	0.265	0.795
GG	0.205	0.308	0.076	0.228		0.121	0.328	0.825	0.825
CC+CG/GG	0.171	0.171	0.107	0.107	0.107	0.139	0.139	0.968	0.968
CC/CG+GG	0.819	0.819	0.183	0.183	0.183	0.595	0.595	0.287	0.287
SHROOM3 rs11724031									
GG*					0.346				
GA	0.261	0.784	0.137	0.412		0.079	0.235	0.919	0.919
AA	0.907	0.907	0.815	0.815		0.661	0.661	0.539	0.919
GG+GA/AA	0.767	0.767	1.000	1.000	1.000	0.941	0.941	0.591	0.591
GG/GA+AA	0.347	0.347	0.169	0.169	0.169	0.092	0.092	0.766	0.766
VDR rs1544410									
GG*					0.724				
GA	0.192	0.577	0.466	0.803		0.908	0.908	0.005	0.014
AA	0.418	0.627	0.871	0.871		0.770	0.908	0.243	0.365
GG+GA/AA	0.540	0.540	0.759	0.759	0.759	0.706	0.706	0.524	0.524
GG/GA+AA	0.149	0.149	0.555	0.555	0.555	1.000	1.000	0.004	0.004
VDR rs7975232									
AA*					0.482				
AC	0.565	0.957	0.250	0.751		0.410	0.872	0.660	1.000
CC	0.872	0.957	1.000	1.000		0.717	0.872	0.809	1.000
AA+AC/CC	0.921	0.921	0.742	0.742	0.742	0.805	0.805	0.921	0.921
AA/AC+CC	0.564	0.564	0.267	0.267	0.267	0.399	0.399	0.648	0.648
VDR rs731236									
TT*					0.408				
TC	0.515	0.570	0.237	0.658		0.669	0.669	0.389	0.778
CC	0.509	0.570	0.439	0.658		0.485	0.669	0.778	0.778
TT+TC/CC	0.516	0.516	0.534	0.534	0.534	0.482	0.482	0.935	0.935
TT/TC+CC	0.427	0.427	0.188	0.188	0.188	0.547	0.547	0.382	0.382
LCT rs2164210									
TT*					0.575				
TC	0.262	0.393	0.855	0.855		0.500	0.531	0.189	0.282
CC	0.189	0.393	0.312	0.486		0.400	0.531	0.281	0.282
TT+TC/CC	0.241	0.241	0.284	0.284	0.284	0.396	0.396	0.262	0.262
TT/TC+CC	0.139	0.139	0.823	0.823	0.823	0.363	0.363	0.119	0.119
LCT rs7579771									
AA*					0.448				
AT	0.088	0.132	0.403	0.410		0.258	0.388	0.071	0.106
TT	0.059	0.132	0.303	0.410		0.245	0.388	0.054	0.106

AA+AT/TT	0.140	0.140	0.298	0.298	0.298	0.337	0.337	0.112	0.112
AA/AT+TT	0.024	0.024	0.245	0.245	0.245	0.149	0.149	0.019	0.019

Table 3. 6: Alpha diversity by candidate SNPs in the control group.

	Shannon		**Faith PD**			**Observed OTUs**		**Evenness**	
Candidate SNP	p-value	q-value	p-value1	q-value	p-value2	p-value	q-value	p-value	q-value
FUT2 rs601388									
GG*					0.820				
GA	0.460	0.734	0.588	0.734		0.657	0.657	0.657	0.865
AA	0.700	0.734	0.699	0.734		0.439	0.657	0.657	0.865
GG+GA/AA	0.705	0.705	0.705	0.705	0.705	0.450	0.450	1.000	1.000
GG/GA+AA	0.458	0.458	0.570	0.570	0.570	0.570	0.570	0.631	0.631
NCAM1 rs17115310									
CC*					0.110				
CG	0.288	0.288	0.072	0.171		0.086	0.171	0.683	0.683
GG	0.114	0.288	0.114	0.171		0.114	0.171	0.206	0.576
CC+CG/GG	0.140	0.140	0.340	0.340	0.340	0.193	0.193	0.256	0.256
CC/CG+GG	0.174	0.174	0.041	0.041	0.041	0.049	0.049	0.940	0.940
SHROOM3 rs11724031									
GG*					0.741				
GA	0.805	0.805	0.741	0.741		0.869	0.869	0.741	0.741
AA	-	-	-	-		-	-	-	-
GG+GA/AA	-	-	-	-	-	-	-	-	-
GG/GA+AA	0.805	0.805	0.741	0.741	0.741	0.869	0.869	0.741	0.741
VDR rs1544410									
GG*					0.576				
GA	0.264	0.264	0.576	0.576		0.371	0.371	0.434	0.434
AA	-	-	-	-		-	-	-	-
GG+GA/AA	-	-	-	-	-	-	-	-	-
GG/GA+AA	0.264	0.264	0.576	0.576	0.576	0.371	0.371	0.434	0.434
VDR rs7975232									
AA*					0.498				
AC	0.258	0.466	0.298	0.703		0.587	0.827	0.189	0.466
CC	0.513	0.513	0.827	0.827		0.827	0.827	0.827	0.827
AA+AC/CC	0.361	0.361	0.584	0.584	0.584	0.715	0.715	0.465	0.465
AA/AC+CC	0.353	0.353	0.353	0.353	0.353	0.642	0.642	0.237	0.237
VDR rs731236									
TT*					0.371				
TC	0.264	0.264	0.371	0.371		0.263	0.263	0.434	0.434
CC	-	-	-	-		-	-	-	-
TT+TC/CC	-	-	-	-	-	-	-	-	-
TT/TC+CC	0.264	0.264	0.264	0.264	0.264	0.263	0.263	0.434	0.434
LCT rs2164210									
TT*					0.260				

TC	0.651	0.651	0.910	0.910		0.910	0.910	0.213	0.213
CC	0.172	0.517	0.107	0.236		0.710	0.910	0.172	0.213
TT+TC/CC	0.210	0.210	0.101	0.101	0.101	0.630	0.630	0.148	0.148
TT/TC+CC	0.349	0.349	0.588	0.588	0.588	0.805	0.805	0.104	0.104
LCT rs7579771									
AA*					0.090				
AT	0.514	0.514	0.794	0.794		0.695	0.695	0.240	0.289
TT	0.052	0.155	0.051	0.077		0.139	0.208	0.102	0.289
AA+AT/TT	0.051	0.051	0.028	0.028	0.028	0.086	0.086	0.110	0.110
AA/AT+TT	0.113	0.113	0.189	0.189	0.189	0.257	0.257	0.077	0.077

Table 3. 7: Alpha diversity by candidate SNPs in the efavirenz group.

	Shannon		**Faith PD**			**Observed OTUs**		**Evenness**	
Candidate SNP	p-value	q-value	p-value1	q-value	p-value2	p-value	q-value	p-value	q-value
FUT2 rs601388									
GG*									
GA	0.548	0.554	0.744	0.744	0.696	0.850	0.850	0.503	0.503
AA	0.554	0.554	0.554	0.744		0.693	0.850	0.236	0.503
GG+GA/AA	0.513	0.513	0.433	0.433	0.433	0.769	0.769	0.296	0.296
GG/GA+AA	0.509	0.509	0.834	0.834	0.834	0.809	0.809	0.412	0.412
NCAM1 rs17115310									
CC*									
CG	0.277	0.633	0.727	0.727	0.570	0.531	0.633	0.408	0.946
GG	0.633	0.633	0.375	0.562		0.633	0.633	0.946	0.946
CC+CG/GG	0.536	0.536	0.318	0.318	0.318	0.552	0.552	1.000	1.000
CC/CG+GG	0.502	0.502	0.917	0.917	0.917	0.765	0.765	0.483	0.483
SHROOM3 rs11724031									
GG*									
GA	0.252	0.503	0.674	0.851	0.884	0.268	0.641	0.445	0.662
AA	0.851	0.851	0.851	0.851		0.975	0.975	0.662	0.662
GG+GA/AA	0.619	0.619	0.779	0.779	0.779	0.746	0.746	0.560	0.560
GG/GA+AA	0.425	0.425	0.813	0.813	0.813	0.400	0.400	0.701	0.701
VDR rs1544410									
GG*									
GA	0.900	1.000	0.585	0.853	0.811	0.950	0.950	0.556	0.683
AA	0.795	1.000	0.853	0.853		0.911	0.950	0.353	0.683
GG+GA/AA	0.841	0.841	0.753	0.753	0.753	0.886	0.886	0.548	0.548
GG/GA+AA	0.823	0.823	0.718	0.718	0.718	1.000	1.000	0.363	0.363
VDR rs7975232									
AA*									
AC	0.921	0.921	0.843	0.843	0.759	0.791	0.791	0.643	0.683
CC	0.221	0.653	0.414	0.816		0.307	0.652	0.683	0.683
AA+AC/CC	0.312	0.312	0.470	0.470	0.470	0.360	0.360	0.596	0.596
AA/AC+CC	0.924	0.924	0.949	0.949	0.949	0.924	0.924	0.702	0.702

VDR rs731236									
TT*					0.368				
TC	0.264	0.521	0.313	0.361		0.330	0.520	0.014	0.043
CC	0.347	0.521	0.296	0.361		0.347	0.520	0.210	0.315
TT+TC/CC	0.447	0.447	0.300	0.300	0.300	0.407	0.407	0.581	0.581
TT/TC+CC	0.194	0.194	0.218	0.218	0.218	0.243	0.243	0.009	0.009
LCT rs2164210									
TT*					0.962				
TC	0.447	0.769	0.864	0.864		0.540	0.827	0.607	0.607
CC	0.942	0.942	0.828	0.864		0.718	0.827	0.279	0.419
TT+TC/CC	0.823	0.823	0.823	0.823	0.823	0.823	0.823	0.264	0.264
TT/TC+CC	0.498	0.498	0.821	0.821	0.821	0.498	0.498	0.892	0.892
LCT rs7579771									
AA*					0.896				
AT	0.428	0.902	0.667	0.862		0.541	0.862	0.428	0.428
TT	0.937	0.937	0.813	0.862		0.693	0.862	0.306	0.428
AA+AT/TT	0.823	0.823	0.823	0.823	0.823	0.823	0.823	0.264	0.264
AA/AT+TT	0.473	0.473	0.735	0.735	0.735	0.499	0.499	0.642	0.642

Table 3. 8: Alpha diversity by candidate SNPs in the warfarin group.

	Shannon		**Faith PD**			**Observed OTUs**		**Evenness**	
Candidate SNP	p-value	q-value	p-value1	q-value	p-value2	p-value	q-value	p-value	q-value
FUT2 rs601388									
GG*					0.977				
GA	0.453	0.453	0.895	0.926		0.627	0.627	0.402	0.402
AA	0.116	0.174	0.782	0.926		0.309	0.464	0.021	0.062
GG+GA/AA	0.056	0.056	0.841	0.841	0.841	0.132	0.132	0.021	0.021
GG/GA+AA	0.227	0.227	1.000	1.000	1.000	0.434	0.434	0.118	0.118
NCAM1 rs17115310									
CC*					0.942				
CG	0.247	0.247	0.942	0.942		0.218	0.218	0.192	0.192
GG	-	-	-	-		-	-	-	-
CC+CG/GG	-	-	-	-	-	-	-	-	-
CC/CG+GG	0.247	0.247	0.942	0.942	0.942	0.218	0.218	0.192	0.192
SHROOM3 rs11724031									
GG*					0.184				
GA	0.569	0.862	0.271	0.271		0.425	0.469	0.790	0.862
AA	0.862	0.862	0.117	0.271		0.117	0.352	0.862	0.862
GG+GA/AA	0.869	0.869	0.137	0.137	0.137	0.248	0.248	0.509	0.509
GG/GA+AA	0.619	0.619	0.177	0.177	0.177	0.286	0.286	0.831	0.831
VDR rs1544410									
GG*					0.796				
GA	0.086	0.259	0.806	0.806		0.369	0.369	0.014	0.043
AA	0.206	0.309	0.752	0.806		0.206	0.335	0.206	0.309

GG+GA/AA	0.260	0.260	0.545	0.545	0.545	0.193	0.193	0.435	0.435
GG/GA+AA	0.059	0.059	0.880	0.880	0.880	0.257	0.257	0.010	0.010
VDR rs7975232									
AA*									
AC	0.934	0.934	0.138	0.138	0.138	0.741	0.741	0.621	0.621
CC	-	-	-	-		-	-	-	-
AA+AC/CC	-	-	-	-	-	-	-	-	-
AA/AC+CC	0.934	0.934	0.138	0.138	0.138	0.741	0.741	0.621	0.621
VDR rs731236									
TT*									
TC	0.035	0.106	0.861	0.861	0.331	0.066	0.197	0.054	0.161
CC	0.385	0.578	0.172	0.259		0.535	0.617	0.385	0.578
TT+TC/CC	0.665	0.665	0.140	0.140	0.140	0.795	0.795	0.665	0.665
TT/TC+CC	0.029	0.029	0.782	0.782	0.782	0.063	0.063	0.043	0.043
LCT rs2164210									
TT*									
TC	0.293	0.317	0.930	0.930	0.293	0.188	0.282	0.930	0.930
CC	0.234	0.317	0.089	0.268		0.174	0.282	0.610	0.916
TT+TC/CC	0.231	0.231	0.119	0.119	0.119	0.231	0.231	0.402	0.402
TT/TC+CC	0.169	0.169	0.426	0.426	0.426	0.096	0.096	0.772	0.772
LCT rs7579771									
AA*									
AT	0.048	0.143	0.450	0.450	0.223	0.034	0.061	0.322	0.450
TT	0.097	0.146	0.143	0.236		0.040	0.061	0.378	0.450
AA+AT/TT	0.332	0.332	0.156	0.156	0.156	0.205	0.205	0.709	0.709
AA/AT+TT	0.065	0.065	0.028	0.028	0.028	0.009	0.009	0.522	0.522

3.8.3 Beta diversity

As opposed to alpha diversity, beta diversity measures the species diversity between the samples. It can be measured with Bray-Curtis; Unweighted and Weighted UniFrac, to name a few.

3.8.3.1 Comparison of beta diversity across the three study groups

Like in the alpha diversity, the gut microbiome was clustered according to the drug groups under study, which are untreated control, efavirenz and warfarin groups. Weighted UniFrac distance was calculated to investigate the beta diversity between the three groups. Principal Coordinates Analysis (PCoA) plot was used to show this comparison in Figure3.10.

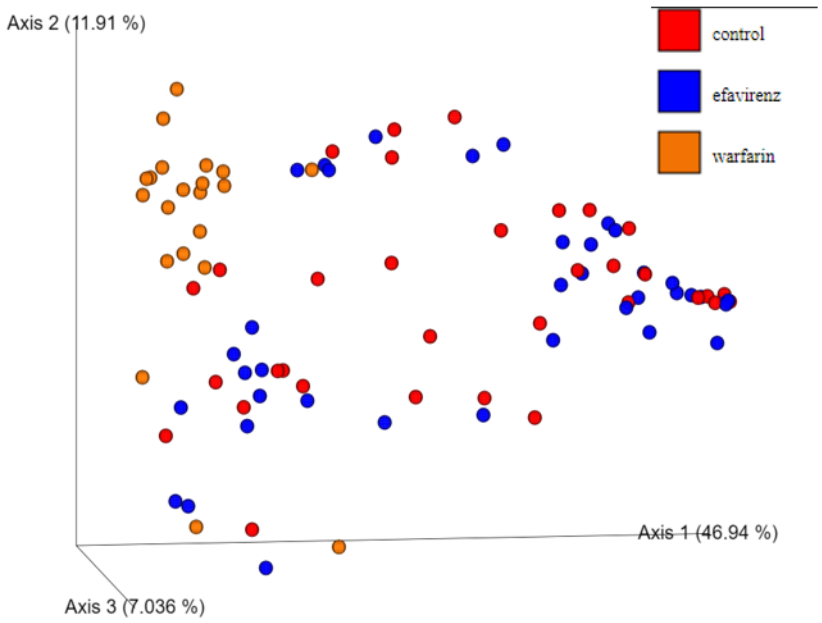


Figure 3. 10: Principal Coordinates Analysis (PCoA) plots of the weighted UniFrac distances within the groups

Not all groups distinctively segregated from each other. Majority of the warfarin group clustered tightly together. The warfarin group segregated from the

untreated control group and there was statistical difference when using the weighted UniFrac distances (p = 0.001, q = 0.002) and so is the case when comparing the warfarin group and the efavirenz one (p = 0.001, q = 0.002). There was no spatial segregation between the untreated control and efavirenz groups (p&q = 0.753).

3.8.3.2 Effect of efavirenz exposure on the beta diversity

To add more meaning to our results, efavirenz plasma concentrations were divided into three groups, those with EFV < 1 µg/mL, 1 ≤ EFV ≥ 4 µg/mL and those with efavirenz plasma concentration above the normal limit, EFV > 4 µg/mL as shown in PCoA below (Figure 3.11) accompanied by weighted UniFrac distance. The three groups did not segregate. The beta diversity by weighted UniFrac distance showed no statistical difference between the gut bacterial composition of participants within normal EFV therapeutic range and those below it, p = 0.347, q = 0.744. The weighted UniFrac distance also showed no statistical difference between the gut bacterial composition of participants within normal EFV therapeutic range and those above it, p = 0.578, q = 0.744.

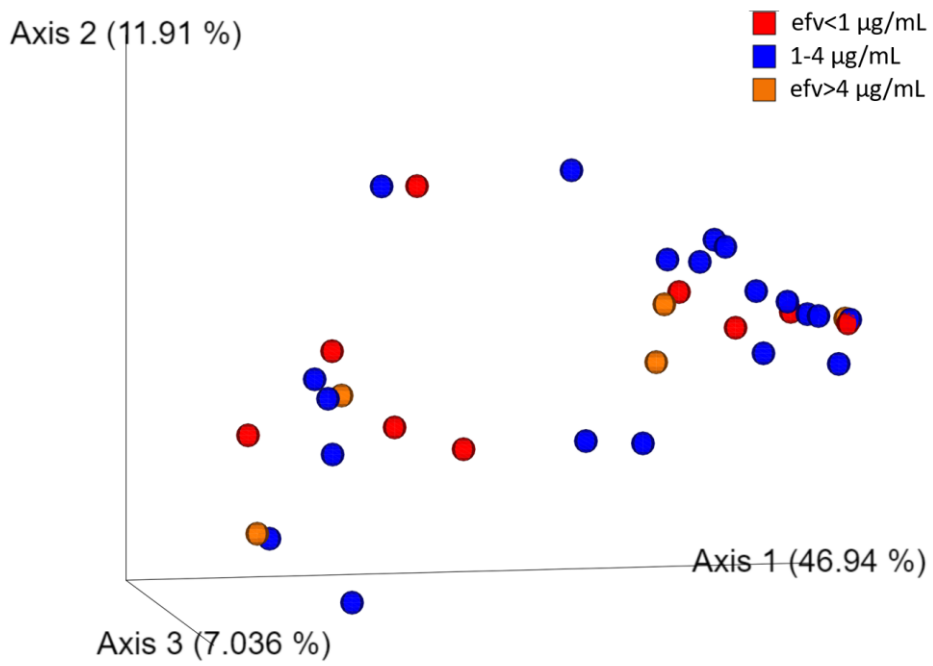


Figure 3. 11: Beta diversity within the efavirenz group.

3.8.3.3 Beta diversity within the warfarin C_{max}

The warfarin C_{max} were grouped on whether they had $C_{max} \leq 1.34$ µg/mL or $C_{max} > 1.34$ µg/mL, of which 1.34 µg/mL was the mean concentration. There was no statistical difference between the samples of two groups (p&q = 0.864), hence there was no distinct separation of samples as shown by the PCoA in Figure 3.12.

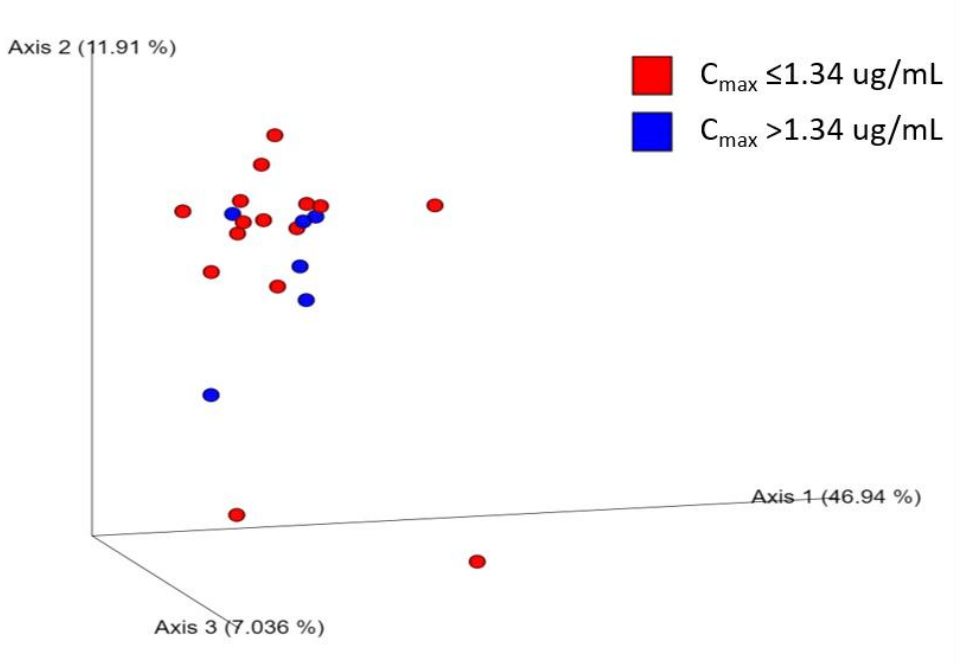


Figure 3. 12: Weighted UniFrac Principal Coordinates Analysis (PCoA) plots for the warfarin exposure.

3.8.3.4 Beta diversity due to host genetics

Eight candidate SNPs from five genes were genotyped as described in section 2.2 and SNP-microbe association analysis was done on QIIME2 using the PERMANOVA statistical test. Three genetic models (additive, dominant and recessive) were used to assess the influence of host genetics on microbiota based on the underlying Unweighted UniFrac distance matrix. Haplotypes of SNPs in the same host genes were also added to analysis to estimate the difference in beta diversity of the gut microbiome. The host genetics-gut microbiome relationship was done in four groups, that is, the untreated control, efavirenz, warfarin and the combination of all three for the reproducibility of results. In this study, weighted UniFrac distance was used to explain the beta diversity in our gut microbiome data.

Table 3.9 below shows the host genetics-gut microbiome association of all the participants without dividing them in groups. All the candidate SNPs were tested in three genetic models. The global p-value2 amongst the genotypes did not reach any statistical significance for each SNP except in the recessive models of VDR rs1544410 (p = 0.039) and LCT rs7579771 (p = 0.038). Genotypes of each SNP were compared with one another in a pair-wise manner (p-value1), there was significant difference in weighted UniFrac distance between the GG and GA (p-value1 = 0.041) genotypes of VDR rs1544410. The gut microbiome abundance difference between AA and TT genotypes of LCT rs7579771 was on the p-value cut-off, p-value1=0.053. Segregation of our gut microbiome data according to the recessive model of LCT rs7579771 resulted in significant difference, p-value1 = 0.040. To control for false discovery rate, p-value1 was adjusted to q-value. The beta diversity between the GG and GA genotypes of VDR rs1544410 changed to q = 0.123, AA and TT of LCT rs7579771 changed to q = 0.159. However, comparison of gut microbiome abundance in the recessive model remained significantly different q = 0.040.

Table 3. 9: Weighted UniFrac for Candidate SNPs with suspected roles in shaping the gut microbiome profiles in general population of this study

Candidate SNP	p-value1	q-value	p-value2
FUT2 rs601388			
GG*			
GA	0.608	0.880	0.782
AA	0.880	0.880	
GG+GA/AA	0.723	0.723	0.723
GG/GA+AA	0.697	0.697	0.673
NCAM1 rs17115310			
C/C*			
C/G	0.857	0.857	0.774
G/G	0.471	0.825	
CC+CG/GG	0.483	0.483	0.525
CC/CG+GG	0.815	0.815	0.796
SHROOM3 rs11724031			

GG*			
GA	0.099	0.297	0.196
AA	0.303	0.455	
GG+GA/AA	0.489	0.489	0.496
GG/GA+AA	0.063	0.063	0.058
VDR rs1544410			
GG*			
GA	0.041	0.123	0.160
AA	0.740	0.754	
GG+GA/AA	0.893	0.893	0.887
GG/GA+AA	0.063	0.063	0.039
VDR rs7975232			
AA*			
AC	0.332	0.525	0.503
CC	0.503	0.525	
AA+AC/CC	0.554	0.554	0.498
AA/AC+CC	0.364	0.364	0.381
VDR rs731236			
TT*			
TC	0.944	0.949	0.870
CC	0.894	0.949	
TT+TC/CC	0.950	0.950	0.944
TT/TC+CC	0.495	0.495	0.524
LCT rs2164210			
TT*			
TC	0.133	0.399	0.400
CC	0.911	0.911	
TT+TC/CC	0.906	0.906	0.893
TT/TC+CC	0.204	0.204	0.209
LCT rs7579771			
AA*			
AT	0.175	0.263	0.072
TT	0.053	0.159	
AA+AT/TT	0.080	0.080	0.079
AA/AT+TT	0.040	0.040	0.038

The untreated control group was analysed separately to assess the impact of genetic variation on gut microbiome, also using the three genetic models and haplotypes as shown in Table 3.10. None of the genotypes reached a statistical significance at a global comparison (p-value2). However, a dominant model for LCT rs7579771 was at p-value2 = 0.055. The pairwise comparisons, p-value1 did not reach statistical significance (p-value1 < 0.05). The adjusted p-value1 (q-value) did not reach significance for any comparison using the weighted UniFrac distance to assess the beta diversity of the gut microbiome.

Table 3. 10: Beta diversity of gut microbiome for candidate SNPs in the untreated control group only.

Candidate SNP	p-value1	q-value	p-value2
FUT2 rs601388			
GG*			0.905
GA	0.778	0.862	
AA	0.862	0.862	
GG+GA/AA	0.787	0.787	0.791
GG/GA+AA	0.861	0.861	0.844
NCAM1 rs17115310			
CC*			0.701
CG	0.437	1.000	
GG	1.000	1.000	
CC+CG/GG	0.912	0.912	0.886
CC/CG+GG	0.529	0.529	0.513
SHROOM3 rs11724031			
GG*			0.728
GA	0.753	0.753	
AA	-	-	
GG+GA/AA	-	-	-
GG/GA+AA	0.707	0.707	0.761
VDR rs1544410			
GG*			0.303
GA	0.317	0.317	
AA	-	-	
GG+GA/AA	-	-	-
GG/GA+AA	0.290	0.290	0.296
VDR rs7975232			
AA*			0.843
AC	0.754	0.908	
CC	0.625	0.908	
AA+AC/CC	0.741	0.741	0.757
AA/AC+CC	0.662	0.662	0.632
VDR rs731236			
TT*			0.492
TC	0.495	0.495	
CC	-	-	
TT+TC/CC	-	-	-
TT/TC+CC	0.509	0.509	0.515
LCT rs2164210			
TT*			0.570
TC	0.362	0.872	
CC	0.872	0.872	
TT+TC/CC	0.889	0.889	0.881
TT/TC+CC	0.423	0.423	0.377
LCT rs7579771			
AA*			0.517
AT	0.514	0.757	
TT	0.278	0.757	
AA+AT/TT	0.375	0.375	0.383
AA/AT+TT	0.278	0.278	0.267

Patients who were on efavirenz were also genotyped for candidate SNPs that shape the gut microbiome to determine the effect of host genetics on gut microbiome. All the three genetic models were also tested to assess the host genetics-gut microbiome relationship as shown in Table 3.11. None of these models reached statistically significant difference in terms of weighted UniFrac distance in all the candidate SNPs.

Table 3. 11: Candidate SNPs variation and association with gut microbiome profiles among the patients that were on efavirenz.

Candidate SNP	p-value1	q-value	p-value2
FUT2 rs601388			
GG*			0.873
GA	0.471	0.979	
AA	0.979	0.979	
GG+GA/AA	0.971	0.971	0.964
GG/GA+AA	0.475	0.475	0.472
NCAM1 rs17115310			
C/C*			0.914
C/G	0.577	0.934	
G/G	0.934	0.934	
CC+CG/GG	0.966	0.966	0.960
CC/CG+GG	0.666	0.666	0.666
SHROOM3 rs11724031			
GG*			0.246
GA	0.520	0.520	
AA	0.090	0.270	
GG+GA/AA	0.120	0.120	0.119
GG/GA+AA	0.188	0.188	0.203
VDR rs1544410			
GG*			0.939
GA	0.972	0.972	
AA	0.665	0.972	
GG+GA/AA	0.754	0.754	0.702
GG/GA+AA	0.829	0.829	0.845
VDR rs7975232			
AA*			0.075
AC	0.120	0.138	
CC	0.104	0.138	
AA+AC/CC	0.095	0.095	0.109
AA/AC+CC	0.111	0.111	0.080
VDR rs731236			
TT*			0.554
TC	0.180	0.540	
CC	0.925	0.960	
TT+TC/CC	0.989	0.989	0.984
TT/TC+CC	0.178	0.178	0.180
LCT rs2164210			
TT*			0.216
TC	0.188	0.282	

CC	0.488	0.488	
TT+TC/CC	0.320	0.320	0.341
TT/TC+CC	0.438	0.438	0.464
LCT rs7579771			
AA*			0.096
AT	0.100	0.300	
TT	0.537	0.537	
AA+AT/TT	0.307	0.307	0.341
AA/AT+TT	0.236	0.236	0.248

In the warfarin group (Table 3.12), genetic determination of gut microbiome profiles was tested on candidate SNPs. Like in the untreated control (Table 3.10) and efavirenz (Table 3.11) groups, the abundance gut microbiome in all the models for all SNPs did not reach statistically significant difference, before or after adjustments for false discovery.

Table 3. 12: Host genetics-gut microbiome association for candidate SNPs in patients on warfarin treatment.

Candidate SNP	p-value1	q-value	p-value2
FUT2 rs601388			
GG*			0.304
GA	0.332	0.433	
AA	0.433	0.433	
GG+GA/AA	0.445	0.445	0.439
GG/GA+AA	0.215	0.215	0.215
NCAM1 rs17115310			
C/C*			0.583
C/G	0.583	0.583	
G/G	-	-	
CC+CG/GG	-	-	-
CC/CG+GG	0.556	0.556	0.612
SHROOM3 rs11724031			
GG*			0.999
GA	0.993	1.000	
AA	0.912	1.000	
GG+GA/AA	0.873	0.873	0.899
GG/GA+AA	0.991	0.991	0.991
VDR rs1544410			
GG*			0.557
GA	0.249	0.747	
AA	1.000	1.000	
GG+GA/AA	1.000	1.000	1.000
GG/GA+AA	0.180	0.180	0.189
VDR rs7975232			
AA*			0.695
AC	0.657	0.657	
CC	-	-	
AA+AC/CC	-	-	-

AA/AC+CC	0.680	0.680	0.638
VDR rs731236			
TT*			0.388
TC	0.340	0.592	
CC	0.592	0.592	
TT+TC/CC	0.479	0.479	0.501
TT/TC+CC	0.299	0.299	0.308
LCT rs2164210			
TT*			0.510
TC	0.312	0.651	
CC	0.651	0.651	
TT+TC/CC	0.667	0.667	0.649
TT/TC+CC	0.269	0.269	0.261
LCT rs7579771			
AA*			0.205
AT	0.192	0.316	
TT	0.316	0.316	
AA+AT/TT	0.513	0.513	0.509
AA/AT+TT	0.104	0.104	0.088

3.8.4 Relative abundance of gut microbiome and potential biomarkers

Both alpha and beta diversity indicated that the gut microbiome composition of individuals in warfarin group differ significantly with those in untreated control and efavirenz group. Linear discriminant analysis Effect Size (LEfSe) (Segata et al., 2011) tool was used to assess the taxonomic biomarkers in each group. When comparing the gut microbiome of participants in untreated control group with those in warfarin one, there were more than 80 bacterial taxa that were significantly abundant in the untreated control than in warfarin group (LDA score ≥2.0, $p < 0.05$). The Linear discriminant analysis (LDA) score was adjusted to 6,7; $p<0.05$ to reduce the density of the significant bacterial taxa shown in Figure 3.13. Bacterial phyla *Actinobacteria*, and *Firmicutes*, were the only significantly abundant phyla in the untreated control group (LDA score ≥ 6.7, $p < 0.05$). *Bacteroidia* and *Erysipelotrichia* classes were found to be abundantly significant in the gut microbiome group of the untreated control group than in the warfarin group (LDA score ≥ 6.7, $p < 0.05$). Bacterial order *Bifidobacteriales* were relatively abundant in the warfarin group than in the untreated control group (LDA score ≥ 6.7, $p < 0.05$), however, orders *Erysipelotrichales*, *Caulobacterales*, *Pseudomonadales* and *Enterobacterales* were more abundant in the untreated control group (LDA score ≥ 6.7, $p < 0.05$). Family *Bifidobacteriaceae* was the only family that was differentially abundant in the warfarin group whereas

Caulobacteraceae, *Moraxellaceae* and *Enterobacteriaceae* were significantly abundant in the gut microbiome of individuals in the untreated control group (LDA score ≥ 6.7, $p < 0.05$). In the genus level, LEfSe indicated that participants in the untreated control group were significantly associated with high abundance of *Hydrogenophilus*, *Staphylococcus*, *Blautia*, *Ruminococcus* and *Enterococcus* than those in the warfarin group. *Eubacterium biforme* and *Faecalibacterium prausnitzii* were the only species significantly abundant in the untreated control group (LDA score ≥ 6.7, $p < 0.05$).

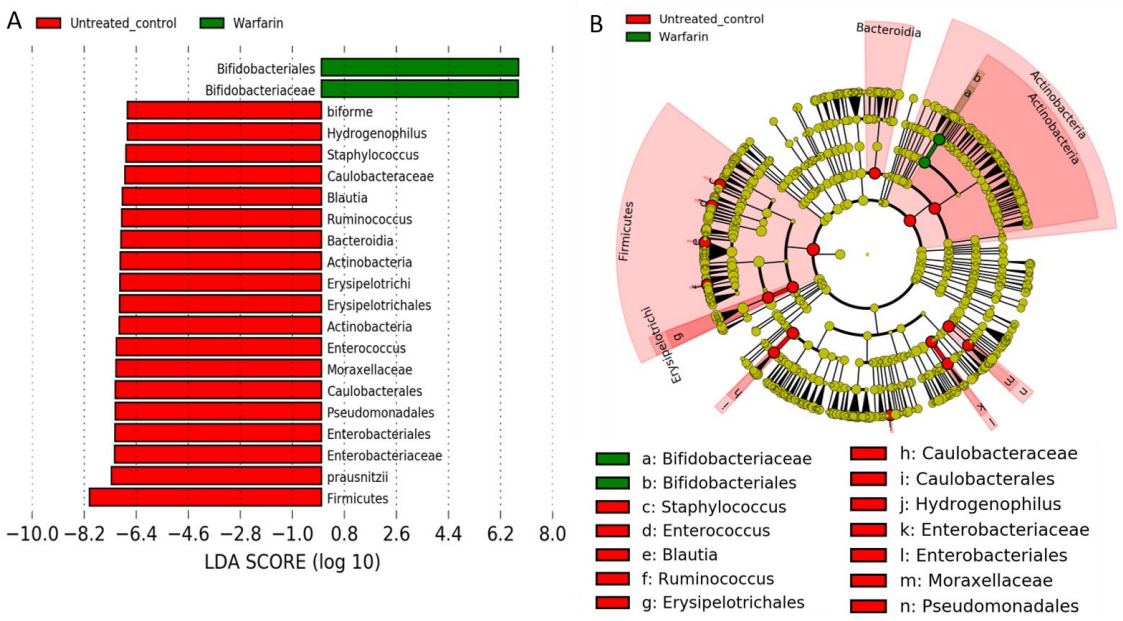


Figure 3. 13: Potential biomarkers for warfarin by LEfSe

When comparing the gut microbiome composition of individuals who were on efavirenz treatment and those who were not in any medication, LEfSe found nine bacterial taxa to be differentially abundant (Figure 3.14). Phylum *Tenericutes* was found to be in high abundance in the untreated control group (LDA score ≥ 2.0, $p < 0.05$). The abundance of unclassified order RF39 was significantly associated with the untreated control group (LDA score ≥ 2.0, $p < 0.05$). Family *Pseudonocardiaceae* was found to be highly abundant in the untreated control group whereas *Symbiobacteriaceae* was significantly plentiful in the efavirenz group (LDA score ≥ 2.0, $p < 0.05$). At genus level, *Saccharopolyspora* and *Slackia* were significantly enriched in the gut microbiome of individuals of the untreated control group, on the other hand, genera *Caldinitratiruptor* and *Erythromicrobium*

were dominating in the efavirenz group (LDA score ≥ 2.0, $p < 0.05$). *Caldinitratiruptor microaerophilus* was the only bacterial taxa that was differentially abundant and was significantly abundant in the gut microbiome of participants of the efavirenz group (LDA score ≥ 2.0, $p < 0.05$). However, none of these bacterial taxa were still significant after FDR correction (LDA score ≥ 2.0, $q < 0.2$).

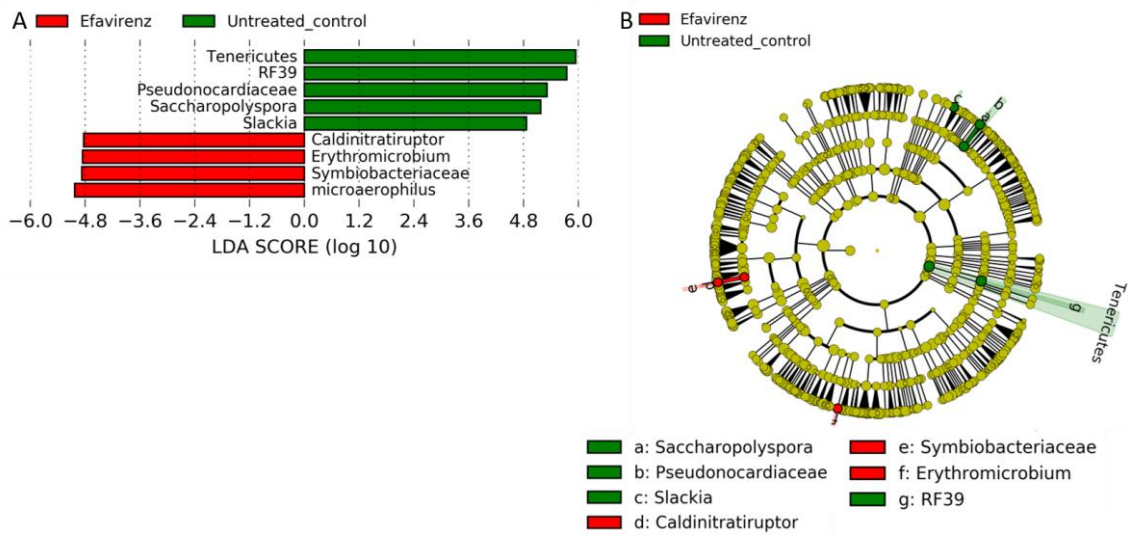


Figure 3. 14: Potential biomarkers for efavirenz

When comparing the gut microbiome between the paricipants with EFV ≤ 4 µg/mL and those with efavirenz plasma concentration above the normal limit, EFV > 4 µg/mL, LEfSe did not identify any bacterial taxa highly abundant in the EFV ≤ 4 µg/mL group (Figure 3.15). All the differentially abundant taxa were found to be significant in the EFV > 4 µg/mL group. These include families Dermacoccaceae and *Pasteurellaceae*; genera *Dermacoccus*, *Finegoldia* and *Actinobacillus*; and species *Streptococcus Iuteciae* and *Eubacterium dolichum*, that were highly abundant in the gut microbiome of participants in the EFV > 4 µg/mL group (LDA score ≥ 5.5, $p < 0.05$).

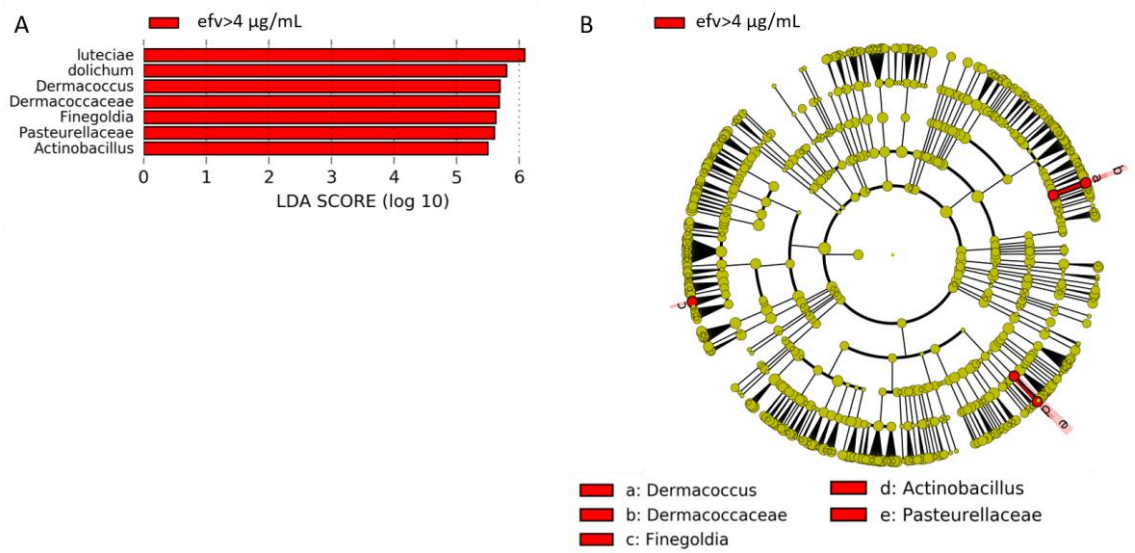


Figure 3. 15: Potential biomarkers for efavirenz drug exposure

Furthermore, only the model test from Tables 3.8 to 3.15 that had p and or q values ≤ 0.05 were further tested for genotype-abundance association using LEfSe.

3.8.4.1 *FUT2* rs601338 potential biomarkers

Comparison between the GG and GA genotypes of the FUT2 rs601338 SNP (Figure 3.16) in the general/combined group (Figure 3.16A), LEfSe identified 12 differential abundant bacterial taxa. Phylum *Planctomycetes* was found to be more abundant in the GA genotype group (LDA score ≥ 2.0, $p < 0.05$). At class level, TM7-3 and *Thermomicrobia* were the only classes that were found to be significantly abundant in the carriers of the GA genotype. Orders JG30-KF-CM45 and *Neisseriales* were also significantly higher in the carriers of the GA genotype (LDA score ≥ 2.0, $p < 0.05$). At family level, participants with the GA genotype were significantly enriched with *Methylobacteriaceae* and *Listeriaceae* (LDA score ≥ 2.0, $p < 0.05$) in their gut microbiome than those with the GG genotype. At genus level, participants with the GG genotype were highly enriched with *Roseburia* whereas those with GA genotype were enriched with *Treponema*, *Listeria* and *Pseudomonas*. *Coprococcus catus* species of phylum *Firmicutes* was significantly abundant in the gut microbiome of participants with the GG genotype of this SNP (LDA score ≥ 2.0, $p < 0.05$). Comparison between the GG and GA genotypes did not yield any significantly abundant bacterial taxa.

Comparing the similarities between the GG and GA genotypes of both untreated control (Figure 3.16B) and efavirenz groups (Figure 3.16C), none of the significantly abundant bacterial taxa were common. Moreover, none of these bacterial taxa were significantly abundant after FDR correction (LDA score ≥ 2.0, $q < 0.2$).

When it comes to the comparison between GG and AA genotypes, the bacterial taxa in the in the untreated control and warfarin groups were not differentially significantly abundant (LDA score ≥ 2.0, $p < 0.05$). In the efavirenz (see Supplementary Figure 1) and the general groups, none of the bacterial taxa were significantly abundant after FDR correction (LDA score ≥ 2.0, $q < 0.2$).

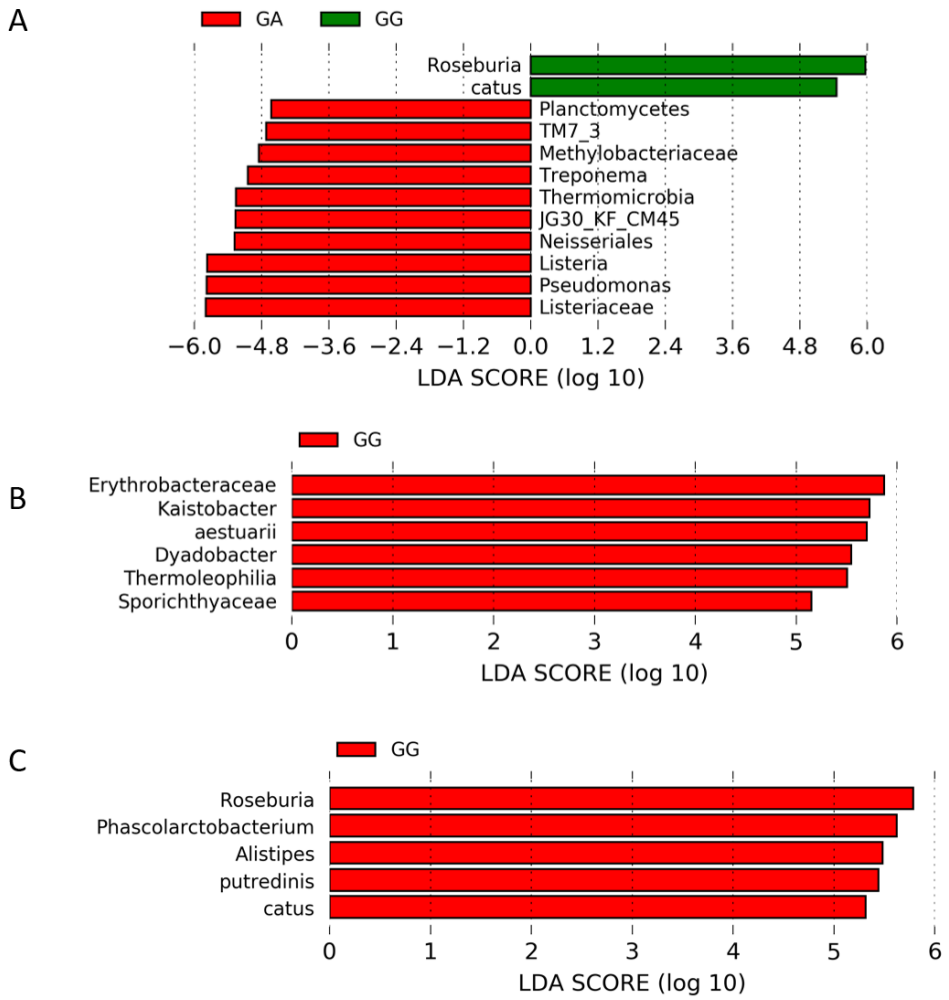


Figure 3. 16: Potential biomarkers for *FUT2* rs601338 SNP by comparing GG vs GA genotypes.

Bacterial composition of individuals with either GG or GA (GG+GA) genotypes were compared against those with the AA genotype for FUT2 rs601338, Supplementary Figure 2. In general (Supplementary Figure 2A), genera *Actinomycetospora* and *Brevibacterium* remained significantly abundant in the gut bacterial composition of non-secretors (AA genotype) after FDR correction (LDA score ≥ 2.0, $q < 0.2$). The gut bacterial composition of non-secretors was also significantly enriched with species *Corynebacterium kroppenstedtii*, *Macrococcus caseolyticus* and *Kocuria kristinae* (LDA score ≥ 2.0, $q < 0.2$) than the secretors (GG+GA genotypes). Looking only onto the untreated control group, none of the bacterial taxa were significantly differentially abundant at LDA score ≥ 2.0, $q < 0.2$ (Supplementary Figure 2B). When it comes to only the efavirenz group (Supplementary Figure 2C), bacterial taxa that remained significantly abundant on the non-secretors include genera *Leucobacter*, *Brevibacterium*, *Sphingobacterium* and *Yaniella* (LDA score ≥ 2.0, $q < 0.2$). At species level, *Kocuria kristinae*, *Ruminococcus albus*, *Geodermatophilus obscurus* and *Mycobacterium aurum* were significantly abundant (LDA score ≥ 2.0, $q < 0.2$) in the gut bacterial composition of individuals with AA genotype for rs601338 SNP as compared to those with GG+GA genotypes. Comparing GG genotype against GA+AA (Supplementary Figure 3), none of the bacterial taxa were significant after FDR correction (LDA score ≥ 2.0, $q < 0.2$).

3.8.4.2 *NCAM1* rs17115310 potential biomarkers

Looking at the comparison between the CC and CG of rs17115310, gut bacterial abundance did not differ significantly after FDR correction (LDA score ≥ 2.0, $q < 0.2$), see Supplementary Figure 4. However, in general (Supplementary Figure 4A), *Tepidimonas*, *Ochrobactrum*, *Alicyclobacillus*, *Amaricoccus*, *Cryocola*, *Actinobacillus*, *Bulleidia*, *Kytococcus*, unclassified N09, *Shewanella* and *Lactococcus* were some of the significantly abundant bacterial genera (LDA score ≥ 2.0, $q < 0.2$) in the gut bacteria composition of individuals with the GG genotype for rs17115310 when compared to those with the CC genotype. In the species level, the gut microbiome participants with the GG genotype were significantly enriched with *Petrobacter succinatimandens*, *Eubacterium dolichum*, *Pseudoclavibacter bifida*, *Brevibacterium paucivorans*, *Pseudomonas fragi*,

Ornithinimicrobium pekingense, *Clostridium citroniae*, *Corynebacterium kroppenstedtii* and *Propionibacterium granulosum* (LDA score ≥ 2.0, q < 0.2). Looking only at the efavirenz group (Supplementary Figure 4B), species *Petrobacter succinatimandens* and *Pseudoclavibacter bifida*; and unclassified order I025 were the only bacterial taxa significant in the GG genotype group after FDR correction (LDA score ≥ 2.0, q < 0.2).

When comparing GG genotype against CC and CG (CC+CG) in the general population, those with the GG genotype were significantly enriched with genera *Tepidimonas*, *Ochrobactrum*, *Alicyclobacillus*, *Amaricoccus*, *Schlegelella*, *Actinobacillus*, *Cryocola*, *Bulleidia*, *Blastomonas* and *Thermomonas* (LDA score ≥ 2.0, q < 0.2) than those with the C allele. In the species level, *Petrobacter succinatimandens*, *Brevibacterium paucivorans*, *Pseudomonas fragi*, *Cellulomonas xylanilytica*, *Ornithinimicrobium pekingense*, *Pseudomonas thermotolerans* and *Pannonibacter phragmitetus* were significantly abundant in the participants with the GG genotype (LDA score ≥ 2.0, q < 0.2). Moreover, in the efavirenz group genera *Ochrobactrum*, *Cryocola*, *Legionella*, *Actinobacillus*, unclassified N09, *Euzebya* and *Geodermatophilus*; and species *Petrobacter succinatimandens*, *Brevibacterium paucivorans*, *Pseudoclavibacter bifida* and *Ornithinimicrobium pekingense* were significantly abundant (LDA score ≥ 2.0, q < 0.2) in the gut bacterial composition of participants with the GG genotype as compared to those with the A allele for rs17115310 SNP.

3.8.4.3 *SHROOM3* rs11724031 potential biomarkers

Family *Caldilineaceae* in the general population (Figure 3.17A) was the only bacteria taxa significantly in the gut bacterial composition of participants with the AA genotype of rs11724031 SNP as compared to those with the GG genotype (LDA score ≥ 2.0, q < 0.2). None of the taxa were still significantly differentially significantly abundant when comparing the participants of GG genotype against those of AA genotype only in the efavirenz group (Figure 3.17B) after FDR correction (LDA score ≥ 2.0, q < 0.2).

When comparing GG+GA against AA genotype in the general population (Figure 3.18A), the latter genotype was significantly associated with abundance of

bacterial family *Caldilineaceae* (LDA score ≥ 2.0, q < 0.2). Looking only in the efavirenz (Figure 3.18B), the AA genotype was significantly enriched with bacterial family *Caldilineaceae* and genus *Sneathia* when compared with the presence of G allele (GG+GA genotype) after FDR correction (LDA score ≥ 2.0, q < 0.2).

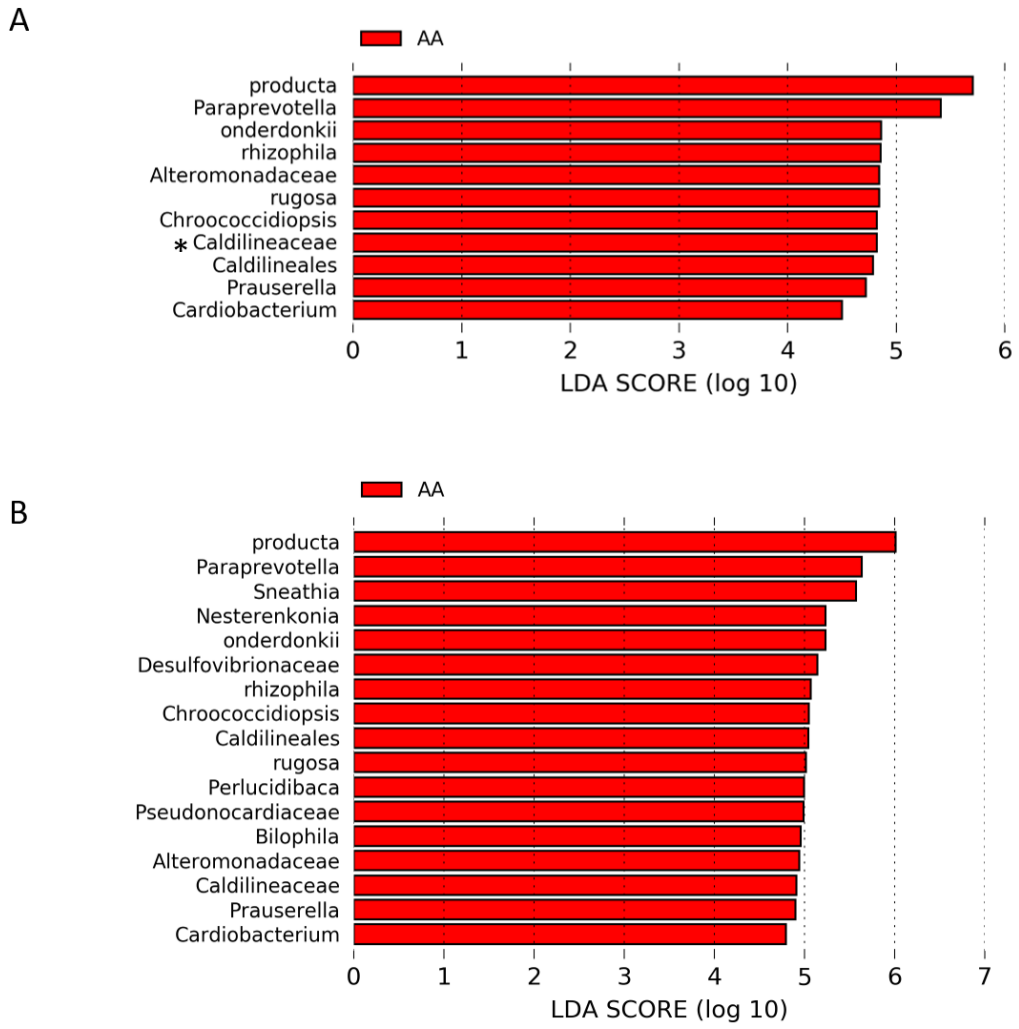


Figure 3. 17: Potential biomarkers for *SHROOM3* rs11724031 by comparing GG vs AA genotypes

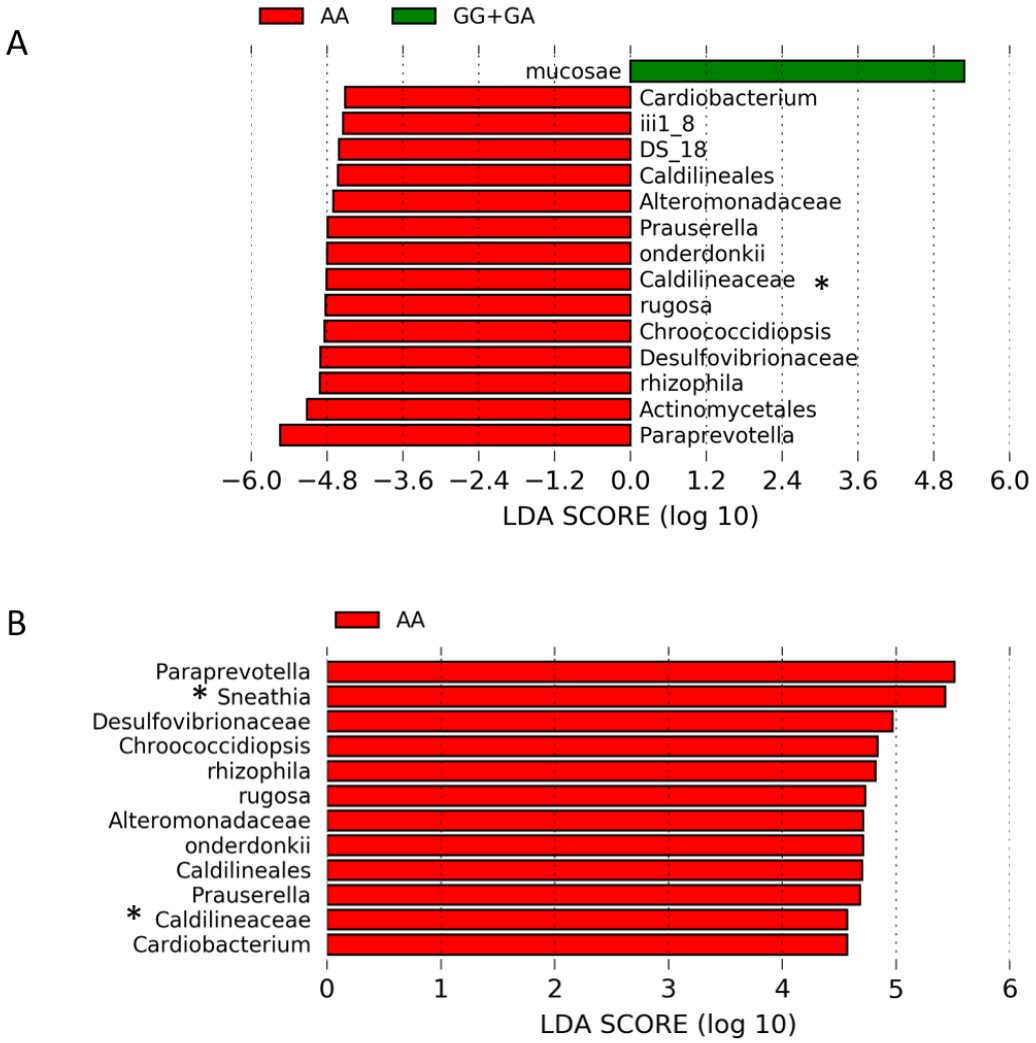


Figure 3. 18: Potential biomarkers for *SHROOM3* rs11724031 by comparing GG+GA vs AA genotypes

3.8.4.4 *VDR* rs1544410 potential biomarkers

When comparing GG against AA genotype in the general population (Figure 3.19A), the AA genotype was significantly associated with abundance of bacterial family *Bartonellaceae* (LDA score ≥ 2.0, q < 0.2). Looking only into the efavirenz group (Figure 3.19B), the AA genotype was significantly enriched with bacterial families *Bartonellaceae* and *Sinobacteraceae* when compared with the GG genotype after FDR correction (LDA score ≥ 2.0, q < 0.2).

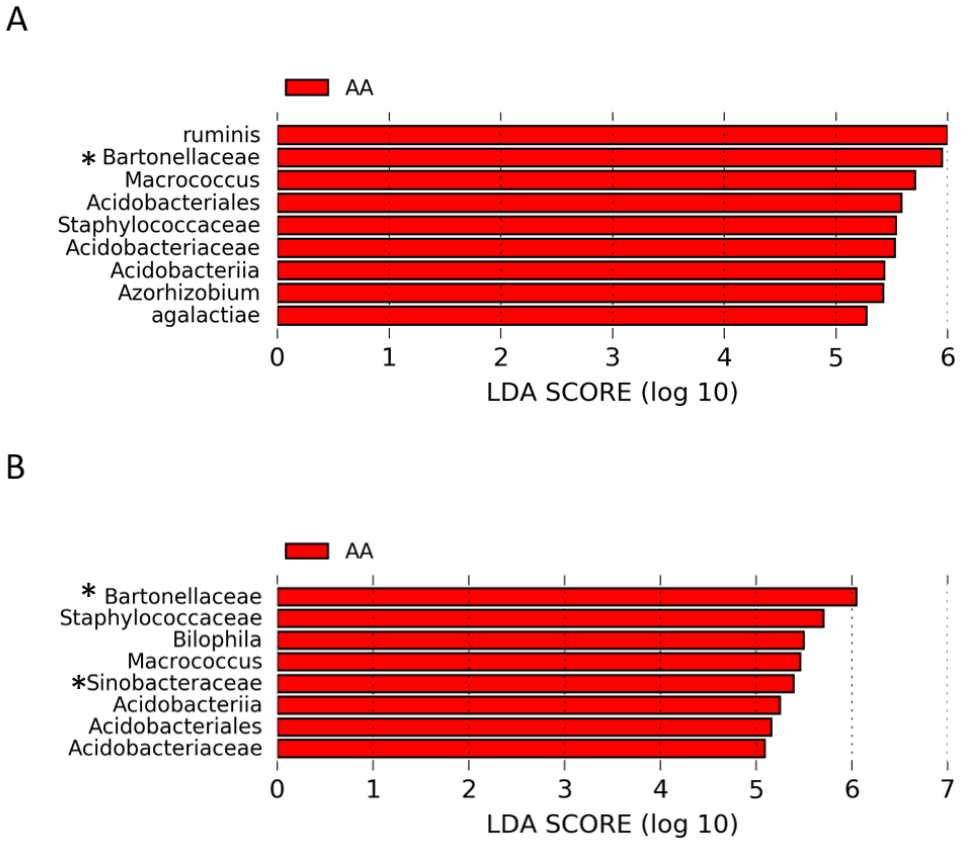


Figure 3. 19: Potential biomarkers for *VDR* rs1544410 by comparing GG vs AA genotypes

When comparing GG+GA against AA genotype, the gut bacterial composition of participants in the general (Figure 3.20A) with AA genotype was enriched with genus *Azorhizobium* and species *Lactobacillus ruminis* (LDA score ≥ 2.0, q < 0.2). Moreover, looking only in the efavirenz group (Figure 3.20B), family *Sinobacteraceae* and species *Lactobacillus ruminis* and *Bifidobacterium* bifidum were significantly abundant in the gut bacterial composition of participants with the AA genotype than those with the G allele (GG+GA genotype) after FDR correction (LDA score ≥ 2.0, q < 0.2).

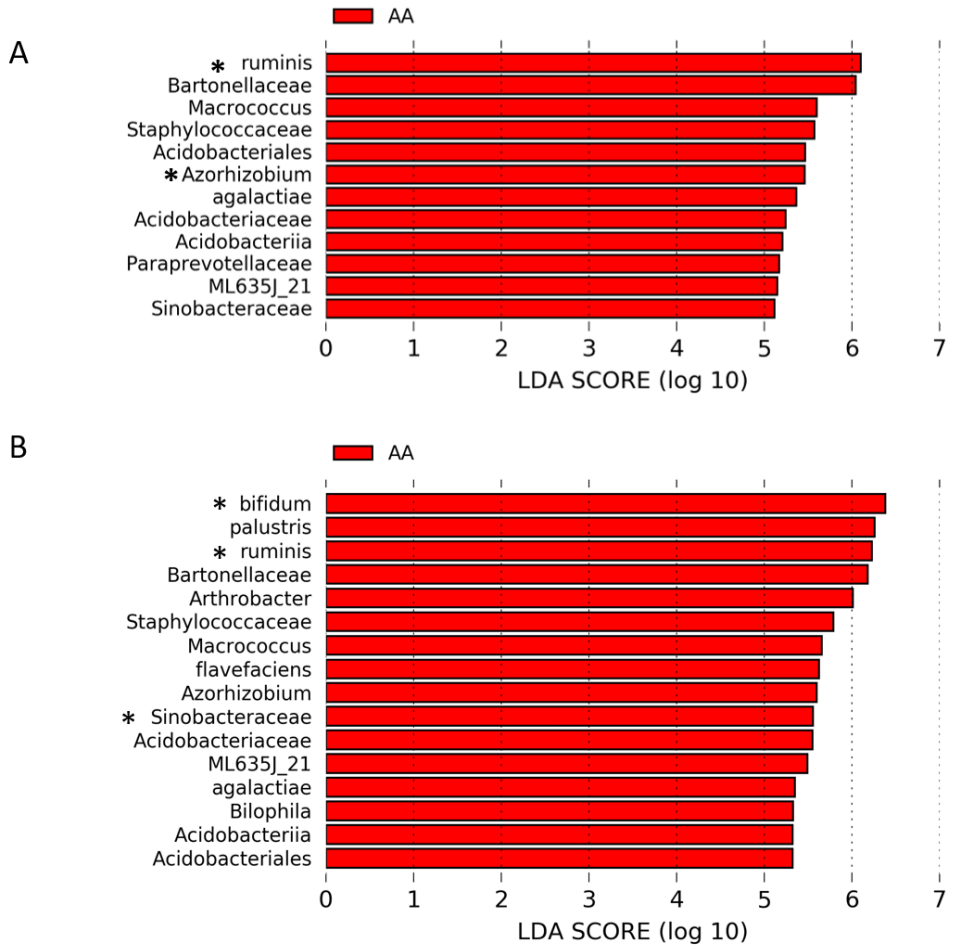


Figure 3. 20: Potential biomarkers for *VDR* rs1544410 by comparing GG+GA vs AA genotypes

3.8.4.5 *VDR* rs7975232 potential biomarkers

For the rs7975232 SNP in general (Figure 3.21), the gut bacterial composition of individuals with the CC genotype were significantly enriched with genera Sneathia, Veillonella, unclassified WAL 1855D and *Flavobacterium* (LDA score ≥ 2.0, $q < 0.2$) than those with the AA+AC genotype (A allele). At species level, *Arcobacter cryaerophilus*, *Methylobacterium adhaesivum* and *Lactobacillus plantarum* were identified to be significantly abundant in participants with the CC genotype (LDA score ≥ 2.0, $q < 0.2$).

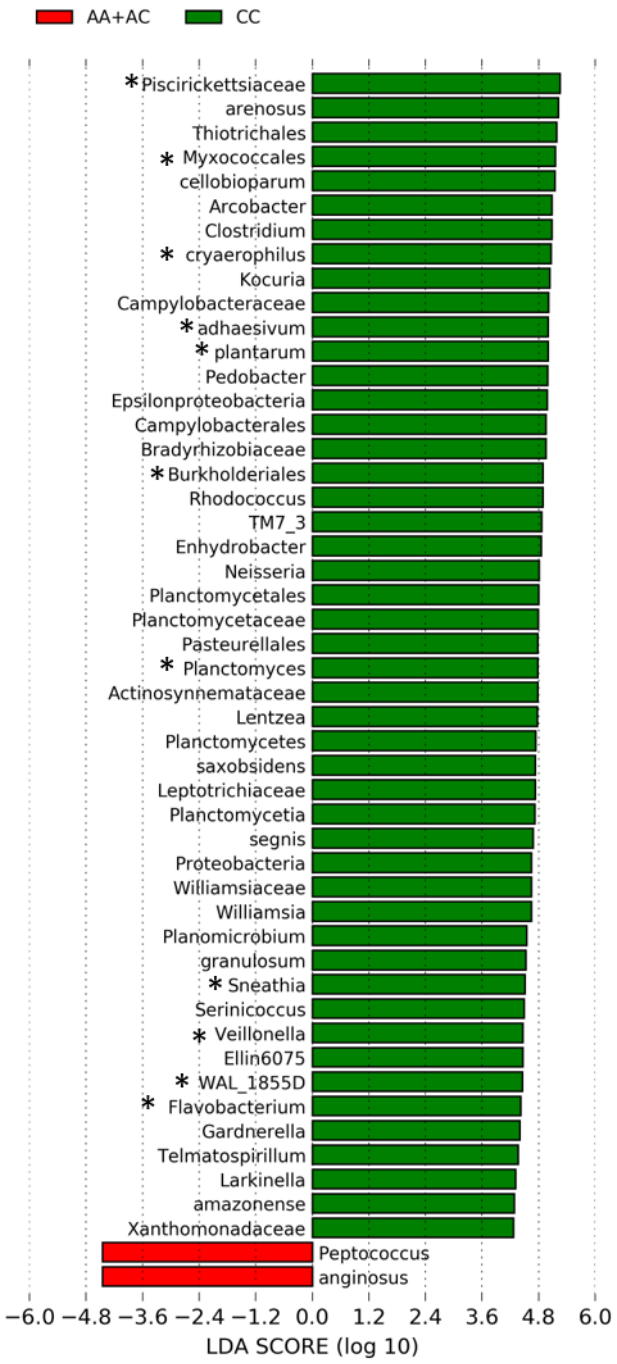


Figure 3. 21: Potential biomarkers for *VDR* rs7975232 by comparing AA+AC vs CC genotypes

3.8.4.6 *LCT* rs2164210 potential biomarkers

As per the rs2164210 SNP, the gut bacteria for participants in general (Figure 3.22A) with the CC genotype were found to be significantly (LDA score ≥ 2.0,

q < 0.2) enriched with genus *Dyella* and species *Clostridium thermopalmarium* and *Capnocytophaga ochracea* when compared with those with the TT genotype. However, when looking only in the efavirenz group (Figure 3.22B), none of the bacterial taxa were significantly differentially abundant after FDR correction (LDA score ≥ 2.0, q < 0.2) when comparing TT and CC genotypes of rs2164210 SNP.

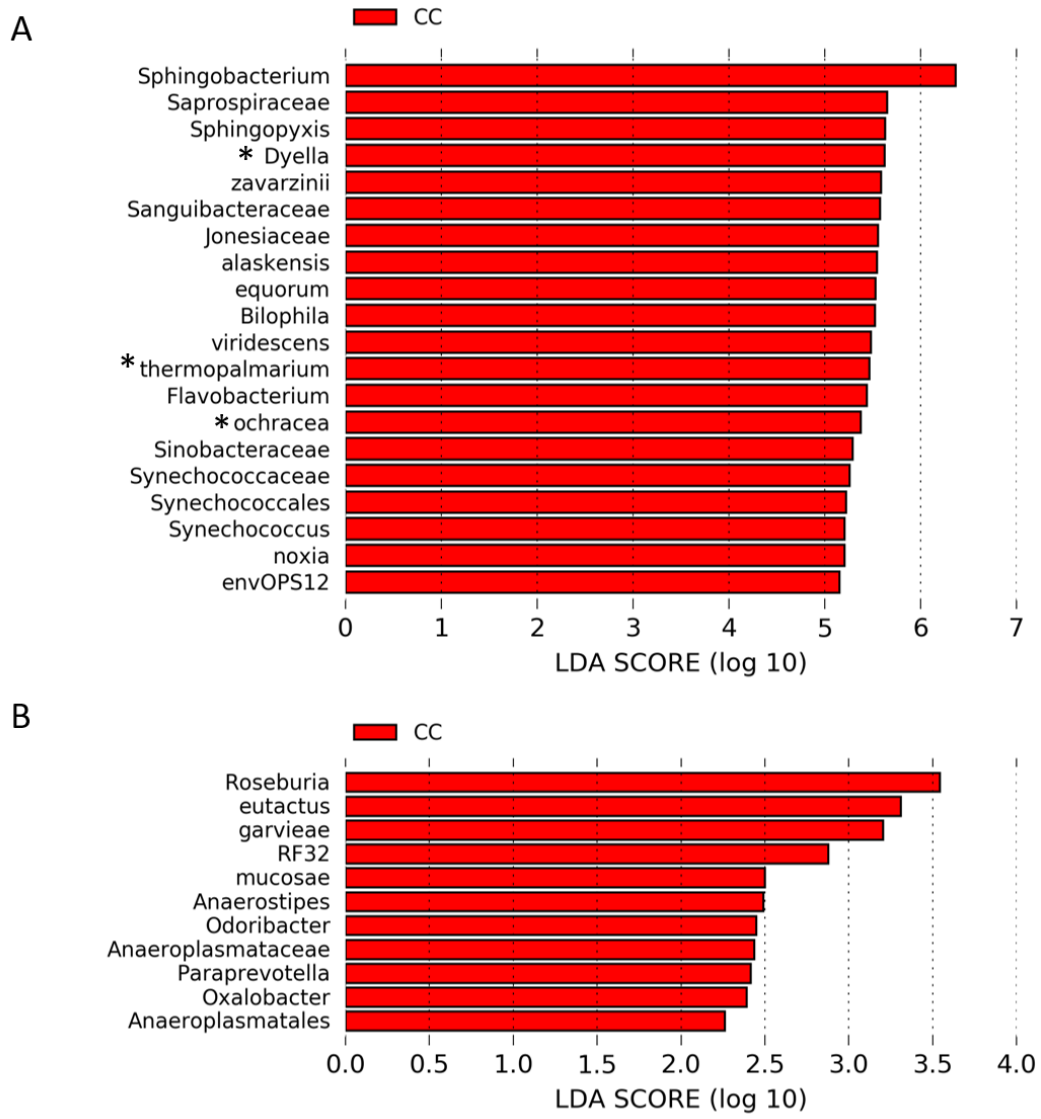


Figure 3. 22: Potential biomarkers for *LCT* rs2164210 by comparing TT vs CC genotypes

When comparing the presence of T allele (TT+TC genotype) and its absence (TT genotype), the general population (Figure 3.23) showed that the TT genotype

was significantly enriched with the genus *Dyella* and species *Clostridium thermopalmarium* and *Capnocytophaga ochracea* (LDA score ≥ 2.0, q < 0.2).

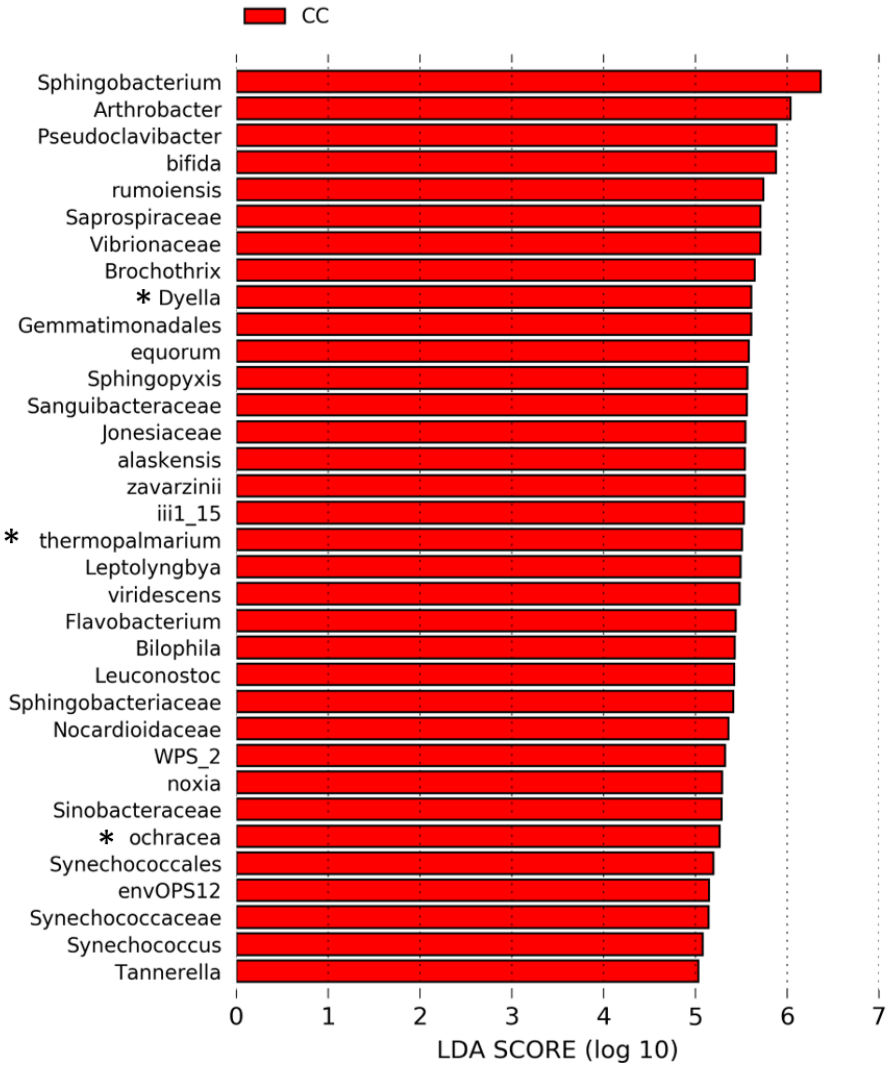


Figure 3. 23: Potential biomarkers for *LCT* rs2164210 by comparing TT+TC vs CC genotypes

3.8.4.7 *LCT* rs7579771 potential biomarkers

When comparing AA+TA and TT genotypes for the general population (Figure 3.24), LEfSe results indicated that species *Treponema amylovorum*, *Sphingobacterium multivorum* and *Cellulomonas xylanilytica* were significantly abundant (LDA score ≥ 2.0, q < 0.2) in gut bacteria of the participants with TT genotype.

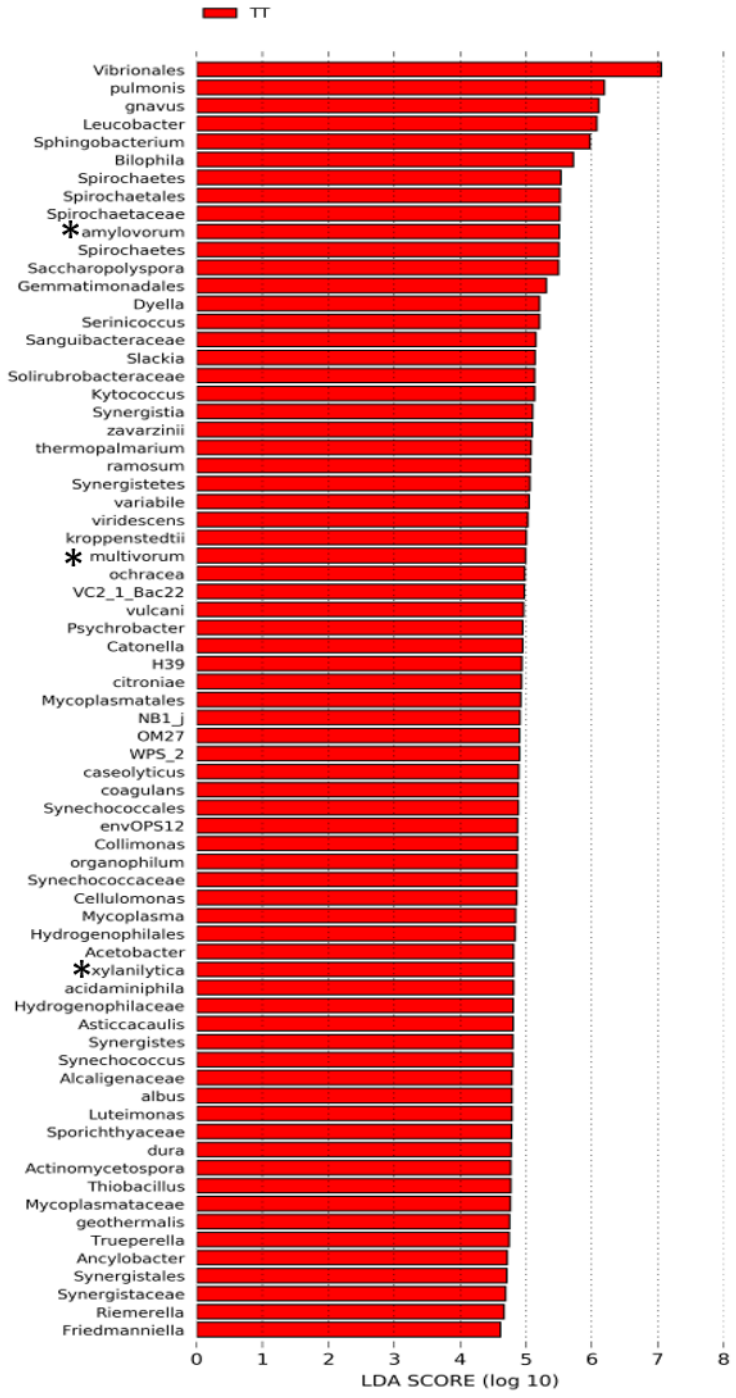


Figure 3. 24: Potential biomarkers for *LCT* rs7579771 by comparing AA+TA and TT genotypes

CHAPTER FOUR: DISCUSSION

This study examined the interplay between host genetics, drug exposure and gut bacterial diversity of Black Zimbabweans. Host candidate SNPs and gut bacterial composition were characterised, and the drug levels were readily available.

4.1 Findings on genetic variants' frequencies

In this study, frequency and distribution of 8 candidate SNPs in five genes that have been shown to influence gut bacterial profiles elsewhere were characterised. The SNPs were selected based on the MAFs reported in African populations.

We report on the frequency of variant rs601338A of *FUT2* (43%), which is comparable with that of other African (Luhya and Mende) and African American populations, since they all have African origin. However, it was not the case with Yoruba, whose rs601338A allele frequency was higher than in our Zimbabwean population. African populations are enriched with significant genetic diversity (Campbell and Tishkoff, 2008). Genetic admixture and gene flow could have led to difference in allele frequencies of this SNP. A study by Oussalah et al. (2012) in West Africa (Benin and Togo) also reported the occurrence of rs601338A = 43%, which is the same as in our study population. Surprisingly, rs601338A frequency was also comparable between the population of this study and that of the European, however, some other studies show a range of A allele from 43% to 53% (Wacklin et al., 2014; Barton et al., 2019; Oussalah et al., 2012). rs601338A is responsible for the non-secretor phenotype in Europeans, Iranians, and Africans but not in Asians. rs601338A allele frequency of this study population was much higher than that of East Asian, probably because positive selection for rs601338 occurred in other populations (Ferrer-Admetlla et al., 2009). Other studies show that rs601338A allele frequency was also less than 1% in the East Asian population (Ferrer-Admetlla et al., 2009; Hu et al., 2014; Chen et al., 2017).

The frequency of NCAM1 rs17115310G (23%) in this study population was comparable with that of African Americans, Luhya and European populations but not with the Mende, Yoruba and East Asian populations. Although there isn't

much literature on this variant, differences in MAF among African populations could be due to geographical distribution (Choudhury et al., 2020).

Literature on *SHROOM3* gene is relatively scarce. rs11724031A frequency (19.2%) in this study population was comparable with other African populations tabled in this study. The frequency of this variant could be close to a fixation in African populations but not in populations out of Africa (Campbell and Tishkoff, 2008).

VDR gene is one of the genes that have been extensively studied. rs1544410, rs7975232 and rs731236 polymorphisms are usually studied in haplotypes and have been found to be in strong LD in a West African (The Gambia, Guinea, and Guinea-Bissau) populations (Bornman et al., 2004); and Egyptians (El-Shorbagy et al., 2017). In this study, the D' is very strong between BsmI and ApaI and between ApaI and TaqI. A similar pattern was observed between TaqI and ApaI among Hispanic/Latina subjects (Mishra et al., 2013).

Variation within the LCT gene has also been studied in haplotypes. In this study, rs2164210 and rs7579771 were studied and the D' is very strong between the two loci. This LD between rs2164210 and rs7579771 has not been reported in any study.

4.2 Comparing bacterial diversity with other populations

In this study, we examined the effect of host genetics on the gut microbiome profiles of the black Zimbabweans. This is the first study to assess the influence of host genetics on the gut microbiome in black Africans.

In the general population, the rs1544410 of VDR gene and rs7579771 of LCT gene influenced the evenness of gut microbiome. In the control group only, rs17115310 of NCAM1 gene and rs7579771 of LCT gene influenced the Shannon index and Faith PD. In the efavirenz group only, rs731236 of VDR gene had an influence in the alpha diversity of the gut bacteriome. The rs601338A of FUT2; rs1544410 and rs731236 of *VDR*; and rs7579771 of *LCT* showed to influence the

alpha diversity of the gut microbiome in the warfarin group. rs7579771 of *LCT* gene in showed to influence the diversity (alpha) of the gut microbiome. When it comes to the beta diversity, rs1544410 of VDR and rs7579771 of LCT showed to influence the alpha diversity of the gut microbiome in the warfarin group. rs7579771 of LCT gene showed to influence the diversity of the gut microbiome.

In this study, *Kocuria kristinae* was found to be more reproducible within the FUT2 gene. FUT2 rs601338 has been associated with an increased risk of Coeliac disease (CD) in infants (Parmar et al., 2012). Sánchez et al. (2013) isolated this species in Spanish children with active CD. Although this study did not look at the CD, this data could indicate that rs601338 of the FUT2 gene play a role in influencing the abundance of *Kocuria kristinae*.

Family *Caldilineaceae* was found to be strongly associated with rs11724031 SHROOM3 gene in this study's population. Species of family *Caldilineaceae* have been isolated from thermophilic ecosystems before (Sekiguchi and Hanada, 2020) but not in human gut microbiome. SHROOM3 gene has been studied and found to influence the cell morphology (Das et al., 2014). This could probably favour the adherence of *Caldilineaceae* family of bacteria in the gut, however this possible relationship has not been established in literature.

Changes in the VDR gene has been associated with shift in the gut microbiome, more especially Parabacteroides (Jin et al., 2015). In this study, abundance of bacterial family *Bartonellaceae* and species *Lactobacillus ruminis* was more associated with the rs1544410 of the VDR gene. On the other hand, rs7975232 of the VDR gene was found to be strongly associated with the abundance of the bacterial family *Piscirickettsiaceae*. The findings in this study differ from those in a German cohort (Wang et al., 2016) where the rs7974353, an intron variant of this gene was found to be associated with beta-diversity and abundance of *Parabacteroides*.

LCT gene has been previously associated with the abundance of *Bifidobactria* (Goodrich et al., 2016). In this study, rs2164210 of the LCT gene has been found to be strongly associated with genus Dyella, species *Clostridium*

thermopalmarium and *Capnocytophaga ochracea*. Conversely, rs7579771 of LCT gene was *Treponema amylovorum*, *Sphingobacterium multivorum* and *Cellulomonas xylanilytica*. None of these bacterial taxa were closely related to the *Bifidobacteria* and these associations have not yet been observed in other studies.

4.3 Significance of this study's findings.

Response to drug treatment shows inter-individual variability. This study looks at the potential influence of host genetics and drug therapy in gut bacterial diversity in African population of Zimbabwe. This could give a more meaningful insight into how the host genetics and gut bacterial diversity can influence drug response. These factors could come to aid in determining a more accurate dosing algorithms in African populations. This could potentially reduce the medical costs as we move from trial-and-error method of prescribing drugs to a more precise dosing driven by both host genetics and gut bacterial diversity among other factors.

Conclusion

Despite a relatively small sample size, this study presented the baseline composition and diversity of the gut bacteria in case-controlled manner in Black Zimbabwean women and identified possible bacterial taxa and host genetics that could be important in influencing drug response. A further research with a larger sample size could validate the findings of this study and compare with those from other populations.

References

Aagaard, K., Petrosino, J., Keitel, W., Watson, M., Katancik, J., Garcia, N., Patel, S., Cutting, M., Madden, T., Hamilton, H., Harris, E., Gevers, D., Simone, G., Mcinnes, P., Versalovic, J., 2013. The Human Microbiome Project Strategy For Comprehensive Sampling Of The Human Microbiome And Why It Matters. FASEB J. 27, 1012–1022. Https://Doi.Org/10.1096/Fj.12-220806

Aceview: Gene:FUT2, A Comprehensive Annotation Of Human, Mouse And Worm Genes With Mrnas Or Estsaceview. [WWW Document], N.D. URL Https://Www.Ncbi.Nlm.Nih.Gov/Ieb/Research/Acembly/Av.Cgi?Db=Human&Term=FUT2&Submit=Go (Accessed 7.16.20).

Adams, J., 2008. Pharmacogenomics And Personalized Medicine | Learn Science At Scitable [WWW Document]. URL Http://Www.Nature.Com/Scitable/Topicpage/Pharmacogenomics-And-Personalized-Medicine-643 (Accessed 11.17.22).

Ahmed, J.H., Makonnen, E., Fotoohi, A., Yimer, G., Seifu, D., Assefa, M., Tigeneh, W., Aseffa, A., Howe, R., Aklillu, E., 2019. Vitamin D Status And Association Of VDR Genetic Polymorphism To Risk Of Breast Cancer In Ethiopia. Nutrients 11, 289. Https://Doi.Org/10.3390/Nu11020289

Ahmed, S., Zhou, Z., Zhou, J., Chen, S.-Q., 2016. Pharmacogenomics Of Drug Metabolizing Enzymes And Transporters: Relevance To Precision Medicine. Genomics Proteomics Bioinformatics 14, 298–313. Https://Doi.Org/10.1016/J.Gpb.2016.03.008

Alessandri, G., Ossiprandi, M.C., Macsharry, J., Van Sinderen, D., Ventura, M., 2019. Bifidobacterial Dialogue With Its Human Host And Consequent Modulation Of The Immune System. Front. Immunol. 10. Https://Doi.Org/10.3389/Fimmu.2019.02348

Auer, L., Mariadassou, M., O'Donohue, M., Klopp, C., Hernandez-Raquet, G., 2017. Analysis Of Large 16S Rrna Illumina Data Sets: Impact Of Singleton Read Filtering On Microbial Community Description. Mol. Ecol. Resour. 17, E122–E132. Https://Doi.Org/10.1111/1755-0998.12700

Aziz, R.K., Saad, R., Rizkallah, M.R., 2011. Pharmacomicrobiomics Or How Bugs Modulate Drugs: An Educational Initiative To Explore The Effects Of Human Microbiome On Drugs. BMC Bioinformatics 12, A10. Https://Doi.Org/10.1186/1471-2105-12-S7-A10

Bader, L.A., Elewa, H., 2016. The Impact Of Genetic And Non-Genetic Factors On Warfarin Dose Prediction In MENA Region: A Systematic Review. Plos ONE 11. Https://Doi.Org/10.1371/Journal.Pone.0168732

Bank, P., Caudle, K., Swen, J., Gammal, R., Whirl-Carrillo, M., Klein, T., Relling, M., Guchelaar, H.-J., 2018. Comparison Of The Guidelines Of The Clinical Pharmacogenetics Implementation Consortium And The Dutch Pharmacogenetics Working Group. Clin. Pharmacol. Ther. 103, 599–618. Https://Doi.Org/10.1002/Cpt.762

Barton, S.J., Murray, R., Lillycrop, K.A., Inskip, H.M., Harvey, N.C., Cooper, C., Karnani, N., Zolezzi, I.S., Sprenger, N., Godfrey, K.M., Binia, A., 2019. FUT2 Gene Variants And Reported Respiratory And Gastrointestinal

Illnesses During Infancy. J. Infect. Dis. 219, 836–843. Https://Doi.Org/10.1093/Infdis/Jiy582
Belle, D.J., Singh, H., 2008. Genetic Factors In Drug Metabolism. Am. Fam. Physician 77, 1553–1560.
Bjørkhaug, S.T., Aanes, H., Neupane, S.P., Bramness, J.G., Malvik, S., Henriksen, C., Skar, V., Medhus, A.W., Valeur, J., 2019. Characterization Of Gut Microbiota Composition And Functions In Patients With Chronic Alcohol Overconsumption. Gut Microbes 0, 1–13. Https://Doi.Org/10.1080/19490976.2019.1580097
Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., Clark, A.G., 2015. Host Genetic Variation Impacts Microbiome Composition Across Human Body Sites. Genome Biol. 16, 191. Https://Doi.Org/10.1186/S13059-015-0759-1
Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K.B., Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., Mcdonald, D., Mciver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., Van Der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., Von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, Interactive, Scalable And Extensible Microbiome Data Science Using QIIME 2. Nat. Biotechnol. 37, 852–857. Https://Doi.Org/10.1038/S41587-019-0209-9
Bonder, M.J., Kurilshikov, A., Tigchelaar, E.F., Mujagic, Z., Imhann, F., Vila, A.V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S.P., Zhernakova, D.V., Jankipersadsing, S.A., Jaeger, M., Oosting, M., Cenit, M.C., Masclee, A.A.M., Swertz, M.A., Li, Y., Kumar, V., Joosten, L., Harmsen, H., Weersma, R.K., Franke, L., Hofker, M.H., Xavier, R.J., Jonkers, D., Netea, M.G., Wijmenga, C., Fu, J., Zhernakova, A., 2016. The Effect Of Host Genetics On The Gut Microbiome. Nat. Genet. 48, 1407–1412. Https://Doi.Org/10.1038/Ng.3663
Bornman, L., Campbell, S., Fielding, K., Bah, B., Sillah, J., Gustafson, P., Manneh, K., Lisse, I., Allen, A., Sirugo, G., Sylla, A., Aaby, P., Mcadam,

K., Bah-Sow, O., Bennett, S., Lienhardt, C., Hill, A., 2004. Vitamin D Receptor Polymorphisms And Susceptibility To Tuberculosis In West Africa: A Case-Control And Family Study. J. Infect. Dis. 190, 1631–41. Https://Doi.Org/10.1086/424462

Campbell, M.C., Tishkoff, S.A., 2008. AFRICAN GENETIC DIVERSITY: Implications For Human Demographic History, Modern Human Origins, And Complex Disease Mapping. Annu. Rev. Genomics Hum. Genet. 9, 403–433. Https://Doi.Org/10.1146/Annurev.Genom.9.081307.164258

Cárcer, D.A. De, Denman, S.E., Mcsweeney, C., Morrison, M., 2011. Evaluation Of Subsampling-Based Normalization Strategies For Tagged High-Throughput Sequencing Data Sets From Gut Microbiomes. Appl. Environ. Microbiol. 77, 8795–8798. Https://Doi.Org/10.1128/AEM.05491-11

Chen, C.T., Liao, W.Y., Hsu, C.C., Hsueh, K.C., Yang, S.F., Teng, Y.H., Yu, Y.L., 2017. FUT2 Genetic Variants As Predictors Of Tumor Development With Hepatocellular Carcinoma. Int. J. Med. Sci. 14, 885–890. Https://Doi.Org/10.7150/Ijms.19734

Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y.J., Ghoorah, A.W., Dareng, E., Odia, T., Falola, O., Adebiyi, E., Hazelhurst, S., Mazandu, G., Nyangiri, O.A., Mbiyavanga, M., Benkahla, A., Kassim, S.K., Mulder, N., Adebamowo, S.N., Chimusa, E.R., Muzny, D., Metcalf, G., Gibbs, R.A., Rotimi, C., Ramsay, M., Adeyemo, A.A., Lombard, Z., Hanchard, N.A., 2020. High-Depth African Genomes Inform Human Migration And Health. Nature 586, 741–748. Https://Doi.Org/10.1038/S41586-020-2859-7

Choudhury, A., Aron, S., Sengupta, D., Hazelhurst, S., Ramsay, M., 2018. African Genetic Diversity Provides Novel Insights Into Evolutionary History And Local Adaptations. Hum. Mol. Genet. 27, R209–R218. Https://Doi.Org/10.1093/Hmg/Ddy161

Chung, M.-I., Nascone-Yoder, N.M., Grover, S.A., Drysdale, T.A., Wallingford, J.B., 2010. Direct Activation Of Shroom3 Transcription By Pitx Proteins Drives Epithelial Morphogenesis In The Developing Gut. Development 137, 1339–1349. Https://Doi.Org/10.1242/Dev.044610

Clarke, G., Sandhu, K.V., Griffin, B.T., Dinan, T.G., Cryan, J.F., Hyland, N.P., 2019. Gut Reactions: Breaking Down Xenobiotic–Microbiome Interactions. Pharmacol. Rev. 71, 198–224. Https://Doi.Org/10.1124/Pr.118.015768

Colombini, A., Brayda-Bruno, M., Lombardi, G., Croiset, S., Ceriani, C., Buligan, C., Barbina, M., Banfi, G., Cauci, S., 2016. Bsmi, Apai And Taqi Polymorphisms In The Vitamin D Receptor Gene (VDR) And Association With Lumbar Spine Pathologies: An Italian Case-Control Study. PLOS ONE 11, E0155004. Https://Doi.Org/10.1371/Journal.Pone.0155004

Cully, M., 2019. Antibiotics Alter The Gut Microbiome And Host Health. Nat. Res. Https://Doi.Org/10.1038/D42859-019-00019-X

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K.J., Gall, A., Girón, C.G., Gil, L., Grego, T., Haggerty, L.,

Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Marugán, J.C., Maurel, T., Mcmahon, A.C., Moore, B., Morales, J., Mudge, J.M., Nuhn, M., Ogeh, D., Parker, A., Parton, A., Patricio, M., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sparrow, H., Stapleton, E., Szuba, M., Taylor, K., Threadgold, G., Thormann, A., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Yates, A.D., Zerbino, D.R., Flicek, P., 2019. Ensembl 2019. Nucleic Acids Res. 47, D745–D751. Https://Doi.Org/10.1093/Nar/Gky1113

Dąbrowska, K., Witkiewicz, W., 2016. Correlations Of Host Genetics And Gut Microbiome Composition. Front. Microbiol. 7. Https://Doi.Org/10.3389/Fmicb.2016.01357

Dalby, A.J., Wessels, P., Opie, L.H., 2013. Warfarin In Non-Valvular Atrial Fibrillation. S. Afr. Med. J. 103, 901-904–904. Https://Doi.Org/10.7196/SAMJ.7172

Das, A., Srinivasan, M., Ghosh, T.S., Mande, S.S., 2016. Xenobiotic Metabolism And Gut Microbiomes. PLOS ONE 11, E0163099. Https://Doi.Org/10.1371/Journal.Pone.0163099

Das, D., Zalewski, J.K., Mohan, S., Plageman, T.F., Vandemark, A.P., Hildebrand, J.D., 2014. The Interaction Between Shroom3 And Rho-Kinase Is Required For Neural Tube Morphogenesis In Mice. Biol. Open 3, 850–860. Https://Doi.Org/10.1242/Bio.20147450

Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., Gilad, Y., 2015. Genome-Wide Association Studies Of The Human Gut Microbiota. PLOS ONE 10, E0140301. Https://Doi.Org/10.1371/Journal.Pone.0140301

De La Cuesta-Zuluaga, J., Escobar, J.S., 2016. Considerations For Optimizing Microbiome Analysis Using A Marker Gene. Front. Nutr. 3. Https://Doi.Org/10.3389/Fnut.2016.00026

Deak, K.L., Boyles, A.L., Etchevers, H.C., Melvin, E.C., Siegel, D.G., Graham, F.L., Slifer, S.H., Enterline, D.S., George, T.M., Vekemans, M., Mcclay, D., Bassuk, A.G., Kessler, J.A., Linney, E., Gilbert, J.R., Speer, M.C., NTD Collaborative Group, 2005. Snps In The Neural Cell Adhesion Molecule 1 Gene (NCAM1) May Be Associated With Human Neural Tube Defects. Hum. Genet. 117, 133–142. Https://Doi.Org/10.1007/S00439-005-1299-7

Dikeocha, I.J., Al-Kabsi, A.M., Miftahussurur, M., Alshawsh, M.A., 2022. Pharmacomicrobiomics: Influence Of Gut Microbiota On Drug And Xenobiotic Metabolism. FASEB J. 36, E22350. Https://Doi.Org/10.1096/Fj.202101986R

Doestzada, M., Vila, A.V., Zhernakova, A., Koonen, D.P.Y., Weersma, R.K., Touw, D.J., Kuipers, F., Wijmenga, C., Fu, J., 2018. Pharmacomicrobiomics: A Novel Route Towards Personalized Medicine? Protein Cell 9, 432–445. Https://Doi.Org/10.1007/S13238-018-0547-2

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 J. Integr. Biol. 18, 402–414. Https://Doi.Org/10.1089/Omi.2014.0018
El-Shorbagy, H.M., Mahmoud, N.H., Sabet, S., 2017. Association Of Vitamin D Receptor Gene Polymorphisms With Breast Cancer Risk In An Egyptian Population. Tumor Biol. 39, 1010428317727738. Https://Doi.Org/10.1177/1010428317727738
Ferrer-Admetlla, A., Sikora, M., Laayouni, H., Esteve, A., Roubinet, F., Blancher, A., Calafell, F., Bertranpetit, J., Casals, F., 2009. A Natural History Of FUT2 Polymorphism In Humans. Mol. Biol. Evol. 26, 1993–2003. Https://Doi.Org/10.1093/Molbev/Msp108
Fischer, H.D., Juurlink, D.N., Mamdani, M.M., Kopp, A., Laupacis, A., 2010. Hemorrhage During Warfarin Therapy Associated With Cotrimoxazole And Other Urinary Tract Anti-Infective Agents: A Population-Based Study. Arch. Intern. Med. 170, 617–621. Https://Doi.Org/10.1001/Archinternmed.2010.37
Fontana, Panebianco, Picchianti-Diamanti, A., Laganà, Cavalieri, Potenza, Pracella, Binda, Copetti, Pazienza, V., 2019. Gut Microbiota Profiles Differ Among Individuals Depending On Their Region Of Origin: An Italian Pilot Study. Int. J. Environ. Res. Public. Health 16, 4065. Https://Doi.Org/10.3390/Ijerph16214065
Forsgård, R.A., 2019. Lactose Digestion In Humans: Intestinal Lactase Appears To Be Constitutive Whereas The Colonic Microbiome Is Adaptable. Am. J. Clin. Nutr. 110, 273–279. Https://Doi.Org/10.1093/Ajcn/Nqz104
Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., Ley, R.E., 2016. Genetic Determinants Of The Gut Microbiome In UK Twins. Cell Host Microbe 19, 731–743. Https://Doi.Org/10.1016/J.Chom.2016.04.017
Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., Spector, T.D., Clark, A.G., Ley, R.E., 2014. Human Genetics Shape The Gut Microbiome. Cell 159, 789–799. Https://Doi.Org/10.1016/J.Cell.2014.09.053
Gupta, R., Ordonez, R.M., Koenig, S., 2012. Global Impact Of Antiretroviral Therapy-Associated Diarrhea. AIDS Patient Care Stds 26, 711–713. Https://Doi.Org/10.1089/Apc.2012.0299
Horner-Devine, M.C., Lage, M., Hughes, J.B., Bohannan, B.J.M., 2004. A Taxa-Area Relationship For Bacteria. Nature 432, 750–753. Https://Doi.Org/10.1038/Nature03073
Hu, D., Shao, X., Xu, C., Xia, S., Yu, L., Jiang, L., Jin, J., Lin, X., Jiang, Y., 2014. Associations Of FUT2 And FUT3 Gene Polymorphisms With Crohn's Disease In Chinese Patients. J. Gastroenterol. Hepatol. 29, 1778–1785. Https://Doi.Org/10.1111/Jgh.12599
Human Microbiome Project Consortium, 2012a. A Framework For Human Microbiome Research. Nature 486, 215–221. Https://Doi.Org/10.1038/Nature11209

Human Microbiome Project Consortium, 2012b. Structure, Function And Diversity Of The Healthy Human Microbiome. Nature 486, 207–214. Https://Doi.Org/10.1038/Nature11234
Ingelman-Sundberg, M., 2001. Genetic And Environmental Causes For Interindividual Variability In Drug Pharmacokinetics. Int. Congr. Ser., Optimal Dose Identification 1220, 175–186. Https://Doi.Org/10.1016/S0531-5131(01)00296-5
Jacobs, A., Bassa, F., Decloedt, E., 2017. A Preliminary Review Of Warfarin Toxicity In A Tertiary Hospital In Cape Town, South Africa. Cardiovasc. J. Afr. 28, 1–4. Https://Doi.Org/10.5830/CVJA-2017-029
Ji, Y., Zhang, F., Zhang, R., Shen, Y., Liu, L., Wang, J., Yang, J., Tang, Q., Xun, J., Qi, T., Wang, Z., Song, W., Tang, Y., Chen, J., Lu, H., 2018. Changes In Intestinal Microbiota In HIV-1-Infected Subjects Following Cart Initiation: Influence Of CD4+ T Cell Count. Emerg. Microbes Infect. 7. Https://Doi.Org/10.1038/S41426-018-0117-Y
Jin, D., Wu, S., Zhang, Y., Lu, R., Xia, Y., Dong, H., Sun, J., 2015. Lack Of Vitamin D Receptor Causes Dysbiosis And Changes The Functions Of The Murine Intestinal Microbiome. Clin. Ther. 37, 996-1009.E7. Https://Doi.Org/10.1016/J.Clinthera.2015.04.004
Josenhans, C., Müthing, J., Elling, L., Bartfeld, S., Schmidt, H., 2020. How Bacterial Pathogens Of The Gastrointestinal Tract Use The Mucosal Glyco-Code To Harness Mucus And Microbiota: New Ways To Study An Ancient Bag Of Tricks. Int. J. Med. Microbiol. IJMM 310, 151392. Https://Doi.Org/10.1016/J.Ijmm.2020.151392
Kachambwa, P., 2017. Influence Of Gut Microbiota On Immune System In Infants.
Kaur, P., Gupta, M., Sagar, V., 2022. FUT2 Gene As A Genetic Susceptible Marker Of Infectious Diseases: A Review. Int. J. Mol. Epidemiol. Genet. 13, 1–14.
Kolde, R., Franzosa, E.A., Rahnavard, G., Hall, A.B., Vlamakis, H., Stevens, C., Daly, M.J., Xavier, R.J., Huttenhower, C., 2018. Host Genetic Variation And Its Microbiome Interactions Within The Human Microbiome Project. Genome Med. 10. Https://Doi.Org/10.1186/S13073-018-0515-8
Kudzi, W., Ahorhorlu, S.Y., Dzudzor, B., Olayemi, E., Nartey, E.T., Asmah, R.H., 2016. Genetic Polymorphisms Of Patients On Stable Warfarin Maintenance Therapy In A Ghanaian Population. BMC Res. Notes 9, 507. Https://Doi.Org/10.1186/S13104-016-2306-X
Laäs, D.J., Naidoo, M., 2018. An Evaluation Of Warfarin Use At An Urban District-Level Hospital In Kwazulu-Natal Province, South Africa. S. Afr. Med. J. 108, 1046-1050–1050. Https://Doi.Org/10.7196/SAMJ.2018.V108i12.13256
Lee, B., Moon, K.M., Kim, C.Y., 2018. Tight Junction In The Intestinal Epithelium: Its Association With Diseases And Regulation By Phytochemicals. J. Immunol. Res. 2018. Https://Doi.Org/10.1155/2018/2645465
Li Z, Zhang Z, He Z, Tang W, Li T, Zeng Z, He L, Shi Y, 2009. A Partition-Ligation-Combination-Subdivision EM Algorithm For Haplotype Inference With Multiallelic Markers: Update Of The Shesis.

Lippi, G., Mattiuzzi, C., Cervellin, G., Favaloro, E.J., 2017. Direct Oral Anticoagulants: Analysis Of Worldwide Use And Popularity Using Google Trends. Ann. Transl. Med. 5, 4. Https://Doi.Org/10.21037/Atm.2017.06.65

Liu, S., Ren, F., Zhao, L., Jiang, L., Hao, Y., Jin, J., Zhang, M., Guo, H., Lei, X., Sun, E., Liu, H., 2015. Starch And Starch Hydrolysates Are Favorable Carbon Sources For Bifidobacteria In The Human Gut. BMC Microbiol. 15. Https://Doi.Org/10.1186/S12866-015-0362-3

Lozupone, C.A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., Vázquez-Baeza, Y., Jansson, J.K., Gordon, J.I., Knight, R., 2013. Meta-Analyses Of Studies Of The Human Microbiota. Genome Res. 23, 1704–1714. Https://Doi.Org/10.1101/Gr.151803.112

Mcburney, M.I., Davis, C., Fraser, C.M., Schneeman, B.O., Huttenhower, C., Verbeke, K., Walter, J., Latulippe, M.E., 2019. Establishing What Constitutes A Healthy Human Gut Microbiome: State Of The Science, Regulatory Considerations, And Future Directions. J. Nutr. 149, 1882–1895. Https://Doi.Org/10.1093/Jn/Nxz154

Mcgovern, D.P.B., Jones, M.R., Taylor, K.D., Marciante, K., Yan, X., Dubinsky, M., Ippoliti, A., Vasiliauskas, E., Berel, D., Derkowski, C., Dutridge, D., International IBD Genetics Consortium, Fleshner, P., Shih, D.Q., Melmed, G., Mengesha, E., King, L., Pressman, S., Haritunians, T., Guo, X., Targan, S.R., Rotter, J.I., 2010. Fucosyltransferase 2 (FUT2) Non-Secretor Status Is Associated With Crohn's Disease. Hum. Mol. Genet. 19, 3468–3476. Https://Doi.Org/10.1093/Hmg/Ddq248

Mekaj, Y.H., Mekaj, A.Y., Duci, S.B., Miftari, E.I., 2015. New Oral Anticoagulants: Their Advantages And Disadvantages Compared With Vitamin K Antagonists In The Prevention And Treatment Of Patients With Thromboembolic Events. Ther. Clin. Risk Manag. 11, 967–977. Https://Doi.Org/10.2147/TCRM.S84210

Mishra, D.K., Wu, Y., Sarkissyan, M., Sarkissyan, S., Chen, Z., Shang, X., Ong, M., Heber, D., Koeffler, H.P., Vadgama, J.V., 2013. Vitamin D Receptor Gene Polymorphisms And Prognosis Of Breast Cancer Among African-American And Hispanic Women. PLOS ONE 8, E57967. Https://Doi.Org/10.1371/Journal.Pone.0057967

Mukonzo, J.K., Okwera, A., Nakasujja, N., Luzze, H., Sebuwufu, D., Ogwal-Okeng, J., Waako, P., Gustafsson, L.L., Aklillu, E., 2013. Influence Of Efavirenz Pharmacokinetics And Pharmacogenetics On Neuropsychological Disorders In Ugandan HIV-Positive Patients With Or Without Tuberculosis: A Prospective Cohort Study. BMC Infect. Dis. 13, 261. Https://Doi.Org/10.1186/1471-2334-13-261

Nagpal, R., Mainali, R., Ahmadi, S., Wang, S., Singh, R., Kavanagh, K., Kitzman, D.W., Kushugulova, A., Marotta, F., Yadav, H., 2018. Gut Microbiome And Aging: Physiological And Mechanistic Insights. Nutr. Healthy Aging 4, 267–285. Https://Doi.Org/10.3233/NHA-170030

Naidoo, P., Chetty, V.V., Chetty, M., 2014. Impact Of CYP Polymorphisms, Ethnicity And Sex Differences In Metabolism On Dosing Strategies: The Case Of Efavirenz. Eur. J. Clin. Pharmacol. 70, 379–389. Https://Doi.Org/10.1007/S00228-013-1634-1

Ndadza, A., Thomford, N.E., Mukanganyama, S., Wonkam, A., Ntsekhe, M., Dandara, C., 2019. The Genetics Of Warfarin Dose–Response Variability In Africans: An Expert Perspective On Past, Present, And Future. OMICS J. Integr. Biol. 23, 152–166. Https://Doi.Org/10.1089/Omi.2019.0018

Ndadzaarinao, Cindizinhle, Makambwaedson, Chimusaemile, Wonkamambroise, P, K., Ntsekhempiko, Dandaracollet, 2019. Warfarin Dose And CYP2C Gene Cluster: An African Ancestral-Specific Variant Is A Strong Predictor Of Dose In Black South African Patients. OMICS J. Integr. Biol. Https://Doi.Org/10.1089/Omi.2018.0174

Olesen, R., Wejse, C., Velez, D.R., Bisseye, C., Sodemann, M., Aaby, P., Rabna, P., Worwui, A., Chapman, H., Diatta, M., Adegbola, R.A., Hill, P.C., Østergaard, L., Williams, S.M., Sirugo, G., 2007. DC-SIGN (CD209), Pentraxin 3 And Vitamin D Receptor Gene Variants Associate With Pulmonary Tuberculosis Risk In West Africans. Genes Immun. 8, 456–467. Https://Doi.Org/10.1038/Sj.Gene.6364410

Onywera, H., Williamson, A.-L., Mbulawa, Z.Z.A., Coetzee, D., Meiring, T.L., 2019. The Cervical Microbiota In Reproductive-Age South African Women With And Without Human Papillomavirus Infection. Papillomavirus Res. 7, 154–163. Https://Doi.Org/10.1016/J.Pvr.2019.04.006

Osman, M.-A., Neoh, H., Ab Mutalib, N.-S., Chin, S.-F., Jamal, R., 2018. 16S Rrna Gene Sequencing For Deciphering The Colorectal Cancer Gut Microbiome: Current Protocols And Workflows. Front. Microbiol. 9. Https://Doi.Org/10.3389/Fmicb.2018.00767

Oussalah, A., Besseau, C., Chery, C., Jeannesson, E., Guéant-Rodriguez, R.-M., Anello, G., Bosco, P., Elia, M., Romano, A., Bronowicki, J.-P., Gerard, P., Paoli, J., Avogbe, P.H., Chabi, N., Sanni, A., Amouzou, E., Peyrin-Biroulet, L., Guéant, J.-L., 2012. Helicobacter Pylori Serologic Status Has No Influence On The Association Between Fucosyltransferase 2 Polymorphism (FUT2 461 G→A) And Vitamin B-12 In Europe And West Africa. Am. J. Clin. Nutr. 95, 514–521. Https://Doi.Org/10.3945/Ajcn.111.016410

Parmar, A.S., Alakulppi, N., Paavola-Sakki, P., Kurppa, K., Halme, L., Färkkilä, M., Turunen, U., Lappalainen, M., Kontula, K., Kaukinen, K., Mäki, M., Lindfors, K., Partanen, J., Sistonen, P., Mättö, J., Wacklin, P., Saavalainen, P., Einarsdottir, E., 2012. Association Study Of FUT2 (Rs601338) With Celiac Disease And Inflammatory Bowel Disease In The Finnish Population. Tissue Antigens 80, 488–493. Https://Doi.Org/10.1111/Tan.12016

Patel, S., Singh, R., Patel, N., 2019. Warfarin, In: Statpearls. Statpearls Publishing, Treasure Island (FL).

Perera, M.A., Cavallari, L.H., Limdi, N.A., Gamazon, E.R., Konkashbaev, A., Daneshjou, R., Pluzhnikov, A., Crawford, D.C., Wang, J., Liu, N., Tatonetti, N., Bourgeois, S., Takahashi, H., Bradford, Y., Burkley, B.M., Desnick, R.J., Halperin, J.L., Khalifa, S.I., Langaee, T.Y., Lubitz, S.A., Nutescu, E.A., Oetjens, M., Shahin, M.H., Patel, S.R., Sagreiya, H., Tector, M., Weck, K.E., Rieder, M.J., Scott, S.A., Wu, A.H., Burmester, J.K., Wadelius, M., Deloukas, P., Wagner, M.J., Mushiroda, T., Kubo, M.,

Roden, D.M., Cox, N.J., Altman, R.B., Klein, T.E., Nakamura, Y., Johnson, J.A., 2013. Genetic Variants Associated With Warfarin Dose In African-American Individuals: A Genome-Wide Association Study. The Lancet 382, 790–796. Https://Doi.Org/10.1016/S0140-6736(13)60681-9

Pinillos, F., Dandara, C., Swart, M., Strehlau, R., Kuhn, L., Patel, F., Coovadia, A., Abrams, E., 2016. Case Report: Severe Central Nervous System Manifestations Associated With Aberrant Efavirenz Metabolism In Children: The Role Of CYP2B6 Genetic Variation. BMC Infect. Dis. 16, 56. Https://Doi.Org/10.1186/S12879-016-1381-X

Pinto-Cardoso, S., Lozupone, C., Briceño, O., Alva-Hernández, S., Téllez, N., Adriana, A., Murakami-Ogasawara, A., Reyes-Terán, G., 2017. Fecal Bacterial Communities In Treated HIV Infected Individuals On Two Antiretroviral Regimens. Sci. Rep. 7, 43741. Https://Doi.Org/10.1038/Srep43741

Pirmohamed, M., 2018. Warfarin: The End Or The End Of One Size Fits All Therapy? J. Pers. Med. 8. Https://Doi.Org/10.3390/Jpm8030022

Quarta, G., Kim, S., Cho, I., Blaser, M., 2015. Mo1817 Warfarin Induces Intestinal Dysbiosis Involving Vitamin K-Expressing Bacteria. Gastroenterology 148, S-718. Https://Doi.Org/10.1016/S0016-5085(15)32447-1

Reay, R., Dandara, C., Viljoen, M., Rheeders, M., 2017. CYP2B6 Haplotype Predicts Efavirenz Plasma Concentration In Black South African HIV-1-Infected Children: A Longitudinal Pediatric Pharmacogenomic Study. Omics J. Integr. Biol. 21, 465–473. Https://Doi.Org/10.1089/Omi.2017.0078

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G.A.D., Gasbarrini, A., Mele, M.C., 2019. What Is The Healthy Gut Microbiota Composition? A Changing Ecosystem Across Age, Environment, Diet, And Diseases. Microorganisms 7. Https://Doi.Org/10.3390/Microorganisms7010014

Robarge, J.D., Metzger, I.F., Lu, J., Thong, N., Skaar, T.C., Desta, Z., Bies, R.R., 2017. Population Pharmacokinetic Modeling To Estimate The Contributions Of Genetic And Nongenetic Factors To Efavirenz Disposition. Antimicrob. Agents Chemother. 61, E01813-16. Https://Doi.Org/10.1128/AAC.01813-16

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., Shilo, S., Lador, D., Vila, A.V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B.C., Avnit-Sagi, T., Lotan-Pompan, M., Weinberger, A., Halpern, Z., Carmi, S., Fu, J., Wijmenga, C., Zhernakova, A., Elinav, E., Segal, E., 2018. Environment Dominates Over Host Genetics In Shaping Human Gut Microbiota. Nature 555, 210–215. Https://Doi.Org/10.1038/Nature25973

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., Tuohy, K., 2018. Gut Microbiota Functions: Metabolism Of Nutrients And Other Food Components. Eur. J. Nutr. 57, 1–24. Https://Doi.Org/10.1007/S00394-017-1445-8

Sánchez, E., Donat, E., Ribes-Koninckx, C., Fernández-Murga, M.L., Sanz, Y., 2013. Duodenal-Mucosal Bacteria Associated With Celiac Disease In Children. Appl. Environ. Microbiol. 79, 5472–5479. Https://Doi.Org/10.1128/AEM.00869-13

Scarpellini, E., Gabrielli, M., Za, T., Lauritano, E.C., Santoliquido, A., Rossi, E., Ojetti, V., Cammarota, G., Stefano, V.D., Gasbarrini, A., 2009. The Interaction Between Small Intestinal Bacterial Overgrowth And Warfarin Treatment. Am. J. Gastroenterol. 104, 2364–2365. Https://Doi.Org/10.1038/Ajg.2009.288

Scepanovic, P., Hodel, F., Mondot, S., Partula, V., Byrd, A., Hammer, C., Alanio, C., Bergstedt, J., Patin, E., Touvier, M., Lantz, O., Albert, M.L., Duffy, D., Quintana-Murci, L., Fellay, J., Abel, L., Alcover, A., Aschard, H., Astrom, K., Bousso, P., Bruhns, P., Cumano, A., Demangel, C., Deriano, L., Di Santo, J., Dromer, F., Duffy, D., Eberl, G., Enninga, J., Fellay, J., Gelpi, O., Gomperts-Boneca, I., Hasan, M., Hercberg, S., Lantz, O., Leclerc, C., Mouquet, H., Pellegrini, S., Pol, S., Rausell, A., Rogge, L., Sakuntabhai, A., Schwartz, O., Schwikowski, B., Shorte, S., Soumelis, V., Tangy, F., Tartour, E., Toubert, A., Touvier, M., Ungeheuer, M.-N., Albert, M.L., Quintana-Murci, L., The Milieu Intérieur Consortium, 2019. A Comprehensive Assessment Of Demographic, Environmental, And Host Genetic Associations With Gut Microbiome Diversity In Healthy Individuals. Microbiome 7, 130. Https://Doi.Org/10.1186/S40168-019-0747-X

Schapkaitz, E., Sithole, J., 2017. Predictors Of Warfarin Dose Requirements In South African Patients Attending An Anticoagulation Clinic. J. Vasc. Nurs. Off. Publ. Soc. Peripher. Vasc. Nurs. 35, 27–30. Https://Doi.Org/10.1016/J.Jvn.2016.05.002

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., Huttenhower, C., 2011. Metagenomic Biomarker Discovery And Explanation. Genome Biol. 12, R60. Https://Doi.Org/10.1186/Gb-2011-12-6-R60

Sekiguchi, Y., Hanada, S., 2020. Caldilineaceae. Pp. 1–3. Https://Doi.Org/10.1002/9781118960608.Fbm00320

Shi YY, He L, 2005. Shesis, A Powerful Software Platform For Analyses Of Linkage Disequilibrium, Haplotype Construction, And Genetic Association At Polymorphism Loci.

Shilaih, M., Angst, D.C., Marzel, A., Bonhoeffer, S., Günthard, H.F., Kouyos, R.D., 2018. Antibacterial Effects Of Antiretrovirals, Potential Implications For Microbiome Studies In HIV. Antivir. Ther. 23, 91–94. Https://Doi.Org/10.3851/IMP3173

Sortino, O., Phanuphak, N., Schuetz, A., Ortiz, A.M., Chomchey, N., Belkaid, Y., Davis, J., Mystakelis, H.A., Quiñones, M., Deleage, C., Ingram, B., Rerknimitr, R., Pinyakorn, S., Rupert, A., Robb, M.L., Ananworanich, J., Brenchley, J., Sereti, I., Teeratakulpisarn, N., Pattanachaiwit, S., De Souza, M., Fletcher, J., Kroon, E., Tantivitayakul, P., Suttichom, D., Rattanamanee, S., Poltavee, K., Intasan, J., Luekasemsuk, T., Savadsuk, H., Tipsuk, S., Puttamsawin, S., Benjapornpong, K., Ratnaratorn, N., Eamyoung, P., Ubolyam, S., O'Connell, R., Akapirat, S.,

Nuntapinit, B., Tantibul, N., Churikanont, N., Getchalarat, S., Vasan, S., Trichavaroj, R., Sajiaweerawan, C., Phuang-Ngern, Y., Jongrakthaitae, S., Sukhumvittaya, S., Saetun, P., Chuenarom, W., Michael, N., Turk, E., Mccullough, C., Butterworth, O., Milazzo, M., 2019. Impact Of Acute HIV Infection And Early Antiretroviral Therapy On The Human Gut Microbiome. Open Forum Infect. Dis. Https://Doi.Org/10.1093/Ofid/Ofz367

Stack, G., Maurice, C.B., 2016. Warfarin Pharmacogenetics Reevaluated Subgroup Analysis Reveals A Likely Underestimation Of The Maximum Pharmacogenetic Benefit By Clinical Trials. Am. J. Clin. Pathol. 145, 671–686. Https://Doi.Org/10.1093/Ajcp/Aqw049

Stafford, D.W., 2005. The Vitamin K Cycle. J. Thromb. Haemost. JTH 3, 1873–1878. Https://Doi.Org/10.1111/J.1538-7836.2005.01419.X

Sugiyama, Y., Gotoh, A., Katoh, T., Honda, Y., Yoshida, E., Kurihara, S., Ashida, H., Kumagai, H., Yamamoto, K., Kitaoka, M., Katayama, T., 2016. Introduction Of H-Antigens Into Oligosaccharides And Sugar Chains Of Glycoproteins Using Highly Efficient 1,2-A-L-Fucosynthase. Glycobiology 26, 1235–1247. Https://Doi.Org/10.1093/Glycob/Cww085

Swart, M., Evans, J., Skelton, M., Castel, S., Wiesner, L., Smith, P.J., Dandara, C., 2016. An Expanded Analysis Of Pharmacogenetics Determinants Of Efavirenz Response That Includes 3′-UTR Single Nucleotide Polymorphisms Among Black South African HIV/AIDS Patients. Front. Genet. 6. Https://Doi.Org/10.3389/Fgene.2015.00356

Swart, M., Ren, Y., Smith, P., Dandara, C., 2012a. ABCB1 4036A>G And 1236C>T Polymorphisms Affect Plasma Efavirenz Levels In South African HIV/AIDS Patients. Front. Genet. 3, Undefined-Undefined. Https://Doi.Org/10.3389/Fgene.2012.00236

Swart, M., Whitehorn, H., Ren, Y., Smith, P., Ramesar, R.S., Dandara, C., 2012b. PXR And CAR Single Nucleotide Polymorphisms Influence Plasma Efavirenz Levels In South African HIV/AIDS Patients. BMC Med. Genet. 13, 112. Https://Doi.Org/10.1186/1471-2350-13-112

Takiishi, T., Fenero, C.I.M., Câmara, N.O.S., 2017. Intestinal Barrier And Gut Microbiota: Shaping Our Immune Responses Throughout Life. Tissue Barriers 5. Https://Doi.Org/10.1080/21688370.2017.1373208

The 1000 Genomes Project Consortium, 2015. A Global Reference For Human Genetic Variation. Nature 526, 68–74. Https://Doi.Org/10.1038/Nature15393

The Integrative Human Microbiome Project: Dynamic Analysis Of Microbiome-Host Omics Profiles During Periods Of Human Health And Disease, 2014. . Cell Host Microbe 16, 276–289. Https://Doi.Org/10.1016/J.Chom.2014.08.014

Torniainen, S., Parker, M.I., Holmberg, V., Lahtela, E., Dandara, C., Jarvela, I., 2009. Screening Of Variants For Lactase Persistence/Non-Persistence In Populations From South Africa And Ghana. BMC Genet. 10, 31. Https://Doi.Org/10.1186/1471-2156-10-31

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M.S., Kevans, D., Smith, M.I., Guttman, D.S., Griffiths, A., Panaccione, R., Otley, A., Xu, L., Shestopaloff, K., Moreno-Hagelsieb, G., Consortium, G.P.R., Abreu, M.,

Beck, P., Bernstein, C., Dieleman, L., Feagan, B., Jacobson, K., Kaplan, G., Krause, D.O., Madsen, K., Marshall, J., Moayyedi, P., Ropeleski, M., Seidman, E., Snapper, S., Stadnyk, A., Steinhart, H., Surette, M., Turner, D., Walters, T., Vallance, B., Aumais, G., Bitton, A., Cino, M., Critch, J., Denson, L., Deslandres, C., El-Matary, W., Herfarth, H., Higgins, P., Huynh, H., Hyams, J., Mack, D., Mcgrath, J., Paterson, A.D., Croitoru, K., 2016. Association Of Host Genome With Intestinal Microbial Composition In A Large Healthy Cohort. Nat. Genet. 48, 1413–1417. Https://Doi.Org/10.1038/Ng.3693

UCT Eresearch HPC, N.D. UCT HPC University Of Cape Town High Performance Computing. UCT Eresearch HPC. URL Http://Hpc.Uct.Ac.Za/ (Accessed 12.2.19).

Uitterlinden, A.G., Pols, H.A.P., Burger, H., Huang, Q., Daele, P.L.A. Van, Duijn, C.M. Van, Hofman, A., Birkenhäger, J.C., Leeuwen, J.P.T.M. Van, 1996. A Large-Scale Population-Based Study Of The Association Of Vitamin D Receptor Gene Polymorphisms With Bone Mineral Density. J. Bone Miner. Res. 11, 1241–1248. Https://Doi.Org/10.1002/Jbmr.5650110908

UNAIDS Data, 2019. URL Https://Www.Unaids.Org/En/Resources/Documents/2019/2019-UNAIDS-Data (Accessed 11.13.19).

Ursell, L.K., Metcalf, J.L., Parfrey, L.W., Knight, R., 2012. Defining The Human Microbiome. Nutr. Rev. 70, S38–S44. Https://Doi.Org/10.1111/J.1753-4887.2012.00493.X

Vasilovici, A.F., Grigore, L.E., Ungureanu, L., Fechete, O., Candrea, E., Trifa, A.P., Vișan, S., Șenilă, S., Cosgarea, R., 2019. Vitamin D Receptor Polymorphisms And Melanoma. Oncol. Lett. 17, 4162–4169. Https://Doi.Org/10.3892/Ol.2018.9733

Velkova, A., Diaz, J.E.L., Pangilinan, F., Molloy, A.M., Mills, J.L., Shane, B., Sanchez, E., Cunningham, C., Mcnulty, H., Cropp, C.D., Bailey-Wilson, J.E., Wilson, A.F., Brody, L.C., 2017. The FUT2 Secretor Variant P.Trp154Ter Influences Serum Vitamin B12 Concentration Via Holo-Haptocorrin, But Not Holo-Transcobalamin, And Is Associated With Haptocorrin Glycosylation. Hum. Mol. Genet. 26, 4975–4988. Https://Doi.Org/10.1093/Hmg/Ddx369

Wacklin, P., Mäkivuokko, H., Alakulppi, N., Nikkilä, J., Tenkanen, H., Räbinä, J., Partanen, J., Aranko, K., Mättö, J., 2011. Secretor Genotype (FUT2 Gene) Is Strongly Associated With The Composition Of Bifidobacteria In The Human Intestine. Plos ONE 6. Https://Doi.Org/10.1371/Journal.Pone.0020113

Wacklin, P., Tuimala, J., Nikkilä, J., Tims, S., Mäkivuokko, H., Alakulppi, N., Laine, P., Rajilic-Stojanovic, M., Paulin, L., Vos, W.M. De, Mättö, J., 2014. Faecal Microbiota Composition In Adults Is Associated With The FUT2 Gene Determining The Secretor Status. PLOS ONE 9, E94863. Https://Doi.Org/10.1371/Journal.Pone.0094863

Wampach, L., Heintz-Buschart, A., Fritz, J.V., Ramiro-Garcia, J., Habier, J., Herold, M., Narayanasamy, S., Kaysen, A., Hogan, A.H., Bindl, L., Bottu, J., Halder, R., Sjöqvist, C., May, P., Andersson, A.F., Beaufort, C. De, Wilmes, P., 2018. Birth Mode Is Associated With Earliest Strain-

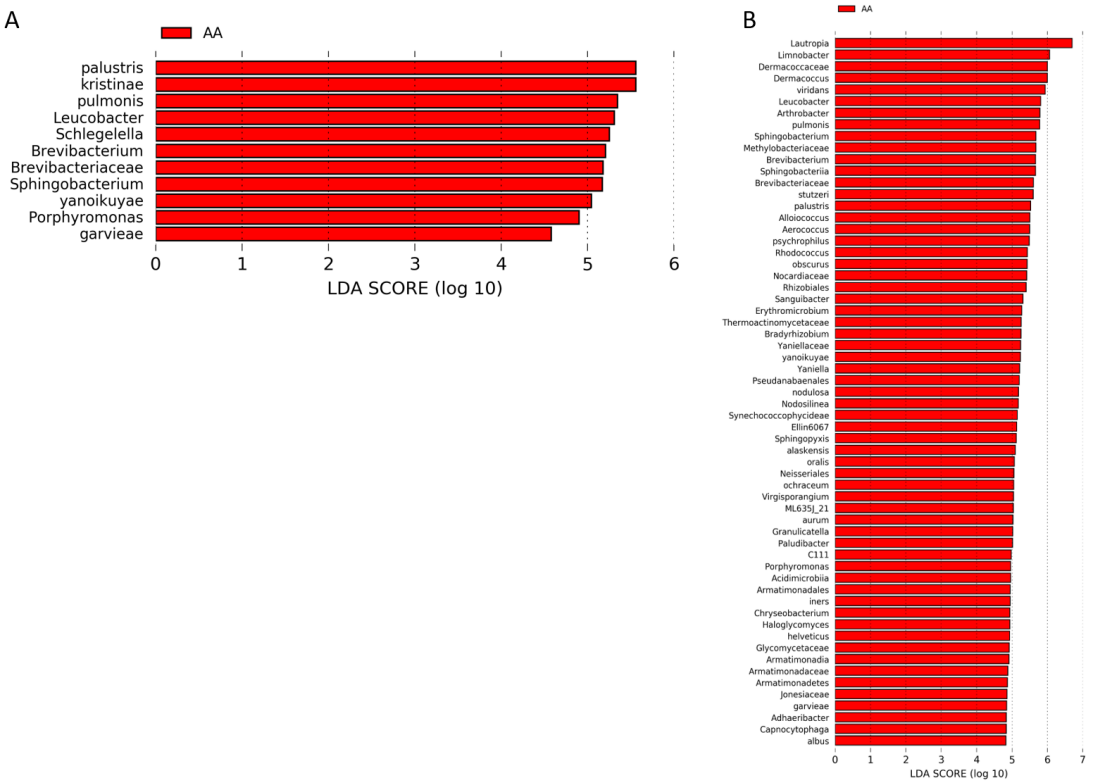
Conferred Gut Microbiome Functions And Immunostimulatory Potential. Nat. Commun. 9, 1–14. Https://Doi.Org/10.1038/S41467-018-07631-X
Wan, Y., Wang, F., Yuan, J., Li, J., Jiang, D., Zhang, J., Li, H., Wang, R., Tang, J., Huang, T., Zheng, J., Sinclair, A.J., Mann, J., Li, D., 2019. Effects Of Dietary Fat On Gut Microbiota And Faecal Metabolites, And Their Relationship With Cardiometabolic Risk Factors: A 6-Month Randomised Controlled-Feeding Trial. Gut 68, 1417–1429. Https://Doi.Org/10.1136/Gutjnl-2018-317609
Wang, J., Thingholm, L.B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J.R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M.C., Szymczak, S., Holm, K., Esko, T., Sun, J., Pricop-Jeckstadt, M., Al-Dury, S., Bohov, P., Bethune, J., Sommer, F., Ellinghaus, D., Berge, R.K., Hübenthal, M., Koch, M., Schwarz, K., Rimbach, G., Hübbe, P., Pan, W.-H., Sheibani-Tezerji, R., Häsler, R., Rosenstiel, P., D'Amato, M., Cloppenborg-Schmidt, K., Künzel, S., Laudes, M., Marschall, H.-U., Lieb, W., Nöthlings, U., Karlsen, T.H., Baines, J.F., Franke, A., 2016. Genome-Wide Association Analysis Identifies Variation In Vitamin D Receptor And Other Host Factors Influencing The Gut Microbiota. Nat. Genet. 48, 1396–1406. Https://Doi.Org/10.1038/Ng.3695
Wang, P.-F., Neiner, A., Kharasch, E.D., 2019. Efavirenz Metabolism: Influence Of Polymorphic CYP2B6 Variants And Stereochemistry. Drug Metab. Dispos. Https://Doi.Org/10.1124/Dmd.119.086348
Wendelboe, A.M., Raskob, G.E., 2016. Global Burden Of Thrombosis. Circ. Res. 118, 1340–1347. Https://Doi.Org/10.1161/CIRCRESAHA.115.306841
Williams, R.C., 1989. Restriction Fragment Length Polymorphism (RFLP). Am. J. Phys. Anthropol. 32, 159–184. Https://Doi.Org/10.1002/Ajpa.1330320508
Wilson, I.D., Nicholson, J.K., 2017. Gut Microbiome Interactions With Drug Metabolism, Efficacy, And Toxicity. Transl. Res., Microbiome And Human Disease Pathogenesis 179, 204–222. Https://Doi.Org/10.1016/J.Trsl.2016.08.002
Wittkowsky, A.K., 2003. Warfarin And Other Coumarin Derivatives: Pharmacokinetics, Pharmacodynamics, And Drug Interactions. Semin. Vasc. Med. 03, 221–230. Https://Doi.Org/10.1055/S-2003-44457
World Health Organization, 2020. The Top 10 Causes Of Death [WWW Document]. URL Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/The-Top-10-Causes-Of-Death (Accessed 6.10.23).
Yao, S., Zhao, Z., Wang, W., Liu, X., 2021. Bifidobacterium Longum: Protection Against Inflammatory Bowel Disease. J. Immunol. Res. 2021, 8030297. Https://Doi.Org/10.1155/2021/8030297
Zhang, T., Yang, Y., Liang, Y., Jiao, X., Zhao, C., 2018. Beneficial Effect Of Intestinal Fermentation Of Natural Polysaccharides. Nutrients 10. Https://Doi.Org/10.3390/Nu10081055
Zhu, B., Wang, X., Li, L., 2010. Human Gut Microbiome: The Second Genome Of Human Body. Protein Cell 1, 718–725. Https://Doi.Org/10.1007/S13238-010-0093-Z

Ziegler, S., Weiss, E., Schmitt, A.-L., Schlegel, J., Burgert, A., Terpitz, U., Sauer, M., Moretta, L., Sivori, S., Leonhardt, I., Kurzai, O., Einsele, H., Loeffler, J., 2017. CD56 Is A Pathogen Recognition Receptor On Human Natural Killer Cells. Sci. Rep. 7. Https://Doi.Org/10.1038/S41598-017-06238-4

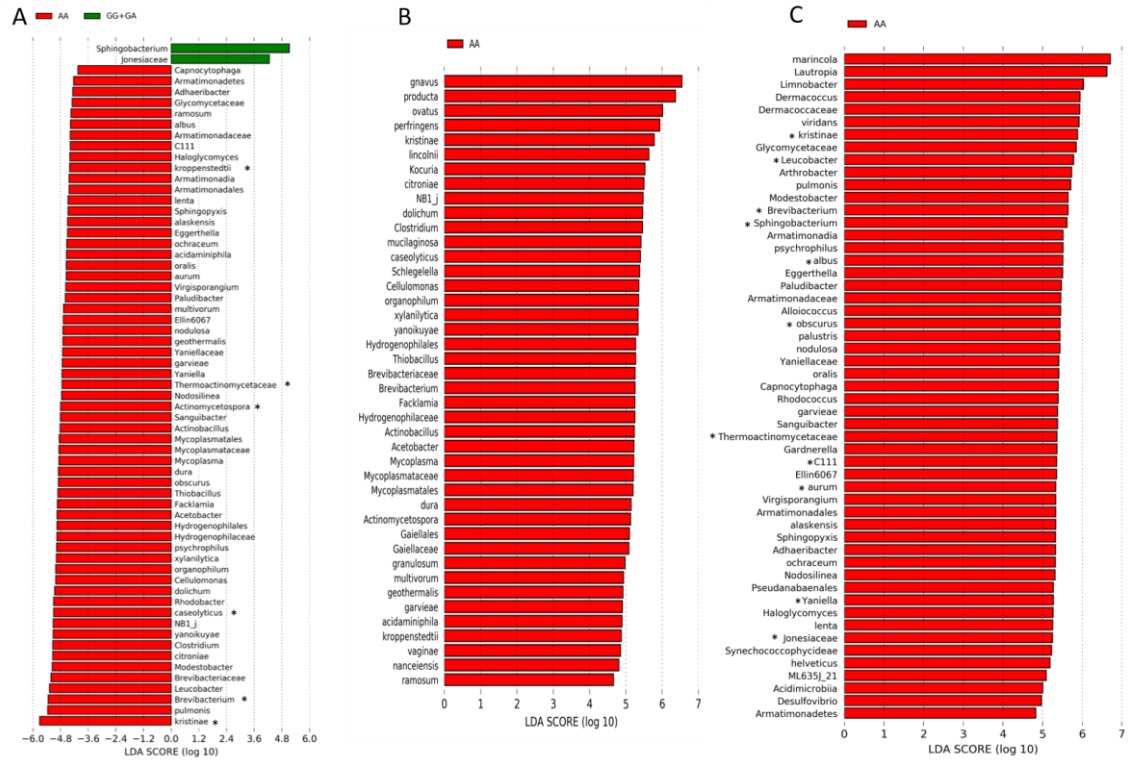
Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., Goodman, A.L., 2019. Mapping Human Microbiome Drug Metabolism By Gut Bacteria And Their Genes. Nature 570, 462–467. Https://Doi.Org/10.1038/S41586-019-1291-3

Zimmermann-Kogadeeva, M., Zimmermann, M., Goodman, A.L., 2019. Insights From Pharmacokinetic Models Of Host-Microbiome Drug Metabolism. Gut Microbes 1–10. Https://Doi.Org/10.1080/19490976.2019.1667724

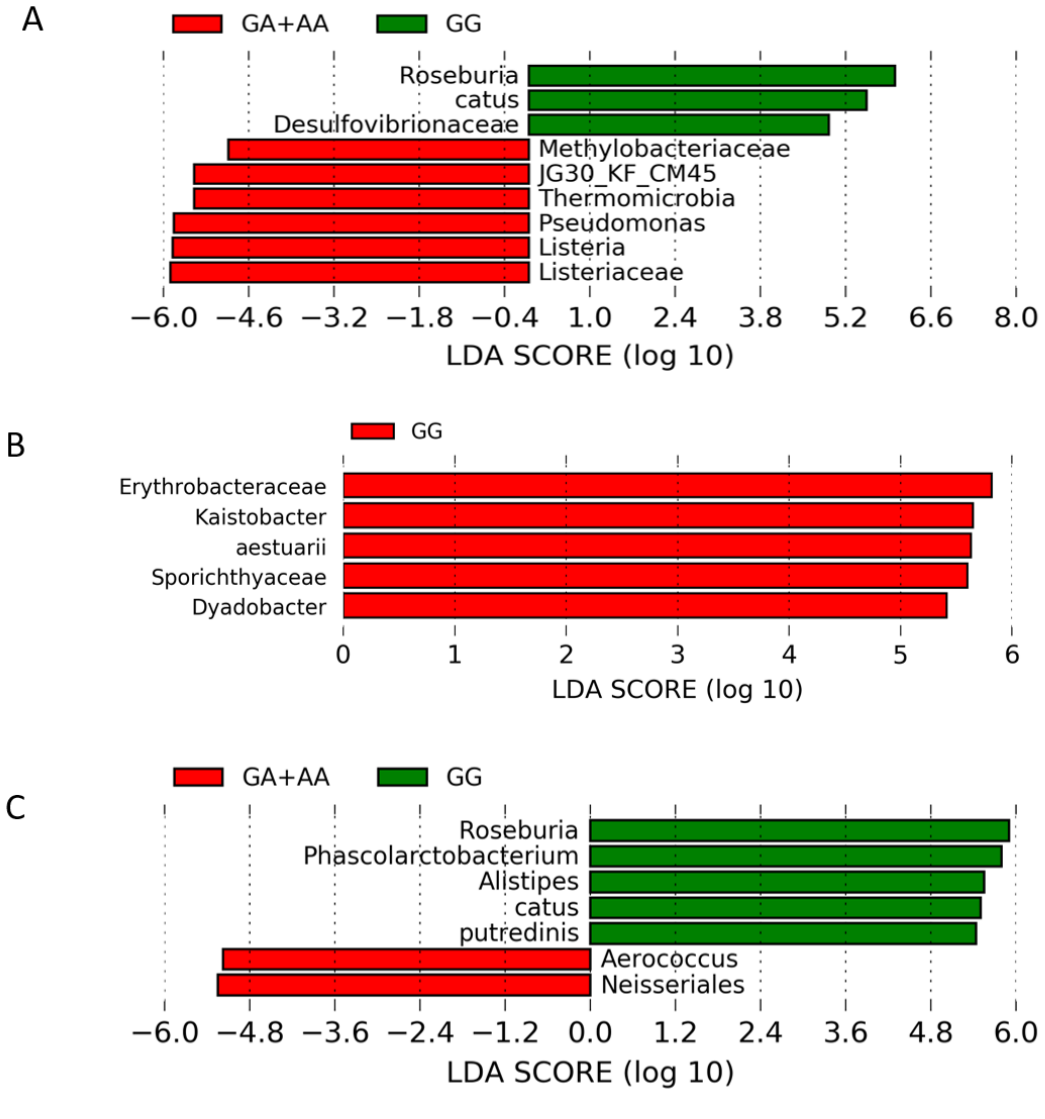
Appendices



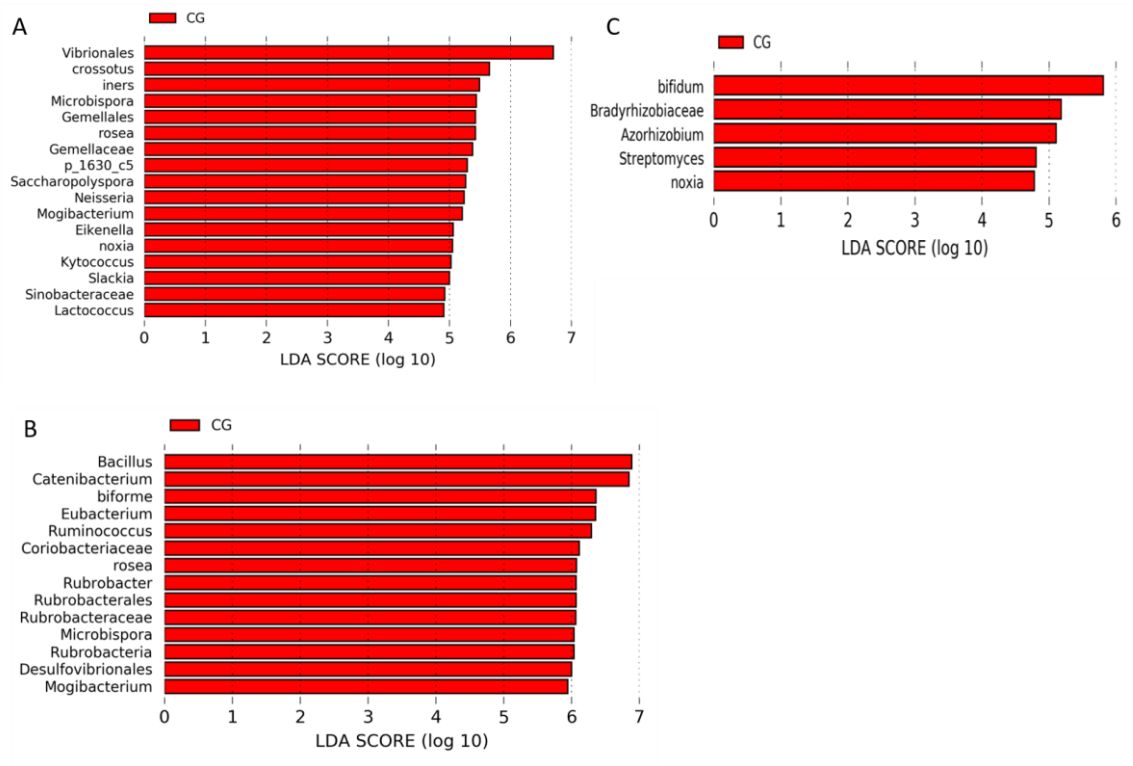
Supplementary Figure 1: FUT2 GG vs AA



Supplementary Figure 2: FUT2 GG+GA vs AA



Supplementary Figure 3: FUT2 GG vs GA+AA



Supplementary Figure 4: NCAM1 CC vs CG