PROFILES OF PHARMACOGENES VARIANTS IN POPULATIONS IN MALARIA ENDEMIC AND NON-ENDEMIC GEOGRAPHICAL AREAS: A BIOINFORMATICS APPROACH

by

HUNDAOL GIRMA HORDOFA

(HRDHUN001)

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree of MSc (Med) in Human Genetics by Dissertation

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

Date of Submission: 05/05/2022
Main Supervisor: Professor Collet Dandara
Co-Supervisor: Professor Emile Chimusa
Department of Pathology, Division of Human Genetics

Declaration

I, Hundaol Girma Hordofa, hereby declare that the work on which this dissertation 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: 05/05/2022

Abstract

Background: Inter-ethnic differences in response to medication can be explained through population structure as observed from genetic variants affecting drug metabolism. Antimalarial drugs are used in populations carrying different profiles of genetic variability. However, there is limited information on the distribution of genetic variants of pharmacogenomics importance in African populations.

Aims: We set out to determine whether there are differences in the pharmacogenetic profiles when comparing populations residing in malaria endemic and non-endemic areas.

Methodology: Genome-wide genotype datasets (n=13, 447) from four malaria endemic African populations, including Mali, Kenya, Gambia, and Malawi, which were part of the MalariaGEN consortium, were accessed. As comparator, reference datasets of global populations comprising of 20 ethnic groups (n~5,000) from African Genome Variation Project (AGVP) and 1000 Genome consortium project were accessed, considering their malaria epidemiology. Pharmacogenes that are verified from databases such as PharmVar and PharmGKB, and nomenclature sites for uridine 5'-diphospho-glucuronosyltransferases (UGT), solute carrier (SLC) transporters, ATP-binding cassette (ABC) transporters, thiopurine S-methyl transferase (TPMT), and N-acetyltransferases (NAT) were used. Pharmacogenes reported to affect antimalarial drugs were noted and their distribution analysed. Single nucleotide polymorphisms (SNPs) that map to pharmacogenes were retrieved from dbSNP databases and used in the comparison of the populations from malaria endemic and non-endemic regions, through population structure evaluation using pharmacogenes variants. As part of confirmation, genotyping for *CYP2C8* SNPs was carried out in native South African population groups, which we have access to our laboratory. *CYP2C8* is particularly interesting because of its central involvement in the metabolism of some of the major antimalarial drugs.

Results: There were quantitative and qualitative differences in the distribution of pharmacogenes variants when comparing populations from malaria endemic regions to those from non-endemic regions. For instance, the analysis of the minor allele frequency (MAF) proportion of all pharmacogenes variants obtained from the databases revealed that a high proportion of common variants (MAF > 0.05) of pharmacogenes were higher in the four African population from malaria endemic areas (MAF proportion > 50%) compared with world populations from malaria non-endemic areas (MAF proportion < 50%). There were a disproportionately higher number of variants in cytochrome P450 enzymes such as *CYP2A13* and *CYP2F1* in the malaria endemic populations (> 80%) compared to non-endemic populations (< 65%). Further analysis based on antimalarial specific pharmacogenes variants showed that *UGT1A9* had the highest proportion of variants (> 80%) in the malaria endemic populations. The ABC drug transporters such as *ABCC4*, *ABCC1* and *ABCC2* were found to have a higher proportion of pathogenic SNPs in populations of the four malaria endemic areas (> 50%) compared with those from non-endemic areas (< 30%). Principal component analysis (PCA) based on pharmacogenes variants also showed clear differentiation into different population. Moreover, four *CYP2C8* variants (*rs11572103T>A (*2), rs11572101A>G, rs11572100T>C* and *rs1926705C>T*), characterized in South Africans, were found to have

statistically significant difference (p-value ≤ 0.05) in allele frequency when compared with global populations.

Conclusions: The findings from this study revealed the profiles of pharmacogenes variants in populations from malaria endemic and non-endemic geographical regions. The variants in *CYP2A13*, *CYP2F1,* and *UGT1A9* were found in higher proportions in populations from malaria endemic areas compared to non-malaria endemic areas. This suggests that pharmacogenetic testing for patients based on antimalarial pharmacogenes having a higher proportion of variants in malaria endemic areas (e.g., *UGT1A9*) may be necessary. Moreover, a high frequency in the functional variant of *CYP2C8*, *rs11572103T>A (*2),* was found in the South African population. This may also have relevance for future consideration of pharmacogenetic testing for a person traveling to malaria endemic areas from non-endemic areas, such as South Africa. Future studies should be done using large scale sequencing datasets from malaria endemic areas and assess the effect of common variants of antimalarial pharmacogenes on the safety and efficacy of currently used antimalarial drugs.

Keywords: pharmacogenes, variants, profiles, malaria, endemic, non-endemic

Table of contents

Acknowledgment

First of all, I would like to thank the almighty God for fiving me the strength to accomplish this master's degree. Then, I would like to thank the Developing Excellence in Leadership and Genetics Training for Malaria Elimination in Sub-Saharan Africa (DELGEME) for funding this research at the University of Cape Town (UCT). Furthermore, I would like to extend my special thanks to UCT for providing me with such a golden opportunity to obtain a Master of Science Degree in Human Genetics.

My special thanks also go to my supervisors Professor Collet Dandara and Professor Emile Chimusa, for their continuous support, encouragement, and guidance throughout this project. Many thanks to my colleague Dr. Delesa Mulisa, who gave me fruitful advice and support during the data analysis for this project.

My thanks are also extended to all the members of the Pharmacogenomics and Drug Metabolism (PharmGX) Group for accepting me into the team and sharing with me the knowledge and experiences they have gained throughout my stay at the University of Cape Town.

Last but not least, I would like to thank my wife, Mrs. Mihiret Yilma, and all other family members for their continuous support and encouragement while doing this MSc.

Abbreviations and acronyms

°C	degrees Celsius
3′	three prime end
5′	five prime end
A	Adenine
ABC	ATP-binding cassette
ABCB1	ATP-binding cassette subfamily B1
ACT	artemisinin-based combination therapy
ADME	Absorption Distribution Metabolism and Elimination
ADQ	Amodiaquine
AGVP	African Genome Variation Project
ARDS	acute respiratory distress syndrome
BLAST	Basic Local Alignment Search Tool
bp	base pair
C	cytosine
CDC	Centers for Disease Control and Prevention
cDNA	complementary DNA
chr	chromosome
COMT	catechol O-methyl transferase
CQ	Chloroquine
CYP	Cytochrome P450
DEAM	Desethylamodiaquine
DHA	Dihydroartemisinin
DHFR	Dihydrofolate reductase

DHPS	Dihydropteroate Synthase
dl	decilitre
DNA	Deoxyribonucleic Acid
dNTP	deoxyribonucleotide triphosphate
ELISA	Enzyme-linked Immunosorbent Assay
EM	extensive metabolizer
ETC	Electron Transport Chain
FCM	flow cytometry
G	guanine
gDNA	genomic deoxyribonucleic acid
GMS	Greater Mekong Subregion
GPI	glycophosphatidylinositol
GRCh37	Genome Reference Consortium Human Build 37
GSTs	glutathione S-transferases
HRP	Histidine-rich Protein
HWE	Hardy-Weinberg equilibrium
IDT	Integrated DNA Technologies
IFA	Immunofluorescence Assay
iHS	Integrated Haplotype score
IM	Intermediate metabolizer
Kb	Kilobase
Kbps	Kilobase pairs
LAMP	Loop-mediated Nucleic Acid Amplification
LDH	Lactate Dehydrogenase

MAF	Minor Allele Frequency
MDR1	multidrug resistance gene 1
MgCl2	magnesium chloride
Min	minute
mM	millimolar
MQ	Mefloquine
MS	Mass spectrometry
MSP1	Merozoite surface protein 1
MWM	Molecular weight marker
NASBA	Nucleic Acid Sequence-based Amplification
NATs	N-acetyltransferases
NCBI	National Center for Biotechnology Information
ng	nanogram
P.f	*Plasmodium falciparum*
P.k	*Plasmodium knowlesi*
P.m	*Plasmodium malariae*
P.o	*Plasmodium ovale*
P.v	*Plasmodium vivax*
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
Pfcrt	*Plasmodium falciparum chloroquine resistant transporter*
PfCytB	*Plasmodum falciparum cytochrome b*
Pfdhfr	*Plasmodium falciparum dihydrofolate reductase*
Pfdhps	*Plasmodum falciparum dihydropteroate synthase*

PfKelch13 *Plasmodium falciparum kelch13*

PfLDH *Plasmodium falciparum* Lactate Dehydrogenase

Pfmdr1 *Plasmodium falciparum multidrug resistance protein*

pfpm2 *Plasmodium falciparum Plasmepsin 2*

PM poor metabolizer

PPQ Piperaquine

PQ Primaquine

*Pv*LDH *Plasmodium vivax* Lactate Dehydrogenase

QC quality control

QN Quinine

RBC Red Blood Cell

RDTs Rapid Diagnostic Tests

RNA Ribonucleic Acid

rRNA Ribosomal RNA

sec second

SLC solute carrier

SNP Single Nucleotide Polymorphism

SP sulfadoxine-pyrimethamine

SULTs sulfotransferases

T thymine

TPMT thiopurine S-methyl transferase

UDP Uridine 5'-diphospho-glucuronosyltransferases

ug microgram

UGTs UDP-glucuronosyltransferases

μL Microliter

μM	micromolar
UM	ultrarapid metabolizer
VCF	Variant Calling Format
WHO	World Health Organization

List of figures

List of tables

CHAPTER 1 : GENERAL INTRODUCTION

1.1 Epidemiology of malaria

Malaria remains a public health concern (1,2). It poses a threat to over half of the world's population, with Africans accounting for the majority (3). According to the 2021 World Health Organization's (WHO) report, there were an anticipated 241 million cases of malaria and 627 thousand deaths worldwide in 2020 (4). The WHO African Region has a disproportionately large percentage of the global malaria burden (**Figure 1.1**).

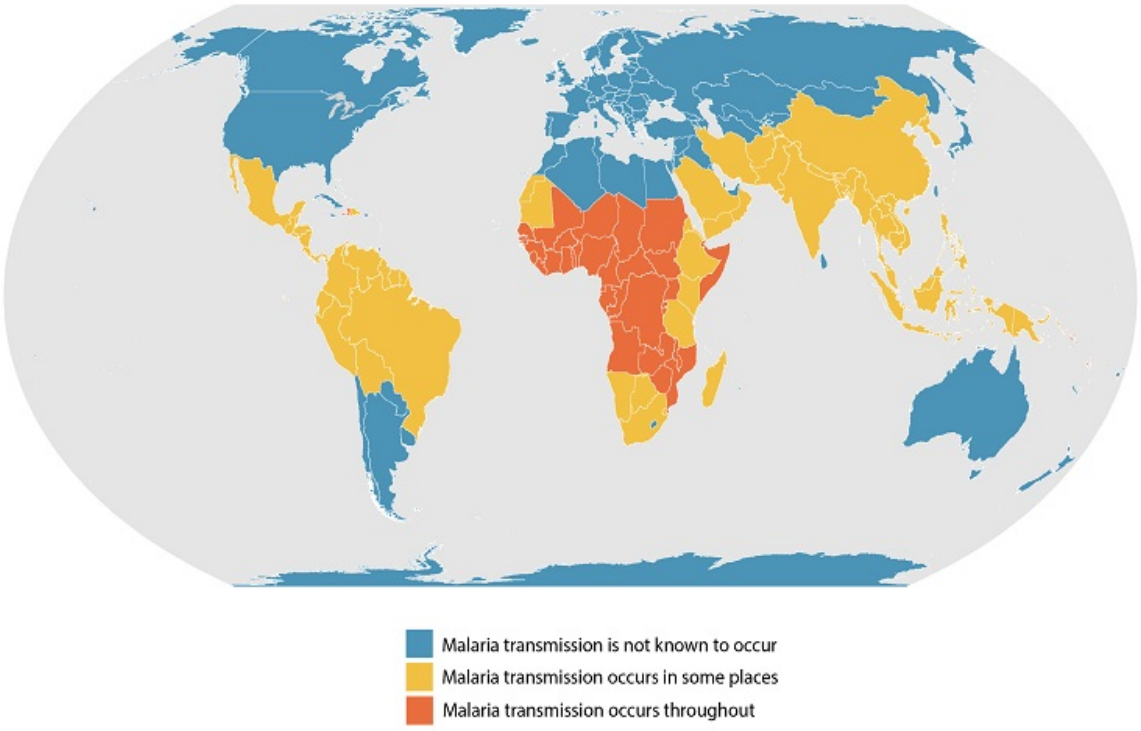

Figure 1. 1 | Map showing Malaria epidemiology. (Adapted from Centers for Disease Control and Prevention (CDC) (5))

Malaria is caused by the *Plasmodium* parasite. *Plasmodium falciparum (P.f), Plasmodium vivax (P.v), Plasmodium malariae (P.m), Plasmodium ovale (P.o)* (6) and *Plasmodium knowlesi (P.k),* are the five *Plasmodium* species known to infect humans (3,7). Two of the species, *P.v* and *P.f*, pose the greatest treat, with the latter being the most lethal (2,8). *P.f* is responsible for most malaria-related deaths in endemic countries (7,9,10). It is the most common malaria parasite in the WHO African region. *P.v* is the most predominant parasite in

the WHO South-East Asia Region, contributes to over half of the global burden, with India bearing the greatest share (47%). It is also common in the WHO Americas Region, accounting for 75% of malaria cases (11).

1.1.1 Transmission

The main transmission of malaria is through the bite of an infective female anopheles Mosquito during a blood meal (12). However, *P.k* is a primate parasite that can infect humans through zoonosis (13). *Anopheles gambiae, Anopheles arabiensis,* and *Anopheles funestus* are the three main Anopheles species that transmit malaria in countries south of the Sahara Desert (14). Transfusions of infected blood, from mother to child before or during delivery (congenital transmission), organ transplantation, and the use of blood-contaminated needles or syringes can all transmit malaria to healthy persons (12,15).

1.1.2 Life cycle

The malaria parasite has a complex life cycle that involves two hosts (human host and female *Anopheles* mosquito vector) (16), (**Figure 1.2**). It undergoes sexual replication in the female *Anopheles* mosquito (the definitive host) and asexual replication in humans (the intermediate hosts). During a blood meal, a malaria infected female *Anopheles* mosquito injects sporozoites along with saliva into the skin of a human host (15). Through gliding motility, the injected sporozoites move within the dermis into the blood stream and are transport to the liver (17). In a process called pre-erythrocytic (hepatic) schizogony, the sporozoites undergo asexual reproduction to form thousands of merozoites and mature into schizonts (15). The merozoites inside the schizonts burst hepatocyte cells and release them into the blood circulation. However, the duration of the hepatic stage differs based on the parasite species. For instance, *P.f* has a shorter duration of liver stage, which is around six days, while other species of the parasite (e.g., *P.v* and *P.o)* have a slightly longer duration, around eight and nine days, respectively. Unless treated, the two species of *Plasmodium* can form a long-lived dormant stage called a "hypnozoite", which can remain in the liver for several months up to years and cause latent infection (relapses) of malaria (15,16).

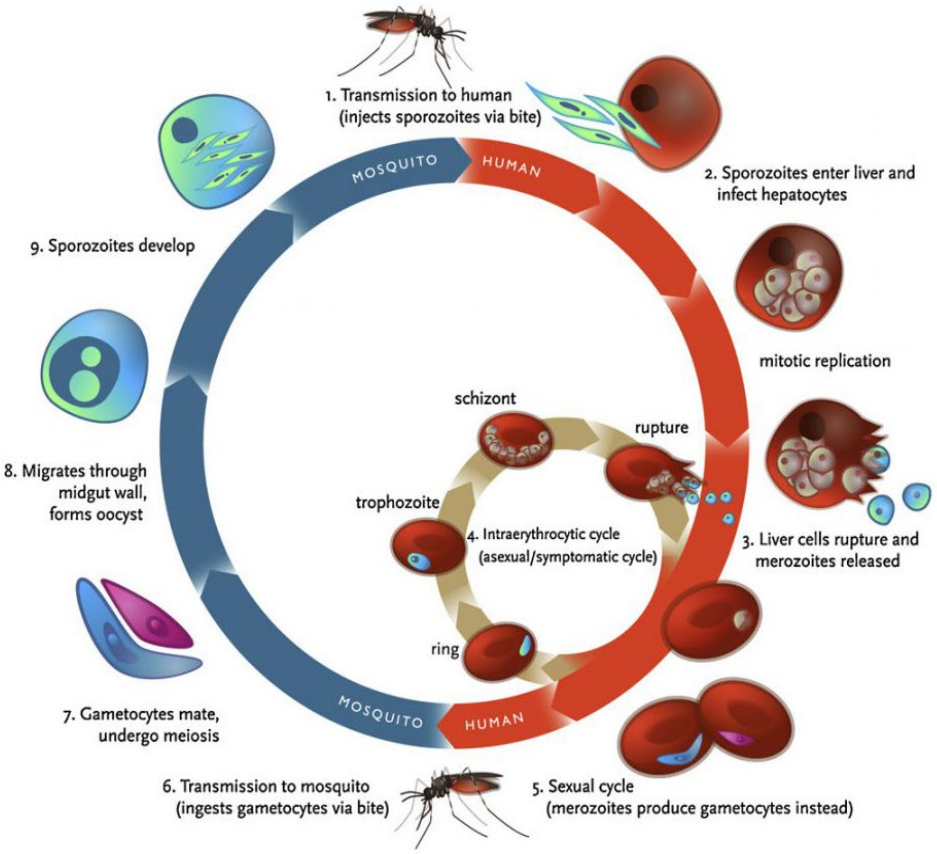

Figure 1. 2 | Life cycle of malaria parasite. (Adapted from (18))

In a process called erythrocytic schizogony, after the release of merozoites from the liver into the bloodstream, the parasites undergo asexual multiplication inside the red blood cells (RBCs). Once the merozoites are released into the blood circulation, they start to invade the erythrocytes in a fast, dynamic, and multi-step process including pre-invasion, active invasion, and echinocytosis that is completed within 2 minutes (19). Merozoite surface protein 1 (MSP1) is the main glycophosphatidylinositol (GPI) related protein on the merozoite surface (20), and it serves as a platform on the merozoite surface for a minimum of three large complexes with diverse extrinsic proteins that bind erythrocytes (21). Merozoites that infect red blood cells thrive and divide inside them for 24 to 72 hours, depending on the parasite species. The parasites transform from early trophozoite (ring form) to late trophozoite inside the RBCs, then go through numerous mitotic divisions to form a mature schizont (22), which is the real replication stage that goes through mitosis and produces a large number of merozoites. The merozoites in the schizont burst the RBC membrane, releasing 16-32 new merozoites that infect new uninfected RBCs in a continual cycle, increasing the parasite count in the affected person (15). It is this stage of the parasite life cycle that causes the clinical symptoms of the

disease, like chills, sweating and fever. A few merozoites mature into gametocytes (male and female) during this stage of merozoite multiplication, which stay on in the blood circulation for several weeks (23).

The gametocytes (male and female) are consumed by the mosquito vector during a blood meal and grow into mature sex cells (gamete) inside the vector (24). The gametes reproduce sexually and form a diploid zygote. The zygotes then mature into actively moving ookinetes, which burrow through the mosquito's midgut wall and create oocysts on the outside. Thousands of active sporozoites form inside the oocyst. The oocyst finally breaks, releasing sporozoites into the body cavity, which go to the salivary glands of the mosquito, and the life cycle continues when the sporozoites are inoculated into a new human host (25).

1.1.3 Pathogenesis and clinical presentations

A malaria parasite infection can cause a wide range of symptoms, ranging from absent or very mild symptoms to severe disease and even death (26). Symptoms of malaria can appear as early as 8 to 10 days or up to a year or more after a person has been bitten and infected with the *Plasmodium* parasite (12). Most *Plasmodium* species cause uncomplicated malaria, which is characterized by mild symptoms. Malaria is considered uncomplicated when symptoms are present but no clinical or laboratory signs of severity or vital organ dysfunction are present, and it is characterized by a flu-like illness with fever, chills, headache, myalgia, and, in some cases, abdominal pain and diarrhoea, which may initially suggest other diagnoses (5,12). Fever starts when *Plasmodium* consumes haemoglobin during its erythrocytic stage (27),and toxic heme is released from erythrocyte breakdown, triggering the cytokine cascade, which is responsible for many of the signs and symptoms of malaria (15).

Among all *Plasmodium* species, the most severe form of malaria, usually known as complex malaria, is caused by *P.f*. Severe malaria develops when infections are complicated by catastrophic organ failures or anomalies in the patient's blood or metabolism. Abnormal behaviour, impairment of consciousness, seizures, coma, or other neurologic abnormalities; severe anaemia caused by haemolysis (massive destruction of erythrocytes); haemoglobinuria (haemoglobin in the urine) caused by haemolysis; acute respiratory distress syndrome (ARDS); abnormalities in blood coagulation; low blood pressure caused by cardiovascular collapse; acute kidney injury; hyperparasitemia; Methemoglobinuria, are symptoms of severe malaria (5).

1.1.4 Diagnosis

Appropriate and accurate diagnosis of malaria is critical to the effective management of malaria (2,12). The diagnosis of malaria entails using one or more clinical, parasitological, or immunological methods to identify individuals with active malaria parasite infection or those whose medical condition is the result of recent malaria parasite exposure (2). The global impact of malaria has urged interest in developing effective diagnostic approaches not only for resource-constricted areas where malaria is a large burden on society but also in developed countries, where malaria diagnostic expertise is often lacking (28). Delays in diagnosis and treatment of malaria are leading causes of death in many countries (12).

Malaria can be diagnosed clinically by a skilled healthcare expert based on the patient's information, such as travel history, signs and symptoms, and physical examination findings (2). However, the overlapping of malaria symptoms with those of other diseases such as common viral infections, bacterial infections, and other febrile illnesses reduces diagnostic specificity, which can lead to inappropriate antimalarial use and compromise the quality of care for non-malarial fever patients in endemic areas (12,15,28,29). As a result, clinical diagnosis is only suggested in high-transmission areas in resource-constrained countries when laboratory testing may be unavailable. However, for making a conclusive diagnosis, verifying the presence of parasites or their components is critical.

Commonly used malaria diagnostic test methods include: Microscopy based diagnosis, Rapid Diagnostic Tests (RDTs), amplification of parasite-specific nucleic acids (e.g., Polymerase Chain Reaction (PCR) and loop-mediated nucleic acid amplification (LAMP)), and Immunological methods (e.g., Enzyme-linked Immunosorbent Assay (ELISA), Immunofluorescence assay (IFA), and flow cytometry) (2). Microscopy using Giemsa-stained blood films is by far the gold standard method for malaria diagnosis (30). In this method, thick and thin blood films on a microscopic slides are used for *Plasmodium* parasite detection and species identification, respectively (2,22). Microscopy's estimated sensitivity limit is 5-10 parasites/ul for expert microscopists, 50-00/µl parasites for average microscopists, and lower for the majority of microscopists in malaria endemic areas (2).

RDTs are immunochromatographic tests that use both soluble and immobilised antibody types on a chromatographic test strip to identify specific antigens (proteins) produced by malaria

parasites in the blood of infected individuals (2). These tests detect the malaria parasite within 2-15 minutes and are approved by the WHO as a substitute for microscopy in situations where microscopic diagnosis is not available (31). Malaria antigens that are currently targeted by RDTs include histidine-rich protein 2 (HRP-2), lactate dehydrogenase (LDH), and aldolase. HRP-2 has been shown to be a highly sensitive and specific marker for identifying *P.f* infections, making it the most targeted *P.f* antigen in current RDTs. Nevertheless, it has a long half-life, and can be detected in the peripheral blood for several weeks after the blood parasite is cleared by antimalarial drugs. This is especially concerning in places with moderate to high transmission, where false-positive testing might lead to antimalarial therapy being given to healthy people (32,33). LDH, on the other hand, has a short half-life and is frequently used in conjunction with HRP-2 to confirm the presence of a current infection. Existing LDH-based RDTs are either pan-*Plasmodium* (detect all five human malaria species) or species-specific (e.g., PfLDH or PvLDH). However, one major limitation of the current RDTs is their lack of sensitivity (detection limits >100 parasites/µl whole blood), which necessitating the use of more sensitive parasitological or molecular approaches, especially in patients with low-density parasitemia (2).

Nucleic acid amplification techniques represent the most accurate and sensitive methods used for detection and confirmatory identification of *Plasmodium* species. These molecular techniques include PCR based methods such as nested, real-time, multiplex, and reverse transcription-based PCR, and isothermal approaches such as nucleic acid sequence-based amplification (NASBA) and LAMP-based methods (34). PCR is based on the use of a Deoxyribonucleic Acid (DNA) polymerase such as Taq polymerase, as well as a pair of sequence-specific primers to amplify target DNA sequence from extracted parasite DNA. This approach involves temperature cycling between template denaturation, primer annealing, and strand synthesis steps. Detection of the resulting amplicon is achieved either by agarose gel electrophoresis or by using a real-time format in which a DNA-binding fluorescent dye or target-specific probe is added to the reaction mix and its fluorescence emission monitored over time (2). PCR based test has high sensitivity and specificity (35), which can detect very few parasites (around 1-5 parasites per µl of blood) as compared with microscopy and RDTs, which detects around 50-100 parasite per µl of blood) (28). Although PCR-based techniques are extremely sensitive and specific when compared with microscopy, they are limited by their high costs, absolute need for electricity and expensive equipment, labour-intensiveness, long turnaround times (2–4 h), as well as susceptibility to common laboratory contamination (36).

As such, the use of PCR-based techniques is limited to reference laboratory settings and for research purposes.

Isothermal nucleic acid amplification techniques such as NASBA and LAMP are quite promising as point-of-care malaria diagnostic approaches, are rapid <30 min for LAMP, and can detect infections of less than 1 parasite/µl of whole blood. NASBA is a nucleic acid amplification method that amplifies ribonucleic acid (RNA) like the replication of retroviruses. This approach generates high numbers of RNA copies per cycle, enabling it to generate detectable amounts of RNA products in a shorter time frame than PCR-based methods. The commonly used diagnostic target for malaria NASBA techniques is the 18S ribosomal RNA (rRNA), with a detection limit of 0.1–0.01 parasites/µl of whole blood. In general, NASBA is a quantitative technique that can produce results in just one hour, and it is highly sensitive and specific. However, it is expensive, labour-intensive, and prone to frequent laboratory contamination and false-positive amplifications. Of the existing molecular tests for malaria diagnosis, LAMP appears to be the most useful in terms of its field applicability, cost-effectiveness, diagnostic performance, and rapidity (2). LAMP is a one-step amplification technique based on the use of strand displacement polymerases, notably best DNA polymerase to amplify target DNA with high sensitivity and specificity under isothermal conditions (65 ^{o}C) (37).

The diagnosis of malaria using immunological methods is based on the detection of antigens or antibodies against malaria parasites in human body fluids. Depending on test formats, immunological diagnostic assays can provide information on past exposure to malaria parasites (antibody tests) or reveal ongoing infections (antibody and antigen detection assays). Among the conventional immunological methods that have been used for malaria diagnosis, those based on indirect IFA and ELISA have received the greatest attention. For IFA-based methods, parasitized erythrocytes are used as capture antigens, permitting the detection of specific antibody antigen reactions by flow cytometry or fluorescence microscopy techniques. However, the use of such techniques is limited to reference laboratory settings, given the need for high technical expertise, their labour-intensiveness, and operational costs. ELISA-based methods rely on the use of crude parasite extracts, heterologously expressed single antigens or peptide antigens for antibody detection, or the use of target-specific antibodies (monoclonal or polyclonal antibodies) to detect circulating parasite antigens in the patient's blood. However, the main limitations of the ELISA-based methods are their labour-intensiveness, long

turnaround times, and lack of standardized reagents. These limit the use of ELISA-based methods mostly to specialized diagnostic laboratories or for research purposes. However, ELISA methods are very useful as primary screening assays in population-based studies, given their medium-throughput potential, and often exhibit high sensitivities, detecting as little as 1–10 parasites per 1000,000 red blood cells (2).

Other methods for malaria diagnosis include mass spectrometry (MS) and flow cytometric (FCM) assay, which have permitted extensive characterization of the malaria parasite and are generating new strategies for malaria diagnosis (28).

1.2 Malaria treatment

Because of the absence of an approved vaccine strategy, antimalarial intervention comprises drug treatment with the primary goal of eradicating malaria parasites totally from the body. However, which antimalarial regimen should be used depends on the clinical status of the patient, the type of *Plasmodium* species that infects the patient, the place where the infection was acquired and its drug-resistance status, pregnancy status, and finally, the history of drug allergies, or other medications taken by the patient (10). **Table 1.1** shows the commonly used antimalarial medications. The erythrocytic stage of malaria infection, which causes symptoms of malaria, is the drug target for most antimalarial drugs. There are different classes of drugs used to treat malaria, and these include quinolines, antifolates, Artemisinin derivatives, etc.

Table 1. 1 | Summary of common antimalarial drugs and their mode of action (compiled from (38–40)**)**

Pharmacological class	Mode of action
Quinolines and relative compounds Chloroquine, Amodiaquine, Primaquine, piperaquine, Mefloquine, Quinine, Quinidine, Tafenoquine, Lumefantrine and Halofantrine	Inhibition of haem polymerase activity of the parasite, yet it still counterseal.
Antifolates Pyrimethamine, Proguanil **Sulpha drugs:** Dapsone, Sulfadoxine, Sulfonamides and Sulfone	Inhibits DHPS and Inhibits DHFR
Naphthoquinones Atovaquone	Inhibition of the mitochondrial electron transport chain (ETC) through inhibition of the malarial cytochrome bc1 complex
Artemisinin derivatives Artemether, Arteether Artesunate and dihydroartemisinin	Heme-mediated decomposition of the endoperoxide bridge to produce carbon-centred free radicals
Antibiotics Doxycycline Clindamycin Tetracycline	Blood schizonticide agent, but the exact mechanism of action is not well defined

In addition to conventional medications, there are also traditional herbal medicines that have been used for the treatment of malaria. Examples of such medicines include *Phyllanthus amarus* and *Momordica balsmamina* (41,42). However, there is a risk of possible drug interactions when herbal and conventional medicines are taken together (43).

1.2.1 Quinoline and related drugs

Quinine (QN), a 4-methanoquinoline, was discovered in the 17th century, and malaria treatment with QN was the first successful use of a chemical substance to treat an infectious disease. Since the 1600s, QN, a component of the bark of the cinchona (quina-quina) tree, has

been used to cure malaria. Quinine is currently on the WHO's essential medical list for the treatment of malaria (44). Quinine's structure, as well as that of chloroquine (CQ), amodiaquine (ADQ), primaquine (PQ), piperaquine (PPQ), and mefloquine (MQ), is based on the quinoline ring system (**Figure 1.3**) (7,45).

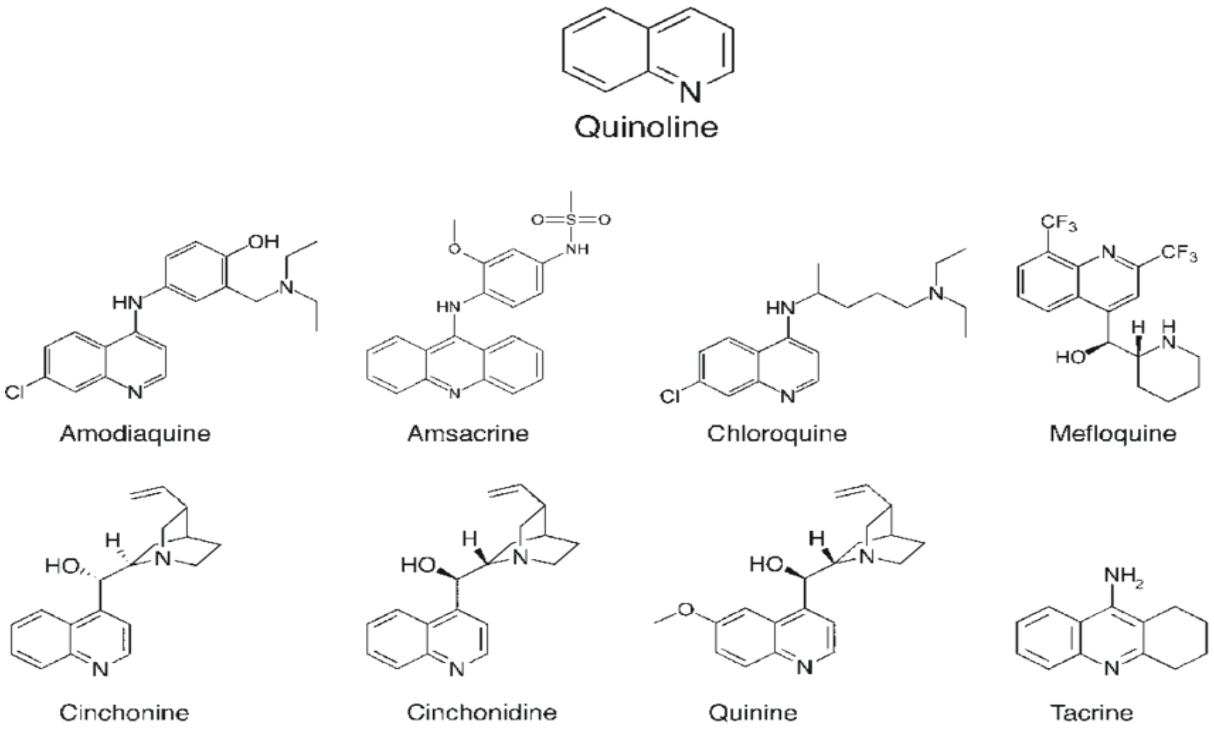

Figure 1. 3 | Chemical structures of eight quinoline derivatives. Adapted from (46)

Except for PQ, which is predominantly used to treat *P.v* infections, all the quinoline derivative drugs are assumed to share a similar target, which is the synthesis of the parasite-specific substance, hemozoin, toxic to the human host (47–49). To minimize the risk of developing resistance to the quinoline family of chemicals, these medications are frequently taken in combination with a complementary drug (e.g., MQ and artesunate). CQ, a 4-aminoquinoline, became clinically available in 1947 as a less toxic and more easily made alternative for QN, and is one of the most long-lasting of the synthetic antimalarials that appeared throughout the last century (7,50).

1.2.2 Antifolates

Malaria prophylaxis and treatment have relied mainly on antifolates. This medication family was found in the 1940s, during the Second World War, and drugs that are currently in therapeutic usage were discovered at that time (51). Antifolates work by inhibiting parasite enzymes involved in folate production as well as nucleic acid synthesis (7). Inhibitors of

dihydropteroate synthase (DHPS), also known as class I antifolates, and inhibitors of dihydrofolate reductase (DHFR), sometimes known as class II antifolates, are two types of antifolates used to treat malaria (51). Proguanil and pyrimethamine are the main inhibitors of DHFR, while sulpha-drugs, the most important of which are dapsone, sulfadoxine, sulphonamides, and sulphone, are the main inhibitors of DHPS (7,51). Since these two kinds of antifolates have a synergistic effect, they are frequently used in combination (51). According to the WHO, Sulfadoxine/pyrimethamine (SP) is one of the 14 essential medications currently used to treat malaria (52).

1.2.3. Artemisinin and its derivatives

For the first time, in 1971, Tu Youyou isolated artemisinin (Qinghaosu) from the herb Artemisia annua (sweet wormwood) (53). Youyou Tu (China) received a portion of the Nobel Prize in Medicine in 2015 for her work in extracting the active ingredient from Artemisia annua. Water-soluble artesunate, oil-soluble artemether, and arteether are three artemisinin derivatives often used in combination therapy against malaria (54). These semi-synthetic compounds are prodrugs that are converted to the active metabolite dihydroartemisinin (DHA) after metabolism. Artemisinin is active against all *Plasmodium* species and works quickly to reduce parasite mass in the bloodstream (3), resulting in a parasite reduction ratio of around 10,000 per erythrocytic cycle and prompt clinical responses (55,56). Despite being the most effective medication against all erythrocytic stages of malaria parasites, artemisinin has a very short elimination half-life (less than 1 hour), which precludes its usage as a malaria prophylactic (54).

Despite massive research efforts on artemisinin since its discovery, there is still a lot of disagreement on how it works against malaria parasites (54,57,58). However, the most widely accepted explanation about their mechanisms of action is that haem activates the molecule, causing it to produce free radicals that damage parasite survival proteins (59,60). Moreover, artemisinin's antimalarial efficacy is dependent on the endoperoxide bridge in the trioxane pharmacophore (54).

For all malaria endemic regions, artemisinin-based combination therapies (ACTs) are currently the mainstay of treatment for uncomplicated *falciparum* malaria (61). There are now 14 drugs listed for the curative treatment of malaria on the WHO Model List of Essential Medicines (52), with the treatments manufactured either as single compounds or as combinations. From

these, the artemisinin-based combinations, which use an artemisinin derivative, are perhaps the most effective (3). Their combination of a fast-acting, short-half-life artemisinin agent with a longer half-life partner medication allows for parasite killing optimization and protection against reinfection (10). The rationale behind ACT is that the risk of parasites developing resistance to two medications with different modes of action at the same time is substantially lower than the risk of parasites developing resistance to a single therapy (54). ACTs are highly successful against *P.f* malaria in most locations, with cure rates surpassing 90% (55). Artemether-lumefantrine, artesunate-amodiaquine, and artesunate-mefloquine are some of the most commonly utilized ACT medications (61). However, it is important to note that there is rampant antimalarial resistance to these mentioned drugs and their combinations.

1.3 Antimalarial drugs resistance due to parasite genetics

Throughout the world, *Plasmodium* resistance to antimalarial drugs is probably the greatest problem confronted by malaria control and eradication programs worldwide (10). Malaria parasites have developed resistance to numerous routinely used antimalarials during the last few decades (44). *P.f* resistance to CQ, for example, was initially documented in the Thai–Cambodian border regions in 1957 and has since spread throughout Sub-Saharan Africa and the rest of the world (62,63). As a result, while CQ is no longer effective against *P.f*, it is still used to treat *P.v* in areas where resistance has yet to emerge (52). Resistance to proguanil was discovered within a year of clinical use (40). SP, which is a replacement drug for proguanil, quickly faced resistance but is currently used primarily for intermittent preventive treatment during pregnancy and for seasonal malaria chemoprevention in combination with AQ and as first-line treatment in combination with artesunate in some SP-sensitive areas (61). Resistance to PPQ monotherapy was reported in China in the late 1980s (40). Artemisinin resistance, which manifests as delayed parasite clearance after treatment with artesunate monotherapy or with an ACT) was first identified in western Cambodia in 2008 (64).

The discovery of molecular resistance markers considerably facilitated the diagnosis of the spread of drug-resistant *Plasmodium* parasites. Resistance to several antimalarials is known to be conferred by amplification of genes and/or Single Nucleotide Polymorphisms (SNPs) in target catalytic enzymes or efflux pumps (38); see **Table 1.2**. For instance, sulfadoxine and pyrimethamine resistance in *P.f* were conferred by dominant mutations in catalytic sites and/or amplification of the *pfdhps* and *pfdhfr* genes (65,66). Resistance to atovaquone is caused by mutations in *P.f* cytochrome b (*PfCytB*), which cause alterations in its catalytic activity (67).

P.f chloroquine resistant transporter (*pfcrt*) has been related to CQ resistance in *P.f* (68). Mutations and amplification of the multidrug resistance protein (*pfmdr1*), cause measurable levels of resistance to many antimalarials, including MQ and lumefantrine, and thus limit effective treatment with ACTs (69). As a result, the widespread amplification of *pfmdr1* in Southeast Asia led to the substitution of DHA, the active metabolite of artemisinin, and PPQ (38). *Plasmodium falciparum kelch13* (*PfKelch13*) gene variants have been identified as molecular markers of partial artemisinin resistance and are ubiquitous in the Greater Mekong Subregion (GMS); they have also been found in Guyana, Papua New Guinea, and Rwanda at a substantial prevalence. Widespread resistance to artemisinin and other antimalarial drugs would jeopardize worldwide malaria eradication and have serious public health and economic consequences (10). As much as antimalarial resistance has been mapped to the parasite genetics, very little has been done to interrogate human host genetics, particularly the variation in genes that metabolize and transport these antimalarial drugs.

Table 1. 2 | Genetic markers of drug resistance for commonly used antimalarial drugs (compiled from (38,40,70,71))

Class of drug	**Common name**	**Target of the parasite stage**	**Genetic markers for drug resistance**	
			P. f	*P. v*
Sesquiterpene lactoneendoperoxides	Artemisinin and its derivatives	All asexual and early sexual parasite stages	*Pfkelch13*	Unknown
4-Aminoquinolines	Chloroquine	Blood stages (trophozoite and schizont)	*Pfcrt*	*pvmdr1*
	Amodiaquine	Blood stages (trophozoite and schizont)	*Pfcrt*, *Pfmdr1*	Unknown
	Piperaquine	Blood stages (trophozoite and schizont)	*Pfplm2*, *Pfcrt*	Unknown
	Pyronaridine	Blood stages (ring, trophozoite and schizont)	*Pfcrt*	Unknown
Amino alcohols	Quinin	Blood stages (trophozoite and stages I to III gametocytes)	*pfcrt*, *pfmdr1*	
	Mefloquine	Blood stages (trophozoite and schizont)	*Pfmdr1* *pfcrt*	*Pvmdr1*
	Lumefantrine	Blood stages (trophozoite and schizont)	*Pfcrt*, *Pfmdr1*	Uknown

	Halofantrine	Blood stages	*Pfcrt*, *Pfmdr1*	Uknown
8-Aminoquinoline	Piperaquine	Blood (gametocyte) and liver (schizont) forms	*pfpm2, pfexo and pfcrt*	Uknown
Antifolates	Pyrimethamine	Blood and liver schizont and mosquito stage (oocysts)	*pfdhfr*	*Pvdhfr*
	Proguanil	Blood stages (schizont and gametocyte) and liver schizont)	*pfdhfr*	Uknown
	Sulfadoxine	Blood stages (Trophozoites and schizont)	*pfdhps*	*Pvdhps*
Naphthoquinone	Atovaquone	Blood stages (schizont and gametocyte) and liver schizont	*pfcytb*	Uknown
Antibiotics	Doxycycline Clindamycin Tetracycline	Blood Stages	Apicoplast target	Uknown

Plasmodium falciparum plasmepsin 2 (*pfpm2*), *Plasmodium falciparum exonuclease* (*pfexo*)

1.4 Antimalarial drugs resistance due to host genetics, focusing on pharmacogenomics

1.4.1 Overview on pharmacogenetics of antimalarial drugs

Genetic variation in drug-metabolizing enzymes and drug transporters is known to influence the pharmacokinetics of various drugs. Such variations are one of the factors that cause variability of drug concentration in the plasma of different individuals, which could result in decreased efficacy or drug toxicity in some patients (10). Antimalarials are known to use genetically variable metabolism and elimination pathways (10,72). Moreover, evidence supports that polymorphically expressed drug metabolizing enzymes, primarily cytochrome P450 (CYP), but also drug transporters, might cause such variations (73). In addition to the CYP phase I enzymes, several phase II enzymes and transporters are known to play an important role in drug response (10).

1.4.2 Cytochrome P450 and role in antimalarial drugs resistance

CYP is a superfamily of enzymes responsible for the phase I hepatic and intestinal metabolism of many endogenous and xenobiotic compounds (74–76), including antimalarials, and they are genetically polymorphic, displaying wide interindividual variations in their enzyme activity (77). CYP enzymes are heme-containing monooxygenases that catalyze a diverse number and types of reactions, including hydroxylations, dehydrogenations, isomerizations, and demethylations of non-activated hydrocarbons at physiological temperature (78). The human genome consists of 57 CYP genes and an equivalent number of pseudogenes, which are grouped based on their sequence similarity into 18 families and 44 subfamilies (75).

Different genetic polymorphisms in CYP genes result in different phenotypes. Basically, there are three main phenotypes, which include: poor metabolizers (PM), intermediated metabolizers (IM), and extensive metabolizers (EM). But, for a few CYP (e.g., CYP2D6), there is an ultrarapid metabolizer (UM) phenotype. Patients who are PM have little or no capacity to metabolize substrate drugs resulting in higher plasma concentration of a drug in the body that might cause drug toxicity. Patients that are EM metabolize drugs expected but much more effectively or faster than PMs and IMs (79). The most clinically relevant CYPs with respect to antimalarial drugs include: *CYP1A2, CYP2A6, CYP3A4, CYP3A5, CYP2B6, CYP2C8, CYP2C9, CYP2C19* and *CYP2D6* (10)(76). **Table 1.3** shows the different types of antimalarials and their CYP metabolizing enzymes.

Table 1. 3 | Primary isoforms of human CYP and the major metabolites identified in the metabolism of antimalarial drugs (80)

Antimalarials drugs	**Primary Isoforms**	**Major metabolite**
Quinine, quinidine	*CYP3A4*	3-hydroxyl derivatives
Chloroquine	*CYP2C8, CYP3A4/3A5, CYP2D6*	Desethyl derivative
Amodiaquine	*CYP2C8*	Desethyl derivative
Mefloquine	*CYP3A4, CYP1A2*	Carboxy derivatives
Primaquine	*CYP3A4,CYP1A2, CYP2D6*	Carboxy derivatives

Halofantrine, lumefantrine	*CYP3A4/3A5*	Desbutyl derivatives
Dapsone	*CYP2C9, CYP3A4*	Hydroxyl/hydroxylamine derivative
Proguanil	*CYP2C19*	Cycloguanil
Artemisinin	*CYP2B6, CYP3A4, CYP2A6*	DHA
Artesunate	*CYP2A6, CYP2B6*	DHA
Artemether	*CYP3A4*	DHA
Arteether	*CYP3A4/3A5, CYP2B6*	DHA

CYP2C8 genetic variation and role in antimalarial drugs resistance

The human CYP2C8 enzyme is an important member of the CYP2C family of enzymes (81) that is coded for by *CYP2C8* gene, localized in the long arm of chromosome 10. *CYP2C8* comprises 9 exons and 8 introns spanning 31 kb (kilobases) (82). CYP2C8 is mainly expressed in the liver (83), and it is involved in the metabolism of a wide range of clinically available drugs (84), including antimalarials such as ADQ, CQ, dapsone, and lumefantrine (7,82,85,86).

Several SNPs have been identified in *CYP2C8*. The most common variants with functional significance include *CYP2C8 *2* and *CYP2C8*3* (84), both associated with a poor metabolizer's phenotype. *CYP2C8*2 (c.805A>T, p.Cys266Phe)* is one of the most prevalent allele in populations of African descents (10%-37%) and is associated with a 50% reduction of the metabolic activity compared to the wild type allele, whereas *CYP2C8*3,* which consists of two highly linked nonsynonymous polymorphisms, *c.416G>A, p.Arg139Lys* and *c.1196A>G, p.Lys399Arg*, is more prevalent among European populations (6.9-19.8%) and it is associated with 85% reduction in the normal enzyme activity (79,84,85).

CYP2C8 involve in the metabolism of AQ to its major metabolite, N-desethylamodiaquine (DEAM) (87). A few studies had shown the effect of *CYP2C8* SNPs on the pharmacokinetics/dynamics of AQ. Parikh and colleagues were the first to reveal the effect of *CYP2C8* polymorphisms on AQ efficacy and safety in 275 Burkina Faso patients in 2007 (72).

In this study, the African-specific *CYP2C8*2* variant was found to have a defect in AQ metabolism. Another study by Adjei *et al.,*(88) in 103 malaria-infected Ghanaian children found a weak association between *CYP2C8*2* allele carriers and lower DEAQ plasma concentrations as compared to wild-type homozygotes (88).

Although the first study showed a significant association of *CYP2C8*2* with impaired AQ metabolism *in vitro*, the allele was not clearly associated with AQ efficacy or toxicity. However, studies by Paganotti *et al*.,(86) and Cavaco *et al*.,(88) suggest that *CYP2C8* genetic variability can influence the occurrence of AQ or CQ resistance in malaria parasites.

Another antimalarial that is mainly metabolized by CYP2C8 is CQ, in which N-desethylchloroquine is the main metabolite (89). However, clinical studies on the impact of CYP2C8 polymorphisms on CQ pharmacokinetics or clinical effects are lacking. Despite this, a recent study identified an association between *CYP2C8*2* and **3* and gametocytaemia and reduced parasitemia clearance rates in chloroquine/primaquine treated patients (90).

CYP2D6 genetic variation and role in antimalarial drugs resistance

CYP2D6 is an essential CYP isoform that metabolizes roughly 20% of clinically used drugs, despite its modest hepatic concentration (2-4%) compared to other CYP enzymes (91), and is encoded by the *CYP2D6* gene (92), which has a length of ~4.3 kbps on the long arm of chromosome 22 (22q13.2) and contains 9 exons (93). It is extensively studied since it metabolizes a wide range of commonly used drugs (92,94), including antimalarials (e.g., PQ and CQ) (89,95,96). *CYP2D6* is highly polymorphic (97–99),(www.pharmvar.org), has phenotypes ranging from total dysfunction to ultra-rapid metabolizers (100).

Nevertheless, there is a profound difference in the distribution of *CYP2D6* allelic variants in different populations. For instance, the poor metabolizer alleles such as *CYP2D6*17* and *CYP2D6*29* are more common in the Sub-Saharan African and African American/Afro-Caribbean populations, whereas the *CYP2D6*10* poor metabolizer allele is the most prominent variant allele in the East Asian population. The malfunctional *CYP2D6*4* allele is more common in the European population, while the *CYP2D6*1xN* duplication/multiplication allele that results in increased function is most often observed in the Oceanian population. Some alleles (e.g., *CYP2D6*2, *4, or *5*) are found at variable frequencies in almost every population

studied, while others have only been found in some populations (e.g., *CYP2D6*44* and **49* in Asians). Finally, populations including South Africans, Caribbeans and others with diverse founding populations and admixture often reveal unique allele frequency patterns (101,102).

In hepatocytes, the CYP2D6 pathway mediates PQ metabolism into an active metabolite(s) (95,103). PQ is the only antimalarial available for the treatment of the hypnozoite stage of *P.v* malaria and it helps to prevent relapse caused by this species of *Plasmodium*. In this regard, studies have been done to evaluate the effect of *CYP2D6* polymorphism on *P.v* recurrence in patients who were taking PQ, and the studies have shown a high recurrence incidence of *P.v* in individuals with reduced activity of *CYP2D6* (97,104).

CYP3A4 and CYP3A5 and their role in antimalarial drugs resistance

CYP3As are the most abundant CYP in the human liver and small intestine (105). CYP3A4 and CYP3A5 are the two main enzymes in the CYP3A subfamily and are responsible for the metabolism of almost 50 percent of clinical drugs. The *CYP3A4* gene, which is situated in the 231-kb area of the long arm of chromosome 7 (7q21.1) and consists of 13 exons, encodes for the CYP3A4 protein (106). *CYP3A4*1B* (105), which is an A to G transition at 392 of the gene's proximal promoter region, is one of the most common allelic variants of *CYP3A4*. This variant allele has been linked to higher *CYP3A4* expression compared with the wild type allele (107)(108), and is associated with poor metabolism of artemether, lumefantrine (109,110) and QN (107).

CYP3A5 is a 502 amino-acid protein and is encoded by the *CYP3A5* gene. *CYP3A5* has nine exons and is found on the long arm of chromosome 7 (7q21-q22.1) (111,112). Due to genetic polymorphisms, the enzymatic activity of CYP3A5 differs between populations (113). The most well-studied *CYP3A5* allelic variants are *CYP3A5*3/*6/*7*, and they are prevalent in African populations (114). The splice acceptor variant *CYP3A5*3 (c.6986A>G)* causes alternative splicing and protein truncation, resulting in the lack of CYP3A5 in some people's tissues (115). *CYP3A5*6 (g.14690G>A)*, a variant found on exon 2, causes CYP3A5 alternative splicing (exon 7 skipping due to a G-to-A transition) and protein truncation, resulting in the absence of the CYP3A5 protein. *CYP3A5*7 (g.27131 27132insT)* is an insertion polymorphism that is found between codons 345 and 346 and causes a reading frame shift. This change causes a premature termination codon (D348) at position 348, resulting in a shortened and non-functional protein (116).

CYP3A5 genetic polymorphisms are known to affect artemether-lumefantrine pharmacokinetics and treatment outcomes (77). According to recent studies by Mutagonda *et al*. (77,113), there was a significant association between *CYP3A5* genotype (**1/*1*) and day 7 plasma lumefantrine concentration among pregnant women in Tanzanian populations.

CYP2C19 genetic variation and their role in antimalarials resistance

CYP2C19 is the main enzyme that converts proguanil and chlorproguanil, the biguanide derivatives, into their major active metabolites, cycloguanil and chlorcycloguanil. This enzyme is encoded by the *CYP2C19* gene, located on the long arm of chromosome 10 (10q23.33) (117). The gene is highly polymorphic (10), and several SNPs have been identified, including *CYP2C19*2, *3, and *17. CYP2C19*2* and *CYP2C19*3* are null alleles with no protein in the CYP2C19 poor metabolizer phenotype, whereas *CYP2C19*17* is a promoter gene variant associated with enhanced substrate turnover in the CYP2C19 UM phenotype (73).

Kaneko *et al.* (10) investigated the impact of *CYP2C19* genetic polymorphisms on proguanil pharmacokinetics in the Malasian population. As a result, proguanil concentrations were found to be greater among PM in this study. The treatment efficacy of PM and EM, however, did not differ significantly.

1.4.3 Phase II enzymes and their role in antimalarial drugs resistance

Phase II drug metabolizing enzymes play a significant role in the biotransformation of endogenous compounds and xenobiotics into more easily excretable forms as well as the metabolic inactivation of pharmacologically active substances (118). They perform conjugation reactions, which include glucuronidation, sulfonation, methylation, acetylation, glutathione and amino acid conjugation. Phase II drug enzymes are mostly transferases that include: Uridine 5'-diphospho-glucuronosyltransferases (UDP-glucuronosyltransferase, UGTs), sulfotransferases (SULTs), N-acetyltransferases (NATs), glutathione S-transferases (GSTs) and various methyltransferases (mainly thiopurine S-methyl transferase (TPMT) and catechol O-methyl transferase (COMT)) (118). From these, the UGTs and NATs were found to have a role in antimalarial drugs' phase II metabolism (119–121) while for the others, concrete evidence was not found.

1.4.3.1 UGT variation and their role on antimalarial drugs resistance

The primary phase II conjugation pathway in the biotransformation and removal of endogenous and xenobiotic chemicals is glucuronidation, which is catalyzed by UGTs. By transferring the glucuronic acid moiety of UDP glucuronic acid to the substrate, the UGTs make the substrate more polar and make it easier for it to be excreted in bile or urine (122). In humans, UGT1A9 and UGT2B7 catalyze the glucuronidation of DHA, which is a derivative of the parent compound artemisinin and the major metabolite of artemisinin derivatives such as artesunate, artemether, and β-arteether (119). In Vietnamese adult malaria patients given artesunate, α-DHA-β-glucuronide was found to be the major urinary product of DHA glucuronidation, and, using V79 cells expressing human UGT1A1, 1A6, 1A9, or 2B7, α-DHA-β-glucuronide formation was observed with UGT1A9 and UGT2B7, but not with UGT1A1 and UGT1A6 (123).

The whole *UGT1A* sub-family is situated on the long arm of chromosome 2 (2q37) and spans approximately 210 kb (124). All *UGT1A* isoforms, including *UGT1A9*, have a unique first exon and four common exons (124). The length of *UGT1A9* cDNA is 1593 base pairs (bp). The gene has so far been linked to 6 nonsynonymous SNPs (*c.8G>A, c.9C>G, c. 387A>T, 98T>C, 766G>A, c.794T>A*) and a SNP at *726T>G* that results in a premature termination codon TAG (125).

Nevertheless, there is limited information on the effects of UGT1A isoform polymorphisms on antimalarial drug treatment outcomes (119,120). One of the studies was by Phompradit *et al.* (119) in 2014 that assessed the effect of *UGT1A9* polymorphisms on artesunate-mefloquine treatment response in Burmese patients (n=71) with *Plasmodium falci*parum malaria from Tak Province, Thailand. In this study, 3 polymorphisms (*UGT1A9*1, *4, and *5*) were genotyped using the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method, and correlation analysis with plasma concentrations of artesunate/dihydroartemisinin and treatment response was performed. However, no concrete evidence on the effect of the variants on treatment outcome was obtained from the study since all patients had the wild-type genotype of *UGT1A9.*

UGT2B7, along with other members of the *UGT2B* subfamily, is localized to chromosome 4q13, spans a region of approximately 16 kb, and is composed of 6 exons (126). So far, 10

nonsynonymous SNPs have been reported in this gene. These SNPs are: *c.122C>T*, *c.137T>C, c.211G>T c.214C>T, c.536C>T, c.802C>T, c.896C>T, c.1133A>G, c.1226G>A,* and *c.1288A>C* (127). The frequencies of these variant alleles vary between populations. For instance, allele frequencies of *c.211G>T* and *802C>T* are (14.8–18.5%, Japanese; Asian-Americans (9%); Hispanic-Americans (2%)) (128,129) and (24.4–29.3%, Japanese; 48.9–53.7%, Caucasians) (129,130), respectively. Despite the polymorphic nature of *UGT2B7,* only one study was found that evaluated the effect on antimalarial drugs treatment outcomes (120).

1.4.3.2 NATs genetic variation and role in antimalarial drugs resistance

NATs are cytosolic enzymes present in a wide range of tissues from different species. They catalyze the acetyl transfer from acetyl coenzyme A to an aromatic amine, heterocyclic amine or hydrazine compound (118,131). NATs have a role in the metabolism of a wide range of chemicals to which we are constantly exposed. Many arylamine and hydrazine medications, as well as several known carcinogens found in diet, cigarette smoke, car exhaust fumes, and the environment in general, are biotransformed via acetylation in humans. In humans, two N–acetyltransferase isoenzymes (NATs), N–acetyltransferase 1 (NAT1) and 2 (NAT2), catalyze acetylation processes (132). NAT2 is one of the two enzymes involved in antimalarial metabolism (110,121)

Human NAT2 is a xenobiotic–metabolizing enzyme that plays a key role in the detoxification of medications such as isoniazid, hydralazine, and endralazine, as well as sulphonamides, procainamide, aminoglutethimide, nitrazepam, and dapsone (132). It is encoded by the human *NAT2* gene, located on the short arm of chromosome 8 (8p22) (132). *NAT2* is highly polymorphic, and the most investigated allelic variants include *NAT2*4*, *NAT2*5*, *NAT2*6*, *NAT2*7*, and *NAT2*14* (110,121). Moreover, the frequency distribution of their variant alleles varies between populations. For instance, the variant allele frequency for *NAT2*5* is 6.6% in Cambodians, 36.4% in Tanzanians, and 48.5% in Ethiopians; for *NAT2*6* it is 40% in Cambodians, 27% in Tanzanians, and 30.1% Ethiopians; for *NAT2*7* it is 21% in Cambodians, 2.4% in Tanzanians, and 5.5% Ethiopians); and for *NAT2*14* it is 2.5% in Cambodians, 17% in Tanzanians, and 1.8% in Ethiopians (110,133).

NAT2 genetic polymorphisms have been associated with the incidence of variable adverse drug reactions between individuals (122). In a study conducted by Kengneet *et al*., (121) a

significant association was found between *NAT*5, *6,* and *7* variants (slow acetylators phenotype) and adverse drug reactions in a longitudinal study involving children under the age of ten from Cameron who were taking ACT such as artesunate-amodiaquine, dihydroartemisinin-piperaquine, and artemether-lumefantrine.

1.4.4 Drug transporters and their role in antimalarial drugs resistance

Transporters are membrane-bound proteins that facilitate the movement of compounds in and out of cells (134). Generally, there are two transporter superfamilies: the solute carrier (SLC) transporters and the ATP-binding cassette (ABC) transporters. The SLC drug transporters are uptake transporters. They are generally involved in the uptake of small molecules into cells and do not rely directly on ATP hydrolysis. By contrast, ABC drug transporters utilize ATP hydrolysis and function as efflux transporters (135,136). A good example of an antimalarial that is transported by an ABC transporter is lumefantrine (79).

1.4.4.1 ABCB1 genetic variation and role in antimalarial drugs resistance

ATP-binding cassette subfamily B1 (ABCB1), also called P-glycoprotein, is a member of the superfamily of ATP-binding cassette (ABC) transporters (73). It is encoded by the *ABCB1* gene (137). The gene has 32 exons and is located on the long arm of chromosome 7 (7q21.12) (138). ABCB1 is widely recognized as a major determinant for the low or variable oral absorption of several structurally unrelated drugs (137). It is mainly located in the membrane of epithelial cells in the intestine, liver, and kidney, thus contributing to the absorption of orally administered drugs and promoting their biliary and renal elimination (136). ABCB1 is also used as a blood-tissue barrier, protecting various tissues (e.g., the brain, fetus, and testes) from toxic xenobiotics (136).

Mutations in the *ABCB1* gene are associated with several drugs, including antimalarials. Three closely linked *ABCB1* polymorphisms (*c.1236C>T, c.2677G>T/A*, and *c.3435C>T*) have been associated with duodenal and lymphocyte P-glycoprotein expression and function, digoxin systemic availability, and response to anti-HIV and anti-epileptic drugs (79). The three variants are found at high frequency in African populations (134).

Genetic variants of drug transporters in human beings might advance the development of drug resistance by, for instance, lowering drug concentrations in red blood cells (79). A case reported by Kiaco K *et al.,* (139) showed an association between the *ABCB1c.3435C>T* variant and artemether-lumefantrine treatment outcome in 103 uncomplicated malaria patients from

Angola. According to the study, a significant predominance of the *ABCB1c.3435CC* genotype was found among the recurrent infection-free patients ($p < 0.01$), suggesting a role for this transporter in artemether lumefantrine interindividual performance. Although it was not statistically significant, recent studies on pregnant women by Mutagonda *et al*. (113,134) have shown an association of lower lumefantrine concentration in carriers of both homozygous and heterozygous functional *ABCB1c.3435C>T* alleles than with the homozygous mutant carriers.

1.4.4.2 SLC genetic variation and role in antimalarial drugs resistance

Human SLC transporters compose a family of 395 membrane-bound proteins that are mainly involved in the uptake of small molecules into cells. SLC transporters interact with a wide variety of substrates and have varying degrees of specificity. They are highly polymorphic, with different populations having different distributions of allelic variations. *SLCO1B1* rs4149056, for example, was found in 23 populations at a frequency of 0.0–0.99, *SLCO1B1 c.597C>T* at 0.0–0.66 (across 9 populations), and *SLCO1B1 c.571C>T* at 0.0–0.42 (across 9 populations). *SLCO1B1* rs4149056 and *SLC22A6* rs1158626 are the two important allelic variants that are more prevalent in African populations than non-Africans (140).

Polymorphisms in SLC transporters were found to have pharmacogenetic relevance for different xenobiotics, including antimalarials (140)(140). Genetic variations in these genes were associated with *Plasmodium* parasitemia and/or gametocytaemia clearance. A study by Sortica *et al*.(141) was the first to evaluate the role of polymorphisms in the *SLCO1A2*, *SLCO1B1,* and *SLCO2B1* genes in CQ and PQ treatment outcomes in *P.v* malaria patients.

1.5 Problem statement

Malaria remains a major public health challenge in sub-Saharan African countries. A viable vaccine for malaria applicable on a large-scale to endemic regions is still not available; thus, the use of antimalarial drugs represents the long-term main strategy for the control of the disease. In Africa, artemisinin derivative-based combination therapy (ACT) is massively applied as a response to the rise of parasite resistance to monotherapy, particularly chloroquine (124). However, such large-scale use of ACTs across endemic areas in Africa might lead to the development of drug resistance by putting high selection pressure on the circulating parasite strains. Resistance to ACTs can potentially result in slow parasite clearance times by putting pressure on partner drugs and thus leads to increased treatment failure rates and the emergence of multidrug-resistant parasites (142). The emergence and spread of artemisinin-resistant *P.f*

is becoming a major obstacle for malaria control in malaria endemic areas in Southeast Asia (e.g., Cambodia) (113). Currently, there is a huge concern that artemisinin-resistant parasites might also emerge and spread in endemic areas in Africa, which would threaten the global malaria eradication efforts (143).

Much of the research on malaria treatment resistance has focused on parasite genetics; several studies, however, have shown wide variations in treatment outcomes among patients (144–146). One of the factors that can explain the observed variability is host genetic variation, particularly, differences in the pharmacogenomics profiles of patients (77,90,113,147). For instance, genetic polymorphisms in genes encoding proteins and enzymes involved in drug Absorption ,Distribution, Metabolism and Elimination (ADME) can result in interindividual differences in drug pharmacokinetics and pharmacodynamics parameters (110). The lack of efficacy of antimalarials in some groups of individuals is known to be the primary driver of drug resistance in endemic countries. Therefore, it is important to devise and implement strategies that minimize the devastating problems. One such strategy is personalizing medications depending on the genotypes of the patients; which can potentially minimize the emergence of drug resistance and improve treatment outcomes (79).

In this regard, inter-ethnic genetic differences among populations can be explained by accessing different profiles of pharmacogenes variants, including the population structure of genetic variants that are involved in drug metabolism (ADME). Due to the interpolation of genetic variations, drugs as well as markers used for pharmacogenotyping in one population may not be appropriate for another population. Furthermore, owing to the heritable characteristics of human genetics, inter-individual genetic differences are manifested at population levels (148). The role of investigating the population structure of drug response genes in the world population has been described previously (149). Nevertheless, there is no study found that assess the profiles of antimalarial pharmacogenes variants among different world populations from malaria endemic and non-endemic areas. Analysis of population specific genetic structure has many applications in medical and population genetic research, as well as ensuring drug efficacy and the development of pharmacogenetic tests (149).

AIM: Here, we aim to determine the pharmacogenomic profiles of populations in malaria-endemic and non-endemic areas, to enable identification of genomic markers for antimalarial drug response and individualised treatment.

1.6 Objectives

1. To access genomic datasets covering pharmacogenomic loci from populations in malaria endemic and non-endemic regions
2. To perform quality control of different datasets of malaria endemic and non-endemic populations
3. To determine the variant allele frequency distribution of pharmacogenes relevant for malaria response in populations from malaria endemic and non-endemic areas
4. To perform the principal component analysis (PCA) using data on pharmacogenes variants and comparing the two groups, malaria endemic, and non-endemic.
5. To characterize 7 SNPs in *CYP2C8* gene among South Africans to validate data obtained from datasets and compare with world populations from malaria endemic and non-endemic areas
6. To come up with possible recommendations on the utility of the pharmacogenomics knowledge generated

1.7 Significance of the study

The distribution of antimalarial pharmacogenes variants differs among geographical populations (150). Unfortunately, there is limited information on the population genetics of these variants in the world population in general and particularly among Africans. In addition to this, most of the genetic studies are limited to non-African populations, and it is important to study the diverse and heterogeneous African population. In this study, we analysed large scale genetic data using bioinformatics approaches to determine the population genetics of antimalarial pharmacogenes variants among populations of malaria endemic areas in Africa and world populations of non-malaria endemic areas.

The findings from this project will have implications for the pharmacogenomics of antimalarials. One of the most important uses of this study will be providing baseline information for future studies related to the pharmacogenomics of antimalarial drugs. Since most of the studies on the pharmacogenetics of antimalarial drugs are limited to candidate gene approaches (77,113,151). However, this study may provide additional, broader information that can be used in future studies. Moreover, the findings from this study may also have implications for the pharmacovigilance of ACT drugs in Africa, to assess the safety of these drugs. Understanding the pharmacogenomic profiles of populations in malarial endemic regions could assist in rational drug design, while the knowledge, when applied to non-endemic regions, could assist in cases where individuals travel to regions endemic with malaria and get infected.

CHAPTER 2 : MATERIALS AND METHODS

2.1 Description of the study datasets and quality control

Genome-wide genotype datasets of four malaria endemic African populations that include Gambia, Kenya, Malawi, and Mali (n=13,447), were obtained from the MalariaGEN consortium (152). These included cases who were admitted to hospitals and fulfilled the WHO case definition for severe malaria, and controls were obtained from the general populations (153). The samples were genotyped on Illumina Omni 2.5M array as described in (154).

As a comparator group, we accessed reference datasets of global populations composed of 20 ethnic groups (n~5000) from the African Genome Variation Project (AGVP) (155) and the 1000 Genome Consortium Project (156). Datasets were merged on common SNPs using PLINK software (157) and quality control (QC) filtering including Hardy-Weinberg equilibrium (p=0.01), SNP-missingness proportion (0.05), MAF (0.05), sample missingness proportion (0.05), sample relatedness proportion (0.05) and heterozygosity rate were performed using PLINK software (157) (**Annex A**). The QC-filtered VCF files of each population were aligned to the forward strand of the human reference sequence (GRCh37) using genotype harmonizer (158) and phased using the SHAPEIT2 method (159) (**Annex B**).

2.2 Genetic structure of pharmacogenes variants

The pharmacogenes were obtained from databases including Pharmacogene Variation Consortium (PharmVar) (www.pharmvar.org), The Pharmacogenomics Knowledge Base (PharmGKB) (www.pharmgkb.org/), UDP-Glucuronosyltransferase (UGT) nomenclature (www.pharmacogenomics.pha.ulaval.ca/ugt-alleles-nomenclature/), membrane transporters databases such as solute carrier (SLC) transporters (www.guidetopharmacology.org/GRAC/FamilyDisplayForward?familyId=863) and ATP binding cassette (ABC) transporters (www.genenames.org/data/genegroup/#!/group/417), Thiopurine methyltransferase (TPMT) Nomenclature (www.liu.se/en/research/tpmt-nomenclature-committee) and Arylamine N-acetyltransferase (NAT) Allele Nomenclature Database (www.nat.mbg.duth.gr/). Antimalarial pharmacogenes reported thus far in

published literature were also noted and included. SNPs that map to the aforementioned pharmacogenes from the dbSNP database were retrieved and used to compare the population structure of the pharmacogenes in malaria endemic and global populations. Briefly, datasets containing only pharmacogenes SNPs from the merged reference data composed of the 20 global populations and malaria endemic datasets of the four malaria endemic populations were created. Allele frequencies of the SNPs residing in the pharmacogenes across the populations were computed and binned into different allele frequency spectrums using PLINK software (157). The proportion of pathogenic SNPs in each pharmacogene was also computed to understand the burden of pathogenicity in antimalarial pharmacogenes across global populations using ANNOVAR software (160). Furthermore, principal component analysis was performed based on the SNPs residing only in the pharmacogenes to assess the differentiation levels of these genes across populations using smartpca software.

2.2.1 Positive selection analysis

Positive selection metrics were computed from genotype datasets of populations from malaria endemic areas, mainly Kenya, Gambia, and Malawi, to check if positive selection is attributing to the high proportion of pharmacogenes variants in the populations. Therefore, to identify selection signals in each population, an integrated haplotype score (iHS) (161), implemented in REHH packages v.3 (162), was applied and an iHS score was calculated.

Then, candidate iHS regions were computed by splitting the genome into segments of 50 and setting the significance threshold at 4 using 'calc_candidate_regions()' utility in the REHH package (162). For each window, a summary statistic including a minimum number of markers, a minimum number of extremal markers, a minimum percentage of extremal markers, and an average of extremal scores were obtained. The candidate identified regions were then annotated using the UCSC table browser (https://genome.ucsc.edu). From the candidate regions, we extracted genomic regions that contain the identified pharmacogenes.

2.3 Network analysis of antimalarial pharmacogenes

A network analytic approach was applied to evaluate the functional relationships between pharmacogenes. The Multiple Association Network Integration Algorithm (geneMANIA) tool (http://genemania.org) was used to get the interaction profiles of antimalarial pharmacogenes.

2.4 Characterization of 7 *CYP2C8* SNPs in South Africans using Sanger Sequencing

2.4.1 Sample recruitment and quality check

One hundred and fifty (n=150) samples of native South Africans were recruited from the biorepository of the Pharmacogenomics and Drug Metabolism Research Laboratory, Division of Human Genetics, University of Cape Town. A NanoDrop spectrometer measurement of the samples was performed to check the quality and concentration of genomic DNA (gDNA); a minimum concentration of 10 ng/dl gDNA was considered for Sanger sequencing. Additionally, a DNA integrity test, using 1% agarose gel electrophoresis was also performed to check the intactness of the gDNA. Lastly, dilution of the selected samples was performed to get a final concentration of 50 ng gDNA with a final volume of 50 μL (using nuclease-free water as a diluent) and used for a PCR reaction. Dilution was not performed for samples with an initial concentration of <50 ng/dl; diluted samples were stored at 4 ^{o}C during genetic analysis.

2.4.2 Primer designing

The targeted region for Sanger sequencing consisted of 7 SNPs of the CYP*2C8* gene, which are in proximity with the African specific *CYP2C8*2* variant. The 7 SNPs include *rs1926705 G>A, rs11572100 A>G, rs11572101 T>C, CYP2C8*14 c.712G>C (rs188934928), CYP2C8 c.792C>G (rs1058930), CYP2C8 c.805A>T (rs11572103) & rs11572105 C>A/G/T*. The targeted PCR product length was about 798 base pairs of nucleotides (see **Annex C** for more primer information). Primers were designed using the Integrated DNA Technologies (IDT) PrimerQuest tool (Integrated DNA Technology, Iowa, USA) and were blasted on the National Center for Biotechnology Information (NCBI) Primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_ LOC=BlastDescAd). Finally, the primers were purchased from Inqaba biotec (Inqaba biotech, Pretoria, South Africa).

2.4.3 Preparation of primers working solution

The working solutions of *CYP2C8* primers were prepared according to the manufacturer's protocol (Inqaba biotec). Briefly, we took 10 μL of 100 μM stock primer solution (for each forward and reverse primer, separately) and mixed it with 90 μL of nuclease-free water, to make a final concentration of 10 μM primer working solution. Finally, the primers were aliquoted into 50 μL and kept in -20^{o}C refrigerator until further analysis was performed.

2.4.4 PCR and agarose gel electrophoresis

PCR was performed in a 25 μL reaction mix containing 0.4 μM forward primer: 5′-CCATCGTTCTCAGCATACTATCAC-3′ and reverse primer: 5′-CTATGCATTCTAGCCATTGGACAAT-3′each, 0.4 mM of dNTPs, 0.65 U of GoTaq DNA polymerase, 1X Go-Taq buffer, 50 ng DNA template, 1 mM $MgCl_2$, and 13.87 μL of distilled H_2O. BIO-RAD T100 Thermal Cycler (Bio-Rad Laboratories, California, USA) was used for PCR amplification and cyclic conditions involved an initial denaturation at 95^{o}C for 3 minutes, followed by 30 cycles of denaturation at 95^{o}C for 30 second, annealing at 65^{o}C for 30 seconds, and extension at 72^{o}C for 1 minute, followed by a final incubation at 72^{o}C for 5 minutes, to allow extension. Then, we checked the presence of DNA products using agarose gel electrophoresis.

2.4.5 Clean-up PCR product and sequencing

For clean-up, a 20 μL reaction mix composed of 5 μL of PCR product, 4 U of Exo I, 1 U of FastAP, and 13.8 μL of distilled H_2O, was used**.** Then, the reaction mix was incubated at 37^{o}C for 1 hour and at 75^{o}C for 15 minutes using the BIO-RAD T100 Thermal Cycler (Bio-Rad Laboratories, California, USA).

Following clean-up, products were sequenced using the Big-Dye Terminator sequencing, according to the manufacturer's protocol (Applied Biosystems, California, USA). The sequencing reaction included a 10 μL reaction mix, which comprised of 5 μL of clean-up PCR product, 5U Big-Dye Terminator mix, 1U Big-Dye Terminator buffer, and 1 μM of reverse primer. The reaction was carried out in an Applied Biosystems SimpliAmp Thermal Cycler (Applied Biosystems, California, USA) with the following conditions: an initial denaturation at 95^{o}C for 5 min, followed by 25 cycles of denaturation at 96^{o}C for 30 sec, annealing at 50^{o}C for 15 sec, and extension at 60^{o}C for 4 min.

2.4.6 Post-sequencing clean-up and capillary electrophoresis

The sequencing products were cleaned-up using the ethanol precipitation method, according to the protocol in (**Annex D**). For capillary electrophoresis, the cleaned-up products were analysed using the ABI 3130 DNA Analyzer (Applied Biosystems, California, USA), and results were analysed using ChromasPro sequence assembly software version 2.0.1 (Technelysium, South Brisbane, Australia).

2.4.7 Statistical analysis

Microsoft Excel 365 v.2109 (Microsoft Corporation, Redmond, WA, USA) was used to calculate Hardy-Weinberg Equilibrium (HWE), genotypes and allele frequencies of the seven *CYP2C8* SNPs. STATA software for Windows version 15 (StataCorp, College Station, Texas, USA), was used to perform a chi-square to compare the seven *CYP2C8* allele frequencies with world populations, and p-value ≤ 0.05 was considered statistically significant.

CHAPTER 3 : RESULTS

3.1 Description of the datasets and quality control

We accessed genome-wide genotype datasets of four malaria endemic African populations including, Mali, Kenya, Gambia, and Malawi (n=13,447) from the MalariaGEN consortium and reference panels of the global populations from AGVP (155) and the 1000 Genome Consortium Project (156) as described in the Materials and Methods section. We performed basic quality controls on each dataset (**Table 3.1**). We used the quality filtered datasets for the population genetics analyses. However, since the Mali dataset was poor in quality, as shown by a massive loss of SNPs (511,497 SNPs) after QC, it was excluded from most of the analysis.

Table 3. 1 | Quality control for genome-wide datasets of populations from malaria endemic areas

Population	**Sample(N)**		**SNPs (N)**	
	Before QC	After QC	Before QC	After QC
Kenya	3865	3142	1.6 million	1.5 million
Gambia	5594	4921	1.6 million	1.5 million
Malawi	3088	2516	1.6 million	1,502,608
Mali	900	900	1,550,514	1,039,017

3.2 Genetic structure of non-antimalarial specific pharmacogenes variants

Gene variants of pharmacogenomics importance were obtained from the sources indicated in the methods section. Based on our analysis, we found the MAF proportion of common variants (MAF > 0.05) in the pharmacogenes were generally higher in the four malaria endemic populations compared to the 20 ethnic groups (**Figure 3.1**), while a disproportionally higher number (> 80%) of rare variants were obtained mainly in the European and Asian populations compared to the four malaria populations. Further analysis of pharmacogenes' specific allele frequency showed that the majority of the variants were located on the CYP-isoenzymes, mainly *CYP2A13* and *CYP2F1* (**Figure 3.2**).

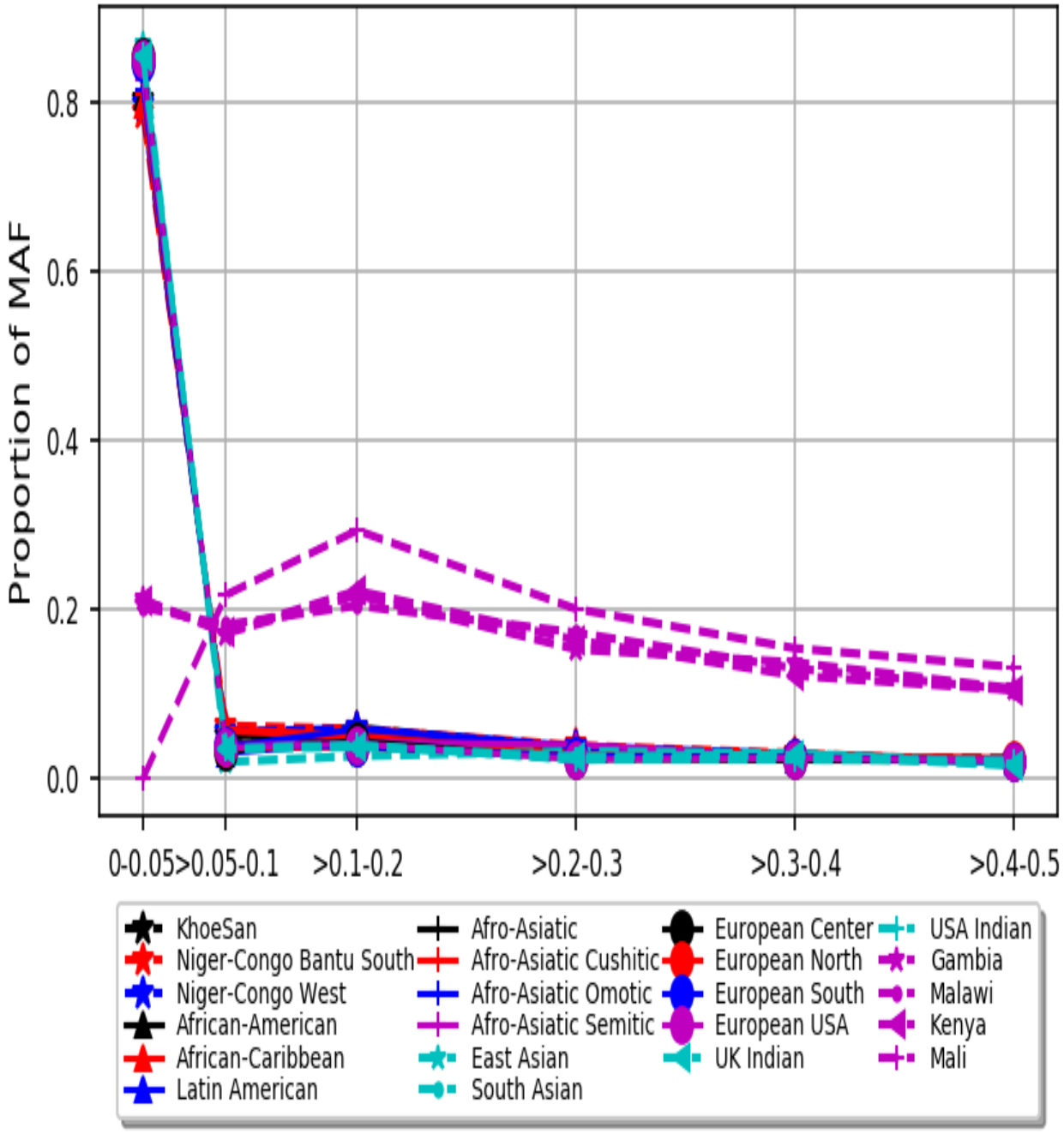

Figure 3. 1 | MAF proportion of SNPs mapped to non-antimalarial specific pharmacogenes in the four malaria endemic populations and world population of 20 ethnic groups. The x-axis represents different MAF bins, and the y-axis represents allele frequency proportion. The different colours and symbols represent populations.

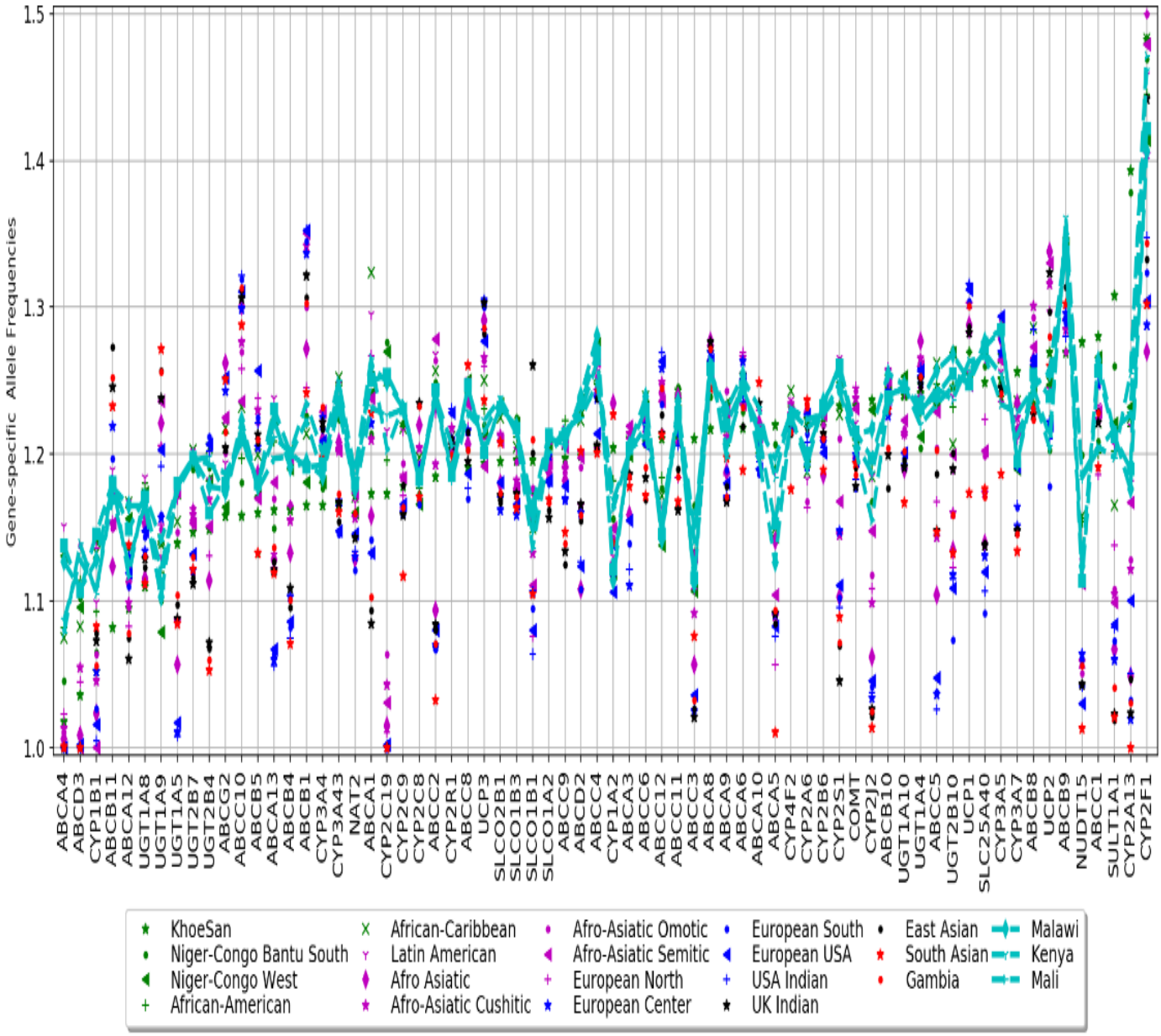

Figure 3.2 | Gene specific allele frequency spectrum of prioritized pharmacogenes from all non-antimalarial specific pharmacogenes

3.3 Genetic structure of antimalarial specific pharmacogenes variants

Like the non-antimalarial specific pharmacogenes variants, analysis based on variants of antimalarial specific pharmacogenes showed a high proportion of common variants (MAF > 0.05) in the four malaria endemic populations compared to the 20 ethnic groups (**Figure**

3.3). On different analysis, we found most of the variants were located in the *CYP2C19* and *UGT1A9* genes (**Figure 3.4**). However, *CYP2C19* variants were higher in proportion in the populations from malaria non-endemic areas, while *UGT1A9* variants were higher in proportion in the populations from malaria endemic areas.

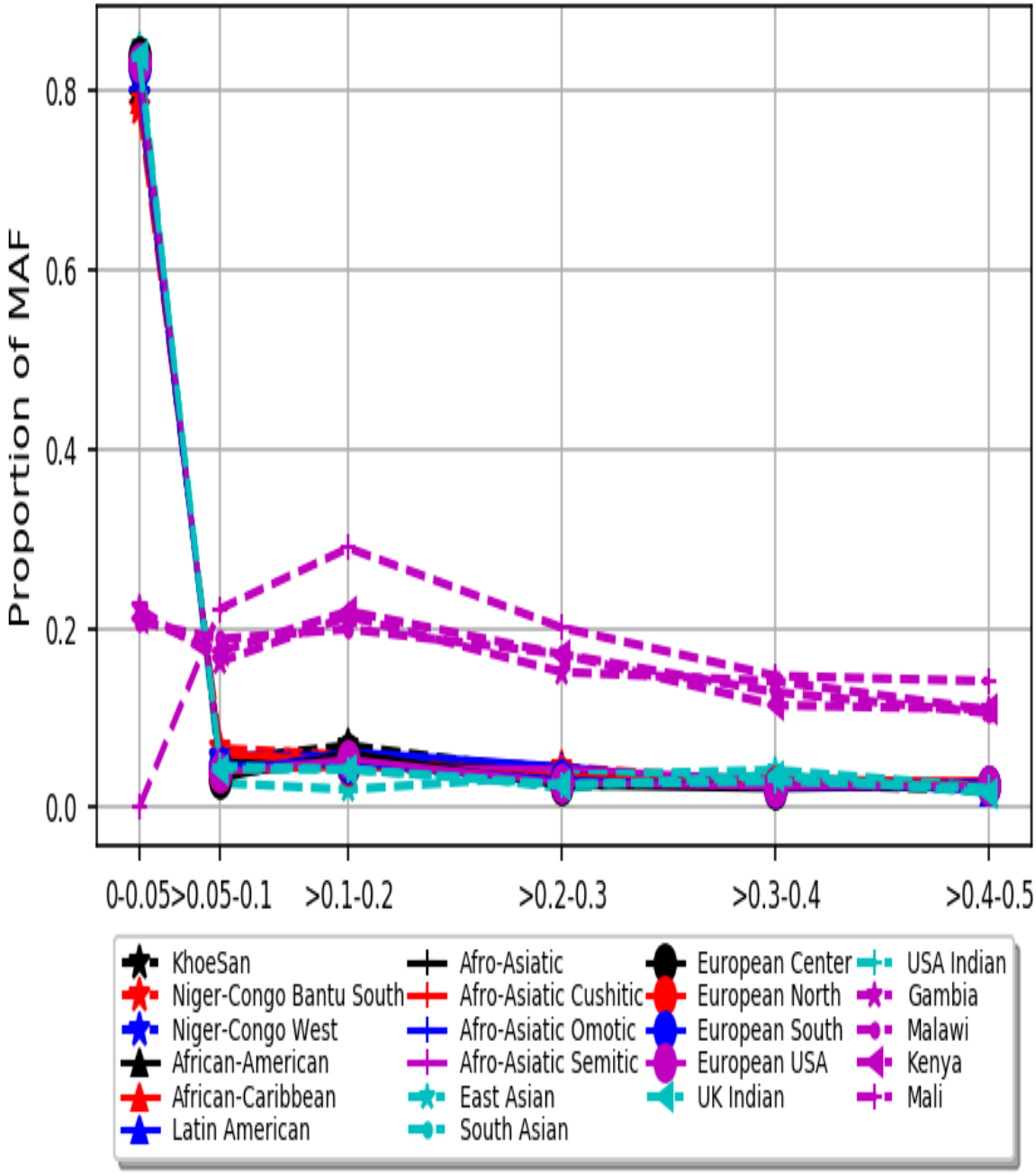

Figure 3.3 | MAF proportion of antimalarial specific pharmacogenes. The x-axis shows the frequency spectrum (bin) of pharmacogenes variants, and the y-axis shows the proportion of variants along the frequency spectrum.

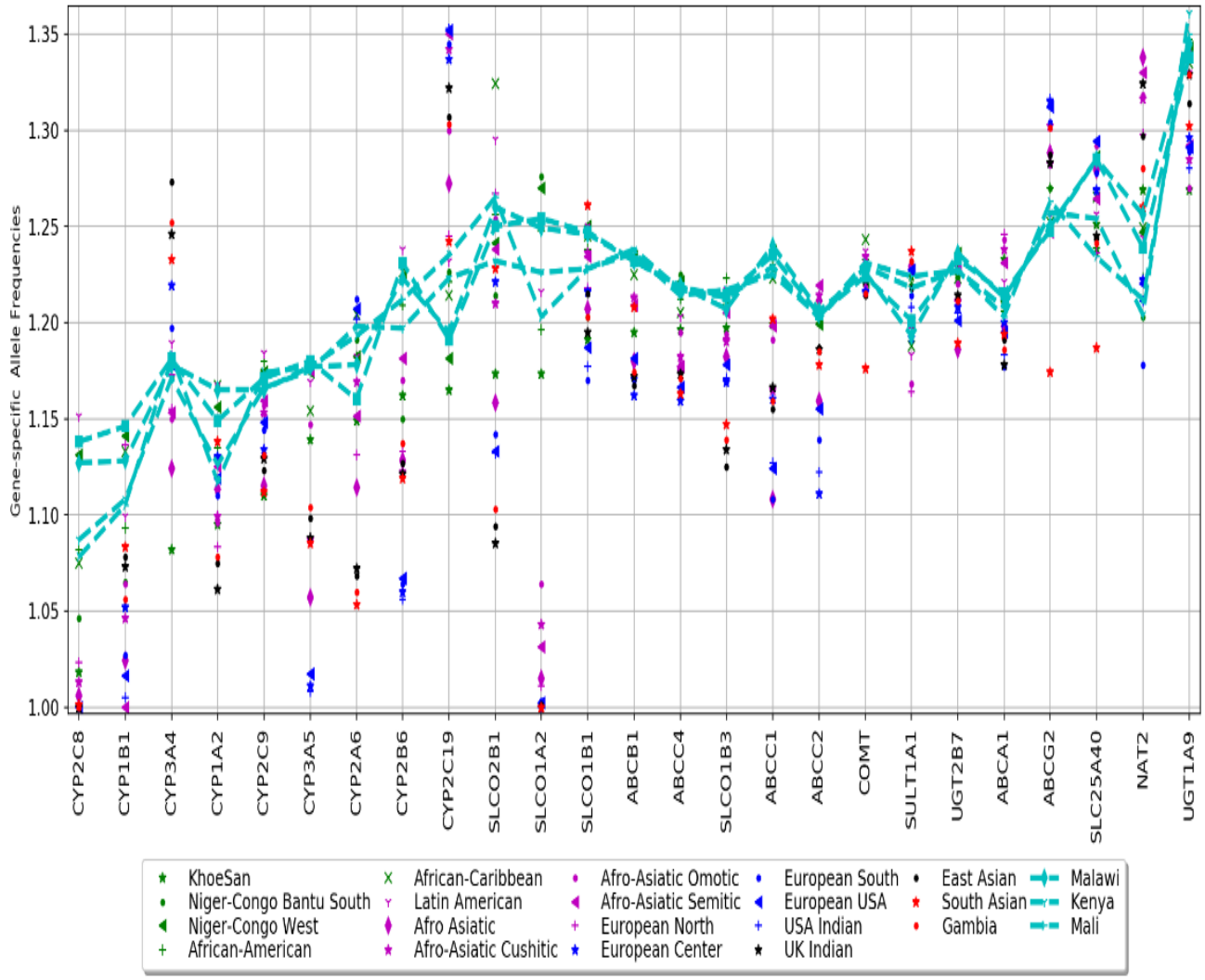

Figure 3.4 | Gene specific allele frequency spectrum of antimalarial specific pharmacogenes variants

3.4 Positive selection analysis for populations from malaria endemic areas

3.4.1 Quality control

Quality control of the three malaria datasets (Gambia, Kenya, and Malawi) was performed prior to the identification of positively selected pharmacogenes. The quality control showed that all three populations' datasets (Gambia, Kenya, and Malawi) had normal distribution graphs and normal quantile-quantile plots (q-q plot), respectively (**Figure 3.5**). The Mali dataset, however, was not included in the analysis due to the quality issue indicated in the quality control section of the method.

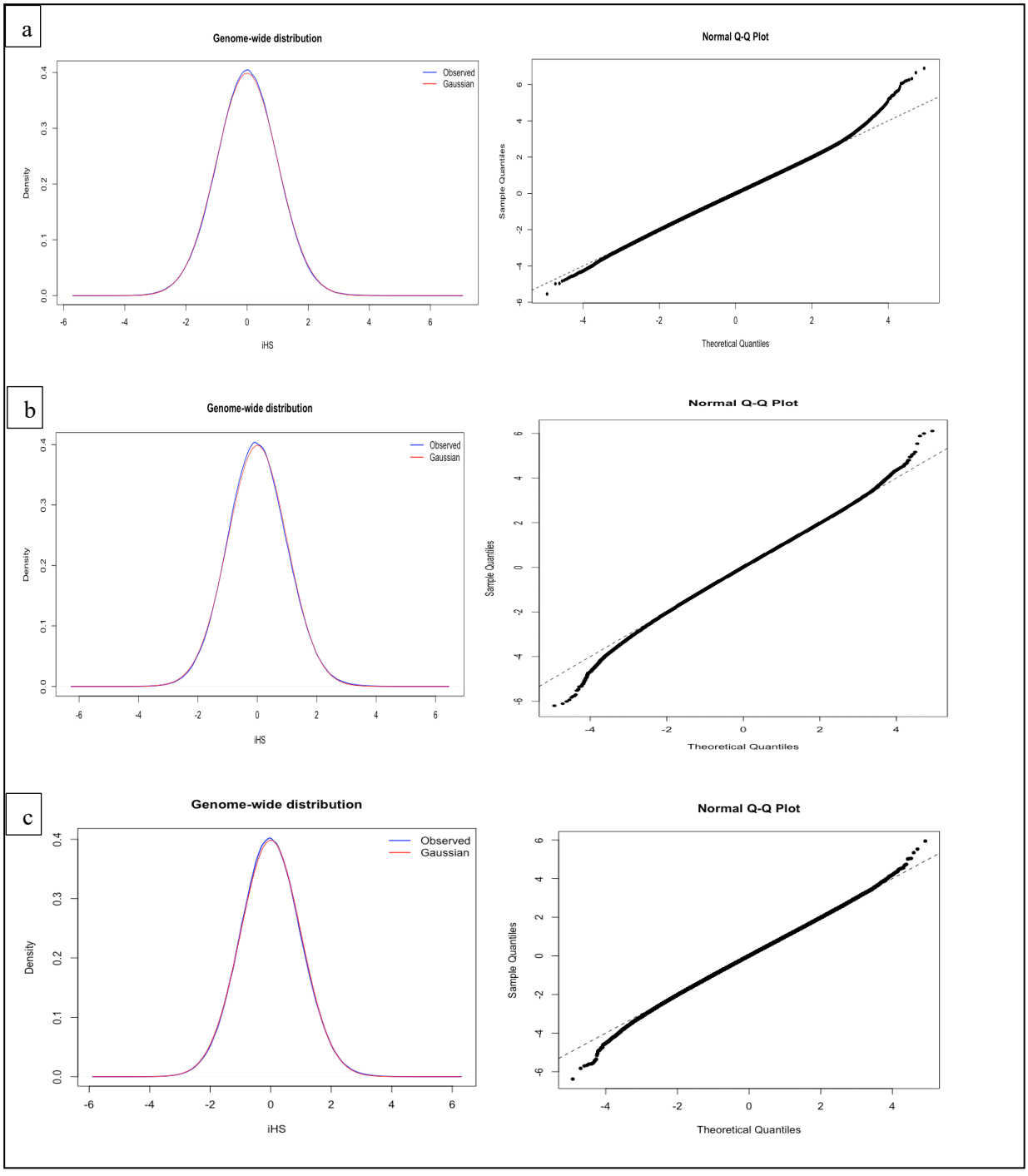

Figure 3. 5 | Quality control of the three-malaria datasets (Gambia, Kenya, and Malawi). (a), (b) and (c) represents normal distribution graph and qq-plot for Gambia, Kenya and Malawi populations, respectively.

3.4.2 iHS scores

Genome-wide SNP datasets of the three populations from malaria endemic areas were used to identify genomic regions that have been under natural selection. As described in the materials and methods section, we used the concept of an 'Extended Haplotype Homozygosity' (EHH) introduced by (161), and the iHS function from the "rehh" package was applied to detect positive or 'Darwinian' selection within a single population. **Figure 3.6** shows Manhattan plot of the genome-wide distribution of iHS scores for the autosomes in the three malaria populations (Gambia (a), Kenya (b), and Malawi (c). The figure illustrates the general view of positively selected SNPs across the genome for each population. Positively selected SNPs in all autosomal chromosomes are those with an iHS score of > 4 or < -4.

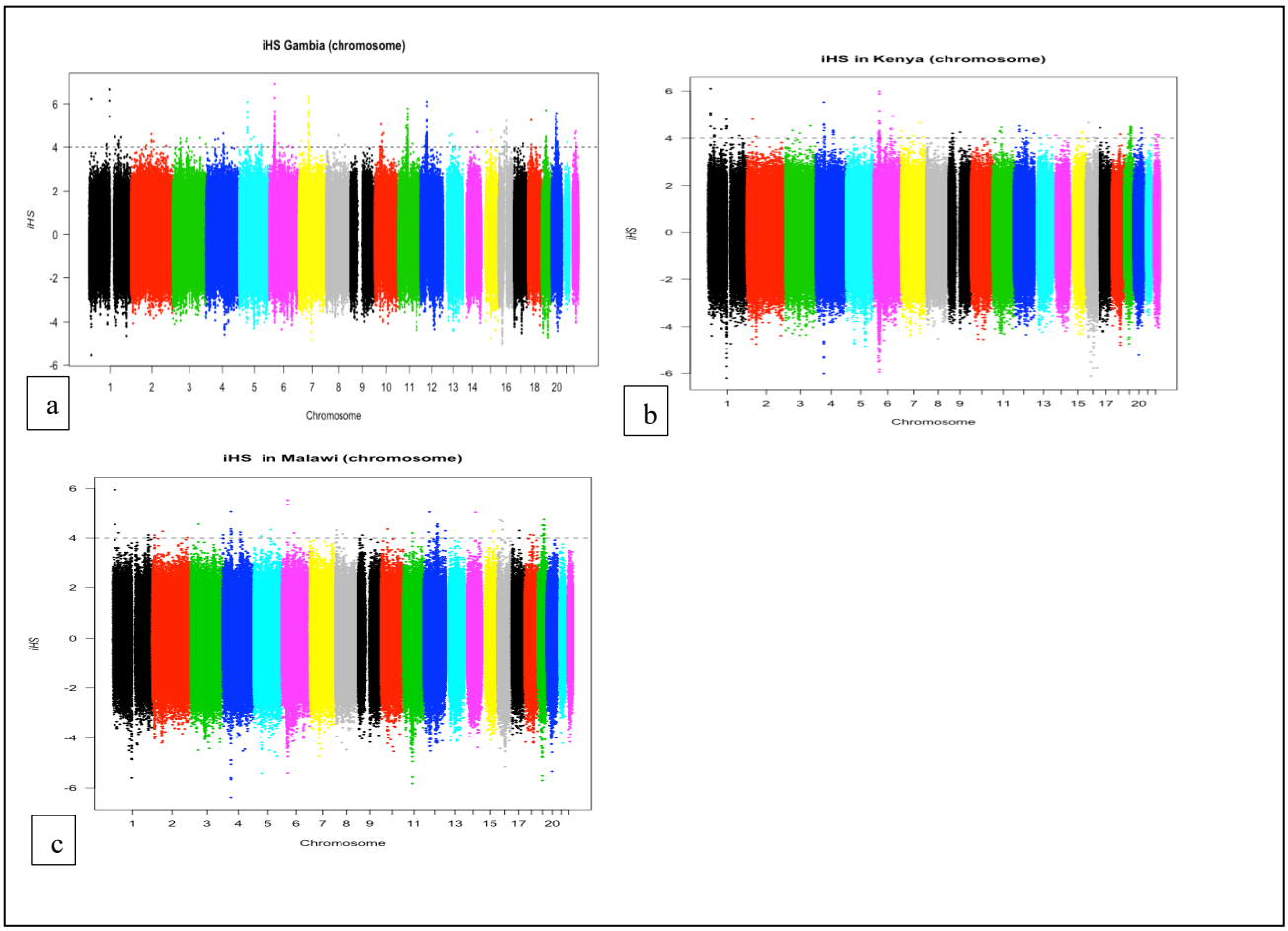

Figure 3.6 | Manhattan plot of the genome-wide distribution of iHS scores for the autosomes in the three malaria populations (Gambia (a), Kenya (b) and Malawi (c)). iHS scores greater than 4 and less than -4 are considered under positive selection.

3.4.3 Pharmacogenes in candidate iHS regions

Table 3.2 shows pharmacogenes found in the candidate iHS regions that were positively selected and previously represented by variants in the Manhattan plot with an iHS score of > 4 or < -4 for the three malaria populations. As a result, a total of seven pharmacogenes were found in the candidate iHS regions in the genomes of the three populations from malaria endemic areas (Gambia, Kenya, and Malawi). From these, four pharmacogenes, such as *CYP19A1, SLC14A2, ABCD4, and CALU* on chromosomes 15, 18, 14, and 7, respectively, were found in positively selected regions in the genome of the Gambian population. One pharmacogene, *ABCC1* on chromosome 16, was the only pharmacogene found in the Kenyan population. Two pharmacogenes, *ABCA11P* on chromosome 4 and *ABCC11* on chromosome 16, were found in the Malawi population. Genes found in the candidate iHS regions other than pharmacogenes are provided in **Annex E**.

Table 3. 2 | Pharmacogenes in candidate iHS regions in the three-malaria endemic populations (Gambia, Kenya and Malawi)

POPULATION	CHR	START	END	N_MRK	MEAN_MRK	MAX_MRK	N_EXTR_MRK	PERC_EXTR_MRK	MEAN_EXTR_MRK	GENES
Gambia	14	74600000	74800000	113	2.24832698	7.4728723	7	6.19	6.815842261	ABCD4
	7	1,28E+08	1,28E+08	114	1.485742114	6.73593313	4	3.51	6.467404811	CALU
	15	51400000	51750000	251	2.531047214	12.54386554	31	12.35	8.089890417	CYP19A1
	18	42700000	42900000	150	2.746025998	6.998441635	12	8	6.622799245	SLC14A2
Kenya	16	16200000	16350000	77	0.982031488	4.941657606	4	5.19	4.536583355	ABCC1
Malawi	4	450000	600000	17	2.25824513	4.352817323	3	17.65	4.18004572	ABCA11P
	16	48050000	48250000	73	1.722489209	4.700112881	5	6.85	4.408716573	ABCC11

3.5 Pathogenic SNPs proportion

The proportion of pathogenic SNPs in each antimalarial pharmacogenes was computed to understand the burden of their pathogenicity across global populations using ANNOVAR software (160), as described in materials and methods. As a result, the three ABC drug transporter families such as *ABCC1*, *ABCC2,* and *ABCC4* had a higher proportion of

pathogenic SNPs (45% to 100%) in populations from malaria endemic areas compared to non-malaria endemic areas (**Figure 3.7**).

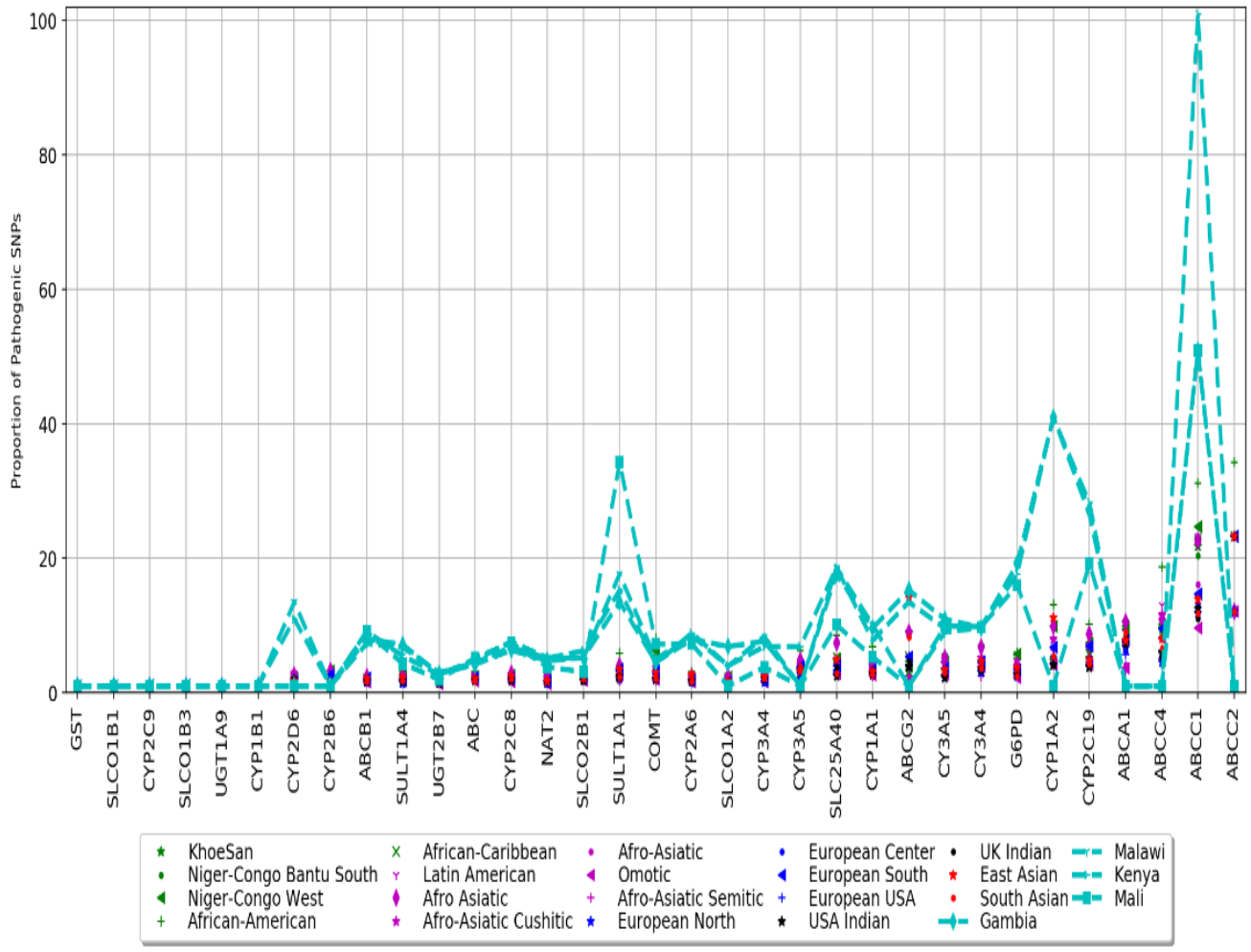

Figure 3.7 | Pathogenic SNPs proportion of antimalarial specific pharmacogenes in the four malaria populations and 20 world populations. The x-axis represents the pharmacogenes and the y-axis represent the pathogenic SNPs proportions in each pharmacogenes. The different colours represent populations.

3.6 Principal Component Analysis (PCA)

3.6.1 PCA based on non-antimalarial specific pharmacogenes variants

PCA was performed using non-antimalarial specific pharmacogenes SNPs that we obtained from databases. As a result, the PCA analysis successfully clustered the majority of the populations based on their ancestry (**Figure 3.8**).

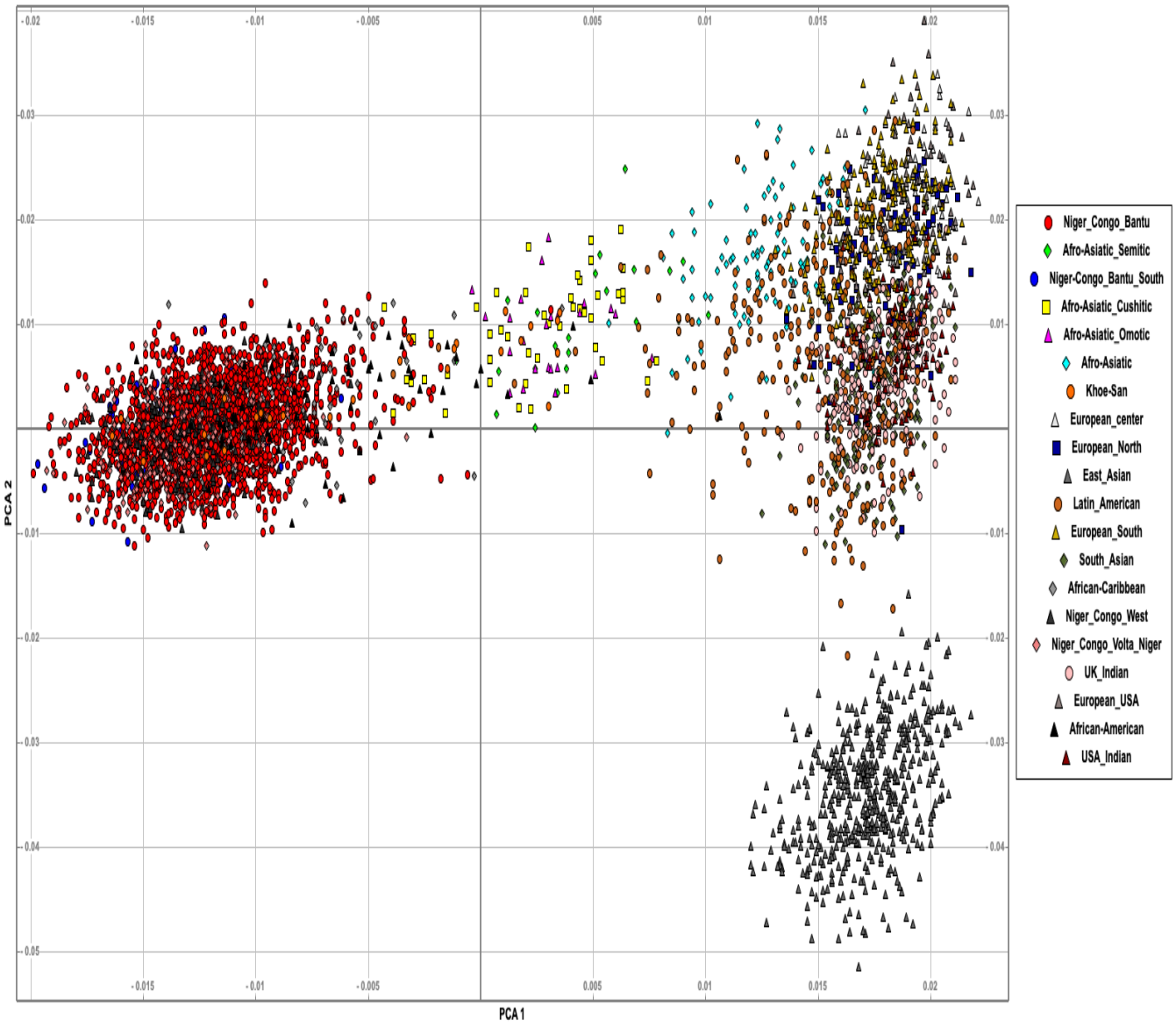

Figure 3.8 | PCA based on non-antimalarial specific pharmacogenes SNPs for the twenty populations.

However, PCA based on non-antimalarial specific pharmacogenes SNPs of the four malaria endemic populations (Gambia, Kenya, Mali, and Malawi) barely clustered the SNPs in to the four populations (**Figure 3.9**).

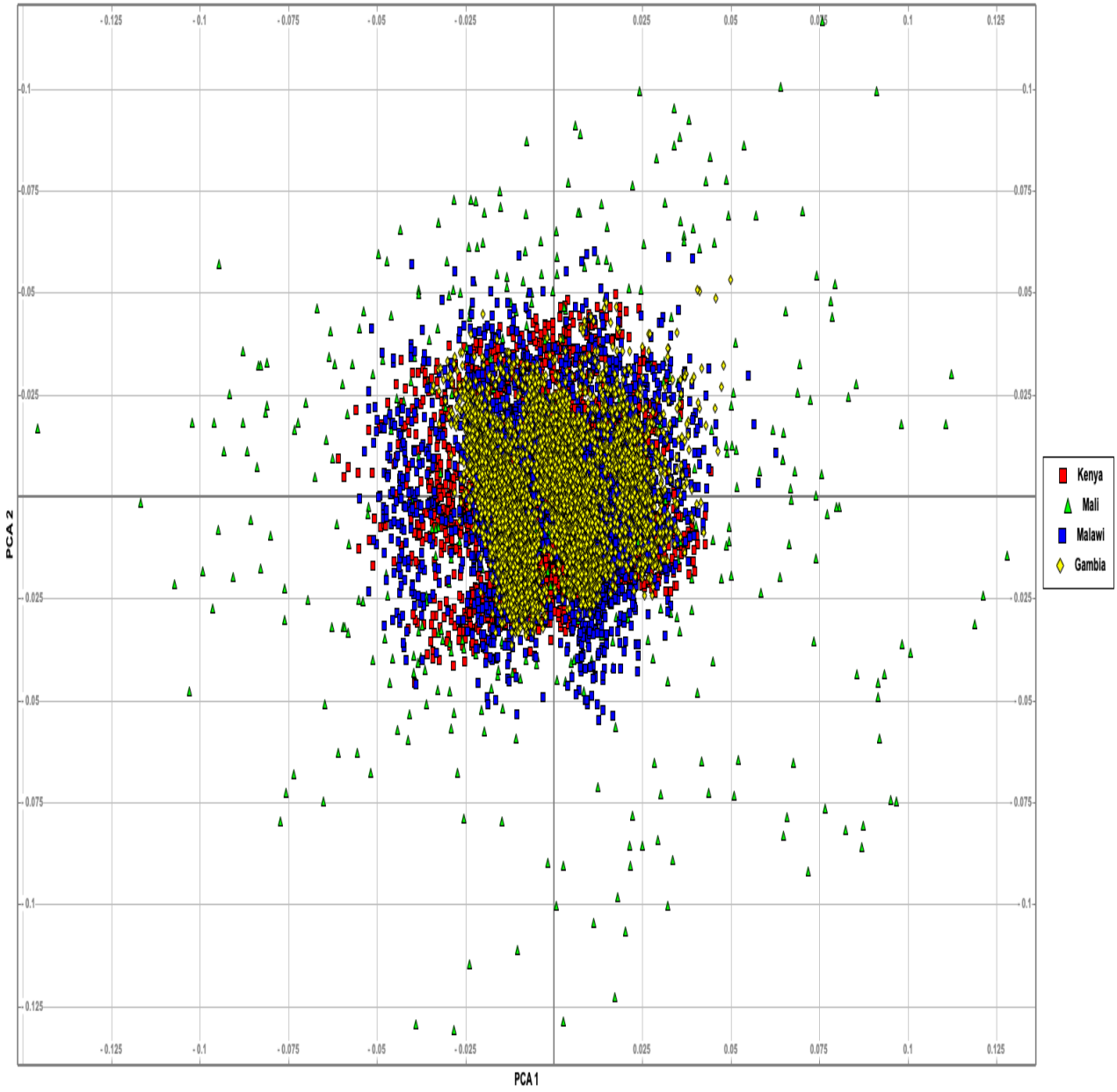

Figure 3.9 | PCA based on non-antimalarial specific pharmacogenes SNPs for the four malaria endemic populations (Gambia, Kenya, Mali and Malawi).

3.6.2 PCA based on antimalarial specific pharmacogenes variants

Further analysis of PCA was performed based on antimalarial specific pharmacogenes SNPs, to assess the differentiation levels of these SNPs across populations. Like the non-antimalarial specific pharmacogenes SNPs, the PCA analysis effectively clustered the majority of the populations according to their ancestry (**Figure 3.10**).

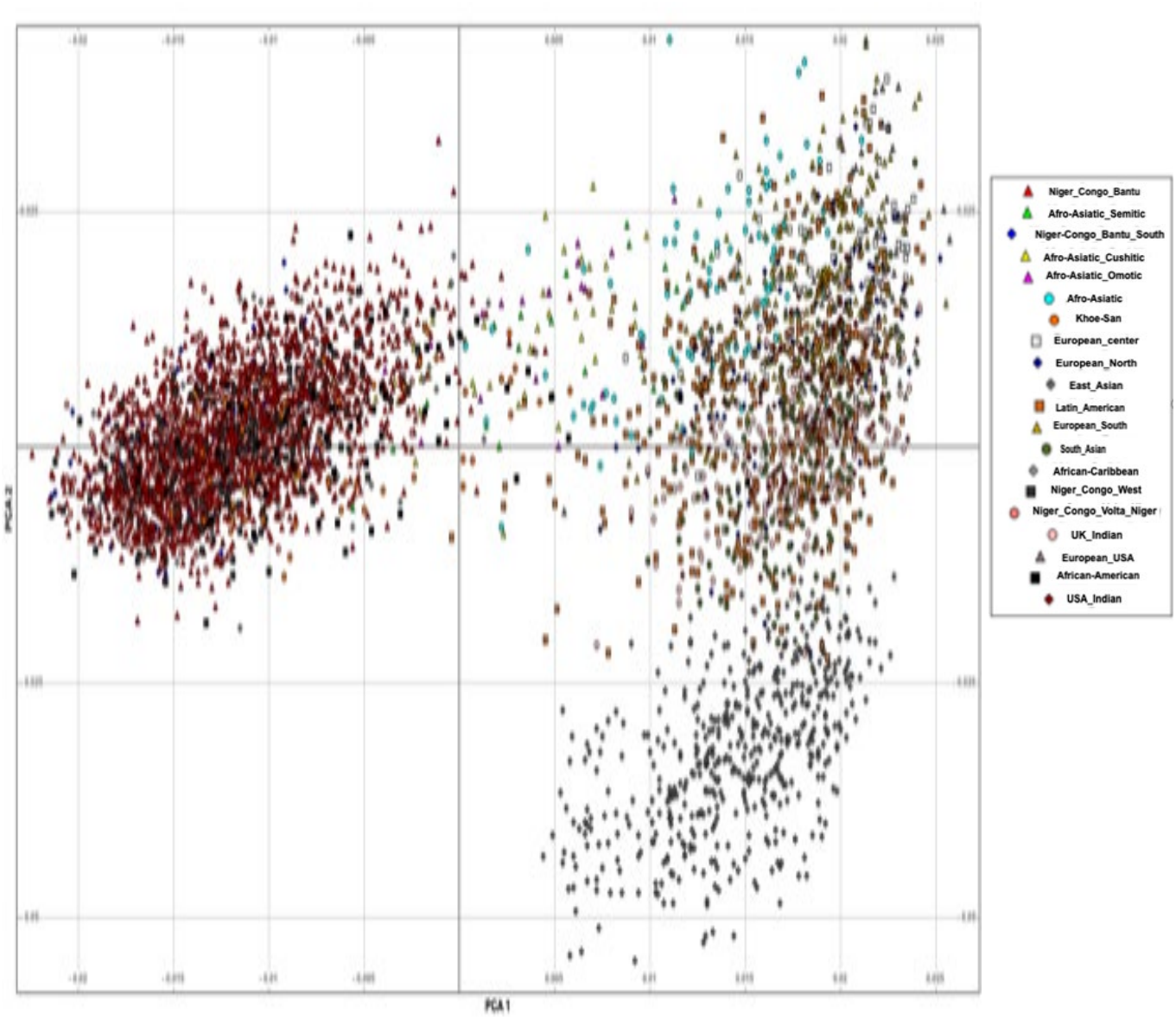

Figure 3.10 | PCA based on antimalarial specific pharmacogenes SNPs of the twenty populations

However, PCA based on antimalarial specific pharmacogenes SNPs of the three malaria endemic populations (Gambia, Kenya, and Malawi) barely clustered the SNPs in to the three populations (**Figure 3.11**).

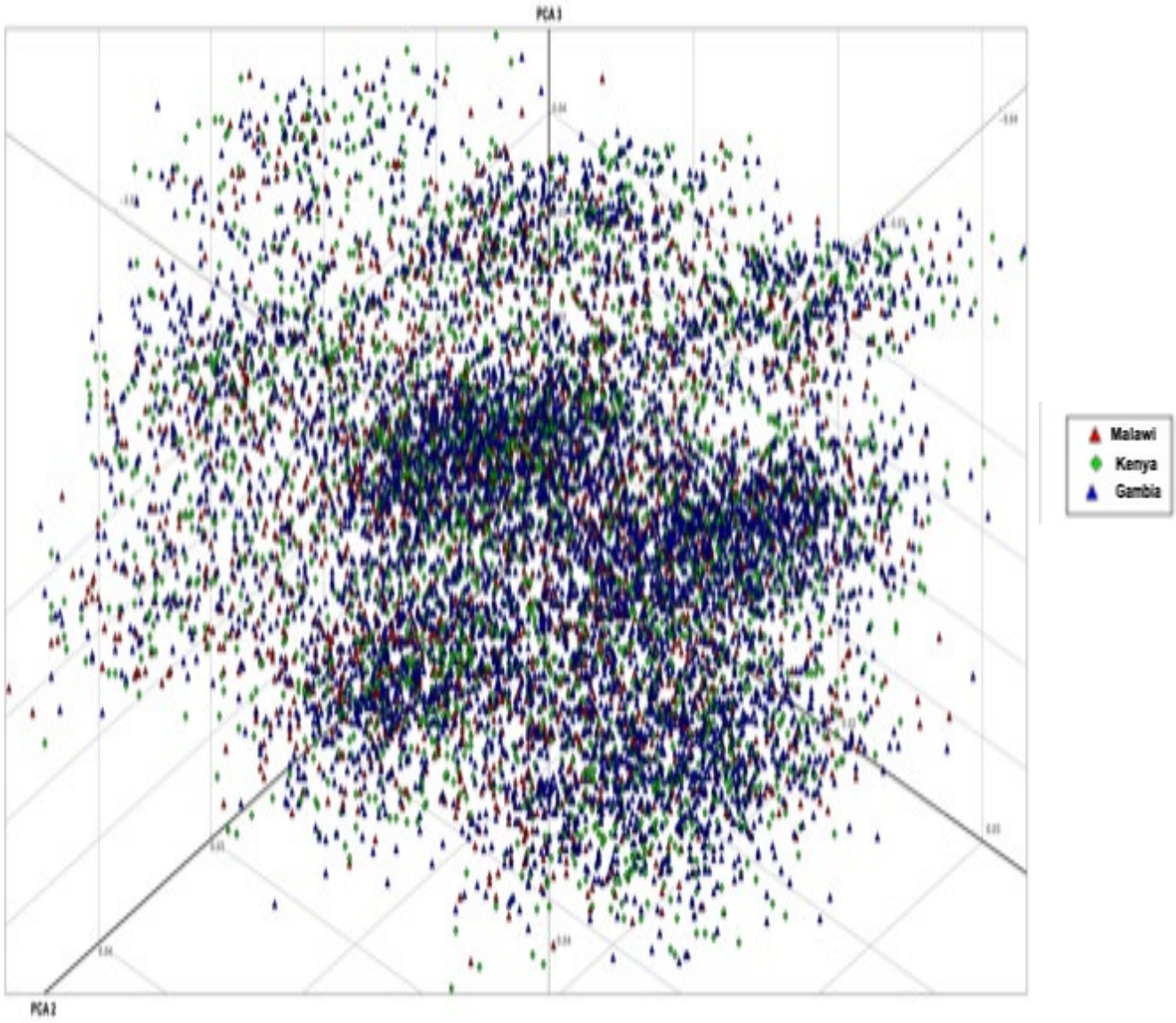

Figure 3.11 | PCA based on antimalarial specific pharmacogenes variants of the three malaria endemic populations (Malawi, Kenya, and Gambia)

3.7 Network analysis based on antimalarial specific pharmacogenes

The network analysis based on 29 antimalarials specific pharmacogenes shows (**Figure 3.12**) possible interactions between the pharmacogenes, and 20 more extra genes were found in the network. The interactions are based on co-expression, shared protein domains, physical interactions, co-localization, and functions**.**

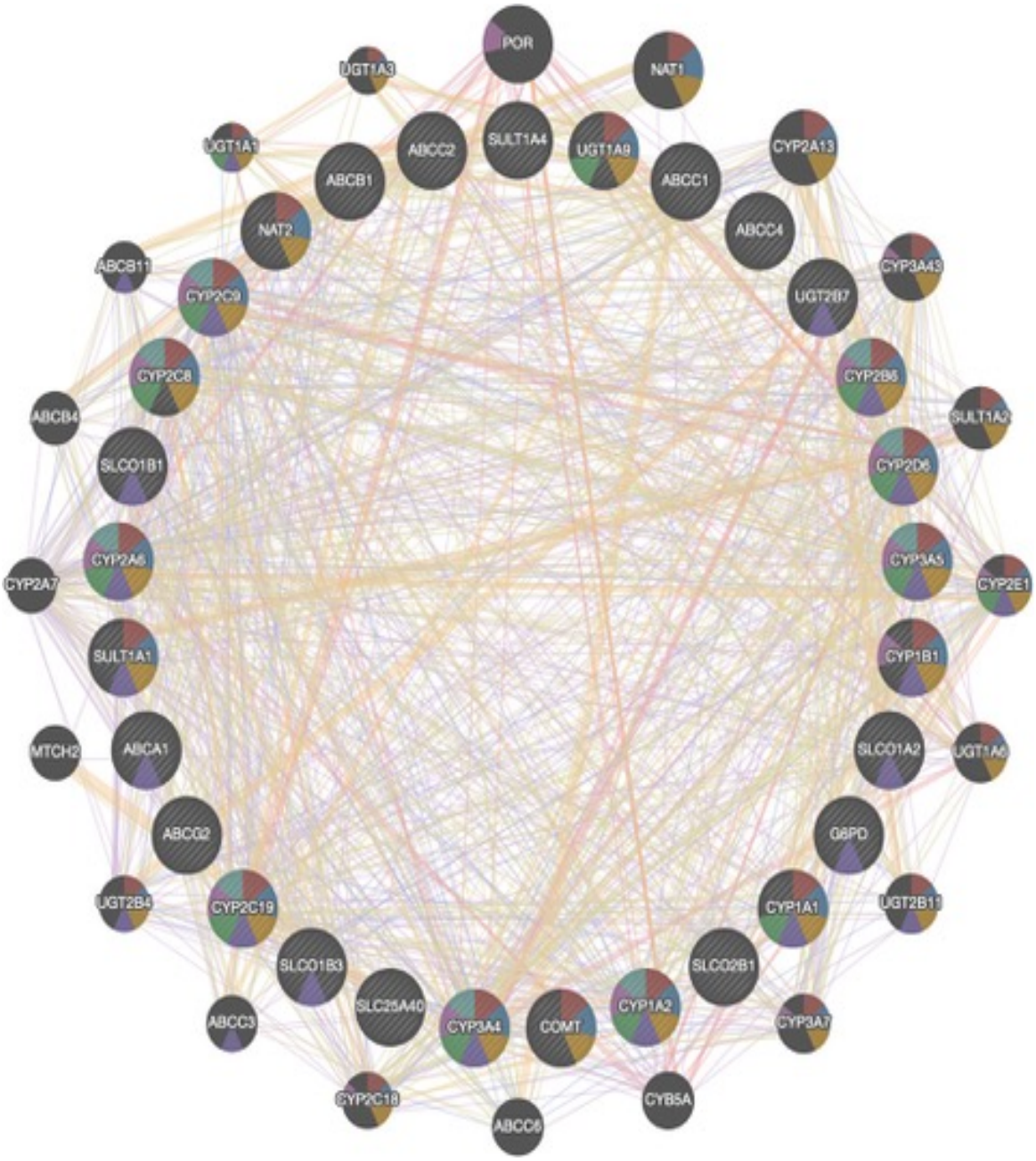

Figure 3.12 | Network analysis based on antimalarial specific pharmacogenes using geneMANIA software.

3.8 Characterization of seven *CYP2C8* SNPs using Sanger sequencing

3.8.1 Quality check for gDNA

NanoDrop spectrometer measurements of recruited samples were performed to check the quality and concentration of genomic DNA (gDNA) as described in the materials and methods. According to the measurements, samples having a 260/280 absorbance ratio of approximately 1.8 were taken as a good quality gDNA (**Figure 3.13**).

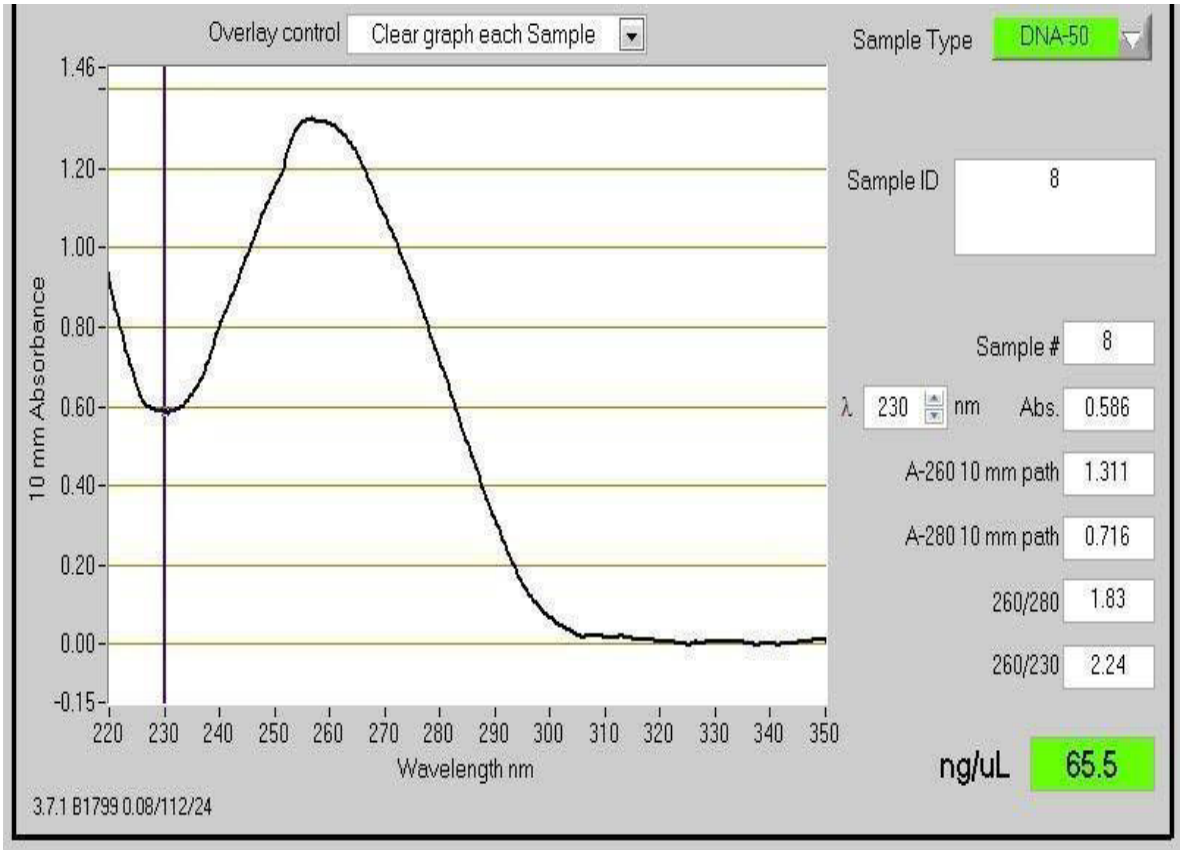

Figure 3.13 | represents of a good quality gDNA using NanoDrop spectrophotometry. The 260/280 ratio shows the quality of gDNA. The 65.5 ng/ul shows the concentration of the gDNA**.**

3.8.2 gDNA integrity test

In addition to the NanoDrop spectrometer measurement, to check the intactness of gDNA, an integrity test was also performed using 1% agarose gel, as described in materials and methods (**Figure 3.14**). In this regard, samples with intact gDNA remained on the top of the gel (e.g., sample ID 4146), while samples with degraded DNA formed a smear on the gel (e.g., sample ID 4109).

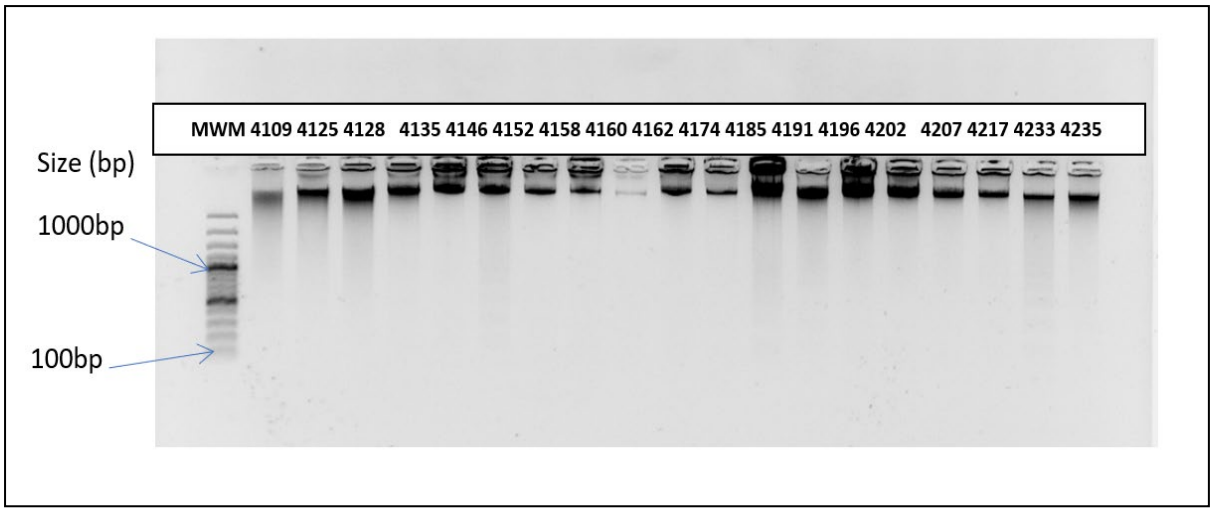

Figure 3.14 | Evaluation of DNA integrity using a 1% agarose gel. The 100bp molecular weight marker (MWM) was used and lanes 1 to 19 indicate variation in DNA concentration and quality of the DNA samples from 19 individuals**.**

3.8.3 Agarose gel electrophoresis

To check for the presence of targeted PCR products, we performed a 1.5% agarose gel electrophoresis, as described in the materials and methods. A typical example of a gel result obtained is depicted in (**Figure 3.15**). The figure shows that almost all the samples with the given sample IDs had PCR products with bp lengths of 798.

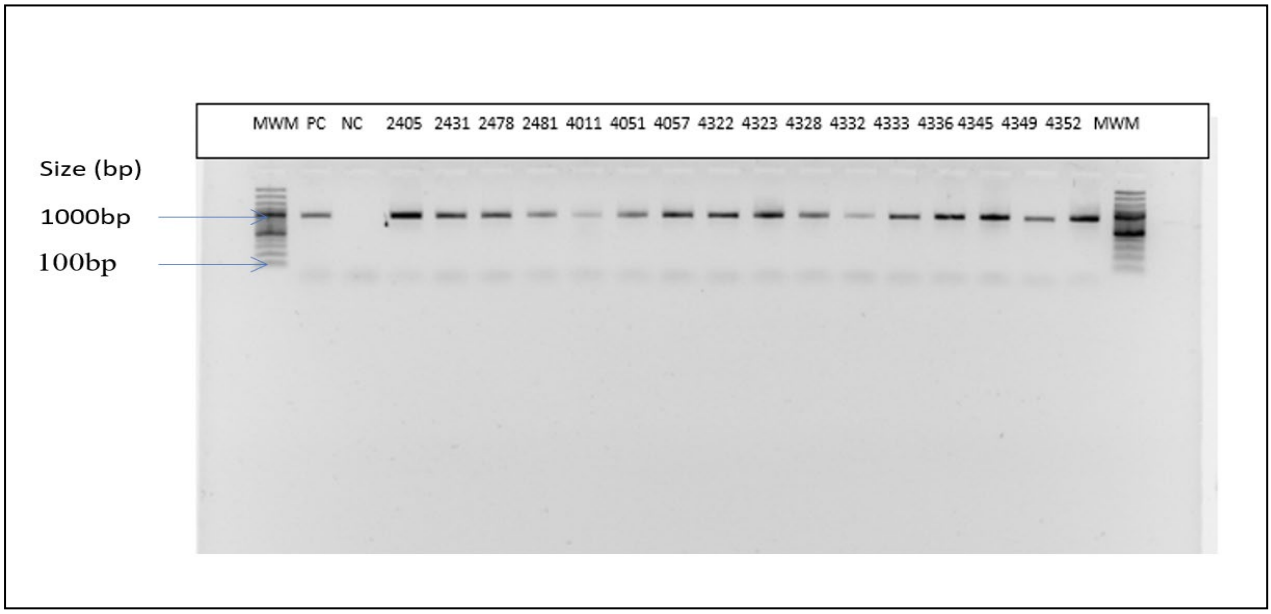

Figure 3.15 | represents result for agarose gel electrophoresis of PCR products. 1.5% agarose gel was used. The targeted PCR product was 798bp of nucleotides of *CYP2C8* and 100bp MWM was used.

3.8.4 Chromatogram of Sanger sequencing

The analysis of capillary electrophoresis results from the ABI 3130 DNA Analyzer was performed using ChromasPro version 2.0.1 sequence assembly software. **Figure 3.16** illustrates a typical example of a chromatogram of Sanger sequencing after analysis using ChromasPro software. The highlighted box indicates genotypes of the targeted SNP, and the colours indicate different nucleotides. Therefore, heterozygous alleles have two peaks with different colours that appear together (e.g., sample IDs 4260 and 4314), while homozygous alleles have a single peak with a colour (e.g., sample ID 4243).

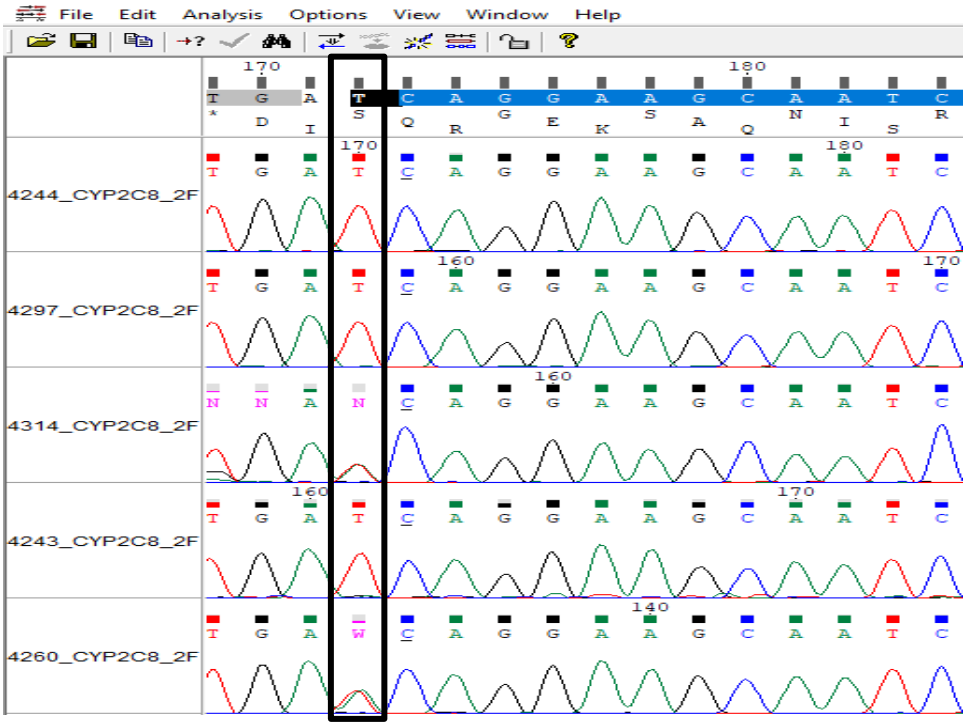

Figure 3.16 | represents a chromatogram of Sanger sequencing after analysis using ChromasPro sequence assembly software. The highlighted box indicates the genotype for the targeted SNP of interest. The different colours indicate the different types of nucleotides (red = Thymine (T); blue = Cytosine (C); black = Guanine (G); green=Adenine (A)). Heterozygous alleles have two peaks with different colours that appear together (e.g., sample IDs 4260 and 4314), while homozygous alleles have a single peak with a single colour (e.g., sample ID 4243).

3.8.5 Genotype and allele frequencies of *CYP2C8* polymorphisms

The genotype distributions of the seven *CYP2C8* SNPs were checked for HWE. **Tables 3.3** and **3.4** show the genotype and allele frequencies of the genetic polymorphisms in *CYP2C8* in South Africans, respectively. We found three variants with high MAF, which include *rs11572101A>G* (0.19), *rs11572103T>A* (0.16) and *rs1926705C>T* (0.14) (**Table 3.4**).

Table 3. 3 | Genotype frequency of genetic polymorphisms in the *CYP2C8*

rsID	Type of variant	Nucleotide change	Genotype: frequency (count)
rs11572105	Intron variant	*G>T*	GG: 0.98 (146) GT: 0.02 (3) TT: 0.00 (0)
*rs11572103 (*2)*	Missense variant	*T>A*	TT: 0.70 (105) TA: 0.28 (42) AA: 0.02 (3)
*rs1058930 (*4)*	Missense variant	*G>C*	GG: 0.97 (146) GC: 0.03 (4) CC: 0.00 (0)
*rs188934928 (*14)*	Missense variant	*C>G*	CC: 1.00 (150) CG: 0.00 (0) GG: 0.00 (0)
rs11572101	Intron variant	*A>G*	AA: 0.66 (99) AG: 0.29 (44) GG: 0.05 (7)
rs11572100	Intron variant	*T>C*	TT: 0.82 (123) TC: 0.17 (26) CC: 0.01 (1)
rs1926705	Intron variant	*C>T*	CC: 0.03 (4) CT: 0.23 (34) TT: 0.74 (109)

Table 3. 4 | Allele frequency of genetic polymorphisms in the *CYP2C8*

rsID	Minor allele	Alleles: frequency (count)
rs11572105	T	G: 0.99 (295) T: 0.01 (3)
*rs11572103 (*2)*	A	T: 0.84 (252) A: **0.16** (48)
*rs1058930 (*4)*	C	G: 0.99 (296) C: 0.01 (4)
*rs188934928 (*14)*	G	C: 1 (300) G: 0.00 (0)
rs11572101	G	A: 0.81 (242) G: **0.19** (58)
rs11572100	C	T: 0.91 (272) C: 0.09 (28)
rs1926705	C	C: **0.14** (42) T: 0.86 (252)

Numbers in bold represent alleles with high frequency

3.9 Comparison of variant allele frequencies of the seven *CYP2C8* SNPs with world populations

Genetic variants that are relevant pharmacogenetically have different distributions among different populations. The frequencies of the seven *CYP2C8* SNPs genotyped in South African population groups, *rs11572105G>T, rs11572103T>A, rs1058930G>C, rs188934928C>G, rs11572101A>G, rs11572100T>C, and rs1926705C>T,* were compared with world populations that include West Africans (e.g., Yoruba in Ibadan, Nigeria (YRI), East Africans (e.g., Luhya in Webuye, Kenya (LWK), African Americans, East Asians and Europeans. As a result, the four variants (*rs11572103T>A*, *rs11572101A>G, rs11572100T>C, and rs1926705C>T)* showed a statistically significant difference (p-value $\leq$ 0.05) in allele frequencies when compared with non-African populations (**Table 3.5**). Exceptionally, however, the variant allele frequency of *rs11572103T>A* variant in African American populations was statistically significant when compared with our study population.

Table 3. 5 | Comparison of variant allele frequencies of the seven *CYP2C8* SNPs from this study with world population from 1000 Genome project

	rs11572105G>T	***rs11572103T>A***	***rs1058930G>C***	***rs188934928 C>G***	***rs1157210 1A>G***	***rs1157210 0T>C***	***rs1926705C>T***
South Africans (this study)	0.01	0.16	0.01	0.00	0.19	0.09	0.86
West Africans (YRI)	0.04	0.20	0.005	0.00	0.19	0.11	0.88
East Africans (LWK)	0.05	0.14	0.00	0.00	0.17	0.14	0.87
African Americans[a]	0.04	0.07*	0.02	0.00	0.19	0.09	0.80
East Asians	0.00	0.00*	0.00	0.00	0.38*	0.00*	0.45*
Europeans	0.04	0.00*	0.06	0.00	0.19	0.04	0.69*

YRI, Yoruba in Ibadan, Nigeria; LWK Luhya in Webuye, Kenya

* Statistically significant differences when compared with South African populations

[a]Adapted from Sarudzai et al. (163)

CHAPTER 4: DISCUSSION

Genetic variations are responsible for a considerable amount of inter-individual and inter-ethnic heterogeneity in drug response (164). However, the development of clinical tests for most drugs currently in use was performed in cohorts accounting for a few limited ethnic diversities, of which most are from European and Asian ancestry. To the best of our knowledge, this is the first study to examine profiles of pharmacogenes variants in malaria endemic areas of heterogeneous African populations and global populations of various ethnic groups. Understanding the profiles of pharmacogenes variants will be important in guiding the extrapolation of antimalarial drug usage in diverse ethnicities like Africa and the rational design of new therapeutics.

In this study, four genome-wide datasets of populations from malaria endemic areas in Africa (i.e., Gambia, Kenya, Mali, and Malawi) and datasets from twenty other ethnic groups comprising different global populations were analysed. We assessed if there were differences in the profiles of pharmacogenes variants in populations from malaria endemic areas compared to those from non-malaria endemic areas. The proportion of rare (frequency < 0.01) variants was generally higher ($> 80\%$) in the populations from non-malaria endemic areas, mainly Europeans, compared to the four African populations from malaria endemic areas. However, a comparison of African populations from malaria endemic areas versus non-malaria endemic areas showed a negligible difference in the proportion of rare variants. This is in agreement with a previous finding by Delesa *et al.,* (165). The high proportion of rare variants in Europeans is expected because most of the genetic data generated using high-throughput sequencing technologies is from European populations (166). However, African populations are expected to have a much higher proportion of rare variants compared to European or Asian populations, in general (167). This indicates that the genetic data for African populations is underrepresented, so there is a need for more sequencing of a diverse cohort of African genomes.

Unlike the rare variants, the proportion of common variants (frequency > 0.05) of non-antimalarials specific pharmacogenes was generally higher in the four African populations from malaria endemic areas compared with other African populations from malaria non-endemic areas. Interestingly, the finding was also the same when using antimalarials specific

pharmacogenes. The possible explanation for this finding might be that most genetic studies are based on a candidate-gene approach, which is based on a particular gene and it's variants have an effect on disease pathogenesis or drug pharmacogenetics (113,168,169). Nevertheless, the high proportion of common pharmacogenes variants is a good indication that it would be necessary to study more about the pharmacogenetics of antimalarial drugs in populations in malaria endemic areas.

We further analysed allele frequencies of pharmacogenes and their distribution to identify the types of pharmacogenes with the highest number of variants in malaria endemic areas using non antimalarial-specific pharmacogenes. The analysis revealed that *CYP2A13* and *CYP2F1* had the highest proportion of variants in the four malaria endemic populations compared to the global populations. The reason for such an observation could be the high level of genetic diversity in the African population that is caused by its complex population history and the dramatic variation in climate, diet, and exposure to infectious disease (170). Since CYPs are involved in the phase I metabolism of various xenobiotics, including antimalarial drugs (100), the high frequency of SNPs in these pharmacogenes in the malaria endemic populations may play a pivotal role in the response to different antimalarials that are currently in use.

Similarly, based on analysis using antimalarial-specific pharmacogenes, a higher proportion of pharmacogenes variants was found in the four malaria endemic populations compared to the global populations. But, here, we found that *UGT1A9,* one of the phase II enzymes, had a higher proportion of gene specific allele frequencies in the four malaria populations. *UGT1A9* is involved in the metabolism of DHA, which is the active ingredient in ACT drugs (119). As a result, the high proportion of *UGT1A9* variants may be relevant, given that ACTs are the first line drugs for malaria treatment in almost all malaria endemic areas (61).

From an evolutionary perspective, SNPs are often used to detect genomic areas that have been subjected to recent natural or artificial selection (162). Therefore, using genome-wide SNPs of genotyped data of populations from malaria endemic areas (Gambia, Kenya, and Malawi), selection analysis was performed to check if natural selection is attributing to the

high proportion of pharmacogenes variants in the populations. As a result, pharmacogenes, such as *CYP19A1, SLC14A2, ABCD4, and CALU* in the Gambian population, *ABCC1* in the Kenyan population, and *ABCA11P* and *ABCC11* in the Malawi population, were found to be positively selected in the genome of each population. Genes involved in drug pharmacokinetics are primarily involved in xenobiotic defence, which could explain the enrichment of selection signals in the pharmacogenes (171).

The proportion of pathogenic SNPs was higher in the four malaria endemic populations compared to the global populations in which the ABC drug transporter families, such as *ABCC1*, *ABCC2,* and *ABCC4*, pose the greatest percent, which ranges from 40% to 100%. This might be caused by the failure of the efflux drug transporters to effectively transport antimalarials prescribed to the patients with malaria, which may result in minimal exposure of the *Plasmodium* parasites to antimalarial drugs in the RBCs and result in the development of parasite drug resistance (79). *Plasmodium* resistance to antimalarial drugs may increase their propagation in the human host, then eventually increase their circulation in malaria endemic areas.

The PCA analysis based on antimalarial and non-antimalarial specific pharmacogenes SNPs effectively clustered the majority of the populations according to their ancestry. This further confirms a previous finding by Delser and Fuselli (172), which showed population level differentiation based on pharmacogenes variants. This finding suggests that the SNPs residing in the pharmacogenes are differentiated across geographical locations and ethnic backgrounds. Moreover, this is a good indication that pharmacogenetic studies may need to be done in different world populations due to the fact that there is population level differentiation based on the pharmacogenes SNPs.

However, PCA analysis based on antimalarial and non-antimalarial specific pharmacogenes SNPs, respectively, for the four malaria endemic populations (Gambia, Kenya, Mali, and Malawi) barely clustered the SNPs for each population. This is in agreement with a previous finding by Delesa *et al.,*(165). The overlap in the clustering of the populations is likely because of the similar malaria endemicity in these populations. The Mali dataset, however, was poorly clustered, and that could be caused by the poor quality of the data.

CYP2C8 is one of the most important CYPs involved in antimalarial drug metabolism (87,89). In this study, a total of seven *CYP2C8* SNPs (*rs1926705, rs11572100, rs11572101, rs188934928, rs1058930, rs11572103,* and *rs11572105),* were characterized in the South African population group. The SNPs were targeted due to their proximity to the African-specific variant (*rs11572103T>A (*2)*). From these, four of them (*rs11572105G>T, rs11572101A>G, rs11572100T>C,* and *rs1926705C>T)* are intron variants, and three of them *(rs11572103T>A (*2), rs1058930G>C, rs188934928 C>G (*14))* are missense variants. However, a statistically significant (p-value ≤ 0.05) difference in allele frequencies was found mainly for the four variants (*rs11572103T>A (*2), rs11572101A>G, rs11572100T>C,* and *rs1926705C>T*) when compared with non-African populations. This is consistent with a previous finding by Sarudzai et al. (163). A possible explanation for such differences in allele frequency could be the interethnic differences between the populations. None of the variants shows a statistically significant difference in allele frequencies when compared with other African populations, except for the African Americans for *rs11572103T>A (*2)* variant. The high prevalence of *CYP2C8* variants with functional significance is an indication for future consideration of pharmacogenetic testing before using antimalarial drugs that are metabolized by CYP2C8, both for the African populations who are living in malaria endemic areas such as YRI and LWK and for those traveling to malaria endemic areas from non-endemic areas such as South Africa.

However, the current study has inherent limitations. In this study, we only focused on SNPs that might influence various antimalarials responses. But, there are other factors that may also influence the variability of a drug's response, including diet, exposure to chemicals in the environment, a person's disease condition, etc. (173,174). Epigenetic regulation of ADME genes and drug targets may also be a key predictor of antimalarial drug responses.

CHAPTER 5: CONCLUTIONS AND PERSPECTIVES

The findings from this study revealed the profiles of pharmacogenes variants in populations from malaria endemic and non-endemic geographical regions. High proportions of variants in *CYP2A13*, *CYP2F1,* and *UGT1A9* were found in populations from malaria-endemic areas compared to non-malaria endemic areas. This suggests that it may be necessary to consider pharmacogenetic testing for patients based on antimalarial pharmacogenes having a higher proportion of variants in malaria endemic populations (e.g., *UGT1A9*). A high frequency in the functional variant of *CYP2C8*, *rs11572103T>A (*2),* was found in the South African population. This may also have relevance for future consideration of pharmacogenetic testing for a person traveling from non-endemic areas, such as South Africa, to malaria endemic areas. Future studies should be done using large scale sequencing datasets from malaria endemic areas to assess the effect of common antimalarial pharmacogenes variants on the safety and efficacy of currently used antimalarial drugs.

REFERENCES

1. WHO. malaria fact-sheets [Internet]. 2021. [cited 2022 Mar 17]. Available from: https://www.who.int/news-room/fact-sheets/detail/malaria
2. Ayong L, Moukoko CEE, Mbacham WF. Diagnosing malaria: Methods, tools, and field applicability. Methods Mol Biol. 2019;2013:73–82.
3. Tse EG, Korsik M, Todd MH. The past, present and future of anti-malarial medicines. Malar J. 2019;18(1):1–21.
4. WHO. World malaria report 2021. World Heal Organazation, Geneva. 2021
5. CDC. Where malaria occurs [Internet]. [cited 2020 Oct 4]. Available from: https://www.cdc.gov/malaria/about/distribution.html.Accessed 4th october 2020.
6. Oguike MC, Betson M, Burke M, Nolder D, Stothard JR, Kleinschmidt I, et al. Plasmodium ovale curtisi and Plasmodium ovale wallikeri circulate simultaneously in African communities. Int J Parasitol. 2011;41(6):677–83.
7. Müller IB, Hyde JE. Antimalarial drugs: modes of action and mechanisms of parasite resistance. Future Microbiol. 2010;5(12):1857–73.
8. Crabb BS, Cowman AF. Plasmodium falciparum virulence determinants unveiled. Genome Biol. 2002;3(11):3–6.
9. Moxon CA, Gibbins MP, McGuinness D, Milner DA Jr, Marti M. New Insights into Malaria Pathogenesis. 2019;12(39):315–43.
10. Elewa H, Wilby KJ. A Review of Pharmacogenetics of Antimalarials and Associated Clinical Implications. Eur J Drug Metab Pharmacokinet. 2017;42(5):745–56.
11. Global Malaria Programme: WHO Global. World malaria report 2019. WHO Regional Office for Africa. 2019.
12. Gerstenlauer C. Recognition and Management of Malaria. Nurs Clin NA. 2019; 54(2):245-260.
13. White NJ. Plasmodium knowlesi : The Fifth Human Malaria Parasite. 2008;46:172–3.
14. Mzilahowa T, Hastings IM, Molyneux ME, Mccall PJ. Entomological indices of malaria transmission in Chikhwawa district, Southern Malawi. Malar J. 2012;11:1–9.
15. Bartoloni A, Zammarchi L. Clinical Aspects of f Uncomplicated and Severe Malaria. 2012;4.
16. Cowman AF, Healer J, Marapana D, Marsh K. Malaria: Biology and Disease. Cell. 2016;167(3):610–24.
17. Celli S, Shorte S, Amino R, Thiberge S, Frischknecht F, Me R. Quantitative imaging

of Plasmodium transmission from mosquito to mammal. Nat Genet. 2006;12(2):1–5.
18. Klein EY. Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread. Int J Antimicrob Agents. 2013;41(4):311–7.
19. Weiss GE, Gilson PR, Taechalertpaisarn T, Tham W, Jong NWM De, Harvey KL, et al. Revealing the Sequence and Resulting Cellular Morphology of Receptor-Ligand Interactions during Plasmodium falciparum Invasion of Erythrocytes. Plos Pathog. 2015;1–26.
20. Holder AA. Proteins on the surface of the malaria parasite and cell invasion. Parasitology. 1994;108:S5–18.
21. Lin CS, Uboldi AD, Epp C, Bujard H, Tsuboi T, Czabotar PE, et al. Multiple plasmodium falciparum merozoite surface protein 1 complexes mediate merozoite binding to human erythrocytes. J Biol Chem. 2016;291(14):7703–15.
22. Basu S, Sahi PK. Malaria: An Update. Indian J Pediatr. 2017;84(7):521–8.
23. Talman AM, Domarle O, Mckenzie FE, Ariey F, Robert V. Gametocytogenesis : the puberty of Plasmodium falciparum. 2004;3(24):1–14.
24. Malaria Vaccine Initiative. Available from: https://www.malariavaccine.org/malaria-and-vaccines/vaccine-development/life-cycle-malaria-parasite. Accessed 4th october 2020.
25. Touray BMG, Warburg A, Laughinghouse A, Krettli AU, Miller LH. Developmentally Regulated Infectivity of Malaria Sporozoites for Mosquito Salivary Glands and the Vertebrate Host. J Exp Med. 1992;175:1607–12.
26. CDC. Disease [Internet]. Available from: https://www.cdc.gov/malaria/about/disease.html.Accessed 04 october 2020.
27. Herraiz T, Guillén H, González-peña D, Arán VJ. Antimalarial Quinoline Drugs Inhibit β -Hematin and Increase Free Hemin Catalyzing Peroxidative Reactions and Inhibition of Cysteine Proteases. 2019; 9(1):15398:1–16.
28. Tangpukdee N, Duangdee C, Wilairatana P, Krudsood S. Malaria Diagnosis : A Brief Review. Korean J Parasitolgy. 2009;47(2):93–102.
29. Lillie PJ, Duncan CJA, Sheehy SH, Meyer J, Hara GAO, Gilbert SC, et al. Distinguishing malaria and influenza : Early clinical features in controlled human experimental infection studies. Travel Med Infect Dis. 2012;10(4):192–6.
30. Gitta B, Kilian N. Diagnosis of Malaria Parasites Plasmodium spp. in Endemic Areas: Current Strategies for an Ancient Disease. BioEssays. 2020;42(1):1–12.
31. WHO. How malaria RDTs work [Internet]. Available from:

https://www.who.int/malaria/areas/diagnosis/rapid-diagnostic-tests/about-rdt/en/. Accessed 4th october 2020.

32. Ali IM, Bigoga JD, Forsah DA, Cho-Ngwa F, Tchinda V, Moor VA, et al. Field evaluation of the 22 rapid diagnostic tests for community management of malaria with artemisinin combination therapy in Cameroon. Malar J. 2016;15(1):1–7.
33. Hendriksen ICE, Mtove G, Pedro AJ, Gomes E, Silamut K, Lee SJ, et al. Evaluation of a PfHRP2 and a pLDH-based rapid diagnostic test for the diagnosis of severe malaria in 2 populations of african children. Clin Infect Dis. 2011;52(9):1100–7.
34. Britton S, Cheng Q, Mccarthy JS. Novel molecular diagnostic tools for malaria elimination : a review of options from the point of view of high-throughput and applicability in resource limited settings. Malar J. 2016;15(88):1–8.
35. B. Morassin, R. Fabre, A. Berry J-FM. one year's experience with the polymerase chain reaction as a routine method for the diagnosis of imported malaria. 2002;66(5):503–8.
36. Johnston SP, Pieniazek NJ, Maniphet V, Slemenda SB, Wilkins PP, Silva AJ, et al. PCR as a Confirmatory Technique for Laboratory Diagnosis of Malaria PCR as a Confirmatory Technique for Laboratory Diagnosis of Malaria. J Clin Microbiol. 2006;44(3):1087–1089.
37. Chander Y, Koelbl J, Puckett J, Moser MJ, Klingele AJ, Liles MR, et al. A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). Front Microbiol. 2014;5(395):1–12.
38. Haldar K, Bhattacharjee S, Safeukui I. Drug resistance in Plasmodium. Nat Rev Microbiol. 2018;16(3):156–70.
39. Gaillard T, Madamet M, Pradines B. Tetracyclines in malaria. Malar J. 2015;14(1):1–10.
40. Blasco B, Leroy D, Fidock DA. Parasite Biology To the Clinic. Nat Med. 2017;23(8):917–28.
41. Onyesom I, Adu F. Phyllanthus amarus possesses malarial curative and pancreatic tonic potentials in experimental mice. J Chem Pharm Res. 2015;7(5):7–15.
42. Grellier P, Abdoulaye A, Moussa I, Ousmane A, Berry A, Ikhiri K, et al. In vitro and in vivo Antiplasmodial Activity of Momordica balsamina Alone. Chemotherapy. 2006;9:288–92.
43. Thomford NE, Dzobo K, Chopera D, Wonkam A. Pharmacogenomics Implications of Using Herbal Medicinal. Pharmaciticals. 2015;8:637–63.

44. Achan J, Talisuna AO, Erhart A, Yeka A, Tibenderana JK, Baliraine FN, et al. Quinine, an old anti-malarial drug in a modern world : role in the treatment of malaria. Malar J. 2011;10(144):1–12.
45. Kumar S, Bawa S, Gupta H. Biological Activities of Quinoline Derivatives. Mini-Reviews Med Chem. 2009;9:1648–54.
46. Henry C. OU, Sarah Keating PW et al. Quinoline Ring Derivatives Protect Against Aminoglycoside-Induced Hair Cell Death in the Zebrafish Lateral Line Henry. J Assoc Res Otolaryngol. 2012;13:759–70.
47. Fitch CD. Ferriprotoporphyrin IX , phospholipids, and the antimalarial actions of quinoline drugs. Life Sci. 2004;74:1957–72.
48. Foley M, Tilley L. Quinoline Antimalarials : Mechanisms of Action and Resistance. Int J Parasitol. 1997;27(2):231–40.
49. Kapishnikov S, Staalsø T, Yang Y, Lee J, Pérez-Berná AJ, Pereiro E, et al. Mode of action of quinoline antimalarial drugs in red blood cells infected by Plasmodium falciparum revealed in vivo. Proc Natl Acad Sci U S A. 2019;116(46):22946–52.
50. Meara WPO, Mangeni JN, Steketee R, Greenwood B. Changes in the burden of malaria in sub-Saharan Africa. Lancet Infect Dis. 2010;10(8):545–55.
51. Nzila A. The past, present and future of antifolates in the treatment of Plasmodium falciparum infection. J Antimicrob Chemother. 2006;57:1043–54.
52. WHO. World health organization model list of essential medicines. World Heal Organazation, Geneva. 2019;21:119–34.
53. White NJ, Hien TT, Nosten FH. A Brief History of Qinghaosu. Trends Parasitol. 2015;31(12):607–10.
54. Cui L, Su X. Discovery, mechanisms of action and combination therapy of artemisinin. Expert Rev Anti Infect Ther. 2009;7(8):999–1013.
55. White NJ. Qinghaosu (artemisinin): The price of success. Science (80-). 2008;320(5874):330–4.
56. Woodrow CJ, Haynes RK, Krishna S. Artemisinins. Postgrad Med J. 2005;81(952):71–8.
57. Olliaro PL, Haynes RK, Meunier B, Yuthavong Y. Possible modes of action of the artemisinin-type compounds. Trends Parasitol. 2001;17(3):122–6.
58. O'Neill PM, Barton VE, Ward SA. The molecular mechanism of action of artemisinin - The debate continues. Molecules. 2010;15(3):1705–21.
59. Wang J, Zhang CJ, Chia WN, Loh CCY, Li Z, Lee YM, et al. Haem-activated

promiscuous targeting of artemisinin in Plasmodium falciparum. Nat Commun. 2015;6:1–11.

60. Shandilya A, Chacko S, Jayaram B, Ghosh I. A plausible mechanism for the antimalarial activity of artemisinin: A computational approach. Sci Rep. 2013;3(2513):1–7.
61. WHO. Guidlines for the treatment of Malaria-Third edition. WHO,Geneva. 2015
62. Payne D. Spread of chloroquine resistance in Plasmodium falciparum. Parasitol Today. 1987;3(8):241–6.
63. Wellems TE, Plowe C V. Chloroquine-Resistant Malaria. J Infect Dis. 2001;184:770–6.
64. Noedl H, Se Y, Schaecher K, Smith BL, Socheat D, Fukuda MM. Evidence of artemisinin-resistant malaria in Western Cambodia. N Engl J Med. 2008;359(24):2619–20.
65. Wang P, Read M, Sims PFG, Hyde JE. Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. Mol Microbiol. 1997;23(5):979–86.
66. Happi CT, Gbotosho GO, Folarin OA, Akinboye DO, Yusuf BO, Ebong OO, et al. Polymorphisms in Plasmodium falciparum dhfr and dhps genes and age related in vivo sulfadoxine-pyrimethamine resistance in malaria-infected patients from Nigeria. Acta Trop. 2005;95(3):183–93.
67. Schwöbel B, Alifrangis M, Salanti A, Jelinek T. Different mutation patterns of atovaquone resistance to Plasmodium falciparum in vitro and in vivo: Rapid detection of codon 268 polymorphisms in the cytochrome b as potential in vivo resistance marker. Malar J. 2003;2(5):1–7.
68. Djimdé A, Doumbo OK, Cortese JF, Kayentao K, Doumbo S, Diourté Y, Coulibaly D, Dicko A, Su XZ, Nomura T, Fidock DA, Wellems TE, Plowe CV. A molecular marker for chloroquine-resistant falciparum malaria. New Engl J Med. 2001;344(4):257–63.
69. Veiga MI, Dhingra SK, Henrich PP, Straimer J, Gnädig N, Uhlemann AC, et al. Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. Nat Commun. 2016;7:1–12.
70. Korsinczky M, Fischer K, Chen N, Baker J, Rieckmann K, Cheng Q. Sulfadoxine resistance in Plasmodium vivax is associated with a specific amino acid in dihydropteroate synthase at the putative sulfadoxine-binding site. Antimicrob Agents

Chemother. 2004;48(6):2214–22.

71. Boonyalai N, Vesely BA, Thamnurak C, Praditpol C, Fagnark W, Kirativanich K, et al. Piperaquine resistant Cambodian Plasmodium falciparum clinical isolates: In vitro genotypic and phenotypic characterization. Malar J. 2020;19(1):1–13.
72. Jose Pedro Gil. Amodaquine pharmacogenetics. Pharmacogenomics. 2008;9(10):1385–90.
73. Kerb R, Fux R, Mörike K, Kremsner PG, Gil JP, Gleiter CH, et al. Pharmacogenetics of antimalarial drugs : effect on metabolism and transport. Expert Opin Drug Metab Toxicol. 2011;9(12):760–74.
74. Abdullahi ST, Soyinka JO, Olagunju A, Bolarinwa RA, Olarewaju OJ, Bakare-Odunola MT, et al. CYP2B6*6 Genotype Specific Differences in Artemether-Lumefantrine Disposition in Healthy Volunteers. J Clin Pharmacol. 2020;60(3):351–60.
75. Zanger UM, Turpeinen M, Klein K, Schwab M. Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. Anal Bioanal Chem. 2008;392(6):1093–108.
76. Dandara C, Swart M, Mpeta B, Wonkam A, Masimirembwa C. Cytochrome p450 pharmacogenetics in african populations: Implications for public health. Expert Opin Drug Metab Toxicol. 2014;10(6):769–85.
77. Mutagonda RF, Kamuhabwa AAR, Minzi OMS, Massawe SN, Asghar M, Homann M V, et al. Effect of pharmacogenetics on plasma lumefantrine pharmacokinetics and malaria treatment outcome in pregnant women. Malar J. 2017;16(1):1–10.
78. Fred Peter G. Cytochrome p450 and chemical toxicology. Chem Res Toxicol. 2008;21:70–83.
79. Kerb R, Fux R, Mörike K, Kremsner PG, Gil JP, Gleiter CH, et al. Pharmacogenetics of antimalarial drugs: effect on metabolism and transport. Lancet Infect Dis. 2009;9(12):760–74.
80. Mehlotra RK, Henry-Halldin CN, Zimmerman PA. Application of pharmacogenomics to malaria: A holistic approach for successful chemotherapy. Pharmacogenomics. 2009;10(3):435–49.
81. Lai X-S, Yang L-P, Li X-T, Liu J-P, Zhou Z-W, Zhou S-F. Human CYP2C8: Structure, Substrate Specificity, Inhibitor Selectivity, Inducers and Polymorphisms. Curr Drug Metab. 2010;10(9):1009–47.
82. Peko SM, Ntoumi F, Vouvoungui C, Nderu D, Kobawila SC, Velavan TP, et al.

Distribution of the cytochrome P450 CYP2C8*2 allele in Brazzaville, Republic of Congo. Int J Infect Dis. 2019;85:49–53.

83. Totah RA, Rettie AE. Cytochrome P450 2C8: Substrates, inhibitors, pharmacogenetics, and clinical relevance. Clin Pharmacol Ther. 2005;77(5):341–52.
84. Daily EB, Aquilante CL. Cytochrome P450 2C8 pharmacogenetics: A review of clinical studies. Pharmacogenomics. 2009;10(9):1489–510.
85. Backman JT, Filppula AM, Niemi M, Neuvonen PJ. Role of Cytochrome P450 2C8 in drug metabolism and interactions. Pharmacol Rev. 2016;68(1):168–241.
86. Paganotti GM, Gallo BC, Verra F, Sirima BS, Nebié I, Diarra A, et al. Human genetic variation is associated with plasmodium falciparum drug resistance. J Infect Dis. 2011;204(11):1772–8.
87. Li XQ, Björkman A, Andersson TB, Ridderström M, Masimirembwa CM. Amodiaquine clearance and its metabolism to N-desethylamodiaquine is mediated by CYP2C8: A new high affinity and turnover enzyme-specific probe substrate. J Pharmacol Exp Ther. 2002;300(2):399–407.
88. Adjei GO, Kristensen K, Goka BQ, Hoegberg LCG, Alifrangis M, Rodrigues OP, et al. Effect of concomitant artesunate administration and cytochrome P4502C8 polymorphisms on the pharmacokinetics of amodiaquine in Ghanaian children with uncomplicated malaria. Antimicrob Agents Chemother. 2008;52(12):4400–6.
89. Projean D, Baune B, Farinotti R, Flinois JP, Beaune P, Taburet AM, et al. In vitro metabolism of chloroquine: Identification of CYP2C8, CYP3A4, and CYP2D6 as the main isoforms catalyzing N-desethylchloroquine formation. Drug Metab Dispos. 2003;31(6):748–54.
90. Sortica VA, Lindenau JD, Cunha MG, Ohnishi M Do, Ventura AMR, Ribeiro-Dos-Santos ÂK, et al. The effect of SNPs in CYP450 in chloroquine/primaquine Plasmodium vivax malaria treatment. Pharmacogenomics. 2016;17(17):1903–11.
91. Taylor C, Crosby I, Yip V, Maguire P, Pirmohamed M, Turner RM. A review of the important role of CYP2D6 in pharmacogenomics. Genes (Basel). 2020;11(11):1–23.
92. Zanger UM, Raimundo S, Eichelbaum M. Cytochrome P450 2D6: Overview and update on pharmacology, genetics, biochemistry. Naunyn Schmiedebergs Arch Pharmacol. 2004;369(1):23–37.
93. CYP2D6. Bethesda (MD). National Library of Medicine (US) [Internet]. National Center for Biotechnology Information. [cited 2022 Apr 7]. Available from: https://www.ncbi.nlm.nih.gov/gene/1565.

94. Charity Nofziger, Turner AJ, Sangkuhl K, Whirl-carrillo M, José AG, Black JL, et al. PharmVar GeneFocus: CYP2D6. Clnical Pharmacol Ther. 2019;107(1):154–70.
95. Pybus BS, Marcsisin SR, Jin X, Deye G, Sousa JC, Li Q, et al. The metabolism of primaquine to its active metabolite is dependent on CYP2D6. Malar J. 2013;12(1):1–7.
96. Huang H, Dong Y, Xu Y, Deng Y, Zhang C, Liu S, et al. The association of CYP2D6 gene polymorphisms in the full - length coding region with higher recurrence rate of vivax malaria in Yunnan Province , China. Malar J. 2021;20(160):1–10.
97. Brasil LW, Rodrigues-Soares F, Santoro AB, Almeida ACG, Kühn A, Ramasawmy R, et al. CYP2D6 activity and the risk of recurrence of Plasmodium vivax malaria in the Brazilian Amazon: A prospective cohort study. Malar J. 2018;17(57):1–6.
98. Dong Y, Huang H, Deng Y, Xu Y, Chen M, Liu Y, et al. Prediction of the CYP2D6 enzymatic activity based on investigating of the CYP2D6 genotypes around the vivax malaria patients in Yunnan Province , China. Malar J. 2021;20(448):1–10.
99. Chen N, Dowd S, Gatton ML, Auliff A, Edstein MD, Cheng Q. Cytochrome P450 2D6 profiles and their relationship with outcomes of primaquine anti-relapse therapy in Australian Defence Force personnel deployed to Papua New Guinea and East Timor. Malar J. 2019;18(140):1–10.
100. Dandara C, Swart M, Mpeta B, Wonkam A, Masimirembwa C. Cytochrome P450 pharmacogenetics in African populations : implications for public health. Expert Opin Drug Metab Toxicol. 2014;10(6):769–85.
101. Friedrich DC, Genro JP, Sortica VA, Suarez-Kurtz G, De Moraes ME, Pena SDJ, et al. Distribution of CYP2D6 alleles and phenotypes in the Brazilian population. PLoS One. 2014;9(10):5–12.
102. Nofziger C, Turner AJ, Sangkuhl K, Whirl-Carrillo M, Agúndez JAG, Black JL, Dunnenberger HM, Ruano G, Kennedy MA, Phillips MS, Hachad H, Klein TE, Gaedigk A. PharmVar GeneFocus: CYP2D6. Clin Pharmacol Ther. 2020;107(1):154–70.
103. Pybus BS, Sousa JC, Jin X, Ferguson JA, Christian RE, Barnhart R, et al. CYP450 phenotyping and accurate mass identification of metabolites of the 8-aminoquinoline, anti-malarial drug primaquine. Malar J. 2012;11(259):1–9.
104. Baird JK, Louisa M, Noviyanti R, Ekawati L, Elyazar I, Subekti D, et al. Association of Impaired Cytochrome P450 2D6 Activity Genotype and Phenotype With Therapeutic Efficacy of Primaquine Treatment for Latent Plasmodium vivax Malaria. JAMA Netw open. 2018;1(4):1–12.

105. Marwa KJ, Schmidt T, Sjögren M, Minzi OMS, Kamugisha E, Swedberg G. Cytochrome P450 single nucleotide polymorphisms in an indigenous Tanzanian population : a concern about the metabolism of artemisinin-based combinations. Malar J. 2014;13(11):1–7.
106. CYP3A4. Bethesda (MD). National Library of Medicine (US) [Internet]. National Center for Biotechnology Information. [cited 2022 Apr 7]. Available from: https://www.ncbi.nlm.nih.gov/gene/1576
107. Rodríguez-Antona C, Sayi JG, Gustafsson LL, Bertilsson L, Ingelman-Sundberg M. Phenotype-genotype variability in the human CYP3A locus as assessed by the probe drug quinine and analyses of variant CYP3A4 alleles. Biochem Biophys Res Commun. 2005;338(1):299–305.
108. Amirimani B, Ning B, Deitz AC, Weber BL, Kadlubar FF, Rebbeck TR. Increased Transcriptional Activity of the CYP3A4 * 1B Promoter Variant. Environ Mol Mutagen. 2003;42:299–305.
109. Piedade R, Gil JP. The pharmacogenetics of antimalaria artemisinin combination therapy. Expert Opin Drug Metab Toxicol. 2011;7(10):1185–200.
110. Hodel EMS, Csajka C, Ariey F, Guidi M, Kabanywanyi AM, Duong S, et al. Effect of single nucleotide polymorphisms in cytochrome P450 isoenzyme and N-acetyltransferase 2 genes on the metabolism of artemisinin-based combination therapies in malaria patients from Cambodia and Tanzania. Antimicrob Agents Chemother. 2013;57(2):950–8.
111. Lee JS, Cheong HS, Kim LH, Kim JO, Seo DW, Kim YH, et al. Screening of genetic polymorphisms of CYP3A4 and CYP3A5 genes. Korean J Physiol Pharmacol. 2013;17(6):479–84.
112. Lee SJ, Goldstein JA. Functionally defective or altered CYP3A4 and CYP3A5 single nucleotide polymorphisms and their detection with genotyping tests. Pharmacogenomics. 2005;6(4):357–71.
113. Mutagonda RF, Minzi OMS, Massawe SN, Asghar M, Färnert A, Kamuhabwa AAR, et al. Pregnancy and CYP3A5 Genotype Affect Day 7 Plasma Lumefantrine Concentrations. Drug Metab Dispos. 2019;47(12):1415–24.
114. Adehin A, Bolaji OO, Kennedy MA. Polymorphisms in CYP2C8 and CYP3A5 genes in the Nigerian population. Drug Metab Pharmacokinet. 2017;32(3):189–91.
115. Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5

expression. Nat Genet. 2001;27(4):383–91.

116. Lamba J, Hebert JM, Schuetz EG, Klein TE, Altman RB. PharmGKB summary. Pharmacogenet Genomics. 2012;22(7):555–8.

117. CYP2C19. Bethesda (MD). National Library of Medicine (US) [Internet]. National Center for Biotechnology Information. [cited 2022 Apr 7]. Available from: https://www.ncbi.nlm.nih.gov/gene/1557.

118. Jancova P, Anzenbacher P, Anzenbacherova E. Phase II drug metabolizing enzymes. Biomed Pap. 2010;154(2):103–16.

119. Phompradit P, Muhamad P, Cheoymang A, Na-bangchang K. Preliminary Investigation of the Contribution of CYP2A6 , CYP2B6 , and UGT1A9 Polymorphisms on Artesunate-Mefloquine Treatment Response in Burmese Patients with Plasmodium falciparum Malaria. Am J Trop Med Hyg. 2014;91(2):361–6.

120. Zang M, Zhu F, Zhao L, Yang A, Li X, Liu H, et al. The effect of UGTs polymorphism on the auto-induction phase II metabolism-mediated pharmacokinetics of dihydroartemisinin in healthy Chinese subjects after oral administration of a fixed combination of dihydroartemisinin-piperaquine. Malar J. 2014;13(478):1–9.

121. Kengne JPC, Nji AM, Ali IM, Achonduh OA, Ekollo M, Mbacham WF. Predominance of N-acetyl transferase 2 slow acetylator alleles in children less than ten years experiencing adverse treatment events following artemisinin-based combination therapy in North and South West Regions of Cameroon. African J Biotechnol. 2016;15(25):1285–91.

122. Guillemette C. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. Pharmacogenomics J. 2003;3(3):136–58.

123. Ilett KF, Ethell BT, Maggs JL, Davis TME, Batty KT, Burchell B, et al. Glucuronidation of dihydroartemisinin in vivo and by human liver microsomes and expressed UDP-glucuronosyltransferases. Drug Metab Dispos. 2002;30(9):1005–12.

124. Gong Q-H, Cho Aã JW, Huang T, Potter C, Gholami N, Basu NK, et al. Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus Qi-Hui. Pharmacogenetics. 2001;11:357–68.

125. Commitee UN. UGT Alleles Nomenclature [Internet]. [cited 2022 Mar 27]. Available from: www.pharmacogenomics.pha.ulaval.ca/wp-content/uploads/2015/04/SNP-UGT1A9.htm.

126. Carrier JS, Turgeon D, Journault K, Hum DW, Beélanger A. Isolation and characterization of the human UGT2B7 gene. Biochem Biophys Res Commun.

2000;272(2):616–21.

127. UGT Nomenclature Commitee. UGT Alleles Nomenclature [Internet]. [cited 2022 Mar 27]. Available from: www.pharmacogenomics.pha.ulaval.ca/wp-content/uploads/2015/04/SNP-UGT2B7.htm.
128. Saito K, Moriya H, Sawaguchi T, Hayakawa T, Nakahara S. Haplotype analysis of UDP-glucuronocyltransferase 2B7 gene (UGT2B7) polymorphisms in healthy Japanese subjects. Clin Biochem. 2006;39:303–8.
129. Rajeev K. Mehlotra, Bockarie, Moses J, Zimmerman PA. Prevalence of UGT1A9 and UGT2B7 nonsynonymous single nucleotide polymorphisms in West African, Papua New Guinean, and North American populations. Eur J Clin Pharmacol. 2007;63(1):1–8.
130. Bhasker CR, Mckinnon W, Stone A, Ct A. Genetic polymorphism of UDP-glucuronosyltransferase 2B7 (UGT2B7) at amino acid 268 : ethnic diversity of alleles and potential clinical significance. Pharmacogenetics. 2000;10(8):679–85.
131. Butcher NJ, Sim E, Minchin RF. Pharmacogenetics of the arylamine N-acetyltransferases. Pharmacogenomics J. 2002;2:30–42.
132. Butcher NJ, Sim E, Minchin RF. Pharmacogenetics of the arylamine N - acetyltransferases. 2002;2:30–42.
133. Aklillu E, Carrillo J. A., Makonnen E , Bertilsson L DN. N-Acetyltransferase-2 (NAT2) phenotype is influenced by genotype-environment interaction in Ethiopians. Eur J Clin Pharmacol. 2018;74(7):903–11.
134. Kerb R. Implications of genetic polymorphisms in drug transporters for pharmacotherapy. Cancer Lett. 2006;234(1):4–33.
135. Nigam SK. What do drug transporters really do? Nat Rev Drug Discov. 2014;14(1):29–44.
136. Wolking S, Schaeffeler E, Lerche H, Schwab M. Impact of Genetic Polymorphisms of ABCB1 (MDR1, P-Glycoprotein) on Drug Disposition and Potential Clinical Implications : Update of the Literature. Clin Pharmacokinet. 2015;54:709–35.
137. Wahajuddin, Raju KSR, Singh SP, Taneja I. Investigation of the functional role of p-glycoprotein in limiting the oral bioavailability of lumefantrine. Antimicrob Agents Chemother. 2014;58(1):489–94.
138. ABCB1. Bethesda (MD). National Library of Medicine (US) [Internet]. National Center for Biotechnology Information. [cited 2022 Apr 7]. Available from: https://www.ncbi.nlm.nih.gov/gene/5243

139. Kiaco K, Rodrigues AS, Do Rosário V, Gil JP, Lopes D. The drug transporter ABCB1 c.3435C>T SNP influences artemether-lumefantrine treatment outcome. Malar J. 2017;16(1):1–6.
140. Rajman I, Knapp L, Hanna I. Genetic Diversity in Drug Transporters : Impact in African Populations. Clin Transl Sci. 2020;13(5):848–60.
141. Sortica VA, Lindenau JD, Cunha MG, O Ohnishi MD, R Ventura AM, Ribeiro-Dos-Santos ÂKC, et al. SLCO1A2, SLCO1B1 and SLCO2B1 polymorphisms influences chloroquine and primaquine treatment in Plasmodium vivax malaria. Pharmacogenomics. 2017;18(15):1401–11.
142. Phyo AP, Nkhoma S, Stepniewska K, Ashley EA, Nair S, McGready R, et al. Emergence of artemisinin-resistant malaria on the western border of Thailand: A longitudinal study. Lancet. 2012;379(9830):1960–6.
143. Fairhurst RM, Nayyar GML, Breman JG, Hallett R, Vennerstrom JL, Duong S, et al. Artemisinin-resistant malaria: Research challenges, opportunities, and public health implications. Am J Trop Med Hyg. 2012;87(2):231–41.
144. Byakika-kibwika P, Ssenyonga R, Lamorde M, Blessborn D, Tarning J. Piperaquine concentration and malaria treatment outcomes in Ugandan children treated for severe malaria with intravenous Artesunate or quinine plus. BMC Infect Dis. 2019;19(1025):1–7.
145. Sowunmi A, Adedeji AA, Fateye B, A, Babalola CP. Plasmodium falciparum hyperparasitaemia in children : Risk factors, treatment outcomes, and gametocytaemia following treatment. Parasite. 2004;11:317–23.
146. Pukrittayakamee S, Imwong M, Looareesuwan S, White NJ. Therapeutic responses to antimalarial and antibacterial drugs in vivax malaria. Acta Trop. 2004;89:351–6.
147. Yusof W, Hua GS. Gene , ethnic and gender influences predisposition of adverse drug reactions to artesunate among Malaysians. Toxicol Mech Methods. 2012;22(3):184-92.
148. Bachtiar M, Lee CGL. Genetics of Population Differences in Drug Response. Curr Genet Med Rep. 2013;1(3):162–70.
149. Ahsan T, Urmi NJ, Sajib AA. Heterogeneity in the distribution of 159 drug-response related SNPs in world populations and their genetic relatedness. PLoS One. 2020;15(1):1–26.
150. Reddy S, Ornelle D, Assiana E, Thi L, Linh K, Nchang F, et al. Pharmacogenetic considerations in the treatment of co-infections with HIV/AIDS, tuberculosis and

malaria in Congolese populations of Central Africa. Int J Infect Dis. 2021;104:207–13.

151. Almeida ACG, Maria C. B. Puça EFGF, Laila R. Barbosa YEARS, Emanuelle L. Silva MAMB, André M. Siqueira D, José L. F. Vieira MVGL, et al. Influence of CYP2C8, CYP3A4, and CYP3A5 Host Genotypes on Early Recurrence of Plasmodium vivax. Antimicrob Agents Chemother. 2020;64(7):e02125-19.
152. Parker M, Bull SJ, Vries J De, Agbenyega T, Doumbo OK, Dominic P. Ethical Data Release in Genome-Wide Association Studies in Developing Countries. Plos Med. 2009;6(11):4–7.
153. Band G, Rockett KA, Spencer CCA, Kwiatkowski DP, Si Le Q, Clarke GM, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature. 2015;526(7572):253–7.
154. Band G, Le QS, Jostins L, Pirinen M, Kivinen K, Jallow M, et al. Imputation-Based Meta-Analysis of Severe Malaria in Three African Populations. PLoS Genet. 2013;9(5):1–13.
155. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. Nature. 2015;517:327–32.
156. 1000 Genomes Project Consortium, Altshuler, David L. Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061–73.
157. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ SP. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
158. Deelen P, Bonder MJ, van der Velde KJ, Westra HJ, Winder E, Hendriksen D, Franke L SM. Genotype harmonizer : automatic strand alignment and format conversion for genotype data integration. BMC Res Notes. 2014;7(901):1–4.
159. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10(1):5–6.
160. Wang K, Li M HH. ANNOVAR : functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
161. Voight BF, Kudaravalli S, Wen X PJ. A Map of Recent Positive Selection in the Human Genome. PLOS Biol. 2006;4(3):e72.
162. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB et al. Detecting recent positive selection in the human genome from haplotype

structure. Nature. 2002;419(6909):832–7.

163. Muyambo S, Ndadza A, Soko ND, Kruger B, Kadzirange G, Chimusa E, Masimirembwa CM, Ntsekhe M, Nhachi CFB DC. Warfarin Pharmacogenomics for Precision Medicine in Real-Life Clinical Practice in Southern Africa : Harnessing 73 Variants in 29 Pharmacogenes. OMICS. 2022;26(1):35–50.
164. Van der Weide J SL. Cytochrome P450 enzyme system: genetic polymorphisms and impact on clinical pharmacology. Clin Biochem. 1999;36:722–9.
165. Damena D, Agamah FE, Kimathi PO, Kabongo NE, Girma H, Choga WT, Golassa L CE. Insilico Functional Analysis of Individuals Identifies Candidate Malaria Resistance Genes Enriched in Malaria Pathogenic Pathways. Front Genet. 2021;12(676960):1–18.
166. Popejoy AB FS. Genomics is failing on diversity. Nature. 2016;538(7624):161–4.
167. Raska P, Zhu X. Rare variant density across the genome and across populations. BMC Proc. 2011;5:1–39.
168. Diakite M, Achidi EA, Achonduh O, Craik R, Djimde AA, Evehe MB, et al. Host candidate gene polymorphisms and clearance of drug-resistant Plasmodium falciparum parasites. Malar J. 2011;10(50).
169. Ndila CM, Uyoga S, Macharia AW, Nyutu G, Peshu N, Ojal J, et al. Human candidate gene polymorphisms and risk of severe malaria in children in Kilifi, Kenya : a case-control association study. Lancet Haematol. 2018;5(8):e333–45.
170. Campbell MC TS. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 2008. 2008;9:403–33.
171. Li J, Zhang L, Zhou H, Stoneking M, Tang K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. Hum Mol Genet. 2011;20(3):528–40.
172. Maisano Delser P, Fuselli S. Human loci involved in drug biotransformation : worldwide genetic variation , population structure , and pharmacogenetic implications. Hum Genet. 2013;132(5):563–77.
173. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB K TE. Pharmacogenomics Knowledge for Personalized Medicine. Perspectives (Montclair). 2012;92(4):414–7.
174. Lin YS, Thummel KE, Thompson BD, Totah RA CC. Sources of Interindividual Variability. Methods Mol Biol. 2014;1113:363–415.

ANNEXES

Annex A: Script used for QC (for Mali dataset as example)

```
#!/bin/bash
#PBS -N QC_Mali
#PBS -q serial
#PBS -P CBBI0818
#PBS -l select=1:ncpus=2
#PBS -l walltime=48:00:00
module add chpc/gnu/parallel-20160422
. /mnt/lustre/groups/CBBI0818/GIRMA/config.txt
path=/mnt/lustre/groups/CBBI0818/GIRMA/script/
data=/mnt/lustre/groups/CBBI0818/GIRMA/GenotypeHarmonizer-1.4.23/Mali_Realigned_Harmonizer_PLINK_Data/
data=/mnt/lustre/groups/CBBI0818/GIRMA/DATA/PHASED/PLINK_FILE/
OUT=/mnt/lustre/groups/CBBI0818/GIRMA/RESULT_QC/
#prefix=Mali.${chr}.realigned.final
SOFT=/mnt/lustre/groups/CBBI0818/GIRMA/soft/

for chr in {1..22}
do
${SOFT}plink –bfile ${data}Mali.chr${chr}.realigned.final --geno 0.05 --mind 0.05 --maf 0.05 --hardy --make-bed --out ${OUT}Mali.chr${chr}.realigned.final.qced
done
```

Annex B: Script used for phasing (for Mali dataset as example)

```
#!/bin/bash
#PBS -N selcan
#PBS -q serial
#PBS -P CBBI0818
```

```
#PBS -l select=1:ncpus=2
#PBS -l walltime=48:00:00

. /mnt/lustre/groups/CBBI0818/GIRMA/config.txt
module load  chpc/python/3.5.1

for chr in {1..22}
do

shapeit=/mnt/lustre/groups/CBBI0818/CPGG/SOFTWARE/soft/shapeit
map=${ref_map}genetic_map_chr${chr}_combined_b37.txt
data=/mnt/lustre/groups/CBBI0818/GIRMA/GenotypeHarmonizer-1.4.23/PHASED/
${shapeit} -B ${data}Mali.chr${chr}.realigned.final.qced -M ${map} \
-O ${data}Mali.chr${chr}.realigned.final.phased.qced --thread 8
done
```

qsub script used to the above shell script is as follows

```
#!/bin/bash
#PBS -N PHASE_VCF
#PBS -P CBBI0818
#PBS -q seriallong
#PBS -l select=1:ncpus=3
#PBS -l walltime=144:00:00

module add chpc/gnu/parallel-20160422

. /mnt/lustre/groups/CBBI0818/GIRMA/config.txt

IFS=" "
SUB_JOBS=""
```

```
for Hunda
do
	SUB_JOBS="${SUB_JOBS}
${CHR}"
done
echo ${SUB_JOBS}
echo "${SUB_JOBS}" | parallel -j 23 -u --sshloginfile ${PBS_NODEFILE} --col-sep ' ' '. \
/mnt/lustre/groups/CBBI0818/GIRMA/config.txt;CHR={1}; \
. /mnt/lustre/groups/CBBI0818/GIRMA/script/step_3.PHASE.sh'
```

##**Note**: **The config file script is as follows**

```
#!/bin/bash
############ SOURCE SUPPORTING DATA AND INITIALISE VARIABLES
############
# set PATHS

module load chpc/BIOMODULES
#module load chpc/python/2.7.11
#module load chpc/python/2.7.12
module add chpc/gnu/parallel-20160422
source /apps/user/CBBI0818/venv.python2_7/bin/activate
module add chpc/python/anaconda/2

module add /apps/chpc/scripts/modules/bio/lib/gnu/gsl_2.1
module add chpc/gnu/plot-5.0.4
module add /apps/chpc/scripts/modules/bio/lib/OpenBLAS/12ab180
#module add chpc/R/3.2.3-gcc5.1.0
#module add chpc/R/3.3.3-gcc6.3.0
module add chpc/R/3.5.0-gcc7.2.0-V2
module add chpc/java/1.8.0_73
```

```
EXE_JAVA=java
source /apps/user/CBBI0818/venv.python2_7/bin/activate
#EXE_PY=python

wkdir=/mnt/lustre/groups/CBBI0818/GIRMA/
#data=/mnt/lustre/groups/CBBI0818/GIRMA/
PATH_DATA=/mnt/lustre/groups/CBBI0818/CPGG/PGC_new/IMPG/bochet.gcc.biostat.washington.edu/beagle/1000_Genomes.phase1_release_v3/

###Merge, Phased, Impute
refdata=/mnt/lustre/groups/CBBI0818/CPGG/refdata/
soft=/mnt/lustre/groups/1000genomes/annotation/soft/
GZIP=${soft}bgzip
TABIX=${soft}tabix
emmax=${soft}emmax
kinship=${soft}emmax-kin
minimac=${soft}Minimac3
impute=${soft}impute
mach1=${soft}mach1
mach_admix=${soft}mach-admix
plink=${soft}plink1.9

# Eagle phasing software
softdir=/mnt/lustre/groups/1000genomes/annotation/
eagle=${softdir}soft/eagle
bcftools=${softdir}soft/bcftools
table=/mnt/lustre/groups/CBBI0818/REF/tables/genetic_map_hg19.txt.gz
minimac4=${softdir}soft/minimac4
m3vcftools=${softdir}soft/m3vcftools
beagle=${softdir}soft/beagle.jar
beagle5=${softdir}soft/beagle5.jar
```

refdir=/mnt/lustre/groups/CBBI0818/PETER/REF/

table1=${refdir}tables/

Annotation databases

annot=/mnt/lustre/groups/1000genomes/annotation/annovar/

softclair=/mnt/lustre/groups/1000genomes/annotation/soft/

Annex C: Primer set and it's Information

Product length: 798 bp

Forward primer: CCATCGTTCTCAGCATACTATCAC

Reverse primer: CTATGCATTCTAGCCATTGGACAAT

Properties	**Forward primer**	**Reverse primer**
Length (bp)	22	25
GC Content (%)	45.45	40.00
Melting temperature	56.23	59.52
Hairpin **ΔG (kcal.mole^{-1})**	0.68 to 1.52	-0.27 to 0.61
Self dimer **ΔG (kcal.mole^{-1})**	-3.61 to -0.96	-10 to -0.96
Hetero-dimer **ΔG (kcal.mole^{-1})**	-7.53 to -1.47	

GGATGAGTTCATGTTCTTTGTAGGGACATGGATGAAGCTAGAAACCATCGTTCTCAGCATACTATC
ACAAGGACAAAAAAACCAAACACTGCATATTCTCATTCATAGGTGGGAATTGAACAATGAGAACA
CTTGAACACAGGAAGGGGAACATTACACACTGGGGTCTGTTGTGTGGTGGGGGGAGGGGGAAGG
GATAGCATTGGGAGATATACCTAATGTATATGACGAGTTATTGGGTGCAGTACACCAACCTGGCAC
ATGTATACATATGTGACAAACCTGCACTTTGTGCACATATACCCTAGAACTTAAAGTGTAATAAA
AAATGTATATATGTATAAAAATTTCCCTTCAAAATGGACATGATGTCTTATTCATATTTATAGT
TATAATTTCAATCAGGGCTTGGTGTAAGATACATATATCTTATGACATGTTTATATTTAATATTCTT
TTCTCTTTTAGGTCTGCAATAATTTCCCTCTACTCATTGATTGTTTCCCAGGAACTCACAACAAAGT
GCTTAAAAATGTTGCTCTTACACGAAGTTACATTAGGGAGAAAGTAAAAGAACACCAAGCATCA
CTGGATGTTAACAATCCTCGGGACTTTATCGATTGCTTCCTGATCAAAATGGAGCAGGTAAGAT
ATTAGCAACAGATCAGTATTTTGATTTCTTGTCCATTTTGTGATTCATCGAATCCTTCTGTAATTTAC
TAAGGATGTTTAAATGATCAGGCCAGTAATGCTTGACAAGCATCTTAATTACTTATTGTATTTAT
GGGCCTGCACTAAACATCATGGAAAATACAAAATTGTCCAATGGCTAGAATGCATAGACCCTCTTC

TTGAAATAAAAATAATCTACATGAGTAGAATAATATAGACACTGCAGTCAAATGGAAACAATTAT
AAATAAGTAGAAAATACAAAATATCCTACCACAAACTGAAGATGAATATCAATATTGCACATAGA
TTATGTGACAATGGAAAAGAATGGCTTATCATAAATGGGCTTCC

CTATGCATTCTAGCCATTGGACAATTTTGTATTTTCCATGATGTTTAGTGCAGGCCCATAAATACAA
TAAGTAATTAAGATGCTTGTCAAGCATTACTGGCCTGATCATTTAAACATCCTTAGTAAATTACAG
AAGGATTCGATGAATCACAAAATGGACAAGAAATCAAAATACTGATCTGTTGCTAATATCTTACCT
GCTCCATTTTGATCAGGAAGCAATCGATAAAGTCCCGAGGATTGTTAACATCCAGTGATGCTTGGT
GTTCTTTTACTTTCTCCCTAATGTAACTTCGTGTAAGAGCAACATTTTTAAGCACTTTGTTGTGAGTT
CCTGGGAAACAATCAATGAGTAGAGGGAAATTATTGCAGACCTAAAAGAGAAAAGAATATTAAAT
ATAAACATGTCATAAGATATATGTATCTTACACCAAGCCCTGATTGAAATTATAACTATAAATATG
AATAAGACATCATGTCCATTTTGAAGGGAAATTTTTATACATATATACATTTTTTATTACACTTTAA
GTTCTAGGGTATATGTGCACAAAGTGCAGGTTTGTCACATATGTATACATGTGCCAGGTTGGTGTA
CTGCACCCAATAACTCGTCATATACATTAGGTATATCTCCCAATGCTATCCCTTCCCCCTCCCCCCA
CCACACAACAGACCCCAGTGTGTAATGTTCCCCTTCCTGTGTTCAAGTGTTCTCATTGTTCAATTCC
CACCTATGAATGAGAATATGCAGTGTTTGGTTTTTTTGTCCTTGTGATAGTATGCTGAGAACGATG
G

Order of the SNPs in close proximity: *rs1926705G>A, rs11572100 A>G, rs11572101T>C, CYP2C8*14 c.712G>C (rs188934928), CYP2C8 c. 792C>G (*4) (rs1058930), CYP2C8 c.805A>T (*2) (rs11572103) & rs11572105 C>A/G/T*

Annex D: Post-sequencing clean up protocol (Ethanol precipitation)

Plate method

1. The protocol is for 20 μL sequencing reaction volume. If you have less, add SABAX water to make it 20 μL.
2. Add 5 μL of 125 mM EDTA in 96 well plate. Pipet mix well and add 60 µL of 100% ethanol (room temperature). Seal with foil (cut seal to size) and vortex for 2-3 seconds.
3. Leave on ice block in freezer for 15 minutes. Centrifuge 45 min @1870 rcf
4. Remove foil and turn plate upside down paper towel and place back in centrifuge. Do the same with the balance but keep the seal on. Centrifuge 1 min @ 180 rcf and add 60 μL of 70% ethanol. Seal with foil and centrifuge 15min @ 1870 rcf .
5. Remove foil and turn plate upside down on paper towel and place back in centrifuge. Do the same with the balance plate keep the seal on. Centrifuge 1 min @180 rcf
6. Air-dry the plate in the dark for 10-15 minutes (put in one of the drawers)
7. Add 10 μL of HiDi formamide and cover it with septa
8. Proceed to denaturing the plate at 95^{o}c for 5 minutes and capillary electrophoresis.

Annex E: Genes found in the candidate iHS regions other than pharmacogenes in Malawi, Kenya and Gambia populations, respectively.

Malawi population

CHR	START	END	N_MRK	MEAN_MRK	MAX_MRK	N_EXTR_MRK	PERC_EXTR_MRK	MEAN_EXTR_MRK	GENES
1	103700000	103850000	56	0.982004465	5.240098905	2	3.57	5.061746307	
1	120200000	120500000	136	1.224339523	7.734260668	11	8.09	5.243823259	ADAM30
1	13700000	13850000	65	1.105961959	8.719973556	3	4.62	6.061715192	LRRC38
1	223300000	223450000	56	2.131109602	4.412929438	3	5.36	4.251467804	SUSD4
2	12250000	12450000	157	1.416273524	4.543412632	4	2.55	4.392165129	MIR3681
2	62700000	62800000	68	1.062291827	4.811378439	2	2.94	4.691353212	TMEM17
3	178300000	178450000	69	1.703417097	4.303110711	2	2.9	4.243112693	KCNMB2
4	107850000	108000000	34	1.688852342	4.583764158	2	5.88	4.490471791	DKK2
4	107850000	108000000	34	1.702323939	4.598825606	4	11.76	4.263726525	DKK2
4	450000	600000	17	2.25824513	4.352817323	3	17.65	4.18004572	
4	450000	600000	17	2.265830699	4.326242712	3	17.65	4.155465969	
4	48700000	49050000	46	3.150770823	9.678719645	16	34.78	5.723722386	CWH43
4	48750000	49050000	44	3.202207747	9.634506659	15	34.09	5.825812769	CWH43
4	53050000	53200000	39	1.902661839	4.624853822	2	5.13	4.490707356	
4	53050000	53200000	39	1.910885737	4.498371579	2	5.13	4.364813003	
5	108950000	109050000	52	1.230977489	4.804747752	2	3.85	4.496083367	MAN2A1
5	49800000	49900000	20	1.115680663	7.0189968	2	10	5.709014462	
5	49950000	50100000	28	1.43425298	4.806390819	2	7.14	4.647364117	PARP8
6	130450000	130600000	77	1.210066108	4.223072992	2	2.6	4.128926309	L3MBTL3
6	27300000	27450000	124	0.765212288	4.501311917	2	1.61	4.330395554	ZNF184
6	32100000	32250000	134	0.72219586	4.097914131	3	2.24	4.067112422	AGER,PPT2-EGFL8,PRRT1,RNF5
6	32900000	33100000	280	1.091553535	7.469182443	9	3.21	5.490976902	BRD2
6	33550000	33700000	77	0.912070126	4.438191565	2	2.6	4.393496651	GGNBP1
6	72750000	72900000	36	1.551315858	4.543252261	2	5.56	4.340487799	RIMS1
7	57300000	57500000	86	1.655085686	5.890428035	4	4.65	4.863426346	LOC100653233
8	5700000	5850000	146	1.140198538	5.071320086	3	2.05	4.566104989	
9	71100000	71250000	88	0.908615877	4.827692614	3	3.41	4.478565671	LINC01506
9	8650000	8800000	109	0.894380061	4.149332682	2	1.83	4.089719998	PTPRD
10	38350000	38450000	25	1.298913943	4.667349325	2	8	4.621562105	ZNF33A
10	42900000	43000000	27	1.310846616	4.288275685	2	7.41	4.165444885	CCNYL2
11	48150000	48300000	41	1.354093643	4.599073946	3	7.32	4.348135911	OR4B1
11	56100000	56250000	46	1.575064581	4.609464455	3	6.52	4.503466932	OR5M3
11	56350000	56500000	41	1.398409538	5.531709532	2	4.88	4.806129428	OR5AP2
11	56600000	56750000	60	1.34032891	7.981549858	5	8.33	5.833764127	LOC101927120

11	56950000	57250000	115	1.063569768	5.931992951	7	6.09	4.757375915	APLNR
12	131300000	131450000	75	0.944562967	4.698255527	2	2.67	4.385097415	ADGRD1
12	33050000	33250000	107	0.797533289	6.119821667	4	3.74	4.737973316	
12	39350000	39500000	58	1.294181713	5.085780215	4	6.9	4.543693529	
12	79650000	79800000	44	2.144112903	4.898411781	4	9.09	4.477506792	SYT1
12	83000000	83200000	69	1.870507127	5.135668829	11	15.94	4.424378943	TMTC2
13	30450000	30600000	125	1.365026555	4.597530507	2	1.6	4.316182676	LINC00297
13	69700000	69850000	74	0.639506147	4.656338885	2	2.7	4.502547346	LINC00383
15	64100000	64250000	82	0.702983665	4.138512463	2	2.44	4.079861576	DAPK2
16	16850000	17000000	90	0.758980483	5.034072777	2	2.22	4.947278131	
16	16850000	17000000	90	0.786476338	5.367451859	2	2.22	5.284775058	
16	31500000	31600000	28	1.226269632	4.902847213	2	7.14	4.592349075	AHSP
16	31500000	31600000	28	1.29049446	5.285881371	2	7.14	4.9579067	AHSP
16	46850000	47000000	19	2.354559188	6.367533801	2	10.53	5.544366915	C16orf87
16	46850000	47000000	19	2.446513327	6.523789234	2	10.53	5.636391208	C16orf87
16	48050000	48250000	73	1.722489209	4.700112881	5	6.85	4.408716573	
16	48050000	48300000	91	1.633691087	4.821172561	7	7.69	4.395936583	
16	48400000	48550000	39	0.976948705	5.092935787	2	5.13	4.701992276	SIAH1
16	48400000	48550000	39	1.011815665	5.076874103	2	5.13	4.84585029	SIAH1
17	45250000	45400000	64	0.975194911	5.097058665	2	3.12	4.725648134	CDC27
18	51400000	51500000	42	0.756547259	4.544152527	2	4.76	4.385690107	
18	51700000	51850000	45	1.059302486	4.581312293	2	4.44	4.398534193	MBD2
19	24250000	24400000	22	1.52258822	4.364304224	2	9.09	4.362668815	HAVCR1P1
19	28400000	28750000	176	1.358799842	7.997799522	12	6.82	5.386802318	
19	38650000	38950000	98	1.923760662	5.220943199	10	10.2	4.741993517	C19orf33,SPINT2,SPRED3,YIF1B
20	25000000	25200000	81	1.61656031	4.741950016	7	8.64	4.288741747	ACSS1

Kenya Population

CHR	START	END	N_MRK	MEAN_MRK	MAX_MRK	N_EXTR_MRK	PERC_EXTR_MRK	MEAN_EXTR_MRK	GENES
1	1,2E+08	1,21E+08	244	1.23924441	9.207269478	21	8.61	5.395887065	ADAM30
1	13700000	13900000	112	1.006537535	8.610016039	5	4.46	6.383962662	LRRC38
1	37800000	37950000	81	1.585990212	4.758538654	4	4.94	4.413681454	LINC01137
2	31500000	31650000	149	0.828221253	5.205089649	2	1.34	4.683460612	XDH
2	57600000	57800000	125	0.815782209	5.657840569	4	3.2	4.744581197	
2	57850000	58000000	75	0.793331644	5.02867939	2	2.67	5.017382372	
3	1,35E+08	1,35E+08	99	0.642018182	4.106470577	2	2.02	4.106470577	
4	1,08E+08	1,08E+08	36	1.768351628	4.882067254	3	8.33	4.727681692	DKK2
4	48750000	49050000	44	3.286675462	8.420192971	13	29.55	5.884550137	CWH43
4	53050000	53200000	41	1.877360072	5.014657784	3	7.32	4.458288372	
5	1,19E+08	1,19E+08	193	0.817452197	5.710117227	9	4.66	4.883876453	DMXL1
5	49800000	49900000	22	1.204616041	5.401044628	2	9.09	5.364228026	
6	1,06E+08	1,07E+08	86	1.189903484	4.997095861	4	4.65	4.538065235	
6	1,06E+08	1,07E+08	86	1.194267106	4.992215617	4	4.65	4.569620652	
6	27150000	27450000	184	1.146190392	7.27453985	9	4.89	5.327888193	POM121L2
6	27300000	27450000	139	1.118966706	7.123273567	4	2.88	6.5683838	ZNF184
6	28150000	28300000	82	0.780240544	4.883208327	3	3.66	4.465227787	NKAPL
6	28150000	28300000	82	0.791679416	4.874153327	3	3.66	4.587421891	NKAPL
6	30250000	30400000	151	0.979108519	4.783881699	2	1.32	4.461458103	HCG17
6	30250000	30400000	151	1.006238116	4.828421376	2	1.32	4.525282393	HCG17
6	31150000	31400000	404	0.707813238	5.057786189	6	1.49	4.343723828	HCG27
6	31150000	31400000	404	0.728817405	5.551138144	7	1.73	4.611260025	HCG27
6	32100000	32250000	133	0.68580919	4.981050132	2	1.5	4.929279043	AGER,PPT2-EGFL8,PRRT1,RNF5
6	32100000	32250000	133	0.696067183	5.079565134	2	1.5	5.02686905	AGER,PPT2-EGFL8,PRRT1,RNF5
6	32700000	32850000	261	0.565104279	4.936813133	2	0.77	4.890355491	HLA-DOB
6	32700000	32850000	261	0.585020561	4.939961901	3	1.15	4.685165183	HLA-DOB
6	32900000	33100000	286	1.771573289	8.515552687	16	5.59	5.467084072	BRD2
6	32900000	33100000	286	1.804338534	8.575718992	18	6.29	5.429542555	BRD2
6	33550000	34000000	394	1.171215998	7.865997186	17	4.31	5.042649721	GGNBP1,MLN,UQCC2
6	33550000	34000000	394	1.195961904	8.17140767	22	5.58	4.949935341	GGNBP1,MLN,UQCC2
6	34950000	35050000	31	1.856180891	4.758993006	2	6.45	4.446461579	ANKS1A
6	35300000	35450000	90	1.298359293	6.940862283	4	4.44	5.417754235	FANCE
6	35300000	35450000	90	1.326360095	7.08742115	6	6.67	5.041636441	FANCE
6	35650000	35750000	51	0.851547224	4.22126508	2	3.92	4.162904409	ARMC12
6	35650000	35750000	51	0.873408561	4.313865196	2	3.92	4.310044736	ARMC12
6	62250000	62350000	20	1.497533825	4.354812329	2	10	4.188536426	MTRNR2L9
6	62250000	62350000	20	1.527888587	4.356863569	2	10	4.190447239	MTRNR2L9
7	1,18E+08	1,18E+08	65	0.797459694	4.576833842	2	3.08	4.394297928	

7	57350000	57500000	71	1.737353334	4.808452677	5	7.04	4.48312335	LOC100653233
7	650000	800000	45	1.435611312	4.814608807	2	4.44	4.562343475	DNAAF5
9	21600000	21750000	70	1.007791332	4.095464273	2	2.86	4.086862966	
9	71100000	71200000	54	1.088123676	4.824332258	2	3.7	4.438189806	LINC01506
10	73700000	73850000	123	0.85034793	4.540358049	4	3.25	4.365823459	CHST3
10	85950000	86100000	89	1.065015497	4.488958738	2	2.25	4.244872374	CCSER2
11	15100000	15250000	110	0.707488372	4.298191119	2	1.82	4.251848071	CALCB
11	46250000	46400000	59	1.071807935	4.865558488	2	3.39	4.807695127	CREB3L1
11	48200000	48300000	30	1.497654807	4.704253943	2	6.67	4.581021404	OR4B1
11	56600000	56750000	65	1.11875623	4.359020263	3	4.62	4.215025845	LOC101927120
11	56900000	57000000	54	0.943858016	4.632082761	2	3.7	4.321412229	LRRC55
12	32900000	33000000	74	0.624668265	5.199974587	2	2.7	4.964590503	PKP2
12	32900000	33000000	74	0.636382093	5.182314921	2	2.7	4.949271485	PKP2
12	59500000	59650000	46	0.909945005	4.049408725	2	4.35	4.040007411	
12	59500000	59650000	46	0.945574818	4.232112231	2	4.35	4.182711232	
12	83000000	83150000	47	1.939608571	4.706926596	2	4.26	4.390567641	TMTC2
12	83000000	83150000	47	2.024996019	4.961062141	4	8.51	4.389645066	TMTC2
13	30450000	30600000	133	1.467165002	4.448848954	3	2.26	4.185728205	LINC00297
15	64100000	64250000	88	0.815692091	4.560172125	2	2.27	4.518209064	DAPK2
15	77200000	77350000	44	1.61286437	5.034384558	5	11.36	4.574352347	PSTPIP1
16	16200000	16350000	77	0.982031488	4.941657606	4	5.19	4.536583355	
16	31350000	31500000	70	1.658375387	8.24981815	4	5.71	6.330315248	ARMC5
16	46850000	47000000	24	2.460636233	6.174936059	3	12.5	5.17621632	C16orf87
16	48400000	48550000	42	1.346028745	7.674884537	3	7.14	6.104761694	SIAH1
16	48950000	49050000	57	0.750202831	5.535581332	2	3.51	5.356132463	
17	21050000	21200000	73	0.69983166	4.853697858	3	4.11	4.54260084	DHRS7B
17	3450000	3550000	70	1.815972045	5.315975454	2	2.86	4.783380272	CTNS
18	51350000	51500000	61	0.718635081	5.928819881	3	4.92	4.831491538	
18	51700000	51850000	48	1.227796749	5.797881975	2	4.17	4.9324225	MBD2
19	24300000	24450000	27	1.491572774	5.391354354	2	7.41	4.785436488	HAVCR1P1
19	28400000	28550000	68	1.334231374	5.684126839	3	4.41	5.139440601	
19	38650000	38950000	129	1.744402554	5.28799861	15	11.63	4.626993376	C19orf33,SPINT2,SPRED3,YIF1B
20	47350000	47500000	69	1.268762972	4.797027395	2	2.9	4.663607665	PREX1
22	25850000	26000000	61	1.09419038	4.449646228	3	4.92	4.287596512	CRYBB2P1
22	43000000	43150000	100	1.06289456	4.459378219	2	2	4.377735758	A4GALT

Gambia population

CHR	START	END	N_MRK	MEAN_MRK	MAX_MRK	N_EXTR_MRK	PERC_EXTR_MRK	MEAN_EXTR_MRK	GENES
1	10150000	10300000	56	2.357929035	6.548680181	4	7.14	6.290775922	KIF1B
1	1,45E+08	1,45E+08	17	9.025520482	116.5747897	3	17.65	43.3734618	PDE4DIP
1	1,46E+08	1,47E+08	65	0.453824975	8.436417528	2	3.08	8.264273642	LOC728989
1	1,48E+08	1,48E+08	44	1.336311272	7.274513257	2	4.55	6.760272984	LOC101927468
1	1,89E+08	1,89E+08	84	1.031883117	7.536266889	2	2.38	6.942904951	
1	2,08E+08	2,08E+08	62	2.266725031	8.629966449	9	14.52	6.706432422	CR1
1	2,26E+08	2,26E+08	91	0.611045061	9.791997212	2	2.2	8.114546517	DNAH14
1	43200000	43350000	89	2.517511134	7.763341638	3	3.37	7.575278992	C1orf50
1	43950000	44100000	79	2.875751981	6.192392716	4	5.06	6.08462017	PTPRF
1	85850000	86000000	106	2.554984692	10.64388409	11	10.38	8.144738679	DDAH1
2	1,05E+08	1,05E+08	112	2.76403186	10.43757004	9	8.04	8.114805512	LINC01102
2	1,11E+08	1,11E+08	67	2.38130589	14.32820927	5	7.46	10.50946077	MALL
2	1,54E+08	1,54E+08	88	1.210566149	6.570622842	2	2.27	6.336803052	
2	1,62E+08	1,62E+08	92	3.315705239	12.39245623	19	20.65	8.128600386	TANK
2	2,12E+08	2,12E+08	297	2.691691955	10.52362498	25	8.42	7.900089785	ERBB4
2	2,14E+08	2,14E+08	62	2.075780492	7.114037392	5	8.06	6.586199188	LOC100130451
2	81550000	81700000	85	2.021515039	11.58992342	12	14.12	8.005132471	LINC01815
2	88300000	88450000	119	1.474909136	14.40820473	5	4.2	11.12711543	FABP1
3	1,17E+08	1,17E+08	114	1.252074207	13.56521349	2	1.75	10.15150747	
3	1200000	1350000	246	0.829278078	6.750525494	2	0.81	6.703066246	CNTN6
3	1,31E+08	1,32E+08	289	4.398071521	34.252135 5	65	22.49	13.03569461	CPNE4
3	1400000	1600000	224	1.995818257	8.938953575	19	8.48	6.840585793	CNTN6
3	1,45E+08	1,45E+08	61	4.240522491	7.224164131	6	9.84	6.492603854	
3	14750000	14900000	104	1.288090367	9.389030575	3	2.88	7.429181861	C3orf20
3	24350000	24550000	160	2.323520827	8.181894157	22	13.75	6.933704205	THRB
3	62450000	62600000	131	2.142608061	8.382202897	3	2.29	7.217557259	CADPS
3	63000000	63150000	112	3.800891412	6.962035498	7	6.25	6.486678328	LINC00698
4	1,08E+08	1,08E+08	91	1.374078806	10.87167432	7	7.69	8.637963501	
4	52900000	53150000	79	6.326274145	42.00306092	21	26.58	15.626897	SGCB
4	8600000	8750000	86	0.946384569	30.70157985	2	2.33	29.94064029	CPZ
5	11250000	11400000	131	0.529850381	17.59880174	2	1.53	11.81943904	CTNND2
5	1,2E+08	1,2E+08	102	0.67668529	7.060880371	2	1.96	6.822122517	
5	1,69E+08	1,7E+08	223	1.046266263	8.927585976	11	4.93	7.433954924	DOCK2
5	1,78E+08	1,79E+08	94	1.650807734	6.628901047	2	2.13	6.371483667	GRM6
5	25600000	25750000	103	1.539712115	6.174929579	2	1.94	6.161689488	
5	46050000	46250000	42	7.928237544	30.27191205	15	35.71	15.75413837	
5	49750000	49950000	46	11.64957579	101.1791448	21	45.65	21.36009746	
5	50100000	50250000	50	1.481453799	12.71871203	3	6	10.63923224	PARP8

5	67700000	67850000	118	1.592225667	7.618474883	3	2.54	7.032764265	
5	70700000	70850000	37	2.085682938	12.15595381	5	13.51	9.754704753	BDP1
5	87000000	87200000	79	7.911129834	20.5016422	43	54.43	12.16109756	
5	9100000	9250000	150	0.885813672	10.30032723	7	4.67	7.150539763	SEMA5A
6	1,11E+08	1,11E+08	149	1.403392948	20.19184432	5	3.36	15.33892081	DDO
6	29650000	29800000	195	3.124936642	9.983239915	15	7.69	6.880434358	HCG4
6	29850000	30150000	383	3.847364001	25.65620512	59	15.4	16.476902	HCG4B,TRIM31-AS1, TRIM40, ZNRD1, ZNRD1ASP
6	32100000	32250000	172	2.624680412	26.138532	20	11.63	11.75057895	AGER,PPT2-EGFL8, PRRT1, RNF5
6	32750000	33000000	440	2.128380023	9.092284352	16	3.64	6.823175199	BRD2
7	1,1E+08	1,1E+08	77	1.428588102	7.670319875	2	2.6	7.039283009	
7	1,28E+08	1,28E+08	114	1.485742114	6.735933133	4	3.51	6.467404811	
7	1,31E+08	1,31E+08	85	1.957522008	9.138680158	8	9.41	7.242712683	MKLN1
7	1,39E+08	1,39E+08	98	1.834640287	11.71857055	7	7.14	7.720585534	HIPK2
7	1,5E+08	1,5E+08	122	0.775323949	9.297198667	2	1.64	7.696832898	GIMAP8
7	1,58E+08	1,59E+08	92	1.081535421	12.89747111	2	2.17	9.917717426	ESYT2
7	45100000	45250000	98	1.126025532	7.423944855	2	2.04	7.085850572	CCM2
7	45450000	45600000	82	1.999414151	7.5414827	2	2.44	7.122299938	
7	57450000	57600000	35	1.934770316	6.410066455	3	8.57	6.35103569	LOC100653233
8	1,05E+08	1,06E+08	76	1.772420523	6.644624757	3	3.95	6.240644787	DPYS
8	1,09E+08	1,09E+08	92	3.139622825	10.77585982	7	7.61	7.825004383	RSPO2
8	1,09E+08	1,09E+08	80	3.920852914	9.685846269	13	16.25	8.058756097	EIF3E
9	1,19E+08	1,2E+08	115	0.596671309	9.433197184	2	1.74	9.09031821	ASTN2
9	29350000	29500000	101	2.336256706	7.231047896	2	1.98	7.155703293	
9	38650000	38800000	150	0.597115412	19.06503162	3	2	17.24476926	
9	8350000	8500000	200	1.192341597	8.326434418	3	1.5	8.025658139	PTPRD
10	1,29E+08	1,29E+08	92	0.52098301	9.209427049	3	3.26	8.184531591	DOCK1
10	1,29E+08	1,29E+08	235	0.568309784	6.434869579	2	0.85	6.364937801	DOCK1
10	33200000	33350000	71	3.078676286	7.556451185	8	11.27	6.942302889	ITGB1
10	47600000	47750000	112	1.465446353	23.37813259	2	1.79	17.61865715	ANTXRL
10	49350000	49550000	97	5.666777162	32.86084586	32	32.99	14.2677563	FRMPD2
10	66450000	66650000	120	3.262366568	11.59110743	22	18.33	7.505105123	ANXA2P3
10	87300000	87450000	86	2.222177728	6.749389778	2	2.33	6.436935647	GRID1
10	97250000	97400000	133	3.456503337	23.08619452	29	21.8	9.193154505	ALDH18A1
11	3300000	3450000	61	2.466502925	26.73740774	3	4.92	15.20845286	OR7E12P
11	4450000	4650000	230	3.137507762	9.017928686	28	12.17	7.436629962	C11orf40
11	4700000	4900000	196	2.681789784	10.32162725	40	20.41	8.378871645	OR51E2
11	47200000	47350000	44	4.799768446	6.784867281	7	15.91	6.393425121	ACP2
11	48050000	48250000	89	4.59803849	9.065322426	14	15.73	7.149066579	MIR3161
11	5200000	5550000	463	1.450578941	7.386797319	7	1.51	6.46753622	BGLT3,OR51B6, OR51I1,OR51I2, OR51M1,OR51Q1, OR51V1,OR52D1,UBQLN3

11	55100000	55250000	40	3.709707338	6.758671764	5	12.5	6.534995778	OR4A15
11	5650000	5850000	227	1.788825113	11.62447122	19	8.37	7.399837013	OR52N1
11	87150000	87300000	104	1.250232255	7.736235032	5	4.81	7.035422461	
12	12350000	12500000	111	1.594612727	15.36096498	2	1.8	11.6879037	LRP6
12	38300000	38450000	30	2.307143897	8.293754561	2	6.67	7.592349187	
12	94600000	94750000	124	1.383557465	11.26793538	9	7.26	7.299577781	CEP83
12	94800000	94950000	119	1.497291405	18.05071338	6	5.04	10.46096298	CEP83
13	30550000	30700000	167	2.373124915	7.308299746	3	1.8	6.968007696	LINC00365
13	36350000	36500000	159	0.873368572	12.41015491	6	3.77	9.937549184	DCLK1
13	84000000	84200000	135	2.163767179	6.727405696	12	8.89	6.276571337	
13	84250000	84400000	76	1.549065415	8.809281687	2	2.63	7.709166531	
13	86950000	87150000	108	2.190711557	7.883990129	15	13.89	6.836277554	
14	23150000	23300000	106	0.641596242	7.598184568	2	1.89	7.148305533	MRPL52
14	32350000	32500000	158	2.21907035	12.77916225	12	7.59	10.26850731	
14	40600000	40750000	72	2.891928702	6.982265824	2	2.78	6.573059584	
14	55500000	55650000	62	1.983655297	6.216060674	3	4.84	6.177892971	DLGAP5
14	55750000	55900000	65	3.609475973	6.361186028	2	3.08	6.297761141	ATG14
14	71300000	71450000	79	2.198900508	9.635727473	5	6.33	8.914250941	PCNX1
14	71550000	71700000	107	1.101558328	8.012310772	4	3.74	7.146178483	PCNX1
14	74600000	74800000	113	2.24832698	7.4728723	7	6.19	6.815842261	
14	76600000	76750000	95	0.845955179	9.887026292	2	2.11	8.519003086	GPATCH2L
15	45900000	46050000	137	0.591646656	7.362542769	3	2.19	6.966586144	BLOC1S6
15	50200000	50350000	98	2.240122028	7.622272705	4	4.08	6.879384257	ATP8B4
15	51400000	51750000	251	2.531047214	12.54386554	31	12.35	8.089890417	
15	55050000	55200000	144	0.886865314	17.53010606	4	2.78	9.400886734	
15	62050000	62200000	117	1.067603909	10.82638474	3	2.56	8.870209008	VPS13C
15	80550000	80700000	154	0.974185469	8.036127841	2	1.3	7.034874191	ARNT2
16	58950000	59100000	147	1.639608147	7.891982068	5	3.4	7.204692931	
16	69450000	69600000	75	2.69819737	11.31313356	8	10.67	7.361004709	CYB5B
16	70600000	70750000	67	2.087537727	6.680917382	2	2.99	6.350971027	IL34
16	8000000	8150000	165	0.615302438	12.09912072	2	1.21	9.614762268	
16	81750000	82000000	348	1.051091982	7.960494817	4	1.15	7.268503226	PLCG2
17	46000000	46150000	79	1.565860243	6.394402273	3	3.8	6.216300302	CBX1
17	49850000	50150000	167	3.679481817	9.927968485	46	27.54	7.176787182	CA10
18	26500000	26650000	69	1.710334876	10.96543064	3	4.35	8.927183023	
18	36450000	36600000	70	3.613783253	6.934262709	4	5.71	6.657875791	
18	42700000	42900000	150	2.746025998	6.998441635	12	8	6.622799245	
18	5050000	5200000	169	0.531649736	7.593466029	2	1.18	7.452562626	AKAIN1
18	62950000	63100000	100	1.500052396	11.60262383	4	4	8.715733941	
18	8900000	9050000	155	1.774659472	6.920954161	2	1.29	6.528682453	
19	24300000	24500000	50	3.701459545	17.14509528	15	30	8.990637721	HAVCR1P1
20	36950000	37100000	109	0.904887331	7.43885515	2	1.83	7.162426691	BPI

20	41000000	41250000	241	2.496157096	8.509566399	21	8.71	6.891185439	PTPRT
20	47450000	47600000	77	1.628112508	8.400972853	8	10.39	7.260609438	ARFGEF2
22	26550000	26700000	139	0.842186462	10.66951579	4	2.88	7.670038555	SEZ6L
22	35900000	36200000	227	4.058392992	31.55254939	31	13.66	16.17585197	APOL5
22	36750000	37000000	189	2.547913991	11.13002418	18	9.52	8.56005454	CACNG2