**GENOME-WIDE SCREENING FOR GENETIC ASSOCIATIONS WITH SEVERE MALARIA IN THREE MALARIA ENDEMIC REGIONS IN CAMEROON**

**ESOH KUM KEVIN**

**A Research Proposal Submitted in partial Fulfillment of the Requirements for the Degree of Master of Science in Molecular Biology and Bioinformatics of Jomo Kenyatta University of Agriculture and Technology**

**May 2018**

**DECLARATION**This research proposal is my original work and has not been presented elsewhere for a degree award

Signature………………………………………… Date: ………………………………………..

Esoh Kum Kevin (HSB316-6872)

**Declaration by Supervisors**

This Research Proposal has been submitted for examination with our approval as supervisors.

1. Dr. Steven Ger Nyanjom, Department of Biochemistry, JKUAT, Kenya

Signature: …………………………………………… Date: …………………………………….

2. Dr. Tobias Apinjoh, Department of Biochemistry and Molecular Biology, University of Buea, Cameroon

Signature: …………………………………………… Date: …………………………………….

3. Prof. Ambroise Wonkam, Division of Human Genetics, Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town (UCT), Cape Town, South Africa

Signature: …………………………………………… Date: ……………………………………

4. Dr. Emile Chimusa, Division of Human Genetics, Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town (UCT), Cape Town, South Africa

Signature: …………………………………………… Date: ……………………………………

**Table of Contents**

[**DECLARATION** i](#_Toc513211104)

[**List of Tables** iv](#_Toc513211105)

[**List of Abbreviations** v](#_Toc513211106)

[**Abstract** vi](#_Toc513211107)

[**Chapter One: Introduction** 7](#_Toc513211108)

[**1.1.** **Background Information** 7](#_Toc513211109)

[**1.2.** **Problem Statement** 8](#_Toc513211110)

[**1.3.** **Justification (Rationale)** 9](#_Toc513211111)

[**1.4.** **Research Questions** 10](#_Toc513211112)

[**1.5.** **Hypothesis** 10](#_Toc513211113)

[**1.6.** **Objectives** 10](#_Toc513211114)

[**1.6.1.** **General Objective** 10](#_Toc513211115)

[**1.6.2.** **Specific Objectives** 10](#_Toc513211116)

[**Chapter Two: Literature Review** 11](#_Toc513211117)

[**Chapter Three: Materials and Methods** 14](#_Toc513211118)

[**3.1.** **Study Area** 14](#_Toc513211119)

[**3.2.** **Research Design** 15](#_Toc513211120)

[**3.3.** **Ethical Clearance** 15](#_Toc513211121)

[**3.4.** **Sampling design and sample size determination** 16](#_Toc513211122)

[**3.4.1.** **Inclusion Criteria** 16](#_Toc513211123)

[**3.4.2.** **Exclusion Criteria** 16](#_Toc513211124)

[**3.4.3.** **Sample size determination** 16](#_Toc513211125)

[**3.5.** **Experimental techniques** 17](#_Toc513211126)

[**3.5.1.** **Data collection** 17](#_Toc513211127)

[**3.5.2.** **Molecular Analysis** 17](#_Toc513211128)

[**3.6.** **Data Analysis** 18](#_Toc513211129)

[**3.6.1.** **Genotyping Quality Control** 18](#_Toc513211130)

[**3.6.2.** **Assessment of population structure** 18](#_Toc513211131)

[**3.6.3.** **Haplotype phasing and genotype imputation** 18](#_Toc513211132)

[**3.6.4.** **Association testing** 18](#_Toc513211133)

[**Time Frame/Work Plan** 19](#_Toc513211134)

[**Budget** 19](#_Toc513211135)

[**REFERENCES** 20](#_Toc513211136)

# **List of Tables**

[Table 1. Work plan 19](#_Toc513211137)

[Table 2. Budget 19](#_Toc513211138)

# **List of Abbreviations**

ACT: Artemisinin-based Combination Therapy

AMA: Apical Membrane Protein

CD: Cluster of Differentiation

CDC: Center for Disease Control and Prevention

CRISPR:  Clustered Regularly Interspaced Short Palindromic Repeats

CSP: Circomsporozoite Protein

DBR: Duffy-binding Receptors

EBL: Erythrocyte-binding Ligand

GDP: Gross Domestic Product

GPI: Glycosylphosphatidylinositol

GWAS: Genome Wide Association Study

HSPG: Heparan Sulphate Proteoglycans

ICAM: Intercellular Cell Adhesion Molecule

LD: Linkage Disequilibrium

MSP: Merozoite Surface Protein

OR: Odds Ratio

PC: Principal Components

PCA: Principal Component Analysis

PfEMP: *Plasmodium falciparum* Erythrocyte Membrane Protein

RON: Rhoptry Neck Protein

SNP: Single Nucleotide Polymorphism

TGF: Transforming growth factor

WHO: World Health Organization

# **Abstract**

Malaria claims the lives of about half a million people yearly. Severe *Plasmodium falciparum* malaria accounts for more than 90% of these deaths. Africans, particular children below the age of 5 are the most vulnerable. With increasing antimalarial and insecticide resistance, and no effective vaccine against all parasite strains, the disease threatens to escalate.

Understanding the mechanisms of disease resistance and susceptibility in humans is crucial to uncovering the molecules involved in host-parasite interaction. And understanding the host-parasite interface is crucial to developing new intervention strategies (drugs and vaccines). This study therefore seeks to screen for genetic determinants of resistance to severe malaria in the genome of individuals living in three malaria endemic regions of Cameroon including the Center, Littoral and South West regions. A genome wide association analyses will be carried out using linkage disequilibrium-based mapping, haplotype phasing and imputation using the 1000 Genomes reference panel. Population structure will be assessed by principal component analysis (PCA) and association testing of phased and imputed genotypes will be performed frequentist and Bayesian approaches.

At the end of the study, we expect to find novel genetic variations associated with severe malaria in Cameroon. This would be particularly important in gaining understanding into host-parasite interaction which will in turn be crucial in informing development of novel intervention strategies.

# **Chapter One: Introduction**

## **Background Information**

Malaria is a complex disease that accounts for many deaths worldwide annually. It is caused by protozoans of the genus *Plasmodium*, which are transmitted by the female *Anopheles* mosquito to humans. More than 90% of world malaria burden lies in Africa, including Cameroon where about 90% of the ~23 million people are at risk of the disease (Mbenda *et al.,* 2013). Although worldwide incidence and death rates reduced in 2016, the number of cases recorded increased in a number of countries, including Cameroon, from 211 million in 2015 to 216 million in 2016 (WHO, 2017). Parasite resistance to antimalarials and mosquito resistance to insecticides are major contributing factors to this increase (WHO, 2017). Children below 5 years old are the most vulnerable, recording the highest number of deaths. *Plasmodium falciparum* is the most virulent species among the five parasite species that are known to infect humans. It is the most common in Africa, including Cameroon where it accounts for about 99.7% of all malaria cases (Kwenti et al., 2017).

Severe malaria (SM), asymptomatic malaria and uncomplicated malaria (UM) are major outcomes of the disease that are due to the interaction of *P. falciparum* with the human immune system (Acharya *et al.,* 2017). SM is life threatening and claims the lives of hundreds of thousands of children yearly. It manifests predominantly as severe malaria anemia (SMA), cerebral malaria (CM) and respiratory distress (RD) (Acharya *et al.,* 2017; Wassmer *et al.*, 2015; WHO, 2014; Achidi *et al.,* 2012). The fatal outcomes of the disease arise from several complex processes. These include the release of hemozoin and other proinflammatory molecules like glycosylphosphatidylinositol (GPI) when the parasites feed on hemoglobin in human red blood cells (erythrocytes). Clogging of blood vessels due to adherence of infected erythrocytes on vascular tissues (cytoadherence) and adherence of infected erythrocytes with uninfected erythrocytes (rosseting) lead to impaired blood flow to tissues and organs of the body like the brain, which can lead to coma (Acharya *et al.,* 2017; Smith *et al.,* 2013).

Prompt diagnosis and effective treatment are important components of malaria control and elimination strategies. Prompt diagnosis is achieved by the use of rapid diagnostics test (RDT) kits while microscopy is the gold standard for diagnosis and confirmation of diagnosis following RDT (CDC, 2018; WHO, 2015). Effective treatment is based on rational use of antimalarial agents. The artemisinin-based combination therapies (ACTs) are the first line of action against malaria. Severe malaria is treated by intravenous or intramuscular injection of artesunate for at least 24 hours and until the patients can tolerate oral medication (WHO, 2015).

Malaria remains a global public health burden as no licensed vaccine against the disease exists to date and there are increasing concerns of drug resistance. Mosqirix (RTS,S), the most advanced vaccine candidate in the vaccine development pipeline has limited efficacy (56% in children aged between 5-17 months) and a few cases of adverse events like meningitis and seizures were recorded during vaccination periods (Wilby *et al.*, 2012). The difficulty in developing a vaccine that is highly efficacious against all *P. falciparum* strains is largely due to the enormous parasite genomic diversity, which allow them to evade the host immune system (Dinko and Pradel, 2016). There is limited understanding of the parasite interactions with host cells that allow them to evade the host immune system (Dundas *et al.*, 2018).

An in-depth understanding of the interactions that parasite proteins make with the host is crucial to developing more effective intervention strategies. This study aims at screening for variations in the genome of Cameroonian case and control subjects by a genome-wide association (GWA) analysis, which may uncover some novel proteins important in host parasite interactions, or enhance characterization of already known but poorly characterized proteins.

## **Problem Statement**

Severe malaria is a major factor in the morbidity and mortality caused by malaria in endemic countries. The magnitude of the burden of the disease is reflected in its impact on the long term economic growth and development of countries affected including Cameroon. The large economic cost of prevention, treatment and loss of productivity due to disease-related morbidity and mortality play significant roles in reducing the gross domestic product (GDP) of these highly burdened countries (Malaney *et al.*, 2004). The disease is highly concerning in these endemic regions of Cameroon whose most affected populations are poor, have low levels of education and engage largely in farming activities (Tonye *et al.*, 2018). They therefore come into frequent contact with disease-causing parasites with little knowledge of prevention and treatment methods. In fact, Tonye *et al.* (2018) in their malaria survey study in Cameroon showed that only 9% of the participants in rural areas had access to insecticide-treated bed nets (ITNs) while only 12% of children below 5 years old with fever in the last two weeks received ACT. Malaria is a major cause of morbidity in these regions.

Malarial drug resistance and the absence of a licensed and highly effective vaccine are major shortcomings to the fight against malaria. *P. falciparum* parasites have incorporated several mutations in their genome to render the effect of antimalarials less effective (*Amato et al.*, 2016; Soe *et al.*, 2016; Nabet *et al.*, 2016). Some mutations have enabled the parasites to elude the host immune system. They have therefore continued to wreak havoc on endemic populations. Parasites have also developed alternative pathways to invade host cells as a result of the enormous diversity in their genome (Acharya *et al.,* 2017). Because the full picture of host and parasite interactions is not well understood, developing a highly effective intervention strategy has been a great challenge.

Therefore, if the gap in knowledge of the host-parasite interactions that are crucial to parasite survival or host resistance/susceptibility to disease is not bridged, then malaria may continue to cause significant damage on the human race.

## **Justification (Rationale)**

*P. falciparum* exerts great pressure on the human genome that is imparting genetic variations. Some variations occurring on genes encoding host proteins involved in interactions with parasites confer resistance while others increase the risk of malaria in some individuals. Information gained from sequencing the entire *P. falciparum* strain 3D7 genome shows that much of the parasite genes and their products are dedicated towards immune evasion and host-parasite interaction (Gardner *et al.*, 2002). However, the larger part of these interactions remains uncharacterized (Dundas *et al.*, 2018). An in-depth understanding of the human genetic factors that are involved in conferring disease susceptibility/resistance is therefore crucial in closing this knowledge gap and to finding better intervention strategies.

Genome wide association studies can permit scanning for such variations. Specific non-randomly occurring genotypes can then be linked with some disease phenotypes. In one such study, the MalariaGEN consortium (2015) found a significant association of a variation with severe malaria resistance in Gambia, Kenya and Malawi. However, the association could not replicate in other populations, including Cameroon and Tanzania, in the replication phase of the study. Since replication involves genotyping a specific region of interest in the genome of samples rather than the whole genome, it is possible that some genetic variations with significant association to severe malaria may have occurred in other regions of the genome of the Cameroon malaria endemic population that were not genotyped in this replication study. This reasoning is supported by a recent study that found some novel significant associations in different regions of the genome of Tanzanian malaria endemic populations (Ravenhall *et al.*, 2018).

Further, the MalariaGEN consortium raised a potential confounder, the very large and diverse dataset in their study, which could negatively impact on imputation accuracy. They therefore recommended regional data analysis. Hence, the need to screen for novel variations in the Cameroon population.

## **Research Questions**

* Is there any difference in structure of the malaria endemic populations in Cameroon?
* Are there novel genetic variations with high signal of association with severe *P. falciparum* malaria in individuals from these zones?
* If present, what are the strengths of their correlation (LD) and their strengths of association?

## **Hypothesis**

There are no novel genetic variations with high signal of association with severe malaria in the genome of individuals from malaria endemic zones in Cameroon.

## **Objectives**

### **General Objective**

To screen for novel genetic variations that may be associated with severe malaria in individuals with *P. falciparum* malaria from Cameroon.

### **Specific Objectives**

1. To determine the structure of the malaria endemic populations from Cameroon.
2. To screen for genetic variations in the genome of the case and control subjects of these populations.
3. To determine variants, if present, that are significantly associated with severe malaria and their strengths of association.

# **Chapter Two: Literature Review**

## **Introduction**

Before the emergence of GWAS, genetic association studies with malaria relied on candidate gene analysis and family linkage studies (Sakuntabhai *et al.,* 2008; Pearson and Maniolo, 2008). These studies are often hypotheses-driven, relying upon careful selection of variants to study, and understanding of the specific biological pathways that relate genes to diseases (Pearson and Maniolo, 2008). Genome-wide association studies have revolutionized the search for new genes that underlie complex human genetic diseases without prior knowledge of the biological pathways they are involved in (Amos *et al.,* 2011; Spencer *et al.,* 2009; Pearson and Maniolo, 2008). Therefore, GWAS may be capable of revealing many previously unknown genetic changes in the human interface of the host-parasite interaction (apparent in the life cycle of *P. falciparum*) in severe malaria cases.

## **Theoretical Review**

## ***Plasmodium falciparum* Life Cycle**

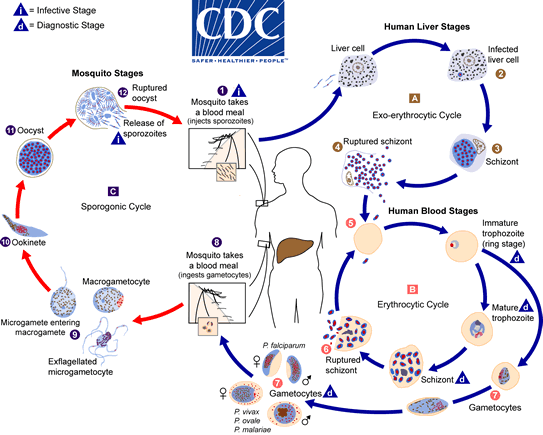


Figure 1. Malaria parasite life cycle

This involves two hosts; the female *Anopheles* mosquito (definitive host) and human host. The mosquito injects the parasite sporozoites into the human blood during a blood meal. Some of the sporozoites deposited in the skin are destroyed while some pass in to the lymphatic system and consequently get presented to T cells in the lymph nodes which mount some protective immune response against them (Acharya *et al.,* 2017). Some sporozoites migrate using their TRAPs which interact with integrins (Dundas *et al.*, 2018) and the cluster of differentiation, CD36 receptors (Soulard *et al.,* 2015; Smith *et al.,* 2013; Kwiatkowski, 2005) on blood vessels and finally reach the liver (Langhorne and Duffy, 2016). The sporozoits interact with heparan sulphate proteoglycans (HSPGs) on liver cells with their circumsprozoit proteins (CSPs), infecting them (Acharya *et al.,* 2017; Langhorne and Duffy, 2016). Contained in protective parasitophorous vesicles, sporozoits differentiate into schizonts containing thousands of merozoites (exo-erythrocytic schizogony) (Soulard *et al.,* 2015). Schizonts later rupture to release merozoites in the blood contained in merosomes (Langhorne and Duffy, 2016), 5-6 days after the hepatic phase (Pounitio *et al.,* 2004). The merozoites spend only 30-60 seconds in the blood circulation before invading red blood cells (Cowman and Crabb, 2006). This presents a short time for proper exposure of parasite antigens to the host immune system, yet another mechanism of evading the host immune system. Some merozoites differentiate into ring stage trophozoites, which mature into schizonts that again rupture releasing merozoites (erythrocytic schizogony). While the erythocytic schizogony cycle continues, a subset of the merozoites differentiate into gametocytes (asexual erythrocytic stages) (CDC, 2018). The *Anopheles* mosquito ingests the gametocytes during a blood meal. The male gametes (microgametes) penetrate the female gametes (macrogametes) in the mosquito stomach generating a zygote. The motile zygote invade the midgut wall of the mosquito and develop into oocysts. The oocysts grow, rupture, and release sporozoites, which migrate to the mosquito’s salivary glands, ready to be injected into a new human host (CDC, 2018).

### **Host-parasite interaction involved in erythrocyte invasion**

Several merozoite proteins interact with erythrocyte receptors during erythrocytes infection. Merozoite surface proteins (MSPs) interact with Duffy binding-like receptors on erythrocyte surface (Jespersen *et al.,* 2016; Kwiatkowski, 2005). Apical membrane protein-1 (AMA-1) interaction with rhoptry neck protein (RON) forms a complex that triggers junction formation and hence invasion (Wright and Rayner, 2014; Cowman and Crabb, 2006). *P. falciparum* erythrocyte binding ligands (EBL) utilize the SA-dependent pathway which involves the erythrocyte binding antigens (EBAs), glycophorins (Kwiatkowski, 2005). EBL-175, EBL-1 and EBL-140 interact with glycophorins A, B and C respectively during parasite invasion (Leffler et al., 2017; Malaria Genomic Network Epidemiology, 2015; Wright and Rayner, 2014). *P. falciparum* erythrocyte membrane protein (PfEMP-1) expressed on the surface of erythrocytes binds CD36 on the surface of endothelial cells, dendritic cells, chondroitin sulfate A (CSA) in the placenta and host of other receptors to cause parasitized erythrocytes to sequester in deep vascular beds and placenta (Dara *et al.,* 2017; Smith *et al.,* 2013). This keeps infected cells away from general circulation hence promoting parasite growth and re-invasion, while shielding parasite from the immune system. PfEMP1 mediated clustering of infected erythrocytes with uninfected erythrocytes (rosetting) promotes re-invasion (Kwiatkowski, 2005). Multiple variations have been reported in these genes which confer resistance to the parasite (Leffler *et al.,* 2017).

### **Approaches employed in the study of host genetic variations and host-parasite interactions**

Severe malaria is a rare outcome in some infected individuals in high malaria transmission zones and several factors have been advanced to this regard including age, immune status and host and parasite genetic makeup (Ravenhall *et al.,* 2018). Approaches that have been employed for over the years to study these genetic changes and host-parasite interactions include; the CRISPR/Cas9 gene editing tool (Dundas *et al.*, 2018), forward genetic screens (Egan *et al.*, 2015), candidate/targeted gene sequencing and analysis (Apinjoh *et al.*, 2014), correlated gene expression studies (Reid and Berriman, 2013), genome-wide association studies (GWAS) (Teo *et al.*, 2010), mass spectrometry-based metabolomic analysis (Olszewski *et al.*, 2009) and immunoscreening (Titanji *et al.*, 2009). Indeed, candidate gene studies have revealed many genetic associations with malaria and severe malaria in Cameroon in addition to the well-established sickle cell trait (HbAS) (Rumaney *et al.,* 2014; Apinjoh *et al.,* 2014, 2013), the ABO blood group system (Apinjoh *et al.,* 2014) and *G6PD* genes (Kavishe *et al.,* 2006). Other variations that have been uncovered in candidate genes studies include transforming growth factor beta (*TGF-beta*)/interleukin-10 (*IL-10*), Toll-like receptor (*TLR*), and Nitrogen Oxide Synthase-2 (*NOS2*) genes that were shown to be associated with protection against malaria in Cameroon (Apinjoh *et al.* 2014, 2013).

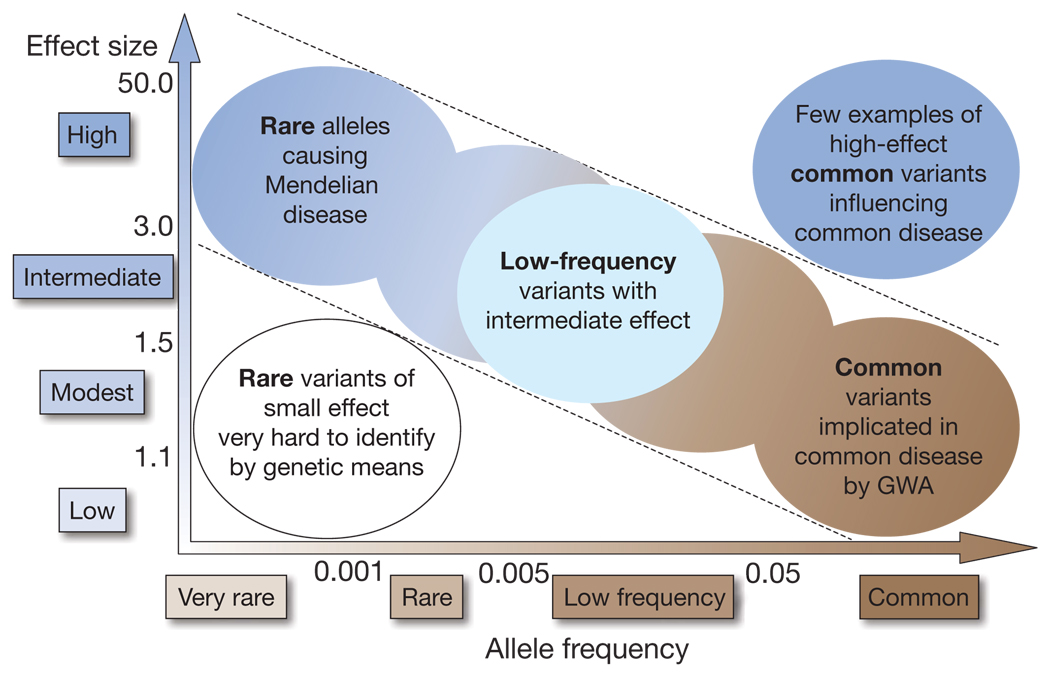
Recent studies have been focused on the genome-wide scale (Teo *et al.*, 2010). GWAS includes three stages; (1) genome-wide scanning of associations, (2) replication of associations and (3) fine mapping of causal variants (*Teo et al.,* 2010). The approach typically uses a case-control study design in which allele frequencies in patients with the disease of interest are compared to those in a disease-free comparison group (Maniolo *et al.,* 2009). DNA samples from subjects are collected and genotyped on high-density genotyping platforms including Illumina and Affymetrix platforms. SNPs that result are quality-checked, and their corresponding chromosomes determined by statistical methods (haplotype phasing) (Delaneau *et al.,* 2013). As no ideal genotyping platform that contains every SNP of the human genome exists, SNPs that were not present and thus were not genotyped are imputed (statistically determined) using a reference haplotype panel including the HapMap and the 1000 Genomes reference panels, maps of human genetic variations (NHGRI, 2015; MalariaGEN, 2015; Band *et al.,* 2013; Howie et al., 2009). GWAS has been used in studying genetic variations of complex diseases of European (Band *et al.,* 2013) and Asian (Teo *et al.,* 2010) ancestries. The power of GWA studies relies on the sample size, linkage disequlibrium (LD), the minor allele frequency (MAF) and effect size (odds ratio - OR) of the SNP that is in association with the disease (Spencer *et al.,* 2009). It has been shown by Spencer *et al.* (2009) that the most powerful GWA studies achieve power between 61% - 83.

## **Empirical Review**

### **Genome wide association study of severe malaria on African Populations**

The enormous genetic diversity of African ancestries pose a great barrier to utilizing GWAS on African populations. The first linkage disequilibrium (LD) mapping-based GWAS of severe malaria in Africa carried out on a sample of Gambian children found considerable population stratification and significant attenuation of association signals due to weak LD (Jallow *et al.,* 2009). However, a multipoint imputation on the *HbS* locus saw a significant increase in the signal of association (*P* = 4 × 10-7 to *P* = 4 × 10-14) (Jallow *et al.,* 2009). Hence the prospects of improving GWAS for African populations by imputation-based analysis was raised. Recent studies of severe malaria on African populations combining LD-mapping and imputation-based approaches have achieved higher resolutions and have uncovered more variants with higher signals of association. Timmann *et al.* (2012) found a locus in the *ATP2B4* gene in Ghanaian samples, while Band *et al.* (2013) found signals in the *HBB* and *ABO* genes following an imputation based meta-analysis. A strong signal of association with severe malaria was found close to the glycophorin gene clusters on chromosome 4 of Gambian, Kenyan and Malawian samples (MalariaGEN, 2015). However, this signal was absent in a replication study on samples from other African Countries including Cameroon, Tanzania, Mali and Burkina Faso (MalariaGEN, 2015). Possible confounding factors for this failure include poor imputation performance of the very large sample size and the enormous population structure and genetic diversity of the populations. Therefore, a regional analysis of the data was recommended. In this light, a more recent study in North-Eastern Tanzania found novel polymorphisms in the interleukin receptor genes *IL-23R* and *IL-12RBR2*, and the kelch-like protein *KLHL3* gene (all P < 10-6) with strong association with resistance to severe malaria (Ravenhall *et al.,* 2018).

### **Conceptual framework**



Genome-wide association studies are usually employed in the study of complex genetic disease and traits. Complex genetic diseases result from the effect of more than one gene affecting a single trait as opposed to those that follow the classical Mendelian ‘one gene, one trait’ pattern (Maniolo *et al.,* 2009).

Clearly, combining LD mapping and imputation in a single GWA analysis increases the power of GWAS on African populations. Therefore, we aim to use this approach to screen for novel genetic variations in populations from Cameroon.

# **Chapter Three: Materials and Methods**

## **Research Design**

## This is an unmatched case-control secondary analysis of human genetic data generated between two different time periods, 2003/2005 and 2007/2008 from three malaria endemic regions in Cameroon (Achidi *et al.,* 2012). The samples were contributed to the MalariaGEN Consortial Project 1 (CP1) alongside data from 11 other African counties. Case-control study designs are typically employed in GWA studies, whereby genetic variations are screened in case and control subjects, and differences in allele frequencies are investigated for association with specific disease phenotypes.

## **Study Area**

The data was collected from four towns across three different malaria endemic regions of Cameroon including Buea and Limbe (South West Region), Douala (Littoral Region) and Yaounde (Central Region).

Yaounde (central region) has fairly constant temperatures of 17- 30oC (mean = 23.1oC), abundant rainfall (1,500–2,000 mm), an average relative humidity index 85% to 90%, characteristic of the rainforest belt of central Africa. It has four distinct seasons: two rainy seasons (March–May/June and September–November) and two dry seasons (December–February and June/July–August) (Mbenda *et al.*, 2014). The maximal malaria transmission period is during and immediately after the two rainy seasons. The Mother and Child Hospital in the region receives patients from neighboring towns and villages including Simbok and Etoa, rural communities whose major activity is farming with fields irrigated by water from the Mefou and Biyeme Rivers. This hospital serves a referral hospital for children and mothers. The Ewondo tribe and part of the Bantu ethnic group are major inhabitants of this region.

The South Western and Littoral regions have fairly constant temperatures and two seasons: a short dry season (November–March), abundant rainfall (2,000–10,000 mm) in a long rainy season (March-November) typical of Cameroon’s equatorial climate. The Mount Cameroon region of the South West has mean annual rainfall of 2625 mm, relative humidity from 75%–80%, with temperature varying from 18oC in August to 35oC in March. The period from July-October characterized by heavy rains sees a peak in incidence of Human malaria. The disease is hyper-endemic during this period while it remains largely meso-endemic during the dry season. The low-altitude areas have prevalence of malaria parasitaemia ranging from 30% in the dry season to 65% in the rainy season with P. falciparum as the major cause (96%). *Anopheles gambiae* is the predominant vector in this area. The Bantu and Semi-Bantu are the major ethnic groups in these areas.

Cameroon is an enormously diverse country not only in population, but also in geographical and epidemiological strata (Bigoga *et al.*, 2007). These may considerably influence the course of malaria infection. In the Buea Sub-division, towns like Bonduma, Bokwaongo, Great Soppo and Buea Town at higher altitudes (700-1000m above sea level) may have considerable difference in transmission patterns as compared to towns like Bolifamba, Molyko and Muea at relatively low altitudes (400-650m above sea level).

#### **Study Population**

The Achidi et al, (2012) study that led to the generation of data to be analyzed herein involved the malaria endemic populations of Cameroon.

#### **Sampling Frame**

The sampling frame of the primary study to identify the factors that account for differences in clinical outcomes of malaria consisted of three regions (South West, Littorale and Centre) and three ethnic groups (Bantu, Semi-bantu and Foulbe). The three regions were selected to assess differences in intensity of transmission and parasite density between the regions, and how they affect the differences in clinical outcomes of malaria. The three ethnic groups were selected to assess how ethnicity affects differences in clinical outcomes of malaria. The three regions are among several malaria hotspots in Cameroon.

#### **Sampling Technique**

This was a stratified sampling technique in which the sampling frame consisted of three strata (three regions, and three ethnic groups within the three regions) from the entire malaria endemic populations. Cases were recruited from hospitals (Bota District Hospital – Limbe, Lanquintinie Hopital – Douala, Mother and Child Hospital – Yaounde, Regional Hospital – Limbe and Buea Regional Hospital Annex) and health centres (Bokova Health Centre, Mount Mary Health Centre - Buea and PMI Down Beach - Limbe). All health facilities were government institutions except for Mount Mary Health Centre. They all receive patients from neighboring towns. Controls were recruited from primary schools, including Catholic School (CS) Buea Station, CS Great Soppo, CS Muea, Government School (GS) Bolifamba, GS Bonduma, Government Practising School (GPS) Molyko I and II, GPS Muea I and II, HOTPEC Primary School Mile 15 Buea, Oxford Primary School Muea and Government Bilingual Primary School Muea as well as the Blood bank in Yaounde

## **Ethical Clearance**

Access to the data for regional analysis was granted by MalariaGEN to the Cameroon regional partners, Professor. Eric. A. Achidi and Doctor Tobias. O. Apinjoh of the University of Buea. Approval for analysis of the data was granted by the two regional partners. Access and analysis of the data will be done in strict adherence to MalariaGEN Data Sharing Policies (MalariaGEN, 2008).

Ethical clearance for data collection towards the primary study was obtained from the Institutional Review Board of the University of Buea (proposal number: ID D7.1.A/MPH/SWP/PDPH/PS.CH/2340/811) and the South West Regional Delegation of Public Health. Authorization to conduct the surveys in primary schools was obtained from the Regional Delegation of Basic Education or the Catholic Education Secretariat.

Informed consent was obtained from each case or their caregiver following a clear explanation of the content of the information sheet for the cases and blood bank donors. Only subjects/caregivers who volunteered to participate by signing a written informed consent were enrolled.

## **Sampling design and sample size determination**

### **Inclusion Criteria**

Of the case samples, only those that meet the criteria for the three major clinical prognostic features; cerebral malaria (CM), severe malaria anemia (SMA) and respiratory distress (RD) will be included in the analysis.

All control samples will be employed in the analysis.

### **Exclusion Criteria**

Case samples that are not classified as either CM, SMA or RD will be excluded in the analysis.

### **Sample size determination**

Generally, a GWAS relies on a large sample size and strong LD of marker (tag) and causal SNPs to attain a high power since it deals with genetic variations of small effect sizes (ORs - typically 1.1 – 1.5) and MAFs ≥ 5% (typically 5%). With relatively weak LD observed in African populations, the power of GWAS is greatly confined to the sample size as shown by the equation (Spencer *et al.*, 2009).

***Power ≈ N* β*2 f (1-f) r2***

N = sample size (case + control)

β = effect size

f = allele frequency of the risk variant (MAF)

r2 = linkage disequilibrium (Pearson’s correlation coefficient) between marker and causal SNPs (measured from 0 – 1, 0 for complete equilibrium and 1 for complete LD)

Therefore, total case samples that meet inclusion criteria = 350 (CM = 49; SMA = 248; RD = 53)

Control samples = 914

Hence total sample that will be employed in the analysis = 350 + 914 = **1,264**

And setting **β** at 1.5, **f** at 0.6 and **r** at 0.07 (7%), the analysis would attain a power of 66.7%

## **Experimental techniques**

### **Data collection**

Severe malaria was diagnosed according to WHO standard of 2000 i.e. the presence of asexual parasitaemia and at least one of the following: cerebral malaria (impaired consciousness or unarousable coma [(Blantyre coma score ≤ 2, corrected for hypoglycaemia (blood glucose <2.2 mmol/l or <40 mg/dl)] with no record of recent severe head trauma, neurological disease or any other cause of coma), severe malarial anaemia (haemoglobin <5g/dl (or haematocrit <15%), be fully conscious, no cases of severe bleeding or observed convulsions), convulsions before/during admission, respiratory distress (presence of alar flaring, intercostals or subcostal chest recession, use of accessory muscles of respiration, or abnormally deep respiration), hypoglycaemia (blood glucose <2.2mmol/l); hyperpyrexia (axillary temperature ≥40oC), hyperparasitaemia (>250,000 parasites/µl). Uncomplicated malaria was diagnosed by fully conscious with haemoglobin ≥8g/dl and no signs of severity and/or evidence of vital organ dysfunction (Achidi *et al.*, 2012).

Controls were asymptomatic adults (aged 17-52 years) and healthy children aged 1-14 years (afebrile and free from any obvious illness, albeit a fraction had asymptomatic parasitaemia). Adults were recruited from a blood bank in the Centre region of Cameroon (Mother and Child Hospital - Yaounde) between July and August 2007. Children were recruited from primary schools in the South-West Region of Cameroon (Buea Metropolis) between 2004-2005 and 2007-2008. Children with parasitaemia and a temperature of 37.5°C or above were not recruited as controls (Achidi *et al.*, 2012).

### **Molecular Analysis**

#### **DNA Extraction**

Genomic DNA was extracted from whole blood or packed cells at the Malaria Research Laboratory, University of Buea using the Promega Wizard (Promega Corporation, Madison, USA) or Nucleon™ BACC Genomic DNA Extraction (Gen-Probe Life Sciences, Manchester, UK) kits using manufacturer’s instructions and quantified. Aliquots of the DNA samples were shipped to the MalariaGEN Resource Centre in Oxford for further processing and quality control for quantity, quality (by genotyping) and confirming appropriate clinical data was available.) (Achidi *et al.*, 2012).

#### **Sequencing**

**Describe the whole genome sequencing and analysis pipeline**

## **Data Analysis**

### **Genotyping Quality Control**

SNPTEST v2.5.2 (mathgen.stats.ox.ac.uk/genetics\_software/snptest/snptest.html) will be used to assess the quality of the genotype data by assessing missingness, deviations from Hardy-Weinberg equilibrium, Differential missingness between cases and controls and Mendelian errors. SNPs with minor allele frequency < 1%, missing data proportion > 5% and Hardy-Weinberg *P* < 1×10-20 in controls will be excluded for downstream analyses.

### **Assessment of population structure**

The R statistical package ([www.r-project.org](http://www.r-project.org)) will be used to assess the population structure by performing principal component analysis.

### **Haplotype phasing and genotype imputation**

SHAPEIT v2.r904 (mathgen.stats.ox.ac.uk/genetics\_software/shapeit/shapeit.html) will be used to phase the genotype data and IMPUTE v2.3.2 (mathgen.stats.ox.ac.uk/impute/impute\_v2.html) will be used to impute genotypes into the Phase 3 dataset of the 1000 Genomes reference panel from the IMPUTE website.

### **Association testing**

Using the principal components (PCs) from the PCA as covariates, SNPTEST will be used for association testing of the variants genotyped and imputed.

Regional association testing will be performed by computing linkage disequilibrium between lead SNPs and genotyped and imputed SNPs. This will be done by computing the Pearson correlation coefficient r2 using the R statistical tool. Effect size (odds ratio - OR) will be computed by the Chi Square statistic on R.

The quantile-quantile (Q-Q) plot will be computed to test for association inflation.

# **Time Frame/Work Plan**

Table 1. Work plan

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2018** | | | | | | | | | | | | **2019** | | | | |
| Activity | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May |
| 1. Study Design |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Proposal 1st draft |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Proposal 2nd draft |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Final proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Data access/study |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Method selection (optimization/development) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Proposal defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Data analysis |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. -1st Draft manuscript |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| -2nd draft manuscript |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| -Submission of manuscript |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. 1st draft thesis |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. 2nd draft thesis |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. Thesis submission and defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# **Budget**

Table 2. Budget

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Activity** | **Item** | **Cost (£)** |
| 1 | Data Analysis, coding/programming | MacBook Pro (UNIX) | 1088 |
| 2 | Travel to South Africa | Flight ticket (round trip) | 550 |
|  | Publication | - |  |
|  | **Total** |  |  |

# **REFERENCES**

1. **Acharya, P., Garg, M., Kumar, P., Munjal, A., & Raja, K. D.** (2017). Host-parasite interactions in human malaria: Clinical implications of basic research. *Frontiers in Microbiology*, **8**: 1–16. https://doi.org/10.3389/fmicb.2017.00889
2. **Achidi, E. A., Apinjoh, T. O., Anchang-Kimbi, J. K., Mugri, R. N., Ngwai, A. N., & Yafi, C. N**. (2012). Severe and uncomplicated falciparum malaria in children from three regions and three ethnic groups in Cameroon: Prospective study. *Malaria Journal*, **11**: 1. https://doi.org/10.1186/1475-2875-11-215
3. **Amato, R., Miotto, O., Woodrow, C. J., Almagro-Garcia, J., Sinha, I., Campino, S., … Kwiatkowski, D. P.** (2016). Genomic epidemiology of artemisinin resistant malaria. *eLife*, **5**. https://doi.org/10.7554/eLife.08714
4. **Apinjoh, T. O., Anchang-Kimbi, J. K., Njua-Yafi, C., Mugri, R. N., Ngwai, A. N., Rockett, K. A., … Achidi, E. A.** (2013). Association of cytokine and toll-like receptor gene polymorphisms with severe malaria in three regions of cameroon. *PLoS ONE*, **8:** 1–11. https://doi.org/10.1371/journal.pone.0081071
5. **Band, G., Le, Q. S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., … Spencer, C. C. A**. (2013). Imputation-Based Meta-Analysis of Severe Malaria in Three African Populations. *PLoS Genetics*, **9**: https://doi.org/10.1371/journal.pgen.1003509
6. **Bigoga, J. D., Manga, L., Titanji, V. P. K., Coetzee, M., & Leke, R. G. F**. (2007). Malaria vectors and transmission dynamics in coastal south-western Cameroon. *Malaria Journal*, **6**. https://doi.org/10.1186/1475-2875-6-5
7. **Busby, G. B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V. D., … Spencer, C. C.** (2016). Admixture into and within sub-Saharan Africa. *eLife*, **5:** 1–44. https://doi.org/10.7554/eLife.15266
8. Centers for Disease Control and Prevention (2017) “Malaria”. https://www.cdc.gov/dpdx/malaria/index.html. Accessed February 2018
9. **Clarke, G. M., Rockett, K., Kivinen, K., Hubbart, C., Jeffreys, A. E., Rowlands, … Kwiatkowski, Dominic P., MalariaGEN Consortium**. (2017). Characterisation of the opposing effects of G6PD deficiency on cerebral malaria and severe malarial anaemia. *eLife*, **6:** 1–30. https://doi.org/10.7554/eLife.15085
10. **Dara, A., Drábek, E. F., Travassos, M. A., Moser, K. A., Delcher, A. L., Su, Q., … Silva, J. C.** (2017). New var reconstruction algorithm exposes high var sequence diversity in a single geographic location in Mali. *Genome Medicine*, **9**: 1–14. https://doi.org/10.1186/s13073-017-0422-4
11. **De Mendonça, V. R. R., Goncalves, M. S., & Barral-Netto, M.** (2012). The host genetic diversity in malaria infection. *Journal of Tropical Medicine*, *2012*. https://doi.org/10.1155/2012/940616
12. **Dinko, B., Pradel, G., Dinko, B., & Pradel, G**. (2016). Immune Evasion by Plasmodium falciparum Parasites: Converting a Host Protection Mechanism for the Parasite’s Benefit, **6**: 82–95. https://doi.org/10.4236/aid.2016.62011
13. **Dundas, K., Shears, M. J., Sun, Y., Hopp, C. S., Crosnier, C., Metcalf, T., … Wright, G. J**. (2018). Alpha-v–containing integrins are host receptors for the *Plasmodium falciparum* sporozoite surface protein, TRAP. *Proceedings of the National Academy of Sciences*, 201719660. https://doi.org/10.1073/pnas.1719660115
14. **Egan, E. S., Jiang, R. H. Y., Moechtar, M. A., Barteneva, N. S., Weekes, M. P., Nobre, L. V, … Doench, J. G.** (2015). HHS Public Access, **348**: 711–714. https://doi.org/10.1126/science.aaa3526.A
15. **Hedrick, P. W.** (2011). Population genetics of malaria resistance in humans. *Heredity*, **107**: 283–304. https://doi.org/10.1038/hdy.2011.16
16. **Jallow, M., Teo, Y. Y., Small, K. S., Rockett, K. A., Clark, T. G., Kivinen, K., … Modiano, D.** (2010). Genmone-wide and fine-resolution association analysis of malria in West Africa. *Nature Genetics*, **41**: 657–665. https://doi.org/10.1038/ng.388.Genome-wide
17. **Jespersen, J. S., Wang, C. W., Mkumbaye, S. I., Minja, D. T., Petersen, B., Turner, L., … Lavstsen, T.** (2016). *Plasmodium falciparum var* genes expressed in children with severe malaria encode CIDRα1 domains. *EMBO Molecular Medicine*, **8**: 839–850. https://doi.org/10.15252/emmm.201606188
18. **Kavishe, R. A., Koenderink, J. B., McCall, M. B. B., Peters, W. H. M., Mulder, B., Hermsen, C. C., … Van Der Ven, A. J. A. M.** (2006). Short report: Severe plasmodium Falciparum malaria in Cameroon: Associated with the glutathione S-transferase M1 null genotype. *American Journal of Tropical Medicine and Hygiene*, **75**: 827–829.
19. **Kehdinga Titanji, V. P., Amambua-Ngwa, A., Anong, D. N., Mbandi, S. K., Tangie, E., Tening, I., & Yengo, R**. (2009). Isolation and expression of UB05, a plasmodium falciparum antigen recognised by antibodies from semi-immune adults in a high transmission endemic area of the cameroonian rainforest. *Clinical Chemistry and Laboratory Medicine*, **47**: 1147–1158. https://doi.org/10.1515/CCLM.2009.255
20. **Kwiatkowski, D. P.** (2005). How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. *The American Journal of Human Genetics*, **77**: 171–192. https://doi.org/10.1086/432519
21. **Langhorne, J., & Duffy, P. E.** (2016). Expanding the antimalarial toolkit: Targeting host–parasite interactions. *The Journal of Experimental Medicine*, **213**: 143–153. https://doi.org/10.1084/jem.20151677
22. **Leffler, E. M., Band, G., Busby, G. B., Kivinen, K., Si Le, Q., Clarke, G. M., … Kwiatkowski, Dominic P., Malaria Genomic Epidemiology Network**. (2017). Resistance to malaria through structural variation of red blood cell invasion receptors, **356**. https://doi.org/10.1126/science.aam6393
23. **Luzzatto, L.** (2012). Sickle cell anaemia and malaria. *Mediterranean Journal of Hematology and Infectious Diseases*, **4**(1). https://doi.org/10.4084/MJHID.2012.065
24. **Malaney, P., Sielman, A., & Sachs, J.** (2004). The malaria gap. In *American Journal of Tropical Medicine and Hygiene*. **71**: 141–146. https://doi.org/https://doi.org/10.4269/ajtmh.2004.71.141
25. Malaria Genomic Epidemiology Network. (2015). *Nature*, **526**:253–257. https://doi.org/10.1038/nature15390.A
26. **Massoda Tonye, S. G., Kouambeng, C., Wounang, R., & Vounatsou, P**. (2018). Challenges of DHS and MIS to capture the entire pattern of malaria parasite risk and intervention effects in countries with different ecological zones: The case of Cameroon. *Malaria Journal*, **17**: 156. https://doi.org/10.1186/s12936-018-2284-7
27. **Mbenda, H. G. aell. N., Awasthi, G., Singh, P. K., Gouado, I., & Das, A**. (2014). Does malaria epidemiology project Cameroon as “Africa in miniature”? *Journal of Biosciences*, **39**: 727–738. https://doi.org/10.1007/s12038-014-9451-y
28. **Nabet, C., Doumbo, S., Jeddi, F., Konaté, S., Manciulli, T., Fofana, B., … Piarroux, R**. (2016). Genetic diversity of Plasmodium falciparum in human malaria cases in Mali. *Malaria Journal*, **15**. https://doi.org/10.1186/s12936-016-1397-0
29. **Olszewski, K. L., Morrisey, J. M., Wilinski, D., Burns, J. M., B, A., Rabinowitz, J. D., & Llinás, M.** (2010). NIH Public Access. *Cell*, **5**: 191–199. https://doi.org/10.1016/j.chom.2009.01.004.Host-parasite
30. **Ravenhall, M., Campino, S., Sepúlveda, N., Manjurano, A., Nadjm, B., Mtove, G., … Clark, T. G.** (2018). Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genetics*, **14**: 1–15. https://doi.org/10.1371/journal.pgen.1007172
31. **Rumaney, M. B., Ngo Bitoungui, V. J., Vorster, A. A., Ramesar, R., Kengne, A. P., Ngogang, J., & Wonkam, A.** (2014). The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. *PLoS ONE*, **9**: 1–10. https://doi.org/10.1371/journal.pone.0100516
32. **Smith, J. D., Rowe, J. A., Higgins, M. K., & Lavstsen, T**. (2013). Malaria’s deadly grip: Cytoadhesion of Plasmodium falciparum-infected erythrocytes. *Cellular Microbiology*, **15**: 1976–1983. https://doi.org/10.1111/cmi.12183
33. **Soe, T. N., Wu, Y., Tun, M. W., Xu, X., Hu, Y., Ruan, Y., … Cui, L**. (2016). Genetic diversity of Plasmodium falciparum populations in southeast and western Myanmar. https://doi.org/10.1186/s13071-017-2254-x
34. **Soulard, V., Bosson-Vanga, H., Lorthiois, A., Roucher, C., Franetich, J. F., Zanghi, G., … Mazier, D.** (2015). Plasmodium falciparum full life cycle and Plasmodium ovale liver stages in humanized mice. *Nature Communications*, **6**. https://doi.org/10.1038/ncomms8690
35. **Teo, Y., Small, K. S., & Kwiatkowski, D. P**. (2010). Europe PMC Funders Group Methodological challenges of genome-wide association analysis in Africa, **11**: 149–160. https://doi.org/10.1038/nrg2731.Methodological
36. **Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., … Horstmann, R. D**. (2012). Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature*, **489**: 443–446. https://doi.org/10.1038/nature11334
37. **Tobias O Apinjoh1, Judith K Anchang-Kimbi, Clarisse Njua-Yafi, André N Ngwai4, Regina N Mugri, Taane G Clark, Kirk A Rockett, Dominic P Kwiatkowski, E. A. Achidi, MalariaGEN Consortium**. (2014). Malaria Journal \_ Full text \_ Association of candidate gene polymorphisms and TGF-beta\_IL-10 levels with malaria in three regions of Cameroon a case¿control study, **13**:236 doi:10.1186/1475-2875-13-236.
38. **Wassmer, S. C., Taylor, T. E., Rathod, P. K., Mishra, S. K., Mohanty, S., Arevalo-Herrera, M., … Smith, J. D.** (2015). Investigating the Pathogenesis of Severe Malaria: A Multidisciplinary and Cross-Geographical Approach. *Am. J. Trop. Med. Hyg*, **93**: 42–56. https://doi.org/10.4269/ajtmh.14-0841
39. WHO: Severe falciparum malaria. Trans R Soc Trop Med Hyg 2000b, **94** (Suppl 1):S1–S90. World Health Organization, Communicable diseases cluster
40. **Wilby, K. J., Lau, T. T., Gilchrist, S. E., & Ensom, M. H.** (2012). Mosquirix (RTS,S): A Novel Vaccine for the Prevention of Plasmodium falciparum Malaria. *Annals of Pharmacotherapy*, **46**: 384–393. https://doi.org/10.1345/aph.1AQ634
41. World malaria report 2017. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO
42. **Wright, G. J., & Rayner, J. C.** (2014). Plasmodium falciparum Erythrocyte Invasion: Combining Function with Immune Evasion. *PLoS Pathogens*, **10**: 1–7. <https://doi.org/10.1371/journal.ppat.1003943>