Step I: White Paper Application

Application Guidelines

1. *The application should be submitted electronically per requirements via the web site of any of the NIAID Genomic Sequencing Centers for Infectious Diseases. Include all attachments, if any, to the application.*
2. *There are no submission deadlines; white papers can be submitted at anytime.*
3. *GSC personnel at any of the three Centers can assist / guide you in preparing the white paper.*
4. *Investigators can expect to receive a response within 4-6 weeks after submission.*
5. *Upon approval of the white paper, the NIAID Project Officer will assign the project to a NIAID GSC to develop a management plan in conjunction with the participating scientists.*

White Paper Application

Project Title: RTS,S Phase III Malaria Vaccine Trial Genotyping

Authors: Clarissa Valim, Sarah K. Volkman, Dyann F. Wirth, Daniel E. Neafsey, Carlota Dobano, Mary Hamel, Umberto D'Alessandro, Tsiri Agbenyega.

Contributing Authors: Amanda Leach, Marc Lievens Ya P. Shi, Terrell Carter, and Joe Campo.

Primary Investigator Contact:

Name	Clarissa Valim, M.D., ScD
Position	
Institution	Harvard School of Public Health
Address	675 Huntington Ave, Boston
State	MA
ZIP Code	02115
Telephone	
Fax	
E-Mail	cvalim@hsph.harvard.edu

1. Executive Summary *(Please limit to 500 words.)*

The GlaxoSmithKline (GSK) Phase III RTS,S/AS01E malaria vaccine trial will evaluate the efficacy of the RTS,S candidate vaccine against malaria disease caused by *Plasmodium falciparum* infection, across diverse malaria transmission settings in Africa. A portion of this vaccine's antigen consists of sequence from the *P. falciparum* circumsporozoite protein (CSP). Natural selection by the host immune system has given rise to high levels of antigenic variation at the CSP locus within many *P. falciparum* populations. It is hypothesized that administration of a candidate vaccine containing CSP antigens could also induce selection for existing or novel antigenic CSP variants. The goal of the proposed genotyping study is therefore to determine if RTS,S/AS01E selects specific parasite variants or alters the number of parasite types within a vaccinated subject and to gain a better understanding of the mechanism of action of the RTS,S/AS01E candidate vaccine. Selection of specific parasite variants or alteration in the number of parasite types by the vaccine could exert a selective pressure over time and drive the emergence or expansion of specific antigenic variants to allow vaccine evasion.

We have developed an ancillary study to genotype parasites from this vaccine trial, which has been approved by the Clinical Trial Protocol Committee (see attached for the full approved proposal). This trial presents the first and perhaps only opportunity to understand how this vaccine is working, and to understand the processes that underlie vaccine failure to better develop effective malaria vaccines in the future. Given that nearly all candidate malaria vaccines currently under development are monovalent vaccines like RTS,S, this trial has enormous potential benefit to the malaria research community due to its potential to validate

or refute the viability of a monovalent vaccine strategy. From these samples (over 12,000-16,000 individuals will be enrolled), we propose to use high-throughput sequencing approaches to evaluate both the *csp* antigen and other parasite antigens by comparing parasites from subjects vaccinated with the RTS,S/AS01E ("vaccinated") and parasites from subjects vaccinated with a comparator vaccine ("comparator"). We will evaluate parasite variants that successfully infect vaccinated subjects (thus evade vaccine-induced immunity) by seeking associations of vaccination with known T cell epitope (*Th2R*, *Th3R*, *CST3*) haplotypes as well as differences in B cell epitope (NANP) repeat count distributions. We will also evaluate if application of the RTS,S/AS01E vaccine affects genetic diversity (within and outside the *csp* locus) by looking at the number of parasite genotypes to determine if the vaccine is limiting the multiplicity of infection (MOI) instead of selecting for particular parasite variants in the RTS,S/AS01E vaccinated population. We will also aim to better understand the increased efficacy of the vaccine against severe disease detected in some Phase II trials, when compared to mild disease and infection. We will assess whether failure of protection from clinical or severe disease is associated with specific parasite variants or multiplicity of parasite populations. This will be accomplished by comparing the genotypes of parasites from vaccinated subjects who develop clinical or severe malaria compared to parasites from vaccinated subjects who only present with infection.

In the Phase III trial, children of 5-17 months and 6-12 weeks will be randomized to three groups and will be vaccinated at 0, 1, 2 and 20 study months. In the first group, named R3R, children will receive three doses of the RTS,S vaccine with adjuvant AS01 and a booster dose at Month 20. In the second group, named R3C, children will receive three doses of RTS,S-AS01 and a comparator vaccine (Meningitis C) at Month 20. In the third group, named C3C, children with 5-17 months will receive three doses of rabies vaccine followed by a dose meningitis C vaccine at Month 20 and children with 6-12 weeks will receive 3 doses of meningitis C vaccine followed by a dose of meningitis C vaccine at Month 20. In addition to the vaccines defined above, children of 6-12 weeks will also receive vaccines included in the World Health Organization Expanded Program on Immunization (EPI), i.e., three doses of Tritanrix-HepB/Hib+OPV followed by a dose at Month 20 of OPV.

2. Justification

1. *State the relevance to infectious disease for the organism(s) to be studied; for example the public health significance, model system etc.*

Malaria due to infection with *P. falciparum* currently causes almost 800,000 deaths annually, 85% of which occur in children under age five (2010 WHO World Malaria Report). Efforts to develop vaccines that provide immunological protection against this infectious organism have generally failed, largely due to the significant level of genetic diversity in the parasite, which thwarts host immune responses. The RTS,S candidate vaccine has displayed efficacy in Phase II trials varying between 30% and 65% [Abdulla et al, 2008; Alonso et al, 2004; Alonso et al, 2005; Aponte

et al, 2007; Bejon et al, 2008; Bojang et al, 2001; Guinovart et al, 2009; Kester et al, 2009; Lell et al, 2009; Owusu-Agyei et al, 2009; Polhemus et al, 2009; Sacarlal et al 2009; Stoute et al, 1998]. During these Phase II trials, considering the antigenic variability of the malaria parasite, there have been concerns whether large scale vaccination with RTS,S could ultimately select resistant parasite strains. In the case of the RTS,S, the *csp* locus is highly polymorphic and varies according to geographic region and transmission season [Alloueche et al, 2000; Barry et al, 2009; de Stricker et al, 2000; Doolan et al, 1992; Good et al, 1988; Jalloh et al, 2009; Jongwutiwes et al, 1994; Kumkahaek et al, 2004; Kumkhaek et al, 2005, Lockyer et al, 1989; Shi et al, 1992; Tanabe et al, 2004; Waitumbi et al, 2009; Yoshida et al, 1990; Zakeri et al, 2007]. The monovalent vaccine is based on the CSP sequence of the *P. falciparum* reference strains 3D7, but it appears to elicit a cross-reactive response against a variety of CSP haplotypes. If this cross-reactive protection is not widespread, however, the vaccine could select for resistant CSP haplotypes over time. Four molecular studies nested within Phase I/IIb trials could not find any evidence that the vaccine would be selecting non-3D7 genotypes [Alloueche et al, 2003; Bojang et al, 2001; Enosse et al, 2006; Waitumbi et al, 2009]. Results of these studies should be cautiously interpreted, considering the study outcomes, analyses, and limited power. More specifically, these four studies analyzed differences between individual polymorphisms but not haplotypes or polymorphism signatures, did not assess B-cell epitopes, and did not include mixed infections in the analysis (or analyzed only the dominant allele in the case of mixed infections). With this project, we propose to perform studies with more definitive conclusions about the long-term efficacy of the RTS,S vaccine.

The vaccine is now in a Phase III trial, the largest clinical trial that will be conducted for this vaccine and that has ever been conducted in malaria, with an estimated 12,000-16,000 subjects enrolled from multiple geographic locations experiencing a wide range of epidemiological and transmission parameters. Likely never again in history will the opportunity present itself to evaluate the consequences of vaccination with a sub-unit vaccine on parasite genotypes in a population before widespread deployment of this intervention strategy. The Gates Foundation, which is underwriting much of the Phase III trial, has provided crucial political support to enable acceptance of this ancillary genotyping protocol by GSK. Thus, we are presented with a unique and powerful opportunity to understand how the vaccine is working and to understand what processes undermine the effectiveness of this vaccine so that it can be improved.

2. *Are there genome data for organisms in the same phylum / class / family / genus? What is the status of other sequencing / genotyping projects on the same organism including current and past projects of the NIAID GSC? Provide information on other characteristics (genome size, GC content, repetitive DNA, pre-existing arrays etc.) relevant to the proposed study. Have analyses been performed on the raw data already generated/published? If additional strains are proposed for a species, please provide a justification for additional strains?*

Whole genome sequencing data for *P. falciparum* are available from a number of

parasite isolates, and additional genome-wide sequencing is ongoing. The *P. falciparum* genome is approximately 23 Mb with about 20% GC content overall. There are many areas of the genome that contain extensive repetitive sequence or low-complexity sequences and highly polymorphic regions within the chromosome ends where the variant surface antigen (*var*) genes are localized. These areas undergo extensive recombination and generally across the genome there is evidence of high levels of recombination. There have been a number (~20) of full genome sequences at high coverage released to the community and additional genome sequences being generated by the Wellcome Trust Sanger Institute (~200) that have not yet been released to the community. There is an approved white paper for sequencing an additional 24 genomes at high coverage and another 50 genomes at lower coverage for SNP discovery. This white paper is in the sequencing phase at the Broad Institute and data will be released to the community as they are generated. Many analyses of the data generated from these projects have been utilized both for analysis (Volkman et al. 2007; Jeffares et al. 2007; Mu et al 2007) and to generate whole genome genotyping tools (Neafsey et al. 2008; Daniels et al. 2008; Mu et al. 2010; Jiang et al. 2008) for interrogation of additional parasites. This proposal is not to generate additional genome-wide sequencing data, but rather to generate focused sequencing data from a large number of patient samples that contain *P. falciparum* to evaluate the effectiveness of a candidate vaccine for this important human pathogen. In addition to the patient samples from the Phase III study, in preparation for the Phase III study we will be examining approximately 2500 samples collected before starting vaccination in the 11 sites included in the clinical trial.

3. *If analyses have been conducted, briefly describe utility of the new sequencing or genotyping information with an explanation of how the proposed study to generate additional data will advance diagnostics, therapeutics, epidemiology, vaccines, or basic knowledge such as species diversity, evolution, virulence, etc. of the proposed organism to be studied.*

In this proposal we will focus on the single locus targeted by vaccine (*csp*) to evaluate the consequences of vaccination on the number and types of *csp* variants in parasites that evade protective immunity conferred by the vaccine. We will also apply 454 next generation sequencing approaches to estimate multiplicity of infection (MOI) and determine if vaccination reduces the numbers and types of parasites in vaccinated relative to unvaccinated individuals, as Phase II trials have suggested. This proposal will utilize innovative sequencing approaches around a limited number of genetically variable loci within the parasite genome to assess haplotypes at both the locus directly under vaccine pressure (*csp*) as well as additional loci.

The utility of these new sequencing data will be to advance both the development of a successful vaccine, and to better understand the basic biology of the *P. falciparum* pathogen in terms of its response to a host immune response elicited by a vaccine. The outcome of these findings will be critical for the malaria community to understand the successes or failures of this malaria vaccine and to inform the necessary next steps to improve vaccine efficacy to promote the

eradication of malaria.

Given that the RTS,S/AS01E is a confirmatory Phase III study that will be used with regulatory purposes, as a first step, we are proposing to study samples collected from the study sites before starting the randomized trial. These samples will be used as a validation set for the filters and quality control procedures we developed for SNP and haplotype calling using 454, as a training set to all the analyses procedures described below. In addition by using traditional and novel sequencing methods and methods to determine MOI, the study of these samples will bridge the two technologies for the scientific community involved in the trial.

3. Rationale for Strain Selection

4. *Provide the rationale behind the selection of strains and the number of strains proposed in the study. The focus of the program is on potential agents of bioterrorism or organisms responsible for emerging or re-emerging infectious diseases. Non-select agents or non-pathogenic organisms will be considered when they can provide insight into these scientific areas.*

The *P. falciparum* samples for sequencing will come from two sources: a) for the preliminary study, individuals who present with infection before the trial started at each of the trial study sites; b) for the main study, individuals who enrolled in the Phase III RTS,S vaccine trial and present with parasites (infection, mild clinical disease, and severe disease) during the trial period. Strains included in the preliminary study will allow us to gain knowledge about the distribution of haplotypes, individual polymorphisms and MOI in the trial study sites so that we can further plan the analysis of the confirmatory study. These samples will allow validation of filters and quality control procedures for SNP calling. Also these samples will allow us to train all proposed analysis models before we start analyzing the actual Phase III data. The Phase III vaccine trial samples represent all *P. falciparum* positive samples detected during the Phase III trial either during two cross-sectional surveys or at time of evidence of malaria disease during the trial.

4a. Approach to Data Production: Data Generation

5. *State the data and resources planned to be generated. (e.g draft genome sequences, finished sequence data, SNPs, DNA/protein arrays generation, clone generation etc.)*

Sequence data will be generated from immunogenic regions of the *csp* locus from all malaria positive samples. In addition, we propose to sequence several additional highly polymorphic loci to acquire estimates of MOI that are independent of *csp*, to guard against bias caused by vaccine-induced modulation of the polymorphism profile at *csp*.

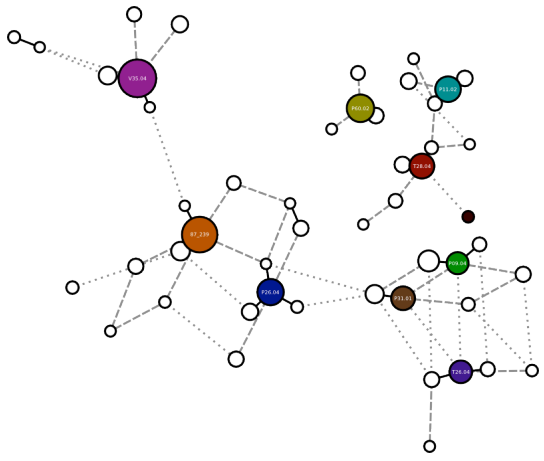
We propose to sequence the highly polymorphic T cell epitope region of *csp* using the 454 platform, as three epitopes lie within a 330 bp region that can be easily

spanned with 454 reads [Good et al, 1991; Malik et al, 1991; Moreno et al, 1993]. We will PCR amplify this region using a barcoded primers, such that the PCR product of every sample will be associated with a unique 8 bp 'barcode' sequence attached to the forward primers. All barcode sequences will differ from each other by at least 2 bp to guard against misidentification due to error. We will pool batches of PCR products for efficient multiplexed sequencing from both ends of the PCR product. We aim to achieve a typical coverage depth of 250-500X per amplicons per sample.

We also propose to sequence the repetitive region of *csp*, which is composed of a variable number of tetra-amino acid repeats (NANP). Immunological studies have shown that these repeats function as a B cell epitope [Calvo-Calle et al, 1993; Munesinghe et al, 1991]. The entire repeat regions spans approximately 600 bp, and consists of a variable number of NANP repeats, followed by a single NVDP amino acid sequence, and then another variable-length block of NANP repeats. The length of the 454 reads will be insufficient to completely span the amplicons, but we will be able to obtain counts of the NANP repeats on either side of the central NVDP oligo from reads that are least 400 bp in length. As with the T epitope amplicons, PCR products will be generated using barcoded primers, and will be pooled and sequenced from both ends using 454 to a coverage depth of 250-500X.

Sequencing of highly polymorphic select regions of the *msp1* and *msp2* loci will be carried out using 454 in a similar fashion to provide independent estimates of MOI for each sample.

We have conducted a small pilot study to validate this sequencing approach at the *csp* locus. Using control mixtures of DNA from up to nine clonal laboratory strains of *P. falciparum*, we have been able to reliably detect all of the expected haplotypes and determine MOI through multiplexed sequence data corresponding to the T epitope region of the protein. The figure below illustrates a haplotype network map constructed from sequencing data for the T epitope region from a control mixture of nine strains.



The colored circles indicate 'correct' haplotypes corresponding to the strains in the

control mixture, with the size of the circles indicating the abundance of particular haplotypes in the sequencing data. Empty circles are haplotypes caused by PCR or sequencing errors. We are evaluating algorithms to remove these error haplotypes. Our approach for error filtering takes into account our prior knowledge of known SNPs and haplotypes in a particular populations and also the observation that error haplotypes are typically a short mutational distance from correct haplotypes, as indicated by the short branch lengths between common error haplotypes and correct haplotypes in the figure above.

As part of the preliminary phase of this study, we will build and validate our error filters using field samples collected right before starting the Phase III study in all the Phase III study sites. The samples used in the preliminary validation study should present with polymorphisms similar to those encountered in the non-vaccinated population of the Phase III trial.

For the preliminary phase of this study, in addition to all sequencing specified above, we will perform MSP typing using a traditionally accepted method described below to establish a parallel between the methods we are proposing here and methods used in genotyping studies performed in the Phase II RTS,S trials. The MSP loci will be typed using nested PCR to discriminate alleles of *msp1* block 2 and *msp2*. At the Harvard School of Public Health, gels will be double scored independently by two investigators blinded to the child's vaccination and malaria outcome and either consensus will be reached or the assay will be repeated. MOI will be based on the maximum number of genotypes across all loci.

The following table indicates the complete scope of data we intend to generate, with each sequenced amplicon approximately 200-400 bp in length

Sample Set	No. Samples	No. PCR amplicons	Total amplicons
Preliminary Phase	2,500	4	10,000
Main Phase III Study	~4,750	4	19,000
TOTALS	7,250	4	29,000

4b. Approach to Data Production: Data Analysis

6. *Briefly describe the analysis (value-add) envisioned to be performed subsequently by the community and the potential to develop hypotheses driven proposals given the datasets and resources produced by this work.*

Raw data processing

We will employ multiple strategies to filter random and systematic errors that will be tested during the preliminary phase of this study. Reads will be mapped to the 3D7 reference sequences for each amplicon and SNP variants with low quality scores will be discarded. Chimeric haplotypes formed during PCR amplification will be identified and discarded using the ChimeraSlayer algorithm. Position-specific,

context-dependent sequencing or base-call errors will be identified via inclusion of appropriate control samples in the pools, and variants at these amplicon positions will be discarded or subjected to higher threshold of evidence for inclusion in subsequent analyses. Prior knowledge of confirmed polymorphisms and haplotypes at the *csp* locus will also be taken into account using a Bayesian haplotype calling approach.

NANP repeat counts to the left and right of the NVDP sequence in *csp* will be tallied through custom scripts. MOI for each sample will be determined by the amplicon exhibiting the maximum number of distinct haplotypes passing all of our data filters.

Analysis of Preliminary Phase Samples

Distributions of haplotypes at each study site in single and mixed infections will be estimated. This analysis will inform our decisions about models to analyze the Phase III study, including restrictions imposed to models and parameters. Associations between haplotypes and candidate confounders of the vaccine study will be assessed through random effects models (multinomial and logistic) to account for correlations within subjects of mixed infections. We will also explore modeling time as a continuous covariate in random effect models. To search for association between specific haplotypes and vaccination or the other predictors, in multivariate multinomial models, we will asses the interaction between each haplotype (i.e., A vs. B vs etc) and the predictor of interest. In addition to traditional interaction tests, we will explore alternative, more powerful interaction tests, such as those proposed by Cai et al (2010); Chatterjee et al (2005 and 2006) and Maity et al. (2009), and Wu et al (2010). MOI will be compared between population subgroups using Poisson regression to adjust for potential confounders and to stratify the analysis.

Main Study – Phase III Ancillary Genotyping Study

Currently the protocol proposes three specific aims, with corresponding outcomes and study comparisons.

Specific Aim 1: To identify *csp* haplotypes that may be targeted by the vaccine through comparison of parasites from subjects vaccinated with RTS,S/AS01E ("vaccinated") with parasites from comparator vaccinated subjects ("comparator").

We will compare subjects who received comparator and subjects who completed vaccination with regards to the following markers of parasite genetic polymorphisms:

- specific amino acids in potentially polymorphic T-cell epitopes, such as Th2R, Th3R and CST3.
- haplotypes within potentially polymorphic T-cell epitopes, such as Th2R, Th3R and CST3.
- number of repeats of NANP B-cell epitope.

Primary comparisons will be restricted to parasites:

- detected by microscopy in infections (defined as prevalent parasitemias) at cross-sectional surveys at Month 20 and Month 32.
- first episodes of clinical malaria who seek care.
- first episodes of severe malaria.

Secondary comparisons will include parasites detected in all episodes of clinical and severe malaria.

For comparison with previous work, we will primarily compare the difference in the proportion of infections (or isolates) with (vs. without) the CSP vaccine haplotype (3D7) at each polymorphic amino acid position within the T cell epitopes separately. In the primary analysis we will assume independence of each subject's observations across time and evaluate statistical significance of differences through Fisher's exact test. Primary analyses will be performed by site, combining infections diagnosed in all age and at all cross-sectional surveys. Secondary analyses will account for correlations across time using logistic regression models with random effects. In those models, we will explore analyzing the data separately for Months 20 and 32. Comparisons at Months 32 will be primarily focused on treatment groups R3R (completed primary vaccination course and received booster) and C3C (comparator arm). In additional analysis, we will compare parasites isolated in clinical cases among vaccinated subjects with parasites from clinical cases among comparator subjects and will compare parasites isolated in first events of severe malaria only (primary case definition) among vaccinated subjects with parasites from first events severe malaria only (primary case definition) in comparator subjects. These comparisons will be performed similarly to the comparisons described above between infections in vaccinated and comparator subjects. Finally, for exploratory purposes, we will analyze: a) clinical and severe malaria based on secondary case definitions; b) all events of clinical and severe malaria (first, second and subsequent episodes). As a second outcome, we will compare the frequency of alternative haplotypes within the three T cell epitopes. Adjustments for covariates will be considered using multinomial regression models (with random effects to account for correlations of multiple infections within a subject).

As a third outcome, we will compare between vaccinated and comparator groups the number of perfect consecutive (NANP) repeat units in the B cell epitope via Wilcoxon tests. Poisson or gamma models will be considered to adjust estimates for covariates and to fit multivariate models accounting for correlations within subjects.

Specific Aim 2: To evaluate whether RTS,S/AS01E has an effect on genetic diversity of the parasite population within and outside the *csp* locus.

We will compare subjects who received comparator and subjects who completed vaccination with regards to MOI. Primary and secondary comparisons will be

performed as described in Specific Aim 1.

We will compare MOI (overall MOI, csp specific MOI, and non-csp related MOI separately) between the two study groups using the same procedures outlined for comparing NANP counts in Specific Aim 1. We will compare MOI between vaccinated and comparator groups via Wilcoxon tests. Poisson or gamma models will be considered to adjust estimates for covariates and to fit multivariate models. Study sites and Months will be handled as described in the analyses of Specific Aim 1.

Specific Aim 3: To assess whether failure of protection from clinical or severe disease in vaccinated subjects is associated with specific parasite variants or multiplicity of infection.

We will compare parasite genetic polymorphisms and diversity of parasites (outcomes) between vaccinated subjects who present:

- with an infection (defined as prevalent parasitemia) at cross-sectional surveys at Month 20 and Month 32 detected by microscopy and with first clinical malaria episode (primary and secondary case definitions), including and excluding severe malaria cases from the comparison.
- with an infection (defined as prevalent parasitemia) at cross-sectional surveys at Month 20 and Month 32 detected by microscopy with severe malaria episode (primary and secondary case definitions 1 and 2).

Results of analysis limited to vaccinated subjects will be contrasted to results of the same comparisons including comparator subjects only. Secondary comparisons will include parasites detected in all episodes of clinical and severe malaria, i.e., in all events presented by each single subject during the whole study period. We will also be comparing parasites of severe malaria cases with parasites of both infection and clinical malaria (excluding severe disease from this group), if parasites of infections and clinical malaria are comparable.

Markers of parasite genetic polymorphism and parasite diversity will be the same as those defined in Specific Aim 1.

Within the T cell epitopes, we will compare the frequency of *csp* vaccine-type alleles at each polymorphic amino acid between clinical malaria (cases) in vaccinated subjects (detected through passive case detection by Month 20) and infections (controls) in vaccinated (prevalent parasitemia) subjects (detected at Month 20) using exact methods or chi-square tests. In secondary analysis, we will include subjects who present with clinical malaria or malaria infection at any time by Month 32. Subjects who develop clinical disease up to one month after the active case detection survey will be excluded from primary and secondary analyses. Subjects in the group R3C will be excluded from the secondary analysis. Although primary analysis will include first event only, secondary analyses will include parasites from all malaria events. Additional analysis will compare parasites detected in severe disease vaccinated subjects with parasites from

infected vaccinated subjects detected at Month 20 and Month 32. We will also explore comparing severe malaria cases to a group comprised by clinical malaria or malaria infection, depending on the appropriateness of collapsing these two last groups.

All analyses of this aim will be adjusted for geographical area and for time of detection of parasites. The idea is that by adjusting the analyses, any differences between detected groups should be attributable to subject's immune status, since compared subjects would have been exposed to the same pool of parasite. If the number of subjects in strata formed by combinations of geographical area and time is scarce, adjustments for these two factors will be performed by matching (within ranges of the matching factors). For example, clinical cases would be paired (within a range) to month of detection of infections (controls), i.e., Months 20 and 32 and neighborhood of the house. Alternatively, if the number of subject in each stratum allows, adjustment for these two factors will be performed in the analysis. We will define the method to adjust for geographical area (in the design by matching or in the analysis by regression) after the data is collected but before the genotype database is linked to the database containing information about the subject. In case we decided to use matching, pairing strata will be defined after data collection so that we simultaneously maximize the number of clinical and severe cases to be included in the analyses and minimize the range of time and geographical distance between cases and controls. Strata will be defined blindly to subject's genotyping results. A detailed analysis protocol will describe the defined strata before the start of analysis.

Differences between clinical malaria in vaccinated subjects and infections in vaccinated subjects will be compared to differences between clinical malaria in comparator subjects and infections in comparator subjects in conditional logistic regression models. In these models, vaccinated and comparator subjects will be included and an interaction term between vaccination status and haplotype (or genotype) will be tested. When analyzing differences in B cell epitope count, we will use exact stratified tests.

Analyses addressing MOI will be carried out using the same procedures and methods described above for Specific Aim 3 but using the number of genotypes (MOI) as a covariate.

Subgroup Analysis

We will repeat all analysis of Specific Aims 1 and 2 comparing only vaccinated subjects to whom the vaccine was immunogenic with all comparator subjects. Immunogenicity of the vaccine will be defined based on anti-CSP antibody levels. Subgroup analysis will also be performed according to levels of parasitemia.

5. Community Support and Collaborator Roles:

7. *Provide evidence of the relevant scientific community's size and depth of interest in the proposed sequencing or genotyping data for this organism or group of*

organisms. Please provide specific examples.

A malaria vaccine would be a key step towards malaria elimination. At present, a high priority and attention has been given to malaria control and elimination by international health and sponsor agencies. Although the RTS,S vaccine has reached only moderate efficacy in Phase II randomized trials (30-60%), this vaccine is considered currently the best vaccine candidate. International agencies, including the Malaria Vaccine Initiative, and the scientific community are now investigating additional antigens or primers that could be used with the RTS,S to increase vaccine efficacy. The whole malaria community and all Health Bureaus of the African continents have high interest in the RTS,S Phase III and its ancillary studies.

Several key scientists in malaria immunology and genomics not associated with the RTS,S vaccine, as well as investigators working with malaria clinical and epidemiological research across Africa have shared with us their enthusiasm with our ancillary study proposed in this White Paper and sent us encouraging words, including.

…"Given the relatively low efficacy of the RTS,S (compared to vaccines against other infectious diseases), it is absolutely essential to investigate if vaccination with RTS,S may select for break-through parasites" (Ivo Mueller)….

…"Genotypic differences in parasite isolates that escape vaccine induced immunity clearly will be an important phenotype as vaccine efforts are increased" (Karl Seydl)…

…"I wholehearted support your study. Your approach to study parasite evasion from RTS,S vaccine induced immunity could bring a definitive answer to the concern as to whether vaccine could select resistant strains." (Herald Noedl)

…"The identification of genetic signatures of malaria parasites that could be associated with the evasion from the RTS,S malaria vaccine induced immunity will help to understand if there is a selection of parasites induced by the accine and the mechanisms that may induce this selection." (John Aponte).

…."I agree that this approach should shed light on the question of whether use of an RTS,S vaccine could select for vaccine variants resistant to the RTS,S vaccine." (Donald Krogstad).

Investigators in the research community who have voiced their support for this project include:

David Conway, MD
Professor of Biology
London School of Tropical Medicine and Hygiene

University of London

Johanna Daily, MD
Associate Professor
Department of Medicine (Infectious Diseases)
Department of Microbiology and Immunology
Albert Einstein College of Medicine

Professor Seydou Doumbia, MD. PhD
Faculty of Medicine
University of Bamako, Mali

Rick Fairhurst, MD, PhD
Chief, Malaria Pathogenesis and Human Immunity Unit
Laboratory of Malaria and Vector Research
National Institute of Allergy and Infectious Diseases

Stephen L. Hoffman, MD,
Chief Executive and Scientific Officer,
Sanaria

Marcelo Urbano Ferreira, MD, PhD
Department of Parsitology
Institute of Biomedical Sciences
University of Sao Paulo, Brazil

Christina Happi, PhD
Senior Research Fellow, Malaria Research Laboratories
Adjunct Lecturer, Department of Pharmacology and Therapeutics
College of Medicine
University of Ibadan, Nigeria

Professor Sanjeev Krishna, MA, BMChB, DPhil, FRCP, ScD, FMedSci
Division of Clinical Sciences
Centre for Infection
St. George's, University of London

Professor Donald Krogstad, MD
Henderson Professor and Chairman of Tropical Medicine
Director, Tulane Center for Infectious Diseases
Department of Tropical Medicine
School of Public Health and Tropical Medicine
Tulane University

Professor Dominic Kwiatkowski, MD, PhD
Supernumerary Fellow in Medicine
St. John's College, Oxford, UK

Myron Levine, MD, DTPH
Professor and Director
Center for Vaccine Development
University of Maryland

Professor Carole Long, PhD
Chief of Malaria Immunity Section
Laboratory of Malaria and Vector Research
National Institute of Allergy and Infectious Disease, NIH

Director, PATH Malaria Vaccine Initiative
Growth Inhibition Assay-Reference Center
Malaria Immunology Section

Bronwyn MacInnis, PhD.
Research Administrator
Centre for Genomics and Global Health
Welcome Trust Sanger Institute
University of Oxford

Professor Soylemane Mboup, PhD
Professor of Head, Bacteriology and Virology Laboratory
Director, AIDS Research Network for Western and Central Asia
University College of London

Ivo Mueller, PhD
Infection & Immunity Division
Walter & Eliza Hall Institute
1G Royal Parade
Parkville, VIC 3052
Australia

Professor Charles Newton, MD, Ph.D
Professor of Tropical Neurosciences and Pediatrics
Institute of Child Health
University College of London

KEMRI
Wellcome Trust Collaborative Programme
Kilifi, Kenya

Harald Noedl, MD, PhD
Department of Parsitology
Institute of Biomedical Sciences
University of Sao Paulo, Brazil

Karl Seydel, MD
Assistant Professor
Michigan State University

Director of Moelecular and Genomic Core
Malawi ICEMR
University of Malawi College of Medicine

Professor Terrie Taylor, DO
Scientific Director
Blantyre Malaria Project
University of Malawi College of Medicine

University Distinguished Professor
Dept. of Internal Medicine
College of Osteopathic Medicine
Michigan State University

Thomas E. Wellems, MD, PhD
Chief of Laboratory of Malaria and Vector Research
Chief of Malaria Genetics Section
National Institute of Allergy and Infectious Diseases

Additionally, Several of the key scientists involved in research in malaria epidemiology, immunology, and vaccine have been involved in the planning of this study, including the Principal Investigators and sub-investigators of the RTS,S Phase III vaccine trial listed below, with corresponding country:

Principal Investigators and Co-Principal Investigators

Burkina Faso
Halidou Tinto, MD
Umberto D'Alessandro, PhD

Gabon
Agnandji Selidji, MD
Peter Kramsner, MD, PhD
Bertrand Lell, PhD

Kintampo, Ghana
Asante Kwaku Poku, MD
Seth Owusu-Ageyei, MD

Kumasi, Ghana
Tsiri Agbenyega, MD, PhD
Daniel Ansong, MD, PhD (also in Kintampo)

Killifi, Kenya
Patricia Njuguna, MD
Chilengi Roma, MD

Kombewa, Kenya
Walter Otieno, MD
Roma Chilengi, MD

Siaya, Kenya
Mary Hamel, MD

Malawi
Francis Martinson, MD, PhD
Irving Hoffman, MD

Mozambique
Jahit Sacarlal, PhD
Pedro Alonso, MD, PhD

Bagamoyo, Tanzania
Salim Addulla, MD, PhD
Marcel Tanner, PhD

Korogwe, Tanzania
John Lusingu, MD
Samwell Gesase, MD

Sub-Investigators of the RTS,S Vaccine Trials

Burkina Faso
Biebo Bihoun, MD
Issa Guiraud, MD
Beranger Kabore, MD
Olivier Sombie, MD
William Kabore, MSc

Kintampo, Ghana
Ruth Owusu, MD

Gabon
José Fernandes, MD
Aurore Houkpatin Bouyoukou, MD
Tidgani Mdai Guirema, MD
Tina Moeller, MD
Issifou Saadou, PhD

Kilifi, Kenya

Trudie Lang, MD

Kombewa, Kenya
Janet Oyieko, MD
Godfrey Otieno, MD
Lucas Otieno, MD
Ben Anadagalu, MD
Stella Apollo, MD

Kintampo, Kenya
Stephen Apanga, MD
Samuel Ayamba, MD
Evans Kwara, MSc
George Adjei, MD
Shalom Abokyi, MD
Samuel Agra, MD

Bagamoyo, Tanzania
Rose Minja, MD
Agnes Mnzava, MD
Mwajabu Loya, MD
Alex Rwezaula, MD
Rockus Ngalela, MSc
Paul Oloo, MD

Korogwe, Tanzania
Anangisye Malabeja, MD
Omari Abdul, PhD
Hassan Kilavo, MD

Mozambique
Miguel Lanaspa, MD
Simon Kariuki, MD
Larry Slutsker, PhD

The protocol proposed for this study was discussed with several members of the RTS,S Clinical Trials Partnership Committee (CTPC) investigators, with GSK investigators, and with the CTPC committee named Genotyping Committee (designated by the CTPC to evaluate the Genotyping Ancillary Study proposed protocol) during several months until reaching a consensual approval. Those individuals are listed below as collaborators. The Malaria Vaccine Initiative (MVI) and several NIH key scientists have supported directly or indirectly this study, including Dr. Lee Hall.

8. List all project collaborators and their roles in the project

Dr. Clarissa Valim and Dr. Sarah Volkman will oversee the conduct of the project. In addition, they and other sub-investigators will be individually responsible for:

- Dr. Clarissa Valim[1,2,3] will be responsible for communications with other investigators, for study design and analysis plan, and for supervision and conduct of all analysis.
- Dr. Sarah Volkman[1,3] will oversee all parasite genotyping and will also be responsible for communications with other investigators.
- Dr. Dyann Wirth[1,3], the leader of the Harvard/MIT Malaria Genome project will oversee the whole project, play a supervisory role and provide scientific guidance.
- Dr. Daniel Neafsey[1,3] will have a key role in development and choice of appropriate technology for evaluating the parasite genetics and in both parasite genotyping and statistical analysis.
- Drs. Carlota Dobano[4,5], Mary Hamel[6,8], Umberto D'Alessandro[7], and Tsiri Agbenyega[9] will have an advisory role and provide scientific and 'from the field' comments to the protocol.
- A computational biologist from the Broad Institute of Harvard and MIT, to be named, will perform all data pre-processing including computer sequence analysis.
- Ms. Terrell Carter[10] will facilitate communications, exchange of material, and data among GSK members, investigators, and the Harvard School of Public Health.
- Dr. Amanda Leach[11], Mr. Marc Lievens[11], and Ms. Myriam Bruls[11] will have advisory roles, be involved in organizing and shipping all blood samples, discuss all statistical analyses, and organize and send the study sponsor the extract of data necessary to perform the association analysis.
- All contributing authors, including Dr. Ya P. Shi[6] and Joe Campo[4,5] provided significant scientific knowledge to this proposal and invaluable suggestions.

Also, this protocol was developed and will be implemented in collaboration with RTS,S Clinical Trials Partnership Committee (CTPC) investigators, including:

1. Pedro Alonso and Jahit Sacarlal – Centro de Investigação em Saúde de Manhiça, Maputo, Mozambique
2. Selidji Todagbe Agnandji (Maxime) and Bertrand Lell – Hopital Albert Schweitzer, Lambaréne, Gabon
3. Salim Abdulla and Marcel Tanner – Ifakara Health Institute, Bagamoyo, Tanzania
4. John Lusingu and Samwel Gesase – National Institute of Medical Research, Korogwe, Tanzania
5. Daniel Ansong and Tsiri Agbenyega – Kumasi Centre for Collaborative Research, Kumasi, Ghana
6. Seth Owusu-Agyei and Kwakupaku Asante – Kintampo Health Research Centre, Kintampo, Ghana
7. Patricia Njuguna and Roma Chilengi – KEMRI Wellcome Collaborative Research Programme, Kilifi, Kenya
8. Walter Otieno and Lucas Otieno – KEMRI/WRAIR, Kombewa, Kenya
9. Francis Martinson and Irving Hoffman – Lilongwe UNC Project, Lilongwe,

Malawi
10. Tinto Halidou and Umberto D'Alessandro – IRSS-Centre Muraz, Nanoro, Burkina Faso
11. Mary Hamel, Simon Karifuki and David Jones – KEMRI/CDC, Siaya, Kenya

Affiliations list:

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health (HSPH), USA
[2] Department of Biostatistics, Harvard School of Public Health, USA
[3] The Broad Institute, USA
[4] Barcelona Centre for International Health Research (CRESIB), Spain
[5] Centro de Investigação em Saúde de Manhiça (CISM), Mozambique
[6] Centers for Disease Control and Prevention (CDC), Atlanta, USA
[7] Institute of Tropical Medicine – Antwerp.
[8] Kenya Medical Research Institute (KEMRI)/CDC, Kisumu, Kenya
[9] Kumasi Centre for Collaborative Research in Tropical Medicine (KCCR), Ghana
[10] Malaria Vaccine Initiative (MVI) – Program for Appropriate Technology in Health (PATH), USA
[11] GlaxoSmithKline Biologicals, Belgium

9. *List availability of other funding sources for the project.*

Collection of samples for this study has been supported and funded by GSK. No funding is available for the proposed Genotyping Ancillary study, i.e., for sequencing or analyzing the sequence data to the scientific community.

6. Availability & Information of Strains:

10. *Indicate availability of relevant laboratory strains and clinical isolates. Are the strains/isolates of interest retrospectively collected, prepared and ready to ship?*

Note: If samples are prospectively prepared the GSC can provide protocols and recommendation based on the Centers past experiences. The samples must however meet minimum quality standards as established by the Center for the optimal technology platform (sequencing/ genotyping) to be used in the study.

The samples for the preliminary study will be shipped by the investigators of the 11 study sites at once. The samples for the Phase III study are currently being collected and archived by GSK and will be shipped in batches sufficient to fill the sample processing and sequencing pipeline, as determined. As part of the quality control procedures carried out before receiving samples from the trial, a small batch of pilot samples of similar material, but outside the trial, will be processed and sequenced to ensure that the quality standards are met and that the data generated is appropriate.

Samples are being collected under ethical review and to ensure patient confidentiality will be coded during data generation. This code will be independent

of any identifying factors or outcome information for the sample processing and sequencing phases. At defined points in the trial, information will be provided by independent third party individuals about which coded samples were obtained from RTS,S vacinees or comparator vacinees. Please see the attached study protocol for full details.

11. *Attach relevant information, if available in an excel spreadsheet for multiple samples:*

Samples are currently being collected at the trial sites. A complete list of samples and associated metadata will be compiled after the conclusion of the trial.

12. *What supporting metadata and clinical data have been collected or are planned on being collected that could be made available for community use?*

For each sample, we will note whether it originated from a member of the vaccinated or comparator group. Subject age, sex, date of sample collection, and village of residence will be recorded. The condition of the subject at the time of sample collection will also be noted (asymptomatic vs. symptomatic), as well as whether the subject presented or progressed to a state of severe malarial disease, and specific clinical characteristics. Additional information to be recorded about each subject will include: parasite density, anti CSP and Hepatites B antibody levels, use of insecticide sprayed bednet and malaria medication. Further metadata on the immunological state of the subjects may be available pending funding of another ancillary study led by Carlota Dobano.

7. Compliance Requirements:

7a. Review NIAID's Reagent, Data & Software Release Policy:

NIAID supports rapid data and reagent release to the scientific community for all sequencing and genotyping projects funded by NIAID GSC. It is expected that projects will adhere to the data and reagent release policy described in the following web sites.

http://www3.niaid.nih.gov/research/resources/mscs/data.htm

http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html

<Each Center to include their website that describes/points to the guidelines>

Once a white paper project is approved, NIAID GSC will develop with the collaborators a detailed data and reagent release plan to be reviewed and approved by NIAID.

Accept ☐ Decline ☐

7b. Public Access to Reagents, Data, Software and Other Materials:

13. *State plans for deposit of starting materials as well as resulting reagents, resources, and datasets in NIAID approved repositories. Sequencing projects will not begin until the strain is deposited into NIAID funded BEI repository (http://www.beiresources.org/). This includes web based forms are completed by the collaborator and received by the NIAID BEI (http://www.beiresources.org/).*

There are no culturable parasite strains that will be available from the bloodspot materials, and only sufficient bloodspot materials will be available to generate the PCR product for sequencing. The remaining PCR products used for sequencing can be made available to the community through the NIAID BEI repository or MR4. All validated PCR product haplotype sequences will be made available through Genbank as well as the Broad website.

7c. Research Compliance Requirements

Upon project approval, NIAID review of relevant IRB/IACUC documentation is required prior to commencement of work. Please contact the GSC Principal Investigator(s) to ensure necessary documentation are filed for / made available for timely start of the project.

Investigator Signature:

Investigator Name: **Date:**

REFERENCES

1. Abdulla, S., et al., *Safety and immunogenicity of RTS,S/AS02D malaria vaccine in infants.* N Engl J Med, 2008. **359**(24): p. 2533-44.

2. Alonso, P.L., et al., *Duration of protection with RTS,S/AS02A malaria vaccine in prevention of Plasmodium falciparum disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial.* Lancet, 2005. **366**(9502): p. 2012-8.

3. Alonso, P.L., et al., *Efficacy of the RTS,S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: randomised controlled trial.* Lancet, 2004. **364**(9443): p. 1411-20.

4. Alloueche, A., et al., *High-throughput sequence typing of T-cell epitope polymorphisms in Plasmodium falciparum circumsporozoite protein.* Mol Biochem Parasitol, 2000. **106**(2): p. 273-82.

5. Alloueche, A., et al., *Protective efficacy of the RTS,S/AS02 Plasmodium falciparum malaria vaccine is not strain specific.* Am J Trop Med Hyg, 2003. **68**(1): p. 97-101.

6. Barry, A.E., et al., *Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, Plasmodium falciparum.* PLoS ONE, 2009. **4**(12): p. e8497.

7. Bejon, P., et al., *Efficacy of RTS,S/AS01E vaccine against malaria in children 5 to 17 months of age.* N Engl J Med, 2008. **359**(24): p. 2521-32.

8. Bojang, K.A., et al., *Efficacy of RTS,S/AS02 malaria vaccine against Plasmodium falciparum infection in semi-immune adult men in The Gambia: a randomised trial.* Lancet, 2001. **358**(9297): p. 1927-34.

9. Cai, T. Tonini, G., Lin, X., Ivo-Calle, J.M., et al., *Kernel machine approach to testing the significance of multiple genetic markers for risk prediction.* Biometrics, 2010. in press.

10. Calvo-Calle, J.M., et al., *Immunogenicity of multiple antigen peptides containing B and non-repeat T cell epitopes of the circumsporozoite protein of Plasmodium falciparum*. J Immunol, 1993. **150**(4): p. 1403-12.

11. Chatterjee, N., et al., *Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions.* Am J Hum Genet, 2006. **79**(6): p. 1002-16.

12. Chatterjee, N., *Semiparametric maximum likelihood estimation in case-control studies of gene-environment interactions.* Biometrika, 2005. **92**: p. 399-418.

13. de Stricker, K., et al., *Conservation and heterogeneity of the glutamate-rich protein (GLURP) among field isolates and laboratory lines of Plasmodium falciparum.* Mol Biochem Parasitol, 2000. **111**(1): p. 123-30.

14. Daniels, R., et al., *A general SNP-based molecular barcode for* Plasmodium falciparum *identification and tracking*. Malar J., 2008. **7**: 223.

15. Doolan, D.L., A.J. Saul, and M.F. Good, *Geographically restricted heterogeneity of the Plasmodium falciparum circumsporozoite protein: relevance for vaccine development.* Infect Immun, 1992. **60**(2): p. 675-82.

16. Enosse, S., et al., *RTS,S/AS02A malaria vaccine does not induce parasite CSP T cell epitope selection and reduces multiplicity of infection.* PLoS Clin Trials, 2006. **1**(1): p. e5.

17. Guinovart, C., et al., *Insights into long-lasting protection induced by RTS,S/AS02A malaria vaccine: further results from a phase IIb trial in Mozambican children.* PLoS ONE, 2009. **4**(4): p. e5165.

18. Good, M.F., J.A. Berzofsky, and L.H. Miller, *The T cell response to the malaria circumsporozoite protein: an immunological approach to vaccine development.* Annu Rev Immunol, 1988. **6**: p. 663-88.

19. Good, M.F., et al., *Parasite polymorphism present within minimal T cell epitopes of Plasmodium falciparum circumsporozoite protein.* J Immunol, 1988. **140**(5): p. 1645-50.

20. Jalloh, A., M. Jalloh, and H. Matsuoka, *T-cell epitope polymorphisms of the Plasmodium falciparum circumsporozoite protein among field isolates from Sierra Leone: age-dependent haplotype distribution?* Malar J, 2009. **8**: p. 120.

21. Jeffares, D.C., et al., *Genome variation and evolution of the malaria parasite* Plasmodium falciparum. Nat Genet, 2007. **39**(1): p. 120-5.

22. Jiang, H., et al., *Detection of genome-wide polymorphisms in the AT-rich* Plasmodium falciparum *genome using a high-density microarray*. BMC Genomics, 2008. **9**: p. 398.

23. Jongwutiwes, S., et al., *Allelic variation in the circumsporozoite protein of Plasmodium falciparum from Thai field isolates.* Am J Trop Med Hyg, 1994. **51**(5): p. 659-68.

24. Kester, K.E., et al., *Randomized, double-blind, phase 2a trial of falciparum malaria vaccines RTS,S/AS01B and RTS,S/AS02A in malaria-naive adults: safety, efficacy, and immunologic associates of protection.* J Infect Dis, 2009. **200**(3): p. 337-46.

25. Kumkhaek, C., et al., *Are extensive T cell epitope polymorphisms in the Plasmodium falciparum circumsporozoite antigen, a leading sporozoite vaccine candidate, selected by immune pressure?* J Immunol, 2005. **175**(6): p. 3935-9.

26. Kumkhaek, C., et al., *A survey of the Th2R and Th3R allelic variants in the circumsporozoite protein gene of P. falciparum parasites from western Thailand.* Southeast Asian J Trop Med Public Health, 2004. **35**(2): p. 281-7.

27. Lell, B., et al., *A randomized trial assessing the safety and immunogenicity of AS01 and AS02 adjuvanted RTS,S malaria vaccine candidates in children in Gabon.* PLoS ONE, 2009. **4**(10): p. e7611.

28. Lockyer, M.J., K. Marsh, and C.I. Newbold, *Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell epitopes of the circumsporozoite protein.* Mol Biochem Parasitol, 1989. **37**(2): p. 275-80.

29. Maity, A., et al., *Testing in semiparametric models with interaction, with applications to gene-environment interactions.* J R Stat Soc Series B Stat Methodol, 2009. **71**(1): p. 75-96.

30. Malik, A., et al., *Human cytotoxic T lymphocytes against the Plasmodium falciparum circumsporozoite protein.* Proc Natl Acad Sci U S A, 1991. **88**(8): p. 3300-4.

31. Moreno, A., et al., *CD4+ T cell clones obtained from Plasmodium falciparum sporozoite-immunized volunteers recognize polymorphic sequences of the circumsporozoite protein.* J Immunol, 1993. **151**(1): p. 489-99.

32. Mu, J., et al., *Genome-wide variation and identification of vaccine targets in the* Plasmodium falciparum *genome*. Nat Genet, 2007. **39**(1): p. 126-30.

33. Mu, J., et al., Plasmodium falciparum *genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs*. Nat Genet, 2010. **42**(3): p. 268-71.

34. Munesinghe, D.Y. , et al., Immunogenicity of multiple antigen peptides (MAP) containing T and B cell epitopes of the repeat region of the P. falciparum circumsporozoite protein. Eur. J. Immunol 1991. **21** (12): p. 3015-20.

35. Neafsey, D.E., et al., *Genome-wide SNP genotyping highlights the role of natural selection in* Plasmodium falciparum *population divergence*. Genome Biol, 2008. **9**(12): R171.

36. Owusu-Agyei, S., et al., *Randomized controlled trial of RTS,S/AS02D and RTS,S/AS01E malaria candidate vaccines given according to different schedules in Ghanaian children.* PLoS ONE, 2009. **4**(10): p. e7302.

37. Polhemus, M.E., et al., *Evaluation of RTS,S/AS02A and RTS,S/AS01B in adults in a high malaria transmission area.* PLoS ONE, 2009. **4**(7): p. e6465.

38. Sacarlal, J., et al., *Long-term safety and efficacy of the RTS,S/AS02A malaria vaccine in Mozambican children.* J Infect Dis, 2009. **200**(3): p. 329-36.

39. Shi, Y.P., et al., *Diversity in the immunodominant determinants of the circumsporozoite protein of Plasmodium falciparum parasites from malaria-endemic regions of Papua New Guinea and Brazil.* Am J Trop Med Hyg, 1992. **47**(6): p. 844-51.

40. Singh, J.P., et al., *Plasmodium falciparum circumsporozoite protein: epidemiological variations among field isolates prevalent in India.* Trop Med Int Health, 2009. **14**(8): p. 957-66.

41. Stoute, J.A., et al., *Long-term efficacy and immune responses following immunization with the RTS,S malaria vaccine.* J Infect Dis, 1998. **178**(4): p. 1139-44.

42. Tanabe, K., N. Sakihama, and A. Kaneko, *Stable SNPs in malaria antigen genes in isolated populations.* Science, 2004. **303**(5657): p. 493.

43. Vashishtha, V.M., *World Malaria Report 2008 A Billion dollar Moment for a Centuries Old Disease?* Indian Pediatr, 2008. **45**(12): p. 985-6.

44. Volkman, S. et al., *A genome-wide map of diversity in* Plasmodium falciparum. Nat Genet, 2007. **39**(1): p. 113-9.

45. Waitumbi, J.N., et al., *Impact of RTS,S/AS02(A) and RTS,S/AS01(B) on genotypes of P. falciparum in adults participating in a malaria vaccine clinical trial.* PLoS ONE, 2009. **4**(11): p. e7849.

46. Wu, M.C., et al., *Powerful SNP set analysis for case-control genomewide association studies.* Am J Human Genet, 2010. **89**(11): 929:942.

47. Yoshida, N., et al., *Plasmodium falciparum: restricted polymorphism of T cell epitopes of the circumsporozoite protein in Brazil.* Exp Parasitol, 1990. **71**(4): p. 386-92.

48. Zakeri, S., et al., *Restricted T-cell epitope diversity in the circumsporozoite protein from Plasmodium falciparum populations prevalent in Iran.* Am J Trop Med Hyg, 2007. **76**(6): p. 1046-51.