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:

Authors:

Primary Investigator Contact:

Name	Stephen L. Hoffman, MD
Position	Chief Executive and Scientific Officer
Institution	Sanaria Inc.
Address	9800 Medical Center Drive, Suite A209, Rockville
State	MD
ZIP Code	20850
Telephone	240-403-2701
Fax	301-770-5554
E-Mail	slhoffman@sanaria.com

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

Malaria afflicts hundreds of millions of people, kills 0.65-1.4 million each year, and greatly impedes economic development in the developing world, especially in sub-Saharan Africa [1, 2]. Powerful tools for eliminating *Plasmodium falciparum* (Pf) from defined geographical areas would be highly effective, long-acting vaccines and/or drugs that target the pre-erythrocytic sporozoite and liver stages of the parasite. The pre-erythrocytic stages are the ideal target, because they do not cause any clinical manifestations or pathology. Thus, effective vaccines or drugs against these stages would prevent infection, and thereby prevent disease and transmission [3].

One of the reasons why it has been difficult to target the liver stages with new drugs and vaccines is the difficulty in culturing Plasmodium liver stages, and even when there is successful parasite invasion and development in hepatocytes, the number of parasite-infected hepatocytes as compared to uninfected hepatocytes is never greater than 5% and generally significantly lower than 1%. Thus, distinguishing between background hepatocyte signals and parasite signals has been challenging. Our advances in axenic (cell-free) culture of *P. falciparum* liver stages and the sensitivity of RNA deep sequencing as compared to microarray technology have provided a unique opportunity to finally do a complete analysis of the genes expressed in the critically important, early liver stages of *P. falciparum*.

We propose two phases for this White Paper. During Phase I, we propose to take advantage of JCVI's RNA deep sequencing capabilities and Sanaria's unique capacity to produce aseptic, purified *Plasmodium falciparum* sporozoites (PfSPZ) and 3 day Pf liver stage parasites in axenic culture to sequence both coding and non-coding RNA and determine the gene expression profile of PfSPZ and PfSPZ that have transformed to liver stage parasites during 3 days in axenic culture. Axenically transformed sporozoites are charachterized by morphology (as they differ from salivary gland sporozoites) and the expression of early liver-stage markers such as PfEXP1 and PfHSP70. By these measures we are sure that these parasites are similar to naturally developed liver stage parasites. We will sequence both coding and non-coding RNA from the same strain of *P. falciparum* as a

comparator for the quality and the quantity of RNA yield and sequencing procedures. Sanaria has the only capability in the world to produce such parasites. RNAseq data will be obtained from three different sets of samples produced during three different production campaigns at Sanaria (biological replicates).
During Phase II, we propose to perform comparative transcriptomics of non-irradiated vs irradiated sporozoites as well as transcriptomics analysis of infected hepatocytes. Phase II will not be discussed in this White Paper, but we expect to develop an addendum as Phase I progresses during year one.
Our expectation is that this work will allow us to define for first time key expression patterns and potentially new genes in this difficult to obtain, and understudied life cycle stage. We expect this to be of great interest to the wide malaria research community. This work will provide critical new information to the community and it will enhance the richness of current databases, such as PlasmoDB, where these results will be released publically in addition to Genbank.

2. Justification

Provide a succinct justification for the sequencing or genotyping study by describing the significance of the problem and providing other relevant background information.

This section is a key evaluation criterion.

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

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?*

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.*

A. Public health Relevance:

Malaria is endemic in much of the developing the world, particularly in Southeast Asia and sub-Saharan African, responsible for more than 200 million infections per year and nearly a million deaths, mostly infants and children. (http://www.who.int/malaria/). Five different species of *Plasmodium* can cause malaria in humans: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, Plasmodium malariae* and *Plasmodium knowlesi*. *P. falciparum* accounts for most of the mortality associated with malaria.

The first step in human infection by a *Plasmodium* parasite occurs in the liver of the host. During a blood meal, a malaria-infected female *Anopheles* mosquito

inoculates sporozoites into the human host. The sporozoites travel through blood vessels and infect liver cells to multiply into thousands of blood stage infective merozoites. In the case of *P. falciparum* this takes a minimum of 5.5 days. The replication in hepatocytes ensures the release of high numbers of merozoites into the blood stream. Thus liver stages of *Plasmodium* have enormous potential for anti-malarial intervention, either by drugs or by vaccine, before any clinical symptoms.

B. Existing Genome Data

PlasmoDB version 9.0 was released on May 2012, and constitutes a complete rebuild of the database, where all genomes, data and analysis were redone and improved. This is one of the main resources for *Plasmodium* and it contains genome information from 17 datasets that correspond to 9 species and strains of the genus *Plasmodium*. Data includes strand-specific, amplification-free transcriptome sequence (**FRT-Seq**) from *P. falciparum* 3D7 parasites (pooled red blood cell stage). It has been shown that the expression of both coding and non-coding RNA is different in different stages of blood-stage Plasmodium parasites. No RNAseq data from sporozoite stages is present. We anticipate to uncover many different both coding and non-coding RNAs that are unique to sporozoites and liver stages of *Plasmodium*. In addition, The Broad Institute, in collaboration with Dyann Wirth, Sarah Volkman and Daniel Hartl from Harvard University, is sequencing two clones of *P. falciparum* used in studies of malarial genetics, HB3 and Dd2, to better understand polymorphism in *P. falciparum*. Other strains and clones described as part of that project are: D10, 7G8, D6, FCC-2, RO-33, SL, K1, Senegal_V34.04, VS-1, IGH-CR14, NF54, and NF135/5.C10. In some instances, nucleus, mitochondrion, and apicoplast were sequenced. Strains correspond to different geographical regions. *Plasmodium vivax* was sequenced at JCVI (TIGR) as part of the MSC, under Jane Carlton direction and Lis Caler project management. Since Sanaria recently succeeded in making aseptic sporozoites, deep sequencing of both coding and non-coding RNA from *P. vivax* sporozoites will be considered during Phase II of this proposal.

C. **Value to the community:**

Successful deep sequencing of the proposed stages of *Plasmodium* will help the scientific community to better understand the biological basis of liver stage infection and help to develop malarial interventions against this stage.

D. **Clinical and scientific value:**

Unexplored liver stages: Despite the obvious importance of the liver stage development in human host and its potential as a vaccine target, our knowledge of this stage of the *Plasmodium* parasite is very limited. Studying liver stage has been difficult due to the technical difficulties of obtaining enough liver stage parasites to perform any "omics" level RNA or protein analysis. Sanaria has the only facility in the world that can produce large quantities of aseptic, purified salivary gland sporozoites. Sanaria also established methods to produce large amounts of liver stage parasites by axenic development of its aseptic, purified sporozoites without any host cells.

Certain genes, pathways or non-coding RNAs may be identified that will show

clear association with parasite development in liver, which will ultimately enable discovery of new therapies or interventions.

Genomic sequences of axenic versus salivary gland stage sporozoites will help us understand the development of liver stage parasites. Successful completion of this project will yield means to extend analogous studies to other stages in the *Plasmodium* life cycle and possibly the host-parasite interactions.

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.*

Sanaria's goal is to manufacture and commercialize a pre-erythrocytic stage vaccine composed of attenuated PfSPZ. Sanaria has succeeded in manufacturing sufficient quantities of an attenuated whole sporozoite vaccine that meets regulatory standards, including purity, sterility, safety, and potency using the NF54 strain of *P. falciparum.* The NF54 strain has been one of the most favored by researchers around the word due to its ease of culturing in laboratory conditions, and the large number of sporozoites per mosquito it yields. The genomic sequence of Sanaria's NF54 parasite has recently been produced at the Broad Institute, and a comparative analysis with the sequence of the 3D7 clone from NF54 is now underway (e.g. 3D7 is a clone of NF54).

A. Sporozoites and 72 Hour Liver Stage Parasites:

Sanaria has unique expertise to produce, extract and purify aseptic PfSPZ based on the proprietary technology developed for Sanaria's sporozoite based human vaccines. These sporozoites are not only free of any mosquito materials but also free of any microorganisms. The manufacturing and release of the sporozoites are described in Epstein et al 2011[4]. Sanaria will use similar methodology to produce 300 million total PfSPZ in three different manufacturing campaigns, 100 million each in three biological replicates, and use 40 million sporozoites for RNA extraction and 60 million sporozoites for axenic transformation into 72 hour liver stage parasites. PfSPZ manufactured will be tested for their ability to invade human hepatocytes using the 3-day hepatocyte potency assay [4]. For axenic development, 60 million sporozoites will be incubated in complete hepatocyte culture medium (10% FBS in 1X HBSS with penicillin and streptomycin) for three days with daily media change. 72 hours later *in-vitro* axenically transformed liver stage parasites will be harvested to make high quality RNA for deep sequencing. We will check the quality of axenic transformation by the change in morphology as well as by the expression of early liver stage specific antigens like PfEXP1.

***B.* Asexual blood stage parasites:**

At Sanaria we routinely produce large number of aseptic blood stage NF54 *Plasmodium* parasites to manufacture PfSPZ Vaccine. We will produce 40 million asexual blood stage parasites for deep sequencing. Coding and non-coding RNA

sequencing of the 3D7 clone of NF54 of blood stage parasites has been reported recently [5]. RNA sequencing of NF54 asexual blood stage parasites will serve as an internal control for RNA extraction, library preparation as well as sequencing. We will compare the sequences that we get from NF54 to that of 3D7.

Table 1. Summary of strain and stages proposed for deep RNA sequencing

Strain	Number and Stage	Purpose of RNA deep sequencing	Time-frame of availability and collaboration
NF54	4×10^7 aseptic, purified salivary gland sporozoites.	**1)** To compare the gene expression profiles of sporozoites and axenically cultured liver stage parasites **2)** To identify novel coding and non-coding RNAs in sporozoites.	3-4 months for material generation 6 months for sequencing, analysis and annotation
NF54	6×10^7 aseptic, purified salivary gland sporozoites axenically transformed to yield 4×10^7 liver stage parasites.	**3)** To compare the gene expression profiles of sporozoites and axenically cultured liver stage parasites **4)** To identify novel coding and non-coding RNAs in liver stages of Pf.	3-4 months for material generation 6 months for sequencing, analysis and annotation
NF54	4×10^7 asexual blood stage parasite as a control	**5)** This will serve as a comparator for deep sequencing as well as sequence analysis	Will be provided to JCVI at the same time as the sporozoites

4a. Approach to Data Production: **Data Generation**

C. 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.)

RNA isolation and proposed libraries from sporozoites and 72 hour axenic liver stage parasites:

Based on preliminary experiments for each set (PfSPZ and 72 hour liver stage parasites), we anticipate requiring $4x10^7$ PfSPZ for the PfSPZ sequencing, and $6x10^7$ PfSPZ to produce the 72-hour liver stage parasites. Thus, $3x10^8$ aseptic, purified PfSPZ in total will be required. Sanaria will produce and prepare the material and JCVI will produce the appropriate libraries, conduct the sequencing, analyze and report the results. Infections /experiments will be performed in triplicates. From these datasets we will perform polyA mRNA enrichment and we will utilize the remaining "flowthrough" RNA to generate ncRNA libraries from pooled material.

We will prepare directional (strand specific) RNA-Seq libraries to enhance the value of our samples for transcriptome annotation, profiling and potential gene discovery[6]. Increasing evidence suggests that noncoding RNAs (ncRNA) perform important regulatory functions. Currently, there is very limited or no high throughput small RNA data on any given Plasmodium strain, and no data at all on non-erythrocyte stages such as sporozoites we are proposing[5]. Therefore, our protocols for RNA processing and library preparation will be performed following the kits indicated below.

Protocols for ncRNA and mRNA extraction. For total RNA isolation, including miRNA, we will use MasterPure™ RNA Purification Kit, followed by miRNA processing using ScriptMiner™ Small RNA-Seq Library Preparation Kit. This kit also generates the

sequencing library so the final product is ready for sequencing. For mRNA, we will use an adaptation of a protocol for RNAseq from Illumina TruSeq RNA Sample Preparation Kits v2.
Bioinformatic analyses to identify differentially expressed genes and pathways will be performed at JCVI and will be compared with existing transcriptomes (data from PlasmoDB) and our internal control sample.

Quality control of RNA sample: In order to ensure the RNA quality, we will make sure that the RNA is not degraded and it is free of DNA (RNA will be treated with DNAse).

RNA Concentration: We will produce RNA samples with a concentration of 100 ng/µl or above and if necessary, we will dilute the RNA solution in 1X TE to achieve appropriate concentration. RNA will be submitted for sequencing as pellets in 70% ethanol.

Documentation: Samples submitted for deep sequencing shall have the following accompanying documentation **-**

Estimation of RNA concentration.
Estimation of total RNA.

In case a different RNA preparation method is used, additional relevant information will also be documented.

4b. Approach to Data Production: **Data Analysis**

D. *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.*

Bioinformatic analyses to identify differentially expressed genes and pathways will be performed at JCVI. Data will be compared with available transcriptomes obtained from other parasite stages. Sporozoite transcriptomic data sets will be an invaluable resource for the Plasmodium community as they will present for first time, a view of this particular stage that will become the first sporozoite reference source.

Identification of differentially expressed, sporozoite-specific genes: JCVI has the capacity to perform RNAseq analysis with a variety of tools that in many instances it is necessary to test, to evaluate the best performance. For this particular project, transcriptome analysis will be performed by comparing the expression profiles of control (erythrocytic stages) and sporozoites immediately extracted from insects and 72 hr axenic culture RNA-seq datasets using the CLC Genomics Workbench package. Reference annotation from most recent PlasmoDB dataset will be imported into CLC, and the number of reads mapped to each gene will be recorded. Expression profiles are normalized for compositional bias in sequenced libraries and for differences between libraries in sequencing depth. CLC offers a variety of statistical methods to compare experiments generating p-values for each gene and providing a number of tools to visualize significant expression differences among control and target samples for each gene of interest. This approach has already been successfully tested in a preliminary study performed at JCVI using the transcriptomes from non-irradiated vs irradiated sporozoites.
Establishing the transcriptional state of the sporozoites and liver-like stage

sporozoites is a fundamental step towards understanding the effects of irradiation in the vaccine targets. This information will be unique, and a stepping stone for future research.

5. Community Support and Collaborator Roles:

E. *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.*

F. *List all project collaborators and their roles in the project*

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

There are several laboratories in the world working on the liver stages of *Plasmodium*, and quite a few of them are interested in knowing the RNA sequencing results of sporozoites and liver stages by deep sequencing.

Names of few investigators who will benefit from the availability of this deep sequencing results:

Dr. David A. Fidock, Ph.D. Professor of Microbiology & Immunology and Medical sciences, Columbia University, New York, NY.

Dr. Sangeeta Bhatia, MD., PhD. Professor and Director, Laboratory of Multiscale Regenerative Technologies, MIT, Cambridge, MA

Dr. Stefan Kappe, PhD. Professor and Malaria Program Director, Seattle Biomedical Research Institute, Seattle, WA.

Dr. Ana Rodriguez, Ph.D. Associate Professor, Department of Microbiology and Parasitology, NYU Medical center, New York, NY.

Dr. Maria M. Mota, Ph.D. Group Leader, Malaria Research Program, Lisbon's Instituto de Medicina Molecular, University of Lisbon, Portugal.

Sanaria will produce all sporozoites and axenically cultured liver stage parasites necessary for deep sequencing. JCVI will be responsible for making high quality RNA, labeling, deep sequencing, analysis, annotation and submission of data to the public data bases.

Collaborators and their role:

Sanaria: Project Co-coordination in collaboration with JCVI Project manager, providing sequencing-quality parasite materials, manuscript preparation.

JCVI: Overall project coordination and management, sample processing and quality control, deep sequencing of RNA, data analysis, data submission to relevant repositories and co-leading manuscript writing.

6. Availability & Information of Strains:

H. *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.

I. *Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g*
   - *Name*
   - *Identifier*
   - *Material type (DNA/RNA/Strain)*
   - *Genus*
   - *Species*
   - *Specimen / Strain*
   - *Isolation source*
   - *Isolated from*
   - *Select agent status*
   - *International permit requirement*
   - *BEIR/ATCC repository accession number*
   - *Other public repository location*
   - *Other public repository identifier*
   - *Sample provider's name*
   - *Sample provider's contact*

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

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://www.niaid.nih.gov/labsandresources/resources/mscs/data.htm
http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html

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:

K. 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/).

Strains. We will use the *Plasmodium falciparum* strain NF54 currently deposited in *Biodefense and Emerging Infections Research Resources Repository (BEI Resources) of* **National** *Institute of Allergy and Infectious Diseases (NIAID) and the ATCC.*

Sequence reads. All original sequence reads will be deposited to the NCBI's Short Read Archive sequence database.

Metadata on samples. Metadata regarding the RNA collection, parasite stage and parasite origin will be appropriately appended to the SRA submission.

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.

References:

1. Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, Sabot O, Rodriguez MH, Abeyasinghe RR, Ghebreyesus TA, and Snow RW. Shrinking the malaria map: progress and prospects. *Lancet* 376: 1566-1578, 2010.
2. Murray CJ, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, Fullman N, Naghavi M, Lozano R, and Lopez AD. Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet* 379: 413-431, 2012.
3. Hoffman SL, Billingsley P, James E, Richman A, Loyevsky M, Li T, Charkravarty S, Gunasekera A, Chattopadhyay R, Li M, Stafford R, Ahumada A, Epstein JE, Sedegah M, Reyes S, Richie TL, Lyke KE, Edelman R, Laurens MB, Plowe CV, and Sim BKL. Development of a metabolically active,non-replicating sporozoite vaccine to prevent *Plasmodium falciparum* malaria. *Human Vaccines* 6: 97-106, 2010.

4. Epstein JE, Tewari K, Lyke KE, Sim BK, Billingsley PF, Laurens MB, Gunasekera A, Chakravarty S, James ER, Sedegah M, Richman A, Velmurugan S, Reyes S, Li M, Tucker K, Ahumada A, Ruben AJ, Li T, Stafford R, Eappen AG, Tamminga C, Bennett JW, Ockenhouse CF, Murphy JR, Komisar J, Thomas N, Loyevsky M, Birkett A, Plowe CV, Loucq C, Edelman R, Richie TL, Seder RA, and Hoffman SL. Live attenuated malaria vaccine designed to protect through hepatic CD8 T cell immunity. *Science* 334: 475-480, 2011.
5. Lopez-Barragan MJ, Lemieux J, Quinones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K, and Su XZ. Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. *BMC Genomics* 12: 587, 2011.
6. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, and Regev A. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* 7: 709-715, 2010.

Investigator Signature:

Investigator Name: **Date**