GENE EXPRESSION PROFILING APPROACH TOWARDS ENHANCEMENT OF MALARIA VACCINE DEVELOPMENT

A Thesis
Presented to
The Academic Faculty

by

Mónica L. Rojas-Peña

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
December 2016

GENE EXPRESSION PROFILING APPROACH TOWARDS ENHANCEMENT OF MALARIA VACCINE DEVELOPMENT

Approved by:

Dr. Greg Gibson, Advisor
School of Biology
Georgia Institute of Technology

Dr. Todd Streelman
School of Biology
Georgia Institute of Technology

Dr. Julia Kubanek
School of Biology
Georgia Institute of Technology

Dr. Robert E. Guldberg
School of Mechanical Engineering
Georgia Institute of Technology

Dr. King Jordan
School of Biology
Georgia Institute of Technology

Date Approved: August 09, 2016

To my mom MamiCeci, viejita te extraño!

ACKNOWLEDGEMENTS

I would like to thank Dr. Greg Gibson for all the help, support and patience during these years, and for giving me the opportunity to join his lab. I thank Dr. Julia Kubanek, Dr. Tood Streelman, Dr. King Jordan and Dr. Robert Guldberg for their willingness to be a part of my committee and for their comments and suggestions.

We thank Dr. Myryam Arevalo-Herrera and Dr. Socrates Herrera for their valuable collaboration and developing the studies to obtain the samples for this research. We would also like to thank all the members of the CAUCASECO Research Center that were part of these studies. We also thank all the volunteers, patients and families for being willing to participate in these studies.

I thank the Schlumberger Faculty for the Future Program for the financial help provided, because it allowed me to focus on my research and it gave me the opportunity to meet such inspiring women that belong to this program.

I would like to thank previous and present Gibson lab members, especially Dalia, Urko, Swetha, and Jing for their friendship and support in the research process; and I would also like to thank Shweta Biliya from Vannberg Lab for all her help and her willingness to always accommodate my sequencing needs.

I also thank my friends from all over the world and in Atlanta, especially the Swordfish Underwater Hockey family. Thanks for the friendship and for making me feel like home, and for keeping me sharp through this process. I'm very glad to know all of you.

Finally, thanks to my family: **Cecilia**, Andrei, Alex, Jefferson, Yenny, Tatiana, Julian, Mildred, Cami, Sammy, Thomy, David, Salome and Adriana for your unconditional love and support, and for teaching me that everything needs time. I would especially like to thank my mom, Cecilia, who taught me to keep smiling and fight even in difficult times, and who I dedicate all my work, and who I deeply miss. A special thanks to my brother Alex, my role model of persistence and dedication. Finally, a special thanks to my husband, Andrei, who empowers me to be better and who supports me no matter what.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

WHO	The World Health Organization
*Pv*RAS	*Plasmodium vivax* Sporozoites
SIVIGILA	Public Health Surveillance System
RBCs	Red Blood Cells
TNF	Tumor Necrosis Factor
PCR	Polymerase Change Reaction
ACT	Artemisinin-based Combination Therapy
GSK	GlaxoSmithKline Biologicals
MVI-PATH	PATH Malaria Vaccine Initiative
CS	Circumsporozoite
HBsAg	Hepatitis B surface antigen
VE	Vaccine Efficacy
G6PD	Glucose-6-phosphate Dehydrogenase Enzyme
DCs	Dendritic Cells
NKs	Natural Killer Cells
TNF-α	Tumor Necrosis Factor α
IFN-γ	Interferon Gamma γ
IL-10	Interleukin 10
TGF-β	Transforming Growth Factor β
iRBC	Infected Erythrocytes
RT-qPCR	Real Time PCR
IRB	Institutional Review Boards
CECIV	Malaria Vaccine and Drug Development Center

RIN	RNA Integrity score
FPKM	Fragments per Kilobase of Exon per Million Aligned Fragments
SNM	Supervised Normalization of Microarray
GEO	Gene Expression Omnibus Archive
PCA	Principal Components Analysis
PVCA	Principal Variance Components Analysis
NLP	Negative log10 of the p-value
BIT	Blood informative transcript
IL	Interleukin
IFN	Interferon
TNF	Tumor Necrosis Factor
CS	Circumsporozoite
CPM	Counts Per mMillion
TMM	Trimmed Mean of the M-values
BTMs	Blood Transcription Modules
TBS	Thick Blood Smear
SLPC	Strongly tend to Differentiate Short-lived
LLPC	Long-lived Plasma Cells
MoH	Colombian Ministry of Health
LT	Low Throughput

SUMMARY

Malaria is one of the most deadly infectious diseases in the world and has an enormous public health burden with significant economic implications. The World Health Organization report (WHO 2015) showed a decrease in mortality of 48% between 2000 and 2014. However, more than 3 billion people are still at risk, and 438,000 deaths are attributed to malaria every year, largely affecting children and pregnant women. Eighty percent of malaria deaths occur in just 15 countries. Among the causative agent *Plasmodium* species, *Plasmodium falciparum* is the most harmful, specifically in Africa, and most of the research is focused on this species. *Plasmodium vivax* is the second most common species and accounts for the 70% of the severe malaria cases and deaths in the Americas. *P. vivax* malaria is a debilitating, occasionally life-threatening, and economically burdensome disease in Central Latin America, where 70% - 80% of the population lives with the endemic risk of infection.

Research progress for this species has been slower than for *P. falciparum* largely because *P. vivax* is difficult to maintain in culture, and this is reflected in the limited information of the species in terms of genome, transcriptome and proteome. Especially since the WHO (2015) reported that prevalence of severe cases of *P. vivax* malaria is emerging, while treatment failure with chloroquine for *P. vivax* malaria is increasing, development of a malaria vaccine is a top priority.

Advances have been made with the development by GlaxoSmithKline Ltd of the RTS,S vaccine (MosquirixTM) which is targeted against the *P. falciparum* circumsporozoite protein. RTS,S Phase 1, 2, and 3 trials have demonstrated efficacy and safety, and the vaccine seems to decrease morbidity and mortality by up to 36% in children and 26% in infants (RTS,S Clinical Trials Partnership, 2015) (Lancet 2015). However, protection with RTS,S is not guaranteed as the vaccine was tested only on a population of African newborns and young children, and is specific to *P. falciparum*. Recognizing that researchers at Emory and Georgia Tech recently used gene expression profiling to identify transcripts that both serve as biomarkers for generation of antibodies against influenza vaccines, and define novel immune-regulatory mechanisms that may contribute to the effectiveness of vaccination (Nakaya et al. 2011), we reasoned that transcriptomics could illuminate mechanisms of vaccine effectiveness in malaria.

This dissertation describes initial gene expression profiling experiments using RNA Sequencing technology (RNASeq) applied to samples collected during clinical trials performed at the CAUCASECO research center in Cali, Colombia. I describe three gene expression-profiling studies.

The first study leveraged an experiment that reported more severe malaria symptoms in subjects who had never experienced malaria before the clinical trial. We found that there is no obvious difference in the transcriptomes of uninfected "naïve" compared with previously exposed "semi-immune" volunteers prior to infection, but several hundred genes showed a stronger response in the naïve individuals at the time that parasites begin

to appear in the peripheral blood. Notably, differential expression of both neutrophil and interferon-related genes was evident at onset of malaria. Also, after interaction analysis, 175 genes showed a significant Timepoint-by-Prior-exposure effect and most of the genes showing this interaction effect were more strongly up- or down-regulated in the naïve than semi-immune individuals, falling into several pathways of interest.

In the second study I report results of gene expression profiling of peripheral blood before and after 7 rounds of immunization with radiation attenuated *P. vivax* sporozoites (*Pv*RAS) in 20 volunteers, as well as after controlled challenge with live *P. vivax*. The most profound changes in gene expression were observed in the contrasts between baseline and post- challenge, with distinct signatures differentiating protected and susceptible individuals. Analysis of transcriptional modules shows that aspects of B-cell signaling are reduced while cell cycle regulation and interferon responses, and likely immunoglobulin production by short-lived plasma cells, are highly elevated in individuals not protected by *Pv*RAS, whereas regulatory T-cell signaling and an inflammatory response are elevated in protected individuals. Furthermore, subtle differences in the protection afforded by Duffy negative status and by *Pv*RAS were observed, and vaccination itself also modified aspects of B and T cell gene expression. Combined with immune cell profiling we expect the systems biology approach to suggest adjuvants that may improve the efficacy of malaria vaccines.

Finally, the third study describes pilot longitudinal profiling of the host transcriptomes of eight complicated malaria cases, each ascertained over four or more days of recovery. I

relate the gene expression changes to clinical features including parasitemia, disease severity, and rate of recovery. Patient CM02, a pregnant woman with symptoms of preeclampsia but no history of fever, who failed to resolve her complicated malaria infection for over a month, showed the most perturbed gene expression. In most of the patients, the profiles are interpreted as providing evidence for inflammation and elevated interferon signaling being resolved several days after hospitalization, and of reticulocyte development associated with recovery from anemia occurring only after the peak of the complex malaria episode. While we observed some commonalities in the responses across patients, there was wide individual variability dominated by baseline differences in immune activity, and I conclude that more detailed analyses of individual cell types linked to cellular and humoral data will be required to resolve the nature of personalized mechanisms of recovery.

The results presented in this dissertation explore how gene expression profiling of the complex mixture of cells present in whole blood can nevertheless reveal the cellular nature and duration of the immune response to *P. vivax* infection, while also highlighting a subset of genes that may mediate adaptive immunity. These results demonstrate the potential value of RNASeq for studying the response of the host transcriptome of a malaria infection, but also show that much more work needs to be done before genomic profiling can be considered for integration as a component of personalized clinical diagnostics.

CHAPTER 1

INTRODUCTION

My thesis utilizes gene expression profiling of the human immune response to malaria infection and vaccination in order to identify mechanisms predictive of vaccine immunogenicity, safety, and clinical recovery from a malaria infection, and to help to define the most effective strategies for vaccine implementation. This research was based on data collected at the CAUCASECO research center in Cali, Colombia, where blood samples from volunteers and patients enrolled in studies towards the development of a *P. vivax* malaria vaccine were used to perform RNAseq analysis.

This thesis consists of three studies. The objective of the second chapter was to characterize the molecular basis for differences in clinical course of disease as a function of prior exposure to malaria. This study was published in *PLOS Neglected Tropical Diseases* in 2015 (**9**(**8**): e0003978). In chapter 3, I adopted a transcriptional profiling approach to characterize signatures of the impact of *Pv*RAS immunization on the response to *P. vivax* malaria challenge and we are currently preparing the manuscript for publication. In chapter 4, I describe a pilot study of the host transcriptomes of eight complicated malaria cases each over four or more days of recovery, and relate the gene expression changes to clinical features including parasitemia, disease severity, and rate of recovery. Another gene expression study not described in this thesis concerned genomic classification of craniosynostosis Rojas-Pena *et al*, (2014) *J. Genomics* **2**: 121-130.

1.1 Malaria epidemiology

Malaria is a protozoan disease transmitted by the female *Anopheles* mosquito. The causative agent of malaria is a parasite of the genus *Plasmodium*. These organisms are single cell eukaryotic protozoa that can infect a wide range of hosts, including humans (Feachem et al. 2010). There are more than 100 species of *Plasmodium* that infect different vertebrates. To date, five *Plasmodium* species are acknowledged to infect humans with malaria: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium knowlesi*. *P. falciparum* is predominantly found in Africa, while *P. vivax* is the most widespread globally (Price et al. 2007), in particular accounting for more than half of the disease in Latin America. *P. vivax* is almost absent in Africa due to the high frequency of the Duffy-negative protective allele (Weppelmann et al. 2013; Liu et al. 2014). There are 124-283 million Malaria infections annually (WHO 2015), which is an enormous burden with significant economic implications (Sanchs and Malaney; 2002). In 2015, the World Health Organization report (WHO 2015) showed a decrease of 48% in malaria mortality between 2000 and 2015 globally. However, 438,000 deaths are still attributed to malaria every year, mostly due to *P. falciparum* infection among children under 5 years old in Africa (WHO 2015).

Currently, malaria is endemic in 97 countries located mainly in tropical and sub-tropical regions, where there are more than 3 million of people at risk of infection (Fig. 1.1). In areas of low transmission, almost every person exposed has a high risk of infection; however, in endemic areas, infection is more restricted to tourists, children and pregnant women, though there are exceptions attributed to a high density of parasites in the blood (Desai et al. 2007; Doolan et al 2009).

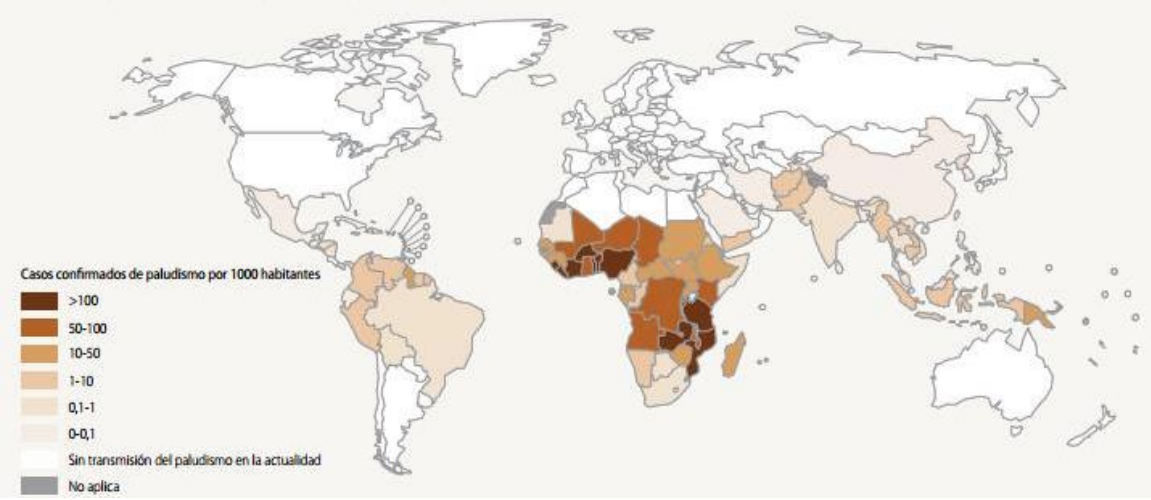

Figure 1.1 Countries where malaria transmission is ongoing, 2013 (WHO 2015)

In Colombia, the Instituto Nacional de Salud (2014) reports the presence of malaria in diverse regions including the Amazon basin, Central valleys, and Caribbean and Atlantic coasts. Malaria represents a serious public health problem due to the fact that 85% of the territory in Colombia is located in environmental, geographical and epidemiological conditions that are suitable for transmission of the disease. It is estimated that 25 million of people (60% of the population) are at risk of infection. The transmission of Malaria in the country is variable and focal. In Colombia there are three areas with active transmission: 1. Urabá, Cauca and south of Córdoba, 2. Pacific Coast (Valle, Chocó, Nariño) and 3. Orinoquia-Amazonia (Fig. 1.2). These zones contribute 45%, 30% and 25% respectively to the total number of cases in the country.

Public Health Surveillance System (SIVIGILA acronyms in Spanish) reported an accumulation of 38,120 cases of non- complicated malaria in 2014, as well as 325 complicated malaria cases and 19 deaths. The most frequent species were *P. vivax* and *P. falciparum* (Fig. 1.2), with a 37% decrease of malaria cases particularly evident for *P vivax* (SIVIGILA 2014).

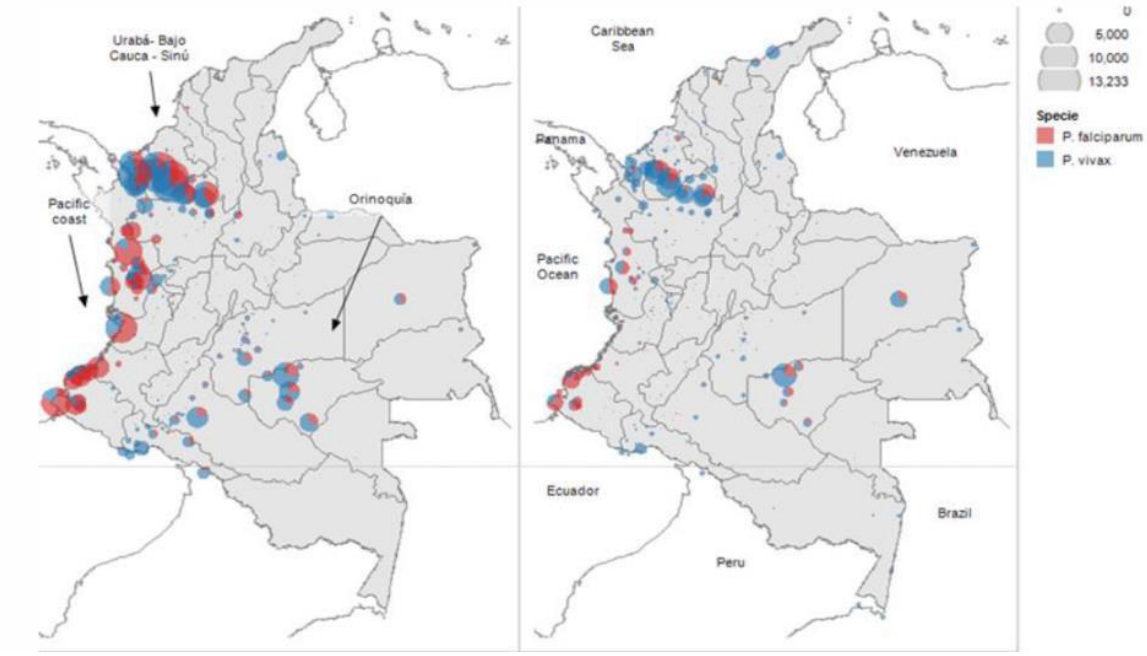

Figure 1.2. Distribution of malaria cases per city and by parasite species during 2003 to 2012. Source: SIVIGILA (2014).

1.2 Malaria Life Cycle

The *Plasmodium* malaria parasite has a complex life cycle involving multiple transformations and two well-differentiated phases. The sexual phase takes places in the midgut of the female of one of many species of Anopheles mosquito, which is the vector. Only females, which feed on blood, are responsible for transmission. Asexual haploid sporozoites are stored in the salivary glands of the vector and then they are transmitted to the human or other mammalian host during a blood meal (Frederich et al. 2002) (Fig. 1.3). Less than 100 sporozoites are deposited and then travel within minutes to the liver,

where they replicate inside the hepatocytes for approximately one week (Rosenberg et al. 1990). Some sporozoites can stay in the dermis tissue, where even a small portion of them can replicate (Gueirard et al. 2010) or enter into the blood stream and migrate to the liver, or into the lymphatic vessels and from there to the regional ganglion (Amino et al. 2006). Some reports suggest that migration of the parasites inside leucocytes and dendritic cells (Wykes et al. 2011) or macrophages (Landau et al. 1999) occurs. Once in the liver, sporozoites mature into merozoites. Following the asymptomatic liver stage, merozoites exit into the bloodstream and initiate the erythrocytic replication cycle, also known as the intra-erythrocyte development cycle (IDC). All primary clinical manifestations of malaria infection (fever, chills etc) occur as a result of blood-phase infection (Schofield and Grau 2005). In the case of *P. vivax* and *P. ovale,* some sporozoites can remain in the latent hypnozoite stage in the liver for weeks, months or even years, from where they can cause relapses (Frevert and Nardin 2005). Invasion of red blood cells (RBCs) occurs when merozoites leave the liver and interact with the RBC membrane. Once it penetrates the RBC, the merozoite abandons its protective shield and develops a specialized vacuole (called the parasitophorous vacuole). Using the erythrocytic membrane as a shelter, the parasite progresses through different stages of growth (Cowman and Crabb 2006). The first stage is called ring, which then transform into trophozoites, which are characterized by increased metabolic activity. Further maturation occurs with the ingestion of the erythrocytic's cytoplasm, giving rise to mature schizont forms (Cowman and Crabb 2006). When schizont rupture occurs, it releases daughter merozoites that restart the parasite cycle. Some of the parasites develop into sexual forms named gametocytes inside of the erythrocyte (Frederich et al. 2002), which the vector can acquire during feeding (Matuschewski 2006).

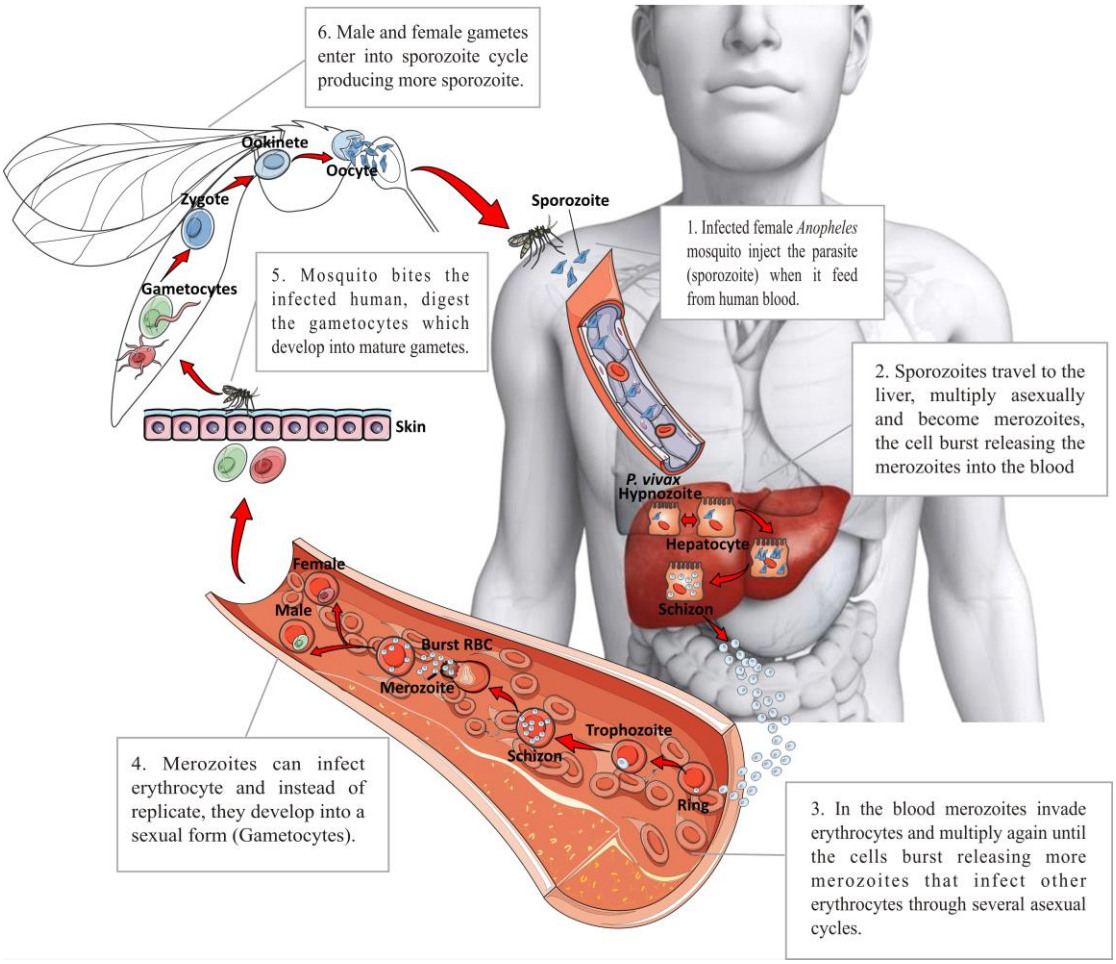

Figure 1.3 Life cycle of the malaria parasite.

1.3 Malaria pathogenesis: Host, parasite, and environmental factors

Malaria has a wide spectrum of clinical manifestations depending on the host and on the stage of the *Plasmodium* parasite. The WHO defines malaria as a febrile disease with an incubation period of less than 7 days (WHO 2015). It is well known that progression to complicated malaria can be influenced by parasite, host and socio-cultural factors. Most malaria infections are uncomplicated and if they are treated appropriately and in time, they can be cured. Approximately 1-2% of infections develop into complicated (severe) malaria cases, exhibiting a variety of clinical and laboratory defined patterns that are associated with an elevated risk of mortality (Tobón et al. 2006; 2009). The most common symptom of malaria is fever, which is classically described as periodic, even though this pattern is not always observed, and has been linked to spiking levels of pro-inflammatory cytokines such as tumor necrosis factor (TNF) (Karunawera et al. 1992).

Other symptoms include chills, cramps, and vomiting. If the disease progresses, it can cause respiratory distress, which reflects severe metabolic acidosis, and severe anemia. Cerebral malaria is an example of how the parasite can affect multiple organs, and it may finally cause coma and death (Miller et al. 2002; Schofield and Grau 2005). Moreover, in endemic areas, chronic *P. falciparum* and *P. vivax* infection are key contributors to anemia in children and pregnant women (Lamikanra et al. 2007; Anstey et al. 2009 and Rogerson et al. 2007).

The most severe (also called complicated) forms due to *P. falciparum* are cerebral malaria and severe malarial anemia, which is the most common syndrome in African children and is the cause of approximately 80% of deaths (Lou et al. 2001, Marsh et al. 1995 and Schellenberg et al. 1999). Malaria produced by other *Plasmodium* species causes high morbidity. Among the common symptoms are fever, headache, muscle pain, chills, vomit and flu-like symptoms; but death is rare. If the symptoms are not treated the infection can evolve and developed into complicated malaria. Complicated malaria is initiated in the erythrocytic stage, at which time it can manifest as a vascular obstruction due to sequestration of infected red blood cells. Subsequent inflammatory processes due to the presence of *Plasmoduim* parasites can cause dysfunction, damage and cell death in different organs (Vasquez and Tobon 2012, Bassat and Alonso 2011). The majority of malaria-related mortality is due to *P. falciparum*. It is now well known that *P. vivax*, previously considered non life-threatening, as well as *P. knowlesi*, can also cause severe and lethal infectivity (Cox-Singh et al. 2008; Tjitra et al. 2008; Genton et al. 2008; Kochar et al. 2009). Furthermore, malaria can lead to miscarriage and pre-term delivery in pregnant women, and low birth weight in infants (Rogerson et al. 2007). Case studies of complicated malaria caused by *P. falciparum* and *P. vivax* are studied in Chapter 4 of this thesis due to its immense public health importance.

1.4 Strategies and Treatments to Control Malaria

Even though in 2015 the World Health Organization report (WHO 2015) showed a decrease of 47% in malaria mortality between 2000 and 2014 globally, there are no efficient, affordable and accessible antimalarial treatments for use in all malaria endemic areas. Use of insecticide-treated bed nets, residual sprays, and reduction of standing water near settlements have had the greatest public health impact. Unfortunately, there is increasing *Plasmodium* strain resistance to the conventional drugs used as anti-malarial treatments, and mosquitoes are building widespread resistance to insecticides (Greenwood et al. 2008). Efforts are currently directed at reducing morbidity and mortality, minimizing the transmission of the disease, decreasing the parasite load in humans, and preventing the increase in resistance to antimalarial drugs. Among the new treatments being considered are: the development of new antimalarial drugs, novel vector control strategies, and introduction of what many consider to be the theoretically most efficient tool, the vaccine, designed to facilitate long-term immunization.

To reduce the adverse effects of malaria treatment, the WHO has recommended diagnosis of the presence of the parasite before administration of prophylactic treatment. The presence of the parasite can be diagnosed through a microscopic "thick smear" blood test and/or, more accurately, by polymerase change reaction (PCR), both of which distinguish malaria from other febrile diseases. This screening will help to reduce resistance to drugs, including chloroquine (first documented in the 1950s) and more recently artemisinin (discovery of which led to the award of the Nobel Prize for Physiology and Medicine in 2015 to the Chinese scientist, Tu Youyou). For this reason the WHO recommended artemisinin-based combination therapy (ACT), in which two or more drugs of different classes are given jointly, for uncomplicated malaria cases. Even though prophylactic treatment does not give complete protection, it will decrease the probability

of a complicated infection. All the drugs used for malaria treatment have secondary effects if used for more than six months, hence more effective strategies are needed.

1.5 Malaria Vaccines

The development of resistance to artemisinin has made it necessary to find alternative drug therapies and molecular targets to treat malaria and to create a vaccine. Progress in technologies toward the development of malaria vaccines has increased in recent years (MVI-PATH 2013), but the development of malaria vaccines remains a complex scientific and logistical challenge. The antigenic variability and complex life cycle of the *Plasmodium* parasite make the production of a vaccine particularly difficult (Engwerda et al. 2005), but at the same time the complexity of this cycle provides several opportunities for different approaches for vaccine development with different effects and kinds of protection (Fig. 1.4). Vaccines designed to attack exoerythrocytic stages (that is, stages outside the red blood cells) should be capable of inducing sterile immunity, namely complete absence of parasite in individuals exposed to infectious mosquito bites. Vaccines designed against the intraerythrocitic stage of the parasite should reduce the symptoms of the infection as well as mortality of affected individuals, generating what is called protective immunity. On the other hand, vaccines that block the transmission of the gametocytes, which developed in human blood, or can prevent the fertilization or development of the parasite in the mosquito, should protect an entire community by eliminating the parasite from endemic areas (Komisar 2007; Moreno and Joyner 2015).

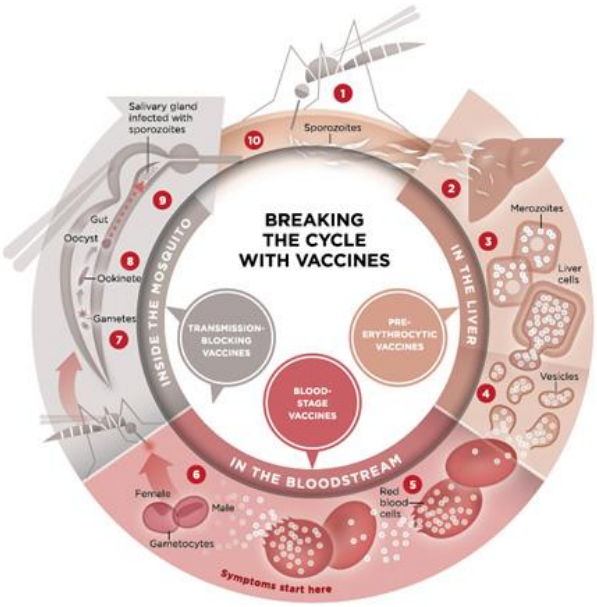

Figure 1.4. Vaccine interruption in the different stages of the *Plasmodium* cycle (MVI-PATH 2013).

Vaccination is one of the key elements of malaria control programs. Progress has been achieved with the development of *P. vivax* pre-erythrocytic subunit vaccines and preclinical trials, using both synthetic peptides and recombinant proteins representing a variety of surface antigens on non-human primates (Herrera et al. 2009; 2011). Further, Phase 1 clinical trials in humans have been conducted with different formulations of *P. vivax* CS-derived subunit vaccines, and these indicate that the vaccines are safe and immunogenic (Herrera et al. 2009; 2011). However, as with most vaccination programs, variability in the capacity to mount an effective immune response is expected.

The *Malaria Vaccine Technology Roadmap* is a global strategy initiated in 2006 with two purposes: (i) develop a vaccine that would provide 50% protection against complicated malaria and death by 2015, (ii) and another that would provide 80% protection in clinical cases by 2025 (MVI-PATH 2013). Currently, four malaria vaccines are in field trials (WHO 2015). Three vaccine candidates are in phase 2B of clinical trials, while RTS,S/ASO1 has completed phase 3 (WHO 2015). RTS,S/ASO1, which is commercially known as “Mosquirix”, has been developed with the support of the Bill and Melinda

Gates Foundation, and through a partnership between GlaxoSmithKline Biologicals (GSK) and the PATH Malaria Vaccine Initiative (MVI-PATH). RTS,S/AS01 is the only vaccine to have entered phase 3 trials. It is an exoerythrocytic vaccine targeting sporozoites, specifically being formulated as a recombinant protein RTS,S between the C-terminal flanking region of the circumsporozoite (CS) protein and the hepatitis B surface antigen (HBsAg). It also contains the adjuvant AS01 which boosts the cellular and humoral immune response (Regules et al. 2011; Garcon et al. 2003; Bojang et al. 2005; Ansong et al. 2011). Final results of phase 3 have recently been released by the RT,S Clinical Trials Partnership (Lancet 2015), for which 15,460 children and young infants from seven sub-Saharan African countries (Burkina Faso, Gabon, Ghana, Kenya, Malawi, Mozambique, and the United Republic of Tanzania) were enrolled. The results showed a moderate vaccine efficacy (VE) of 26-36%, with infants less protected than children between 5 and 15 years of age. One criticism of this extraerythrocytic vaccine approach is that it would be necessary to induce large quantities of antibody with high affinity to effectively trap the sporozoite in the skin (Riley and Stewart 2013), and this is never likely to be completely effective. Furthermore, the RTS,S vaccine offers no protection against *P. vivax* malaria.

Numerous researchers are now evaluating the efficacy of vaccination with whole attenuated live parasites, as opposed to just a handful of surface proteins (Hoffman et al. 2010; Matuschewski et al. 2011; Lindner et al. 2012). Among the various stages of malarial infection, sporozoite (spz) invasion and schizogonic development in the liver cell appear to be ideal therapeutic targets. Killing these parasite stages would prevent parasite development prior to infection of the blood, and therefore prevent febrile disease (Komisar 2007). Early studies showed that immunization with *P. yoelii* and *P. berghei* irradiated sporozoites (irr-spz) can protect mice and rats against sporozoite challenge (Mellousk et al. 1990; Weiss 1990). Similarly, in monkeys, immunization with *P.*

cynomolgi and *P. knowlesi* irr-spz gives partial or total protection (Collins et al. 1972, Gwadz et al. 1979) and full protection in humans with *P. falciparum* or *P. vivax* irr-spz after multiple immunizations over the course of several months (Egan et al. 1993; Clyde 1973; Epstein et al 2011). Therefore, irr-spz is an important model that can induce sterile immunity by killing the pre-erythrocytic stages of the malaria parasite, providing proof-of-concept for this approach and can be used in different species of the parasite. Chapter 3 of this thesis examines the transcriptional response to irradiated *P. vivax* sporozoite vaccination in a small number of Colombian volunteers.

1.6 Innate immune response of the host to malaria

An important factor influencing the clinical course of the disease is prior exposure to malaria. Since there are no completely effective antimalarial treatments that are appropriate for use in malaria endemic areas, the human immune response remains a valuable line of defense against malaria. Adults and older children tend to experience reduced prevalence of malaria infection and have less severe symptoms in endemic areas (Bunn et al 2004; Doolan et al. 2009). Immunity to malaria in endemic areas is hoever generally regarded as short-term and never sterilizing (Doolan et al. 2009; Okell et al. 2009). Recent results using PCR to detect parasitemia have indicated that there are much higher rates of asymptomatic malaria than hitherto appreciated, and antibody arrays have also demonstrated that memory B cells persist in the blood for decades. Consequently, it may be more appropriate to consider the relationship between host and parasite as involving the emergence of tolerance (Ayers et al, 2012) rather than resistance.

The risk of contracting malaria and its clinical manifestations is highly correlated with factors related to the host. Age and genetic variation play an important role in modulating immunity in malaria, as does the microbiome (Idaghdour et al. 2012; Tsang et al. 2014; Yilmaz et al. 2014; Pulendran 2014). Also, the prevalence to erythrocytic polymorphism

and co-infections can influence the immune response (Schofield and Grau 2005; Akpogheneta et al. 2008; Pacheco et al, 2016). Just as the complex life cycle of the parasite can help it to avoid the immune response of the host, the human immune system has exerted strong positive selection on polymorphic immune-dominant antigens (Mackinnon and Marsh 2010). Each stage of the intraerythrocytic development is characterized by the expression of specific proteins, which require immunological mechanisms that have high specificity to eliminate the different parasitic forms (Li et al. 2001). Variation in the Human Leukocyte Antigen complex is an important genetic mediator of natural protective immunity (Lyke et al. 2011).

Protection against malaria is also been associated with selection on several loci in thehuman genome at which the frequency of genetic polymorphisms are correlated with infection. Notable examples are the Hbs allele of the β-blobin gene (the causative agent of sickle cell anemia); regulatorymutations in the alpha and beta globin genes (which cause the thalassemias); alleles leading to deficiency of glucose-6-phosphate dehydrogenase enzyme (G6PD) activity, or the FY*BES allele in the Duffy locus (abbreviated *Fy*-), which in find at high frequency in malaria endemic regions and protects against *P. vivax* infection (Tishkoff and Verrelli, 2003; de Mendonça et al.2012).

It is likely that variation in components of lymphocyte signaling are also important effectors if not of resistance to malaria, then of the course of infection. The first phase of the infection in humans is the liver; this stage is asymptomatic and it lasts for several days (Langhorne et al. 2008). The liver stage barely induces an innate immune response, perhaps due to the lack of antigenic stimuli capable of activating neutrophils to initiate the innate immune response, or simply due to the very low parasite burden (Riley and Stewart 2013). The first barrier to blood-stage parasitemia is nevertheless the innate immune response, which causes many of the clinical symptoms associated with malaria.

This innate mechanism leads to activation of the complement system, monocytes, macrophages, dendritic cells (DCs), natural killer cells (NKs), and natural killer T cells (NKT) (Stevenson and Riley 2004). Also, T cells have an important role as the bridge between the innate immune response and adaptive response, thereby controlling parasite growth (Stevenson and Riley 2004). However, the final stage of control and elimination also depends on the production of antibodies which have increased efficacy as the infection proceeds (Riley and Stewart 2013). The production of pro-inflammatory cytokines such as tumor necrosis factor (TNF-α) and interferon gamma (IFN-γ) are considered key factors in the initial control of the parasite burden and pathophysiology in humans (Clark et al. 2006; Perkins et al. 2011). This can limit the initial replication of the parasite in the erythrocytes, but it has to be controlled by an anti-inflammatory mechanism to avoid tissue damage. The anti-inflammatory molecules include interleukin 10 (IL-10) and transforming growth factor β (TGF-β) expressed by CD4+ T-helper cells (Finney et al. 2010).

The complexity of the mature immune response is emphasized by the engagement of dozens of different immune cell types. Macrophages and monocytes also have a central role against malaria parasites through mechanisms such as phagocytosis of infected erythrocytes (iRBC), and production of nitric oxide, which reduces parasite growth (Stevenson and Riley 2004). Macrophages and monocytes can also mediate cellular cytotoxic responses dependent on antibody-based recognition of merozoites or infected erythrocyte antigens (Chimma et al. 2009), and they can modulate antigen presentation to T-helper cells (Serghides et al, 2003). TNF-α production is also known to block the development of trophozoites (Bouharoun-Tayoun et al. 1995).

One major factor affecting the disease progression is prior exposure to malaria and acquisition of a semi-immune state. More about this topic will be discussed in chapter 2.

The clinical presentation of clinical malaria varies according to the immune status of the person, as high levels of immunoglobulin are positively correlated with survival rates. It is thought that continuous exposure to malaria limits the likelihood of progression of disease severity, also due to development of immunity. However, complicated malaria is still observed in populations highly exposed to malaria (Doolan et al. 2009), so a pilot study was developed to address this process, and is discussed in chapter 4 of this dissertation.

CHAPTER 2

2. TRANSCRIPTION PROFILING OF MALARIA-NAÏVE AND SEMI-IMMUNE COLOMBIAN VOLUNTEERS IN A *PLASMODIUM VIVAX* SPOROZOITE CHALLENGE

2.1 Abstract

Continued exposure to malaria-causing parasites in endemic regions of malaria induces significant levels of acquired immunity in adult individuals. A better understanding of the transcriptional basis for this acquired immunological response may provide insight into how the immune system can be boosted during vaccination, and into why infected individuals differ in symptomology. Peripheral blood gene expression profiles of 9 semi-immune volunteers from a *Plasmodium vivax* malaria prevalent region (Buenaventura, Colombia) were compared to those of 7 naïve individuals from a region with no reported transmission of malaria (Cali, Colombia) after a controlled infection mosquito bite challenge with *P. vivax*. A Fluidigm nanoscale quantitative RT-PCR array was used to survey altered expression of 96 blood informative transcripts at 7 timepoints after controlled infection, and RNASeq was used to contrast pre-infection and early parasitemia timepoints. There was no evidence for transcriptional changes prior to the appearance of blood stage parasites at day 12 or 13, at which time there was a strong interferon response and, unexpectedly, down-regulation of transcripts related to inflammation and innate immunity. This differential expression was confirmed with RNASeq, which also suggested perturbations of aspects of T cell function and erythropoiesis. Despite differences in clinical symptoms between the semi-immune and malaria naïve individuals, only subtle differences in their transcriptomes were observed,

although 175 genes showed significantly greater induction or repression in the naïve volunteers from Cali. Gene expression profiling of whole blood reveals the type and duration of the immune response to *P. vivax* infection, and highlights a subset of genes that may mediate adaptive immunity.

This study was published as Rojas-Peña et al, 2015 (PLoS Negl. Trop. Dis. 9: e0003978).

2.2 Introduction

One of the features of *Plasmodium* species that make them such pernicious parasites is their ability to avoid the host immune system (Wright and Rayner 2014; Zheng et al. 2014). While this is achieved in part by virtue of their complex life cycle that includes intra-erythrocyte cycling and periodic sequestration in various tissue compartments (Zheng et al. 2014), it is also clear that *Plasmodium* infection causes short- and probably long-term modification of host immune function. Molecular methods are shedding some light on the mechanisms behind these modifications. For example, it is now clear that exposed individuals generally do mount an antigen response to *Plasmodium* antigens that persists (Krzych et al. 2014, Stanisic et a. 2013), and that several biochemical pathways are engaged, including interferon and cytokine signaling, membrane lipid modification, and reactive oxygen species metabolism (Gazzinelli et al. 2014). Host factors including genetic variation, both within and between populations, play a role in modulating immunity in malaria, as does the microbiome (Tsang et al. 2014; Yilmaz et al. 2014; Pulendran 2014; Idaghdour et al. 2012).

An important factor influencing the clinical course of disease is prior exposure to malaria. Adults and older children tend to experience reduced prevalence of malaria infection and have less severe symptoms (Bunn et al. 2004; Doolan et al. 2009). Nevertheless the mechanisms responsible for host resistance to malaria are still poorly understood. As a prelude to evaluation of vaccine efficacy in a Colombian population, we recently carried out a challenge experiment in which we evaluated the responses of immunologically naïve and semi-immune individuals to deliberate infection with *Plasmodium vivax* through mosquito bites (Arévalo-Herrera et al. 2014). All nine volunteers from a malaria endemic region near the town of Buenaventura were weakly positive for IgG antibodies to sporozoites or blood stage proteins prior to the experiment, and after challenge eight of them showed increased antibody titers against blood stages. Similarly, five of seven naïve volunteers from the city of Cali converted to sero-positivity that was generally maintained for at least four months. While there was no significant difference in the time to first appearance of blood stage parasite assessed by thick blood smears (12 to 13 days in both groups) or by polymerase chain reaction (PCR) (around 9 days), the naïve volunteers experienced classical early malaria symptoms, whereas the semi-immune volunteers were for the most part nearly asymptomatic, at least at the day of diagnosis when curative prophylaxis was administered (Arévalo-Herrera et al. 2014).

In order to begin to characterize the molecular basis for this difference in clinical course of disease as a function of prior exposure to malaria, we report here two types of transcriptome profiling of peripheral blood samples from the Colombian challenge experiment volunteers. First we used targeted measurement of a set of 96 highly

informative transcripts by nanoscale Real Time PCR (RT-qPCR) (Spurgeon et al. 2008) in order to generate a time course of the infection transcriptional response. Second, we used RNASeq (Cloonan et al. 2008) on a subset of six volunteers contrasting baseline and incident malaria, to ask whether (i) there is a difference in immune profiles between naïve and semi-immune individuals in the absence of infection, and (ii) patent infection results in a differential transcriptional response that may hint at the molecular basis of long-term immunity. We also contrasted our findings with those of cross-sectional studies concluding that history of exposure is just one of many factors mediating host–parasite interactions in malaria.

2.3 Methods

2.3.1 Experimental design and ethics statement

The experimental design protocol of this research was approved by the Institutional Review Boards (IRB) at the Malaria Vaccine and Drug Development Center (CECIV, Cali) and Centro Medico Imbanaco (Cali). It is described in more detail in Arevalo-Herrera et al. (2014), which reports the clinical responses to malaria challenge. Sixteen Duffy-positive (Fy+) male and female volunteers (9 semi-immune, previously exposed to malaria, from Buenaventura and 7 immunologically naïve with respect to malaria, from Cali) were enrolled.

Volunteers where invited to the vaccine center two days (day -2) prior the challenge day (day 0) for physical examination and blood sample collection. Fig 2.1 summarizes the blood sampling strategy. Blood samples used for the RT-qPCR experiment were collected on day -2 (pre-challenge), day 5, day 7, day 9, on the day of first detection of

Plasmodium by thick smear test (day 12-13, Dx), and 4 months later (month 4). RNASeq analysis, also approved by the Georgia Tech IRB, was performed for 12 individuals (six each from Buenaventura and Cali) for two of the timepoints, namely the diagnosis day and baseline (pre-challenge day).

For each sample, approximately 1 mL of blood in 2 mL of buffer was collected into a Tempus tube, which preserves whole blood RNA at 4ºC indefinitely. Whole blood mRNA was extracted using Tempus™ Blood RNA Tube isolation kits provided by the manufacturer Applied Biosystems, and the sample quality was determined based on the Agilent Bioanalyzer 2100 RNA Integrity score (RIN). All samples had RIN greater than 4.0 without meaningful degradation.

2.3.2 RT-qPCR

Reverse Transcription followed by quantitative PCR (RT-qPCR) was performed using Fluidigm 96×96 nanofluidic arrays targeting a set of 96 transcripts that are broadly informative of the major axes of variation for peripheral blood gene expression from Preininger et al. (2013) at six timepoints (Pre-challenge, day 5, day 7, day 9, Diagnosis (Dx) and month 4; Table S3).

The RT-qPCR was completed in three steps: (1) Total whole blood RNA was converted to single stranded cDNA using polyT priming of reverse transcription, (2) the 96 targeted genes were pre-amplified in a single 13-cycle PCR reaction for each sample following conditions outlined in the manufacturer's protocol by combining cDNA with the pooled primers and EvaGreen® Mastermix (Fluidigm BioMark™), and (3) qPCR reactions were performed for each sample and individual gene on each sample on a 96×96 array with 30

amplification cycles. Average Ct value was calculated at a point in which every reaction is in the exponential phase to ensure accuracy and precision of amplification. In order to make the analysis more easily comparable with traditional transcript abundance measures such as those obtained with microarrays or RNASeq, each Ct value was subtracted from 30, setting missing values to 0. Since small Ct values correspond to high transcript abundance, this subtraction yields values ranging from 0 (no expression) to 30 (very high abundance).

2.3.3 RNASeq

Library preparation for RNASeq was performed using the Illumina TruSeq Low Throughput (LT) RNA Sample Preparation Protocol. Short read sequencing was performed in rapid run mode with eight samples per lane on an Illumina HiSeq 2500, generating 100 bp paired-end libraries with an average of 15 million reads per sample,.

The raw RNASeq reads (Fastq files) for each sample were aligned to the reference human genome (hg19) using Bowtie as the short read aligner, and splice junctions were identified using TopHat2 in the Tuxedo protocol (Trapnell et al. 2012). After alignment, estimation of transcript abundance measures as fragments per kilobase of exon per million aligned fragments (FPKM) values was performed using Cufflinks (Trapnell et al. 2012). Genes with an FPKM greater than 2.5 averaged across the 24 samples were retained for downstream analyses, representing 6,154 genes.

FPKM values were then transformed to logarithm base 2 to guarantee that the data were more normally distributed and to simplify the interpretation of the scale of differential

expression (each unit difference corresponds to a two-fold difference in abundance). The supervised normalization of microarray (SNM) procedure was then used to normalize the data with the R package SNM from Bioconductor (Mecham et al. 2010), fitting location and timepoint as the biological variables, and Individual as the adjustment variable (fit but not removed). All downstream analyses were performed on this normalized data set.

The dataset has been deposited into the Gene Expression Omnibus archive (GEO) under accession number GSE67184 and RT-qPCR data accession number GSE67470.

2.3.4 Statistical analyses

Most statistical analyses of both the Fluidigm and RNASeq datasets were performed in JMP Genomics version 5 (SAS Institute, NC), starting with the Basic Expression Workflow, which performs principal components analysis (PCA), and computes a weighted total contribution of the covariates of interest to the axes (principal variance components analysis, PVCA). Linear regression was then used to assess the relationship between the individual covariates and PC, and/or analysis of variance was used to detect differential expression between locations or timepoints. A Benjamini-Hochberg 5% false discovery rate was used to select differentially expressed genes. Volcano plots contrast the significance (negative log10 of the p-value, NLP) against the fold difference (normalized log2 Ct or FPKM units) between specific conditions. Hierarchical clustering was performed using Ward's method.

Blood informative transcript (BIT) axes analysis was performed by generating the first PC for the 10 genes that are most strongly correlated with each of the 9 Axes reported in Preininger et al. (2013). These 9 Axes are consistently conserved in all human peripheral

blood gene expression datasets, and represent coordinated expression of hundreds to in some cases thousands of transcripts. They collectively capture over half of the total transcript abundance, and are thought to reflect gene activity within major cell types (broadly speaking, T cells (Axis 1), reticulocytes (Axis 2), B cells (Axis 3), and neutrophils (Axis 5) or specific immune or physiological responses (Interferon signaling, Axis 7). Principal component one (PC1) for each of these 10 sets of BIT provide a summary axis score, which is then contrasted with respect to the covariates of interest using standard statistical tests.

2.3.5 Comparison with Malarial Gene Expression in a study from Benin, West Africa

In order to further infer whether location influences the axes of variation, we reanalyze data from Idaghdour et al. (2012) who characterized whole blood transcriptomes of infants from the West African Republic of Benin, infected with *Plasmodium falciparum*. They reported on 61 healthy controls from a hospital in the city of Cotonou, and 92 cases drawn approximately equally and without bias with respect to parasitemia levels from Cotonou and the village of Zinvié, located 36 km from Cotonou (GEO accession number GSE34404). They identified parasitemia as the major factor influencing transcript abundance overall, but also described a location effect that is considered with respect to the BIT axes here.

2.4 Results

2.4.1 RT-qPCR Comparison of Naïve and Semi-Immune Responses to Infection

The first objective of this study was to compare the time course of transcriptional changes during response to infection, between naïve and semi-immune volunteers. There were 16 volunteers in all, 7 from Cali who had not previously been exposed to malaria, and 9 from Buenaventura, a village in an endemic region for the disease, all of whom had experienced between 2 and 5 mild bouts of malaria. Fig 2.1 shows results of peripheral blood samples from 14 volunteers at Day 5 following exposure and again at Day 7, from 16 volunteers at Day 9 when PCR later confirmed initial appearance of blood-stage parasites, from 11 volunteers on Days 12 or 13 when parasitemia was diagnosed in thick blood smears, and from 14 volunteers four months after the initiation of the experiment. There were no significant differences between the two groups either in the length of the pre-patent period or the level of parasitemia attained before administration of a curative cocktail of anti-malarial drugs.

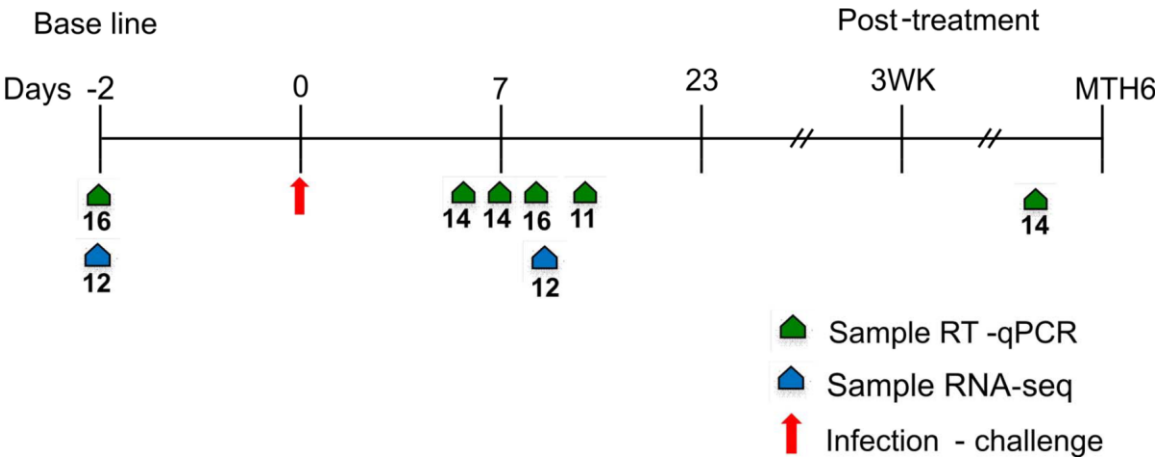

Figure 2.1. Experimental design. The timeline for sample collection. 86 total samples were collected for RT-qPCR analysis and 24 total samples for RNASeq. Green arrows represent timepoints where RT-qPCR was performed: Pre-challenge, 16 samples (7 Cali, 9 Buenaventura); Day 5, 14 samples (6 Cali, 8 Buenaventura); Day 7, 14 samples (5 Cali, 9 Buenaventura); Day 9, 16 (7 Cali, 9 Buenaventura); Diagnosis by thick blood smear day (Day 12-13), 11 samples (5 Cali, 6 Buenaventura) and Month 4, 15 samples (6 Cali, 9 Buenaventura). Blue arrows shows samples used for the RNASeq analysis, 12 per each timepoint Diagnosis day and Pre-challenge (6 Cali, 6 Buenaventura), 24 total.

Whole blood gene expression was monitored in each of the 85 samples using a Fluidigm nanoscale RT-qPCR array targeting 96 genes referred as "blood informative transcripts" (BIT). These BIT consistently capture the covariance of over half of the genes expressed in blood, specifically serving as biomarkers for 10 conserved axes of variation. We confirmed that the genes were also co-regulated in this dataset by observing a strong correlation of expression for each of the 10 BIT for each Axis, and then generated Axis scores as the first principal component of the variance of those 10

BIT. Across all of the gene expression measurements, 30% of the variance was among individuals, and just 6.5% between the timepoints, with very little differentiation between the naïve and pre-immune volunteers (Fig 2.2). The remainder of the variance was due to random biological or technical noise, or to the covariance of gene expression along the Axes.

Only two of the Axes were differentially expressed among timepoints, with the time at Diagnosis most divergent in both naïve and semi-immune individuals (Fig 2.3A, 2.3B). Axis 5 is related to innate immune signaling and neutrophil number, and seems to decline

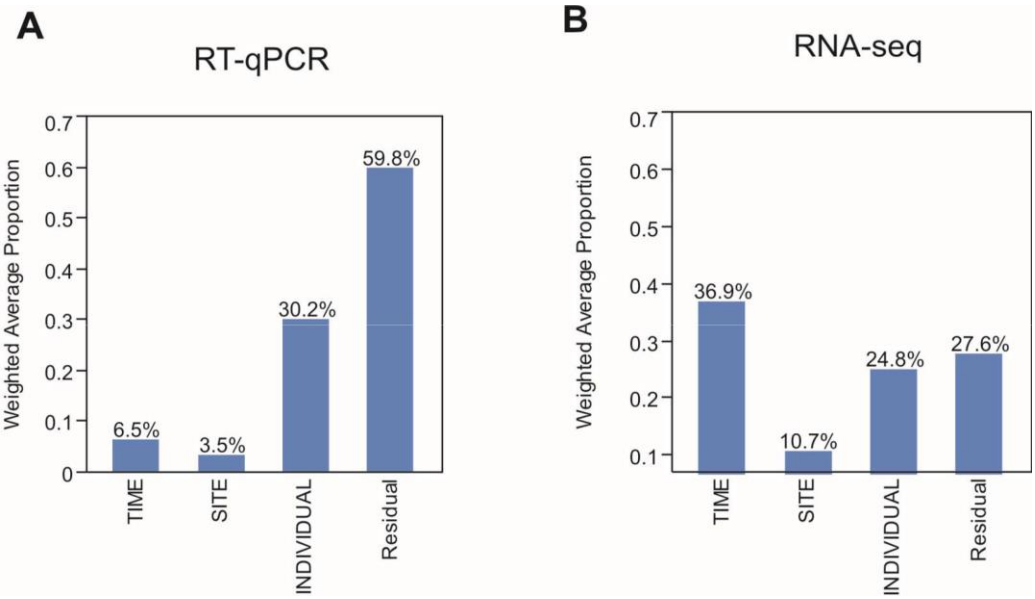

Figure 2.2. Principal component variance component analyses. Bar graphs shows the weighted average weighted average of the variance captured by the first five principal components among samples that is explained by Time (PRE, DAY5, DAY7, DAY9, Diagnosis, MTH4), Site (Cali, Buenaventura) and Individual, Indicating that most the

variability are among Individual for RT-qPCR (A) and Time (PRE and Diagnosis day) for RNASeq data set (B).

at Diagnosis, surprisingly, implying a mild reduction in inflammatory gene activity. Axis 7 represents Type 1 interferon induction and is, as expected, elevated at diagnosis, reflecting a transient specific immune response. Both axes had returned to close to baseline levels three months after recovery. No other gene expression differences detected by this targeted RT-qPCR analysis were associated with time or population. These results are consistent with previously observed stable maintenance of peripheral blood gene expression profiles in healthy adults (Tabassum et al. 2015).

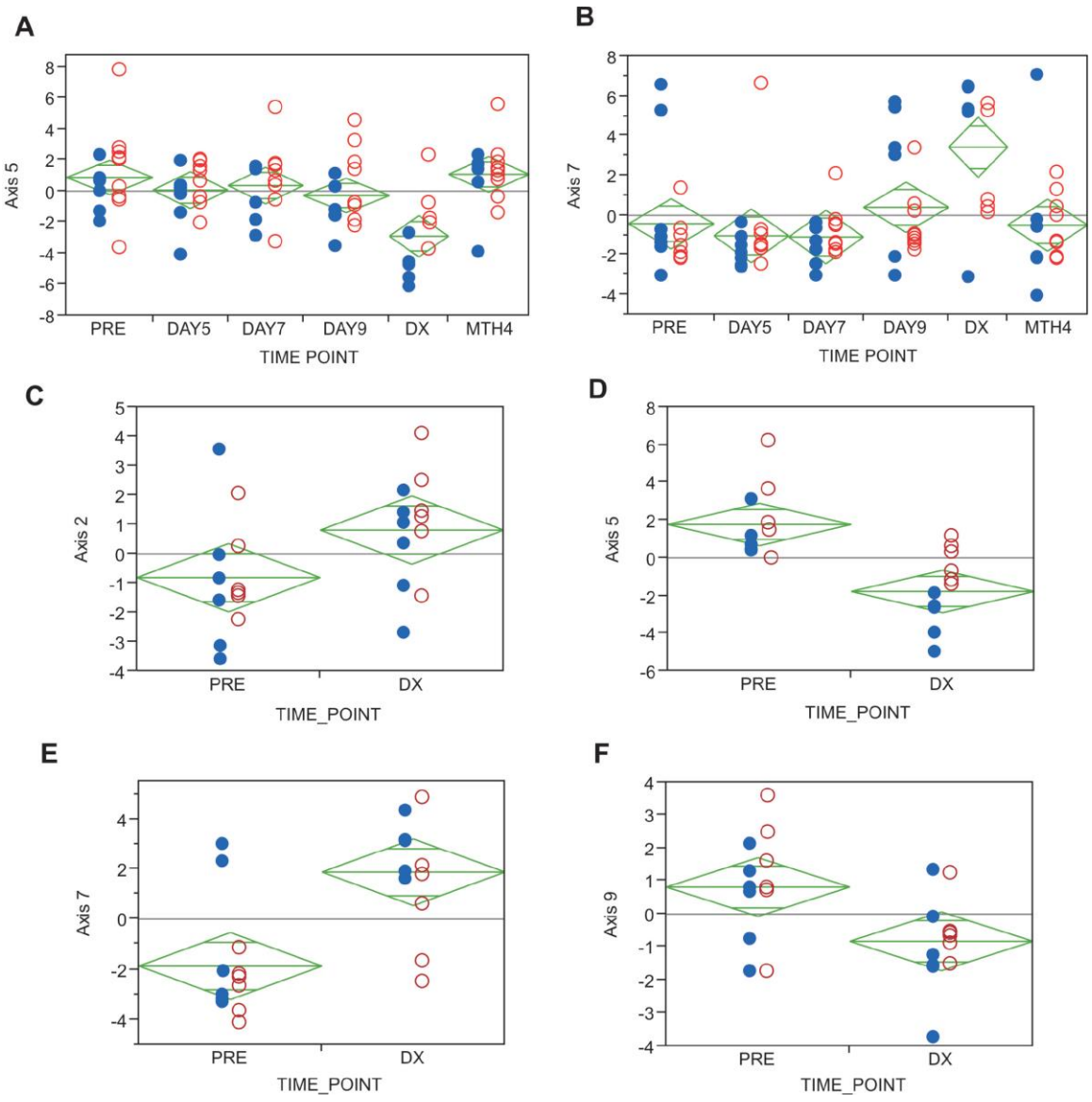

Figure 2.3. Axis of variance analysis. Each plot shows the differences in Axis scores at 6 different timepoints for RT-qPCR (**A and B**) and two for RNASeq **(C-F**); Blue solid point represents Cali, and red open circles represent Buenaventura.

2.4.2 RNASeq Comparison of Naïve and Semi-Immune Responses at time of Parasitemia

In order to obtain a more comprehensive picture of the transcriptome-wide changes in gene expression as parasites first appear in the blood, we performed RNASeq on 6 volunteers each from Cali and Buenaventura, both at Baseline and Diagnosis. An average of 15 million paired-end 100bp short read alignments to the human reference genome were obtained for each sample, allowing us to estimate transcript abundance for each of 6,154 genes. Analysis of variance was used to contrast gene expression relative to population and timepoint, and to assess the interaction between these two factors. Fig 2.2B shows that 25% of the total variance was among individuals, similar to the Fluidigm observation, and that very little differentiation was seen between populations. However, just over one third of the variance was between Baseline and Diagnosis samples, implying a much greater response to infection than suggested by the RT-qPCR data. though it should be noted that only contrasting the two most different timepoints was expected to account for more of the variance.

The differential expression of Axes 5 and 7 was confirmed by the RNASeq data (Fig. 2.3D and 2.3E), which suggested divergence of Axes 2 and 9 (Fig. 2.3C and 2.3F). Up-regulation of Axis 2 is likely to be a sign of elevated erythropoiesis since it is enriched for genes expressed in reticulocytes (Whitney et al. 2003), suggesting a mild physiological response to loss of red blood cell function even in the early stages of malaria. Axis 9 may be reflective of decreased killer T cell function since CD8 expression is correlated with it, but this needs to be independently verified. Interestingly, the increased resolution of RNA-Seq suggests differential responses of Axes 5 and 7

between the naïve and semi-immune populations. Specifically, the reduction of neutrophil and TLR-signaling associated with Axis 5 appears to be much stronger in the naïve individuals (Fig. 2.3D, solid blue points), whereas the induction of interferon signaling is variable in semi-immune volunteers (Fig. 2.3D, open red circles), two of whom showed no response. The directional trends were the same in the Fluidigm data, but less apparent.

Consistent with timepoint rather than Population explaining a large proportion of the variance, gene-specific differential expression analysis revealed more than 250 transcripts up- or down-regulated at the experiment-wide threshold of $p<10^{-5}$ (Fig. 2.4A), but only two transcripts more highly expressed in Buenaventura and none in Cali (Fig. 2.4B). Approximately 50 genes show more than 2-fold up-regulation at Diagnosis relative to Baseline yet are less significant than many of the orange-colored genes (Fig. 2.4A, green-colored genes). The reason is that these genes are even more highly upregulated in a subset of individuals, namely the naïve (Cali) volunteers. In fact, 175 genes show a significant timepoint-by-Population interaction effect at $p<0.05$ (Fig. 2.4C). These are represented in the heat-map in Fig. 2.4C, showing two-way hierarchical clustering of transcripts in samples, two-thirds of the genes are actually down regulated at Diagnosis. Interestingly, there was a marked distinction between the two timepoints (Fig. 2.4C). The Baseline samples were intermingled with respect to whether they were from the naïve or semi-immune populations, whereas the Diagnosis ones showed a near-perfect separation with respect to pre-immune exposure. In other words, most of the genes showing an interaction effect were more strongly up- or down regulated in the naïve than semi-immune individuals. An exception was a Baseline sample from a Cali

volunteer (number 306), which clustered with the Diagnosis set but still showed a robust response to malaria infection along with moderate thrombocytopenia and leukopenia, but so did Cali 310 who was not an outlier.)

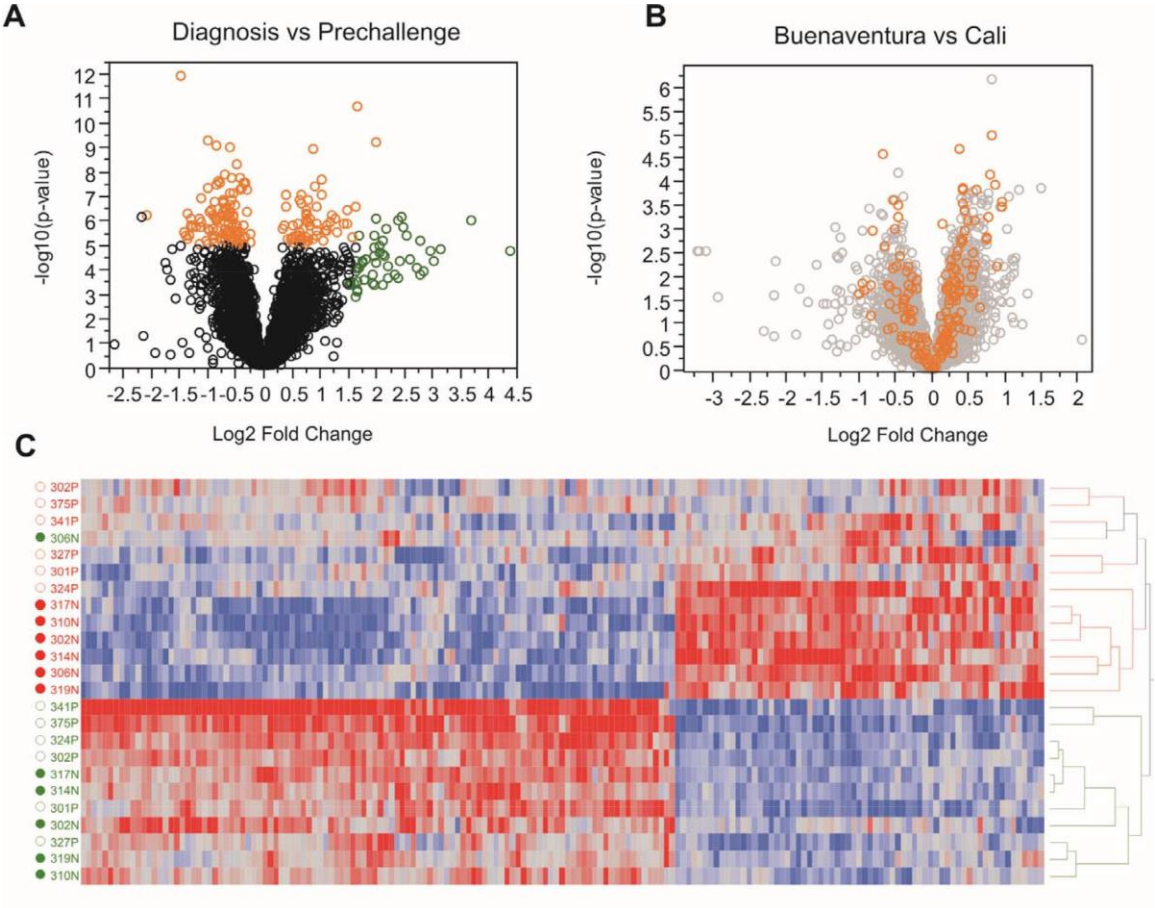

Figure 2.4. Differential expression in whole-blood RNASeq data set. Volcano plots of statistical significance vs. magnitude of differential expression for the contrasts between timepoint and highlights 175 interacting genes by timepoint in orange (A) and in red (B) by location. Y axis shows the significance as –log10 P value, and x-axis shows the magnitude log2. (C) Heat map showing two-way hierarchical clustering of transcripts (columns) in each sample (rows) of 175 genes that show a significant timepoint-by-location interaction effect at $p<0.05$; red represents high expression, blue low, gray

intermediate. Green dots represent pre-challenge and red dots represent Diagnosis day, solid points represent Cali, and open circles represent Buenaventura.

2.4.3 Nature of the Differential Response to Malaria

Given the importance of cytokines to regulation of the immune response, we specifically analyzed the expression of all genes in the RNASeq dataset that are related to Interleukin (IL), interferon (IFN), tumor necrosis factor (TNF), and transforming growth factor (TGF) signaling. This analysis revealed three groups of samples, and three clusters of genes (Fig. 2.5). Once again, the Baseline and Diagnosis samples were separated, excluding the outlier Cali 306 Baseline sample and two others, but in this case there was no clear separation relative to pre-infection malaria status. One cluster of 14 genes, including *IL32* and *IL8*, was not differentially expressed. Another cluster of 23 genes, including the *IL4R*, *IL6R*, and *IL7R* and *IL17R* receptors, was upregulated at Baseline, particularly strongly in three volunteers (Cali 302 and Buenaventura 341 and 375). The third cluster of 19 genes, including *TNF*, *IL1B* and *IL15*, showed the opposite tendency, namely up-regulation at Diagnosis, particularly strongly in two samples (314 from Cali and 324 from Buenaventura). These results imply that there is strong co-regulation of the cytokine response and infection, but that this is not mediating the differential response between naïve and semi-immune individuals. This is somewhat surprising, especially given that the experience of fever was significantly different between the two populations, who might have been predicted to differ with respect to the pyrogenic cytokines IL1, IL6, IL8 and TNF.

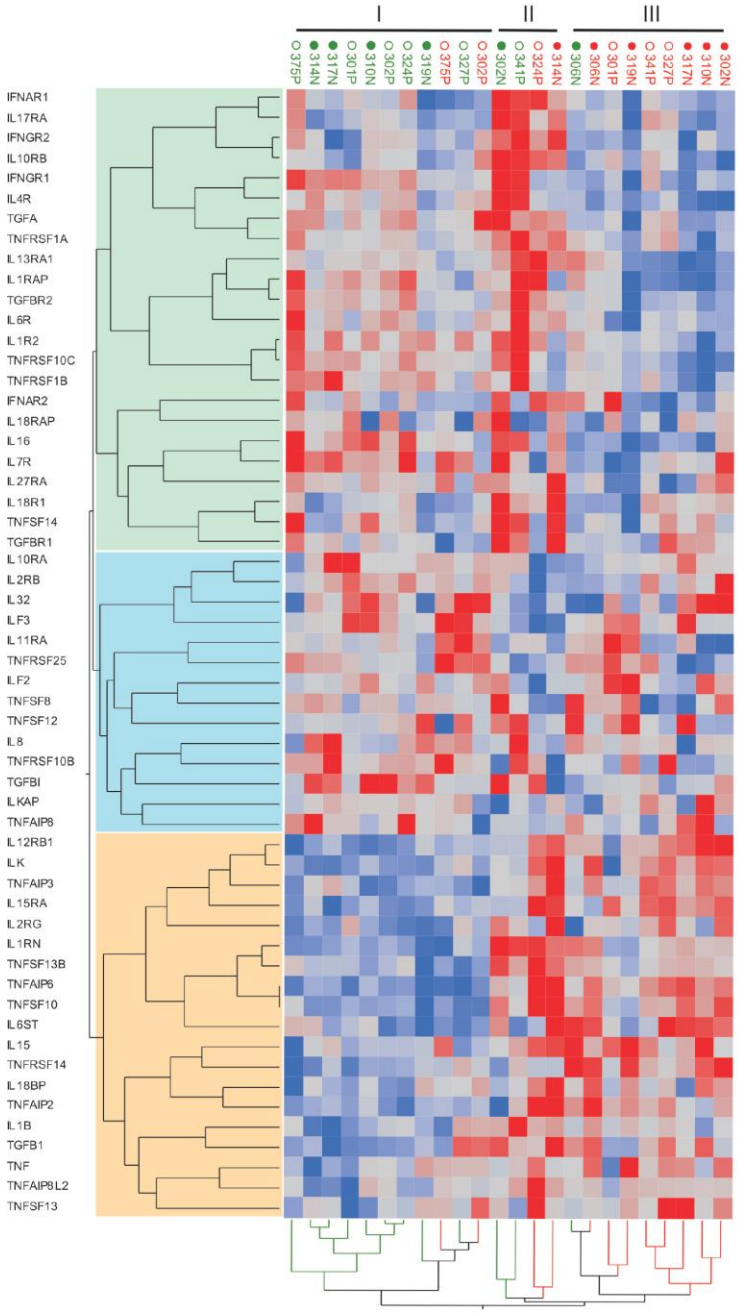

Figure 2.5. Heat map of Interleukins, Interferon (IFN), Transforming Growth Factor (TGF) and tumor necrosis factor (TNF) hierarchical cluster gene expression. Red indicates high expression, blue low expression, and gray intermediate. Three groups of samples (I, II and III)), and three clusters of genes (green, blue and yellow) are

evident. Baseline and Diagnosis samples are separated, excluding the outlier Cali 306N Baseline sample and 327P and 341P. Green points represent pre-challenge and red points represent Diagnosis day; solid points represent Cali, and open circles represent Buenaventura.

Closer examination of the differentially expressed genes between Baseline and Diagnosis suggested a complex network of cross-regulatory interactions. The up- and down-regulated cytokines for example both include pro- and anti-inflammatory peptides and their receptors. Similarly, there appear to be counter-balancing signal transduction profiles: *JAK1* and *RAF1* are both strongly down-regulated in all volunteers at Diagnosis, whereas *IL6ST* and *SOS1* are up regulated.

Among the genes showing a significant interaction effect, namely a stronger response at diagnosis in the immunologically naïve individuals, there are several types of gene functions of interest (Table 2.1). These include lysosomal components (*CTSH*, *RILP*), regulators of macrophage activity (*CD163*, *MMP25*, *SIRPA*, *TBC1D14*, *TNFSF13*), splicing factors (*EIF2C4*, *SNRPB2*, *SNRPG*), lipid biosynthesis (*DGAT2*, *LPPR2*), solute carriers (*S100P*, *SLC6A6*, *SLC11A1*, *SLC7A7*), signal transduction (*G3BP1*, *GAB3*, *MAPK13*, *TLE3*) and Cell Cycle and DNA damage response (*ATM*, *PRKDC*, *ARID4A*). Some genes with an interaction effect showed stronger down-regulation in Cali (Fig. 2.6A, *ATM*), or stronger down-regulation in Buenaventura (Fig. 2.6B, *EIF2C4*), compared with one that showed a similar up-regulation at both locations (Fig. 2.6C, *ATP1B3*).

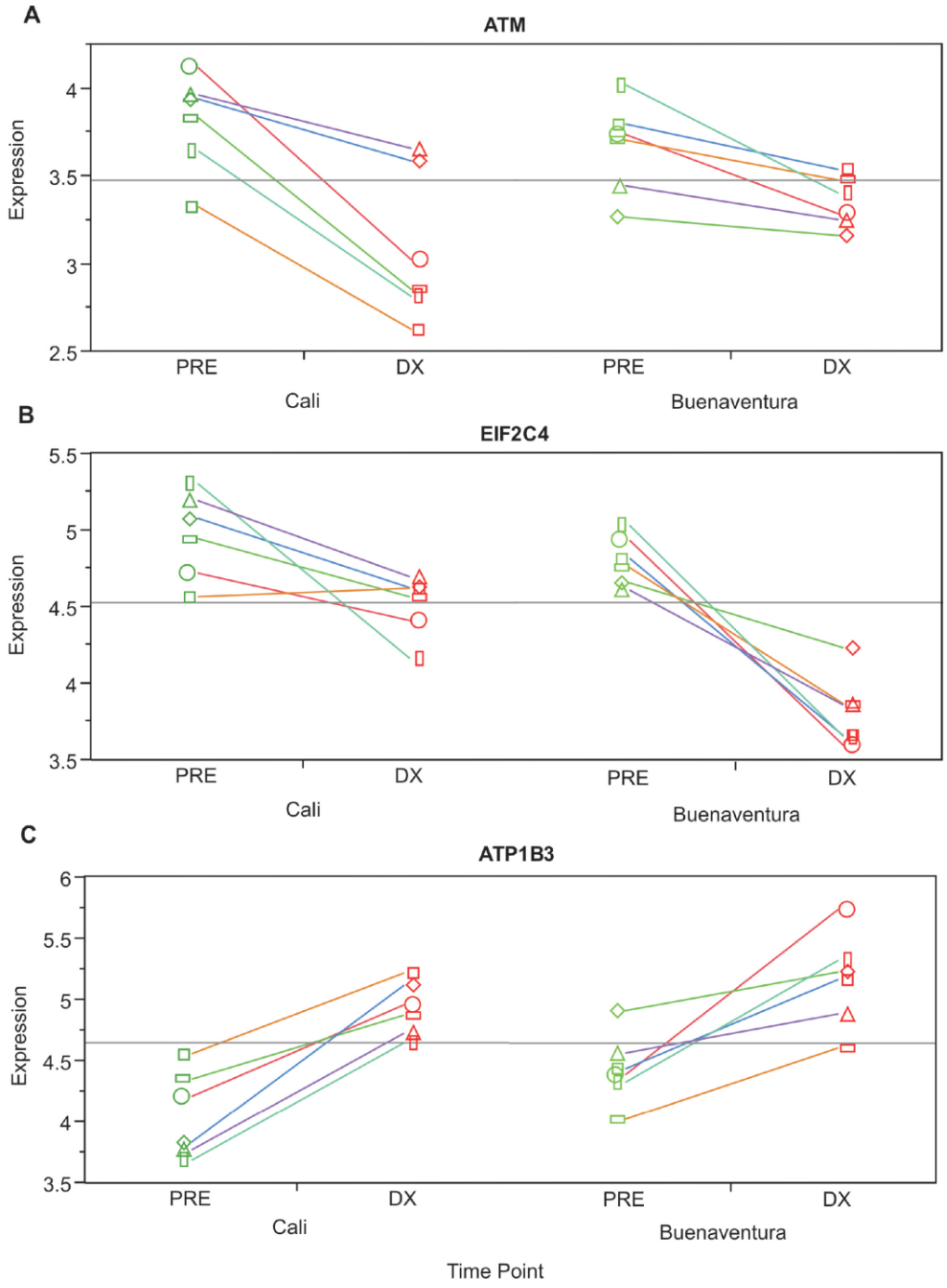

Figure 2.6. Transcriptional interaction effect between location and timepoint. Examples of the interaction effect showing gene ATM down-regulation in Cali (A), gene EIF2C4 down-regulation in Buenaventura (B) and gene ATP1B3 with no significant differnce in degree of up-regulation between the two sets of samples.

Table 2.1. Timepoint-by-Population interaction genes at p<0.05 group function.

Gene	p-value	Function
Immune Regulation		
MME	0.0438	This gene encodes a common acute lymphocytic leukemia antigen that is an important cell surface marker in the diagnosis of human acute lymphocytic leukemia
CXCR2P1	0.0036	Interleukin 8 Receptor, Beta Pseudogene – Non annotated
PECAM1	0.0278	Cell adhesion molecule that is required for leukocyte transendothelial migration (TEM) under most inflammatory conditions.
MAFB	0.0234	Transcriptional activator or repressor. Plays a central role in controlling lineage-specific hematopoiesis repressing ETS1-mediated transcription of erythroid-specific genes in myeloid cells. Is necessary for cell differentiation of monocytic, macrophage, podocyte and islet beta.
Signal transduction		
MAPK13	0.0182	Is one of the four p38 MAPKs which play an critical role in the cascades of cellular responses induced by extracellular stimuli like physical stress controlling the activation of transcription factors like ELK1 and ATF2 or proinflammatory cytokines.
GAB3	0.0092	Is related to numerous growth factor and cytokine signaling pathways.
G3BP1	0.0385	Is a heterogeneous nuclear RNA-binding protein and also an constituent of the Ras signal transduction pathway.
TLE3	0.0086	Transcriptional co-repressor that binds to a diverse number of transcription factors. Constrains the transcriptional activation, which is facilitated by CTNNB1 and TCF family members in the Wnt signaling.
Splicing		
SNRPG	0.0065	Plays an important role in the splicing of the cellular pre-mRNAs.
SNRPB2	0.0173	Encoded protein might play an imortant role in pre-mRNA splicing.
EIF2C4	0.0184	Members of this argonaute protein family are related to RNA silencing and are evolutionarily conserved.
Lysosome activity		
CTSH	0.0074	Protein encoded by this gene is a lysosomal cysteine proteinase; this protein plays an important role in the overall deprivation of lysosomal proteins.
RILP	0.0166	Related to the regulation of lysosomal morphology and distribution.
Cell Cycle, DNA Damage Response		
ATM	0.0154	Cell cycle checkpoint kinase. This genes is involved in signal transduction and cell cycle control. May works as a tumor suppressor.

Table 2.1 (continued)

PRKDC	0.0028	Sensor for DNA damage.
ARID4A	0.0342	Relates with a viral protein-binding domain at the retinoblastoma protein (pRB) this regulates cell propagation.
Phagocytes		
TNFSF13	0.0086	Plays a part in regulation of tumor cell growing. This gen might be involved in monocyte/macrophage-mediated immunological activities.
TBC1D14	0.0101	Adverse regulator of starvation-induced autophagosome formation.
CD163	0.0129	This gene is exclusively expressed in monocytes and macrophages. Functions as a severe phase-regulated receptor related to the clearance and endocytosis of hemoglobin/haptoglobin complexes by macrophage.
SIRPA	0.0227	Facilitates negative regulation of phagocytosis, mast cell stimulation and dendritic cell activation.
MMP25	0.0216	Response to bacterial infection and/or inflammation
Extracellular sensing		
SLC6A6	0.0337	This gene encodes a multi-pass membrane protein that is a member of a family of sodium and chloride-ion related transporters.
S100P	0.032	Might function as calcium sensor and contribute to cellular calcium signaling.
SLC11A1	0.0398	The protein work as a divalent change metal (iron and manganese) transporter involved in iron absorption and host resistance to some pathogens.
SLC7A7	0.0369	Is a transporter that is found in epithelial cell membranes where it transfers large neutral amino acids from the cell to the extracellular area.
Lipid Biosynthesis		
DGAT2	0.0375	This gene encodes one of two enzymes which catalyze the final reaction in the synthesis of triglycerides
LPPR2	0.012	Activity of phosphatide phosphatase
Other		
HAL	0.0249	Histidase converts histidine into ammonia and urocanic acid
TBXAS1	0.0027	Is an enzyme that plays a role in numerous pathophysiological processes that includes hemostasis, cardiovascular disease, and stroke.
POGK	0.0123	Exact function of the protein encoded by this gene is unknown.
FAM212B	0.0365	Uncharacterized Protein.
KIAA0232	0.0428	Uncharacterized Protein.

2.4.4 Comparison with effect of parasitemia on gene expression reported from Benin, West Africa

Finally, we reanalyzed an infant malarial gene expression dataset from Benin (Idaghdour et al. 2012). All samples were collected within a period of 10 weeks in the spring of 2010, and transcript abundance data was generated on Illumina HumanHT-12 BeadChips for 155 individuals (61 controls from Cotonou, 24 high parasitemia from the village of Zinvie, 52 low parasitemia from Zinvie, and 18 from the city of Cotonou). Critical differences relative to our study include (i) comparison with *P. falciparum* rather than with *P. vivax* infection, (ii) infants versus young adults comparison, and (iii) cross-sectional rather than Baseline vs Diagnosis analysis. Nevertheless, a significant correlation (Fig. 2.7A-B) was observed between parasitemia and two Axes of variation, with an apparent interaction effect involving Axis 5, where the stronger effect was again observed in the less endemic location (Fig. 2.7A-B, Cotonou, blue points and regression line). However, in this case there was activation of the innate immunity/inflammation genes as parasite burden increases. Axis 1, which is enriched for T-cell signaling activity (Preininger et al. 2013), was strongly reduced as parasitemia increased, but like Axis 5, not significantly affected in the infants with low parasitemia. From 32 genes showing a significant interaction effect between timepoint and population in our challenge experiment, 12 were nominally differentially expressed between malaria patients in the city of Cotonou and rural village of Zinvie in Benin.

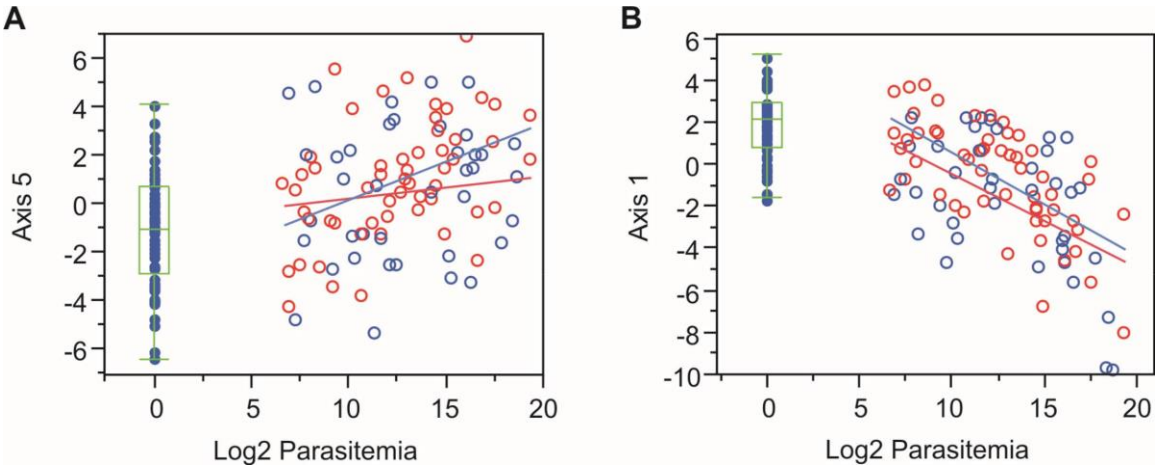

Figure 2.7. Log2 Parasitemia and Axis 2 and 5 Benin Study, West Africa. Significant correlation is evident between parasitemia and axis of variation, even though unrelated to location. **A.** Axis 5 **B.** Axis 1. Each plot shows the correlation among location and axis of variation, blue represents samples from Cotonou, Red represents samples from Zinvié; solid point represent control and open circle represent malaria samples.

2.5 Discussion

The core result of this study was that gene expression was significantly altered at the time of malaria diagnosis, particularly in the immunologically naïve volunteers. Although the targeted expression profiling is less comprehensive and less sensitive than the RNASeq, it suggests that there is minimal transcriptional change in peripheral blood prior to patent infection, and that individual profiles return to baseline within a few months of parasite clearance. No obvious difference in the transcriptomes of uninfected naïve and semi-immune volunteers was seen, but several hundred genes showed a stronger response in the naïve individuals. We cannot however conclude that prior immune exposure is the only reason for this difference as other lifestyle factors that distinguish the inland city of

Cali from the Oceanside town of Buenaventura, (where there is likely a larger proportion of African ancestry) may also play a role, However, the data is strongly suggestive of a long-term modulation of the malaria immune response involving multiple molecular pathways.

By comparison with published cross-sectional studies of gene expression in malaria patients, and although clinically immune individuals infected with *P. vivax* have lower levels of inflammatory and regulatory cytokines than individuals with *P. falciparum* malaria (Gonçalves et al. 2012), there were some surprising observations in this study. Most notably, the down-regulation of multiple genes related to innate immunity, inflammation, and neutrophil abundance, all correlated with Axis 5, was unexpected. The large cross-sectional study of infants with malaria conducted in the West African Republic of Benin (Idaghdour et al. 2012) documented a strong up-regulation of the same genes, although reanalysis of their data shown in Fig. 2.7A suggests that is only true in the presence of high levels of parasitemia. Even more surprisingly, the reduction in inflammatory gene expression was stronger in the naïve than semi-immune volunteers. One possibility is that there is a transient reduction in relative neutrophil counts as the parasite first appears in the bloodstream as the lymphoid cells begin to amplify their response, and this is corrected as parasite levels increase and neutrophilia occurs a few days into the infection (McKenzie et al. 2005; Kotepui et al. 2014).

An observation that is consistent with published data is the strong induction of an interferon response in association with blood-stage malaria (McCall et al. 2010; Jagannathan et al. 2014). It is unclear whether this induction was stronger in Cali or

Buenaventura, since a couple of the Cali volunteers had unusually high baseline interferon-related gene expression captured by Axis 7. It does appear that a few of the semi-immune individuals did not mount an interferon response, consistent with the absence of overt clinical symptoms and implying that their immunological memory was able to deal with at least the early stage of infection without mounting the kind of major immunological response observed in the naïve volunteers. This in turn implies that the presence of blood stage parasites alone is not the only determinant of whether or not an individual mounts an interferon response. The overall cytokine profile shifts reported in Fig. 2.5 did not correlate with the clinical profile differences, which suggests that the level of host immunity can vary due to the degree of acquired immunity through repeated exposure (Laishram et al. 2012). Larger sample sizes and longitudinal profiling during disease may identify associations between gene expression and physiological response, which is also likely to involve other tissues.

On the other hand, multiple classes of gene activity do seem to be differentially activated between naïve and semi-immune volunteers. These include various signal transduction molecules, genes related to macrophage activity, and other cellular processes that are known to influence immune responsiveness including lipid synthesis and lysosomal function, concordant with Portugal et al. (2014) who suggest that as children develop exposure-dependent immunity to *P. falciparum*, the molecular responses reduce pathogenic inflammation and boost anti-parasite mechanisms. The study in Benin again provides a potential comparison, since it included the contrast between children in the city of Cotonou with the rural village of Zinvié. Differences in human peripheral blood

gene expression according to lifestyle are prevalent, but it is nevertheless interesting that, of the 26 genes showing a significant interaction effect between timepoint and population in our challenge experiment, 9 were nominally differentially expressed between malaria patients from the two locations in Benin, compared with no more than three expected. Fig. 2.7B shows that Axis 1 (related to T-cell signaling) is down-regulated with high parasitemia, and consistently reduced in the village of Zinvié. This Axis was not affected in our study, but collectively these observations of context-dependent alterations in gene expression provide further evidence that immune history is an important mediator of the differential clinical profiles observed among individuals.

There is also considerable interest in the use of gene expression profiling to identify genes that may mediate robust vaccine responses. Recent study reports on influenza and yellow fever have highlighted individual genes that are required for vaccine effectiveness, but have also suggested that baseline profiles of immune cell types may provide better predictors of antibody production (Pulendran 2014, Nakaya et al. 2011). Various properties of *Plasmodium* suggest that this organism may present a more difficult scenario for dissecting the molecular basis of vaccine responses, but we consider the results reported here to be an encouraging baseline establishing that differential responses to a malaria challenge can be detected by gene expression profiling. It will be interesting to see whether pre-immune exposure influences the molecular basis of vaccination with irradiated sporozoites in the next phase of this study.

This study shows that gene expression is particularly strong in naïve volunteers in comparison to semi-immune individuals at the time of malaria diagnosis. Gene expression profiling of lymphocytes can thus be used to identify the type and duration of

the immune signals that are biomarkers for vaccine immunogenicity, and establish how semi-immune exposure modifies their activation.

CHAPTER 3

3. PROFILING GENE EXPRESSION OF THE HOST RESPONSE TO AN IRRADIATED SPOROZOITE IMMUNIZATION AND *PLASMODIUM VIVAX* MALARIA CHALLENGE

3.1 Abstract

The development of vaccines that provide sterile protection against pathogenic infection by the *Plasmodium* parasites that cause malaria is a major global public health priority. Development of effective vaccines requires a better understanding of the human immune response to vaccination. Here I report results of gene expression profiling of peripheral blood before and after 7 rounds of immunization with radiation attenuated *P. vivax* sporozoites (*Pv*RAS) in 20 volunteers, as well as after controlled challenge with live *P. vivax*. RNASeq was used to generate whole transcriptome profiles for 3 Controls, 5 protected Duffy Fy-, 5 protected volunteers immunized with RAS, and 7 susceptible volunteers not protected by immunization. The most profound changes in gene expression were observed in the contrasts between baseline and post- challenge, with distinct signatures differentiating protected and susceptible individuals. Analysis of transcriptional modules shows that B-cell signaling is reduced while cell cycle regulation and interferon response, as well as a probable signature of short-lived plasma cell activation, are highly elevated in individuals not protected by RAS, whereas regulatory T-cell signaling and an inflammatory response are elevated in protected individuals. Furthermore, subtle differences in the protection afforded by Duffy negative status and

by RAS were observed, and vaccination itself also modified aspects of B and T cell gene expression. Combined with immune cell profiling we expect the systems biology approach to suggest adjuvants that may improve the efficacy of malaria vaccines.

3.2 Introduction

The development of vaccines that provide protection against pathogenic infection by the *Plasmodium* parasites which cause malaria is a major global public health priority. The most recent World Health Organization report (WHO, 2015) indicates that there was an encouraging 48 percent decrease in mortality between 2000 and 2015. However, more than 3 billion people remain at risk, and 438,000 deaths are still attributed to malaria every year, largely affecting children and pregnant women (WHO 2015). Eighty percent of malaria deaths occur in just 15 countries, predominately due to *Plasmodium falciparum* in Africa, on which most of the malaria vaccine research has been focused. New genomic approaches based on the genome of the species (Gardner et al., 2002), as well as its transcriptome (Bozdech et al. 2003, Le Roch et al. 2003) and proteome (Florens et al. 2002, Sam-Yellowe et al. 2004, Hall et al. 2005), have supported progress in identification of suitable vaccine targets. Most notably, the RTS,S vaccine (Mosquirix^{TM}) which targets the circumsporozoite (CS) protein is now in Phase 3 trials following demonstrated efficacy and safety, decreasing morbidity and mortality by 28% and 18% respectively in young children and newborns (Lancent 2015).

The second most common malaria parasite is *Plasmodium vivax*, which accounts for 70% of the severe malaria cases and deaths in the Americas. Furthermore, WHO (2015) reported a rise in the appearance of severe cases of *P. vivax* malaria, as well as increasing treatment failure with

chloroquine. Consequently, development of a malaria vaccine against this species is a high priority for Latin America. Progress has been slowed by difficulties maintaining *P. vivax* in culture, and, until recently, by relatively limited information of the genome, transcriptome and proteome of the species. Nevertheless, my collaborators (Herrera et al., 2009, 2011) have developed an experimentally controlled *P. vivax* sporozoite challenge protocol for human studies in Colombia, and are now using this approach to evaluate malaria vaccine candidate approaches.

Development of a more effective vaccine also requires better knowledge and understanding of the acquisition of immunity. Gene expression profiling has been used to gain insight into natural clinical protection (Idaghdour et al. 2012, Rojas-Peña et al. 2015), and while it is clear that multiple arms of the immune system are engaged, the precise mechanisms leading to immunity remain to be elucidated. The pre-erythrocytic parasite stage is thought to be the optimal target of malaria vaccine development (Clyde et al 1973). For example, sterile immunity to malaria can be induced by vaccination with radiation-attenuated sporozoites (RAS) (Clyde et al. 1973, 1975; Rieckmann et al. 1974), which likely prevents parasite development in the liver following engagement of various regulatory T cell functions. Sanaria Inc. have developed an injectable *P. falciparum* RAS immunization strategy which is in phase 1 and 2 trials for vaccination of naïve adults (Richie et al 2015), and colleagues Socrates Herrera, Myriam Arévalo-Herrera and their team at Caucaseco in Cali, Colombia, have initiated similar research in relation to *P. vivax*.

Given that *P. vivax* continues to affect millions of people in endemic countries, better understanding of the immune response established by *Pv*RAS immunization is desirable. To this end, we have conducted a Phase 1 trial in which 20 naïve volunteers (5 Duffy negative Fy- and

15 Duffy positive Fy+) from Cali were exposed to 7 immunizations to test the efficacy of immunization with radiation attenuated *P. vivax* sporozoites delivered by mosquito bites (*Pv*RAS)(Arévalo-Herrera et al. in review). This study design allows us to contrast natural (Duffy negative) and induced (PvRAS) protection as well as failure of protection, since it was observed that just 42% (5 of 12) individuals showed sterile immunity in response to vaccination. None of the Duffy negative individuals showed symptoms of parasitemia, while 7 of the Duffy positive subjects (6 men) who received *Pv*RAS immunizations developed parasitemia. One of these volunteers did not develop symptoms, and similarly one of the three non-vaccinated Duffy positive controls was also asymptomatic. There was no difference in the time to first appearance of blood stage parasites assessed by thick blood smears (12 to 13 days in both the non-vaccinated and non-protected groups). The other 5 immunized Duffy positive volunteers (all women) did not show symptoms or parasitemia, and were therefore identified as protected against the *P. vivax* challenge. Samples from this trial provided a unique opportunity to assess the immunological response to vaccination with *Pv*RAS by transcription profiling using RNASeq analysis of peripheral blood samples.

Systems biology approaches based on gene expression profiling have been used to identify gene signatures associated with vaccination response in a variety of viral settings (Nakaya et al. 2011; Li et al. 2014, 2016). These have mostly focused on the process of vaccination itself, rather than its impact on challenge by the pathogen. Here I report an initial understanding of the modular immune response to a *P. vivax* malaria challenge after protection with *Pv*RAS. By contrasting blood samples drawn at baseline, following immunization but pre-challenge, and at the first day of diagnosis of malaria in affected subjects, I ask whether there are differences in the immune

profiles between volunteers (i) before and after the immunization with *Pv*RAS, (ii) after experimental challenge with live *P. vivax*, and (iii) between naturally and vaccine-protected individuals. Multiple arms of the immune system are found to be engaged in the response to live pathogen, and I also document systemic effects of Duffy negative status both before and after parasite challenge.

3.3 Methods

3.3.1 Clinical Study

This study is based on a single-blind clinical trial that included healthy adult volunteers, males and non-pregnant females, between 19-41 years of age, who satisfied inclusion/exclusion criteria as determined by clinical history and serological tests, and who finished the clinical trial (see Arévalo-Herrera et al., submitted). Volunteers were broadly informed about the risks of participation. All volunteers had to pass an oral or written exam related to the trial and its risks. The clinical trial was registered at ClinicalTrials.gov under registry number NCT01082341. The experimental design protocol was approved by the Institutional Review Boards (IRB) at the Malaria Vaccine and Drug Development Center (CECIV, Comité de Ética Centro International de Vacunas, Cali), and Centro Médico Imbanaco (CMI; Comité de Ética en Investigacion, Cali), and subsequently the Georgia Institute of Technology IRB gave approval for the genomic profiling (protocol H13495).

Fifteen malaria-naïve Duffy-positive (Fy+) (12 RAS, and 3 mock-control) and 5 Duffy-negative (Fy-) volunteers from Cali were included. The volunteers visited the clinic in Cali approximately every 8 weeks to receive ~100 bites from *Anopheles albimanus* mosquitoes for a total of 7 visits.

In the case of the Fy+ RAS vaccines, the mosquitoes had been fed radiation-attenuated *P. vivax* sporozoites, whereas the Fy- volunteers were exposed to non-attenuated parasite. Two weeks after the final vaccine dose, the second "pre-challenge" blood sample was drawn, and all participants were given a curative dose of chloroquine and primaquine anti-malarials. A further six weeks later they were challenged with infection by ~100 bites from mosquitoes infected with live, non-irradiated *P. vivax* sporozoites. Approximately fourteen days later, at first diagnosis of infection in susceptible volunteers, the final blood "Diagnosis" blood sample was taken, and where necessary a curative program of anti-malarial medication was given. The volunteers were followed for 60 days to ensure they were clear of malaria. Of the mock-controls (who received bites from parasite-free mosquitoes during the vaccination phase), two contracted malaria after challenge with live parasite but one remained uninfected (parasite-free). Of the RAS-vaccinated individuals, six (all women) were symptom-free, one of whom nevertheless had parasite in her blood. The other six (five men, one woman) were not protected and had mild malaria. All five of the Duffy negative women were naturally protected from infection as expected. Figure 3.1 summarizes the experimental design and blood sampling strategy for the RNAseq.

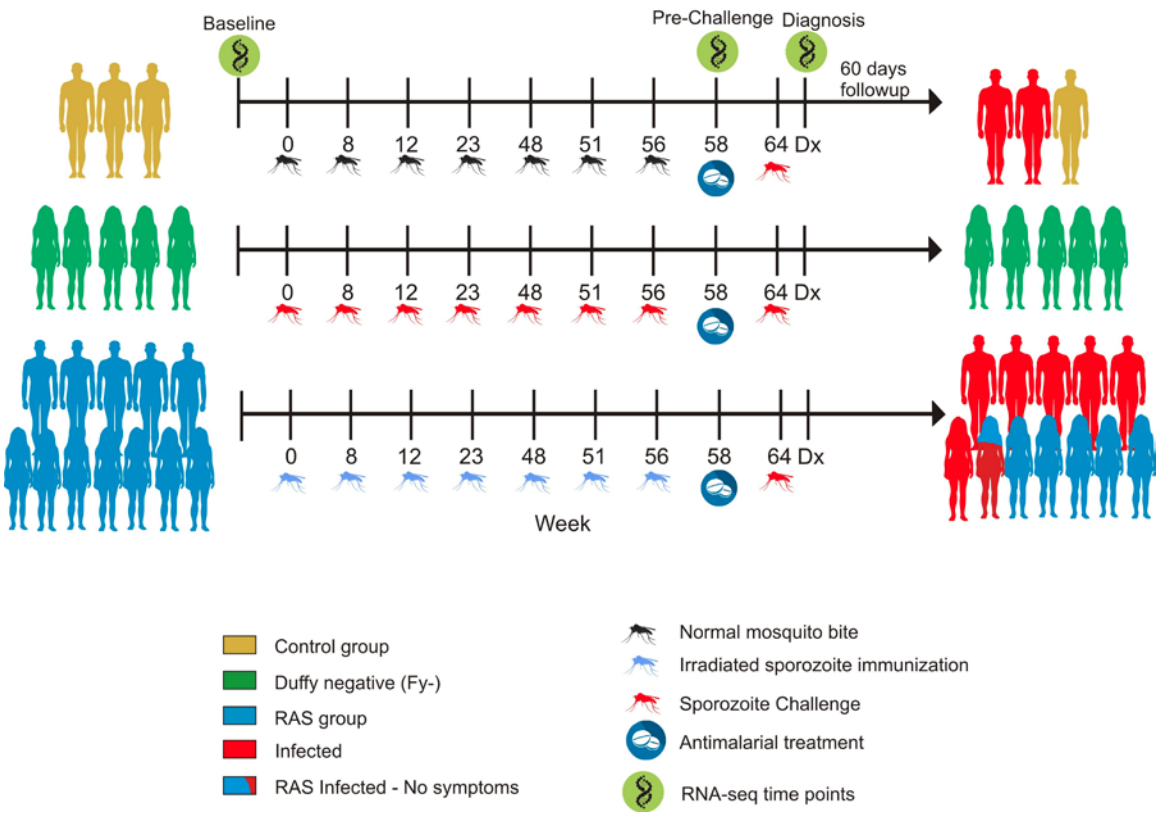

Figure 3.1. Experimental design and sample collection timeline. 58 total samples were used for RNAseq, taken at three time points represented by the green symbols: Baseline (20 samples), Pre-challenge (19 Samples) and day of Diagnosis (19 samples). Controls, Fy- and Fy+ vaccinated individuals are represented by yellow, green and blue silhouettes respectively of representative genders on the left, and red shading on the right indicates which individuals were diagnosed with malaria. One woman (red/blue) was positive for infection but asymptomatic.

3.3.2 RNASeq

Whole blood RNA was prepared for all 20 individuals at each of the three time points, namely Baseline (prefix B), Pre-Challenge (C) and at Diagnosis (D). Approximately 1 ml of blood was taken for each sample, and mixed with 2 ml of buffer in a Tempus tube, which preserves whole

blood RNA at 4°C indefinitely, but depletes the red blood cell and platelet fractions at the RNA extraction step. Whole blood (predominately leukocyte) mRNA was extracted using Tempus Blood RNA Tube isolation kits following the protocol provided by the manufacturer, Applied Biosystems. Sample RNA quality was determined based on the Agilent Bioanalyzer 2100 RNA Integrity score (RIN). Two samples of RNA from one control-pre-challenge and one Duffy-negative-diagnosis sample were severely degraded and were not included for sequencing. A few samples had RIN lower than 6, but these were not found to be outliers in the analysis.

Library preparation for RNASeq was performed using the Illumina Stranded mRNA Sample Low Throughput (LT) RNA Sample Preparation Protocol. Short read sequencing was performed in rapid run mode with eight samples per lane on an Illumina HiSeq 2100 at the Georgia Institute of Technology, generating 100 bp single-end libraries with an average of 31.8 ± 6.1 million single end reads per sample.

The RNASeq dataset has been deposited into the Gene Expression Omnibus archive (GEO) under accession number GSE85263, including short read deposition in the SRA.

3.3.3 Data analysis

Raw RNASeq reads (Fastq files) were evaluated using FastQC software analysis to check the quality of the data for each sample. The overall alignment rate was 93%, and three samples (B025, DGS006 and DGS065) failed QC, so were excluded. The 100 bp single-end reads were then aligned to human genome (hg19/GRCh37 assembly with the UCSC reference annotation) using Bowtie as the short read aligner via Tophat2 (Trapnell et al. 2012). Transcript abundance

was estimated at the level of the whole gene as counts per million (cpm) values using the tool htseq-count of the open source python package HTSeq (Andres et al. 2015). No attempt to assess transcript isoforms was made as the read depth was deemed not sufficient to give reliable estimates, and similarly exon-level data was not analyzed.

To guarantee that the data were more normally distributed and to simplify the interpretation of differential expression, cpm values were transformed to logarithm base 2. Scale Trimmed Mean of the M-values (TMM) normalization was then performed using edgeR from Bioconductor, as described in Robinson and collaborators (2010). All downstream analyses were performed on this normalized data set. Differential expression was also assessed using the linear modeling framework for RNAseq data in edgeR. A Benjamini-Hochberg false discovery rate of 10 percent (approximately corresponding to a p-value of 0.0001) was used to select differentially expressed genes.

All results were verified by a second normalization model using the SNM package in R (Mecham et al, 2010) to fit individual as an adjustment variable with time point and clinical status as biological variables, facilitating analysis of the relative changes in expression after adjustment for inter-individual differences. Since all results were qualitatively the same as the TMM analysis, I only report the TMM. Further analyses to establish the contributions of timepoint, gender, and clinical status to the overall gene expression variation, reported in Figure 3.2, were performed using the Principal Variance Component Analysis (PVCA) routine in the Basic Expression Workflow in JMP Genomics (SAS Institute, Cary NC).

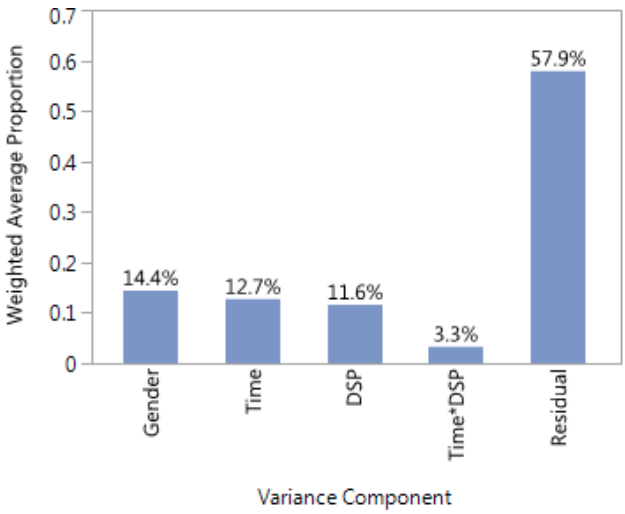

Figure 3.2. Principal Variance Component Analysis of Gene expression. Bar graph shows the weighted average of the variance captured by the first five principal components among samples that is explained by Gender, Time (Baseline, Pre-challenge and Diagnosis), DSP Category (Duffy, susceptible and protected), and time by category.

3.3.4 Blood Transcription Modules (BTMs) and Blood Informative Transcripts (BIT)

Modular analysis of transcript abundance was based on the reconstruction of several hundred gene networks from integrative analysis of over 30,000 transcriptomes in 500 public studies (Li et al. 2014, 2016). These modules are described as the Blood Transcription Modules (BTMs) and disclose distinct aspects of peripheral blood gene expression, including transcriptional indications of antibody responses to vaccination and other immunological functions. In parallel, we also used blood informative transcript (BIT) analysis, which focuses on 10 common axes of variation that were detected in multiple human peripheral blood gene expression datasets of healthy individuals that have been found to consistently co-vary in peripheral blood (Preininger et al. 2013). Each axis includes between one hundred and several thousand genes that gene set enrichment analysis suggests are involved in particular immune functions, broadly speaking, T

cell signaling (Axis 1), reticulocyte number (Axis 2), B cell signaling (Axis 3), inflammation/neutrophil signaling (Axis 5) and Interferon signaling (Axis 7). Axes 4 appears to be related to house-keeping while Axis 6 is as yet undefined, while a newly identified Axis 10 is enriched for regulation of mitosis and cell division. These axes are related to 28 modules of co-expressed genes described initially by Chaussabel and collaborators (2008), and our analyses show that they are embedded within clusters of BTM. In other words, BTM describe refinements of the course patterns of gene expression captured by the BIT Axes.

Both BTM and BIT Axes are computed as first principal components (PC1) of the co-regulated genes, either all of the genes in each module, or the 10 axis-defining BIT for each Axis (note that almost identical scores are obtained with 5 or 20 transcripts: Preininger et al, 2013). PC were generated in JMP Genomics version 8 (SAS Institute, NC), for each BTM, and given the high covariance of representative genes can be thought of as weighted averages of transcript abundance. The strength of covariance is given by the percent variation of the included transcripts that is explained by PC1. These scores were contrasted with respect to the clinical status groups (protected, susceptible, and Duffy negative) across the three time points, using standard parametric t-tests or analysis of variance. Two way hierarchical clustering of the BTM and BIT scores was used to assess the overall functional relatedness of samples and modules, using Ward's method implemented in JMP Genomics.

3.3.5 Variant genotype calling and ethnicity

To avoid the confounding factor of ethnicity across the samples, I calculated the ancestry composition of the volunteers from genotype data inferred from the RNASeq data. The short

read alignments from Tophat were imported into the GATK HaplotypeCaller (McKenna et al., 2010) for each sample at each of the three time points. Given relatively high false positive genotype calls from RNA sequencing, I required that a SNP was called in all three samples for an individual, which was performed by merging the VCF files using the vcf-merge utility from VCFtools. Only common polymorphic sites also found in the 1000G public reference dataset (The 1000 Genomes Project Consortium) were retained for downstream analysis. The curated genotypes were converted to BED/BIM/MAP format using Plink (Purcell et al. 2007), and then merged with data from the 1000G project. Subsequently, the ADMIXTURE program (Alexander et al. 2009) was run on the final merged dataset, using k=3 theoretical populations, which correspond to European, Asian/Native American, and African source populations (this part of the ethnicity analysis was performed by Andrew Conley from Jordan Lab). Figure 3.3 shows that the 20 Colombian individuals in this study tend to have slightly more African ancestry than observed in the Medellin sample in the HapMap collection (International HapMap Consortium, 2003), typically consisting of 40%-60% European ancestry and up to 40% African and 40% Amerindian. The 5 Fy- volunteers are the outliers with between 40% and 80% African ancestry.

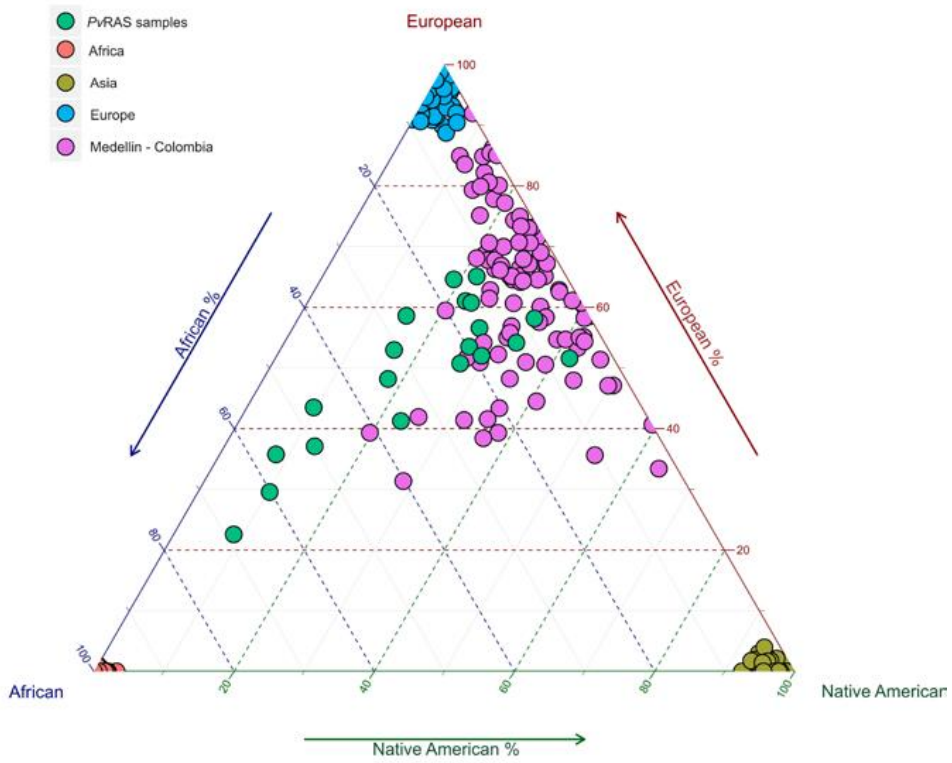

Figure 3.3. Triangle plot. Ancestry proportions inferred from RNASeq polymorphisms.

3.4 Results

Peripheral blood gene expression profiles were obtained for 20 volunteers who completed a single blind randomized clinical trial of *Pv*RAS immunization, as schematized in Figure 3.1. Five individuals were Duffy negative, and 15 were Duffy positive, including 3 mock-vaccination Controls and 12 *Pv*RAS immunized volunteers. After live *P. vivax* challenge, five women showed sterile immunity while one was asymptomatic but positive for parasite, and five men and one woman contracted malaria despite vaccination, so were not protected. The demographics and clinical attributes of the volunteers are shown in Table 3.1. RNASeq was performed for each individual at three different time points (Baseline, Pre-challenge and Diagnosis day), with the exclusion of five samples due either to low quality RNA or failure of quality control of the

RNASeq data. The final dataset consists of 55 RNAseq single end 100bp samples with a total of 13,282 genes that had a cpm of 1 or more in at least 3 of the samples.

Table 3.1. Demographics of the 20 volunteers included in the clinical trial.

Volunteer	Group	Gender	Age	nBites	Infected	TBS
002	Ctrl	M	27	758	NO	-
020	Ctrl	M	41	945	YES	13
065	Ctrl	M	23	963	YES	13
038	Fy	F	24	478	NO	-
058	Fy	F	21	487	NO	-
066	Fy	F	37	358	NO	-
075	Fy	F	19	476	NO	-
084	Fy	F	25	412	NO	-
005	RASNP	M	30	418	YES	13
006	RASNP	M	40	497	YES	13
009	RASNP	F	33	458	YES	13
011	RASNP	M	38	423	YES	13
017	RASNP	M	35	386	YES	13
021	RASNP	M	22	442	YES	12
026	RASNS	F	36	440	YES	12
001	RASP	F	24	440	NO	-
007	RASP	F	21	362	NO	-
010	RASP	F	25	460	NO	-
012	RASP	F	37	428	NO	-
025	RASP	F	21	403	NO	-

RAS, radiation attenuated sporozoites (group infected with irradiated sporozoites); P, protected after immunization; NP, not protected after immunization; NS, asymptomatic Volunteer 026 had *P. vivax* parasitemia >1,000 parasites/µl but no malaria symptoms. Ctrl, control (mock-immunized with non-infected mosquitoes); Fy(-), Duffy negative; F, female; M, male; nBites, estimated number of bites received from infected mosquitoes for the challenge; TBS, day of positive thick blood smear.

To assess the transcriptional signatures and functional content induced by *Pv*RAS immunization and the malaria challenge after immunization, I first combined data from each time point and analyzed them as a single data set consisting of 19 baseline, 19 pre-challenge and 17 diagnosis day samples. Principal variance component analysis indicated that 17 percent of the variation is between the three classes of individual (Duffy negative, protected, susceptible), 12 percent between the three time points, and 3 percent attributable to an interaction effect. Age was not a significant source of variation when included in the model. Post-hoc analysis of the major PC indicates that the Diagnosis day samples tend to be divergent from the Baseline and Pre-challenge ones, and that the susceptible individuals are the most distinct at the Diagnosis day, accounting for the interaction effect. Consequently, I first focus on the question of whether live parasite challenge induces different immune profiles according to clinical outcome.

3.4.1 Differential Gene Expression at Diagnosis Associates with Protection or Susceptibility

To do so, I applied the BTM and axes of variation frameworks to characterize the general patterns of differential gene expression between sample types at Diagnosis day only. Hierarchical clustering analysis of PC1 of the BTMs and BIT axes data revealed three distinct clusters of individual (Fig. 3.4). These three clusters largely distinguish between the clinical groups: RAS volunteers who became protected separated themselves from RAS volunteers who were susceptible, while the Duffy negative volunteers form a distinct cluster more similar to the protected RAS volunteers. Two RAS non-protected (susceptible) individuals have aberrant profiles: one (RASNP 005) has a unique profile dominated by signatures of mitosis (multiple BTM and the new cell division Axis 10) indicating that the infection has temporarily induced

abnormal cell cycle regulation; the other (RASNP 021) is almost identical at Diagnosis and Pre-Challenge and may be a sample mix-up.

Both the nature of the BTM in each of the clusters, as well as of the Axis that is embedded within each cluster, provide evidence that multiple arms of the immune system are engaged in immune-protection. Most apparent is the strong up-regulation of interferon response and mitosis in the non-protected RAS volunteers, contrasting with up-regulation of T-cell and B-cell signaling evident in both the Duffy negative and RAS protected volunteers, to varying degrees. Furthermore, inflammation, as well as platelet and reticulocyte differentiation, are clearly up-regulated in protected RAS volunteers relative to Duffy negative volunteers, implying that natural and vaccine-mediated protection have slightly different immunological consequences. It is particularly noteworthy that volunteer 026, who became serologically positive for the parasite but did not exhibit symptoms, exhibits a combination of characteristics of non-protected (interferon and mitosis response) and protected individuals (platelet and inflammation). It is also apparent that the two controls cluster with their respective clinical groups, namely 002 who did not become infected (but appears to have been exposed since his profile is related to that of the RAS protected volunteers and different from his Pre-challenge sample, not shown), and 020 who has a classical mild malaria profile.

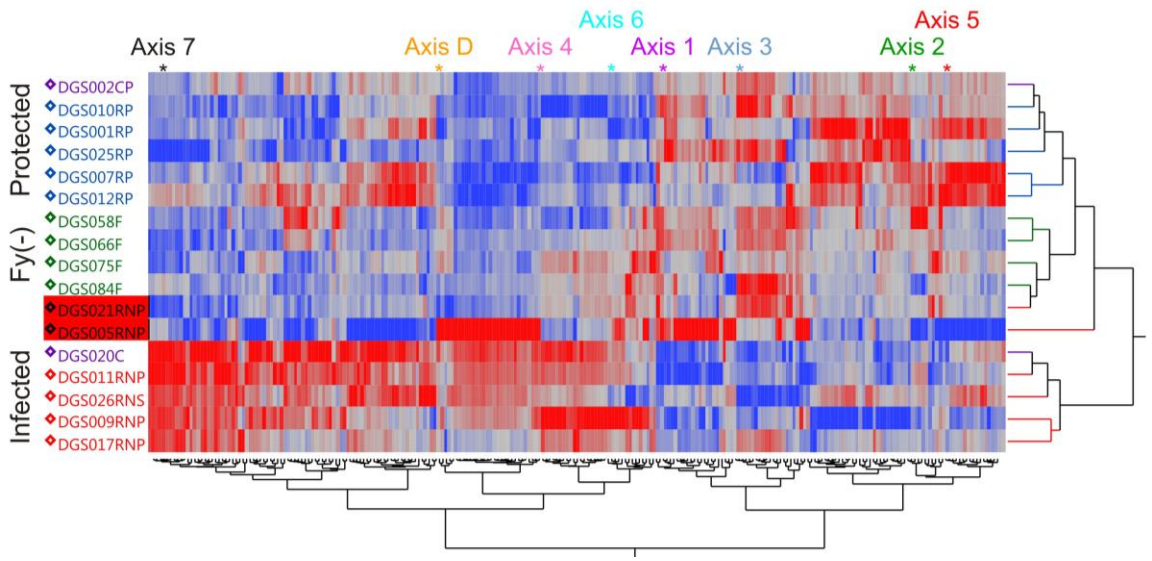

Figure 3.4. Hierarchical clustering of Blood Transcript Modules (BTMs) and Axes of Variation. The heat map shows two-way hierarchical clustering of the PC1 (columns) in each sample (rows) of 250 BTMs and 8 sets of BITs at the day of the diagnosis. Red indicates high expression, blue low, and gray intermediate, of genes in each module. Green sample labels represent Duffy negative volunteers, red labels represent susceptible *Pv*RAS immunized volunteers who became infected during the challenge, blue represent protected *Pv*RAS immunized volunteers, and purple, control volunteers. Nomenclature at the end of the labels means: F, Duffy negative; RNP, RAS non-protected; RP, RAS protected; C, control; CP, control protected and NS, Non symptoms.

In order to obtain a more comprehensive picture of the changes in gene expression as parasites first appear in the blood, I studied the BITs at each of the time points and subdivided the samples into 9 categories, baseline, pre-challenge and diagnosis, each subdivided into Duffy negative, protected (individuals who after the challenge were not infected) and susceptible (individuals who at the time of the challenge were infected). The results shown in Figure 3.5 confirm that T-

cell signaling (Axis 1) is already induced at the pre-challenge in the Duffy negative individuals (Fig. 3.5), but delayed until the diagnosis timepoint in the *pv*RAS protected, while does not become induced in the susceptible individuals at all. Neutrophil activity (Axis 5) is generically reduced in the Fy- even at baseline, and does not change across time within each type, with the exception of a surprising slight reduction in some of the susceptible individuals (consistent with our previous report of the response of non-vaccinated naïve volunteers; Rojas-Peña et al, 2015). Interferon signaling (Axis 7) increases in most of the susceptible individuals at the time of the diagnosis but remains unchanged in the protected Fy- and Fy+ individuals. Similarly, mitotic activity (Axis 10) is dramatically up-regulated only at diagnosis in the susceptible, non-protected vaccines and control.

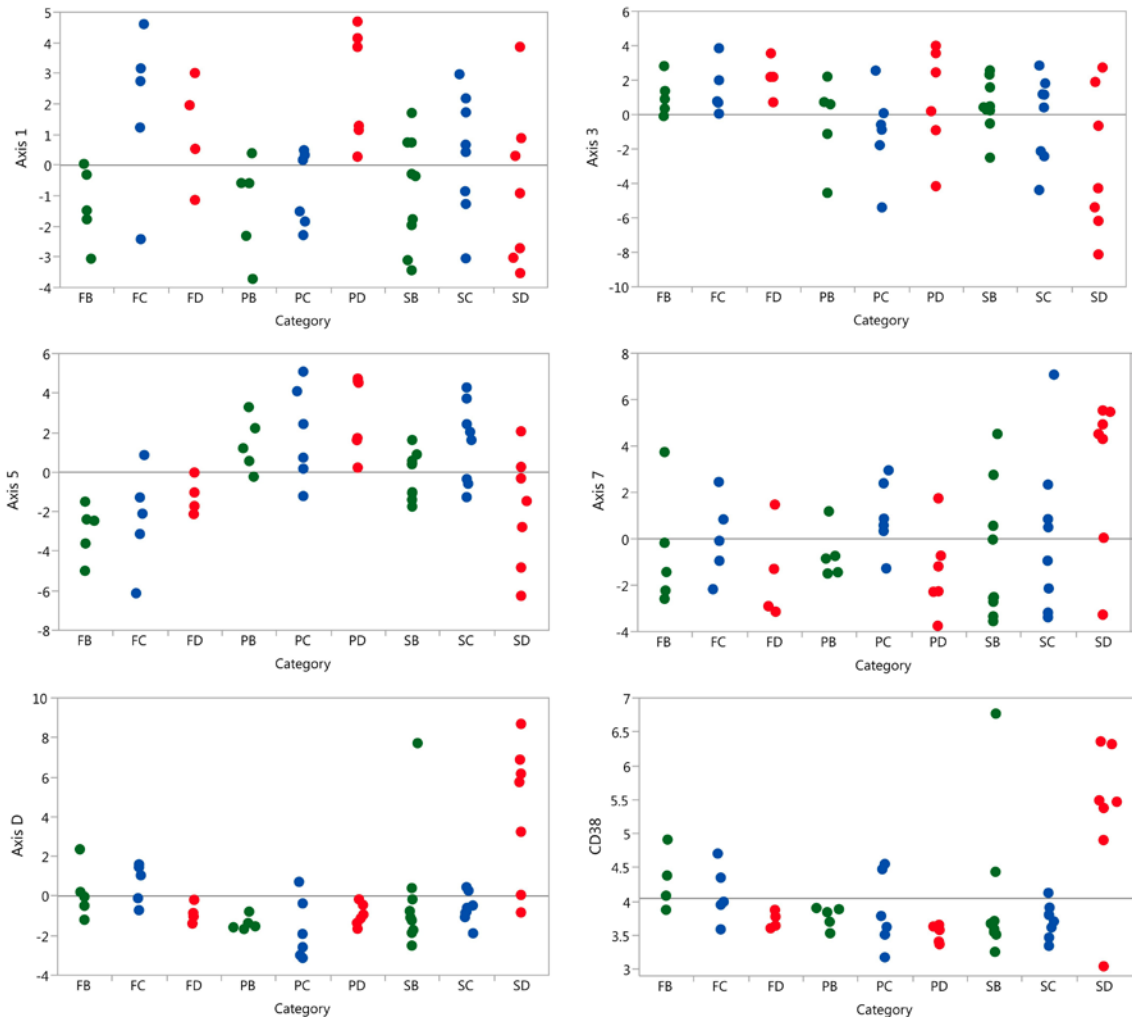

Figure 3.5. Differential expression of axes of variation. Each panel shows the PC1 score for the indicated Axis for individuals at Baseline (green, B), Pre-Challenge (blue, C) or Diagnosis Day (red, D), with from left to right Duffy negative (left, F), Protected (middle, P), or Susceptible (right, S). Panels are top to bottom, left to right: T-cell signaling (Axis 1), B-cell signaling (Axis 3), Neutrophil signaling (Axis 5), Interferon response (Axis 7), Mitosis (Axis D), and expression of the CD38 gene.

Analysis of variance of the PC1 scores of the BTM modules at diagnosis across the three individual categories (susceptible, protected and Duffy negative) identifies 49 modules that are dysregulated with elevation associated with absence of parasite at diagnosis, and 48 modules that are associated with clinical susceptibility, being elevated in the presence of parasites (Appendix A). Examples of these modules are represented in Figure 3.6, which also implies that there are some differences in which modules are related to parasitemia and which to symptomology. The red individual in the "protected" type, in the middle of each panel, is the asymptomatic but infected volunteer 026. Her module score is almost always closer to that of the sick individuals, being more protected-like for some modules (T-cell differentiation/Axis 1, cytoskeletal remodeling) but susceptible-like for others (antigen presentation, CD1 and dendritic cell receptors). The bottom row of panels show three modules which are up-regulated in malarial volunteers and 026, involving chemokine receptors, cell division, and interferon/Axis 7 signaling, implicating these processes more in sickness or recovery than induction of sterile immunity.

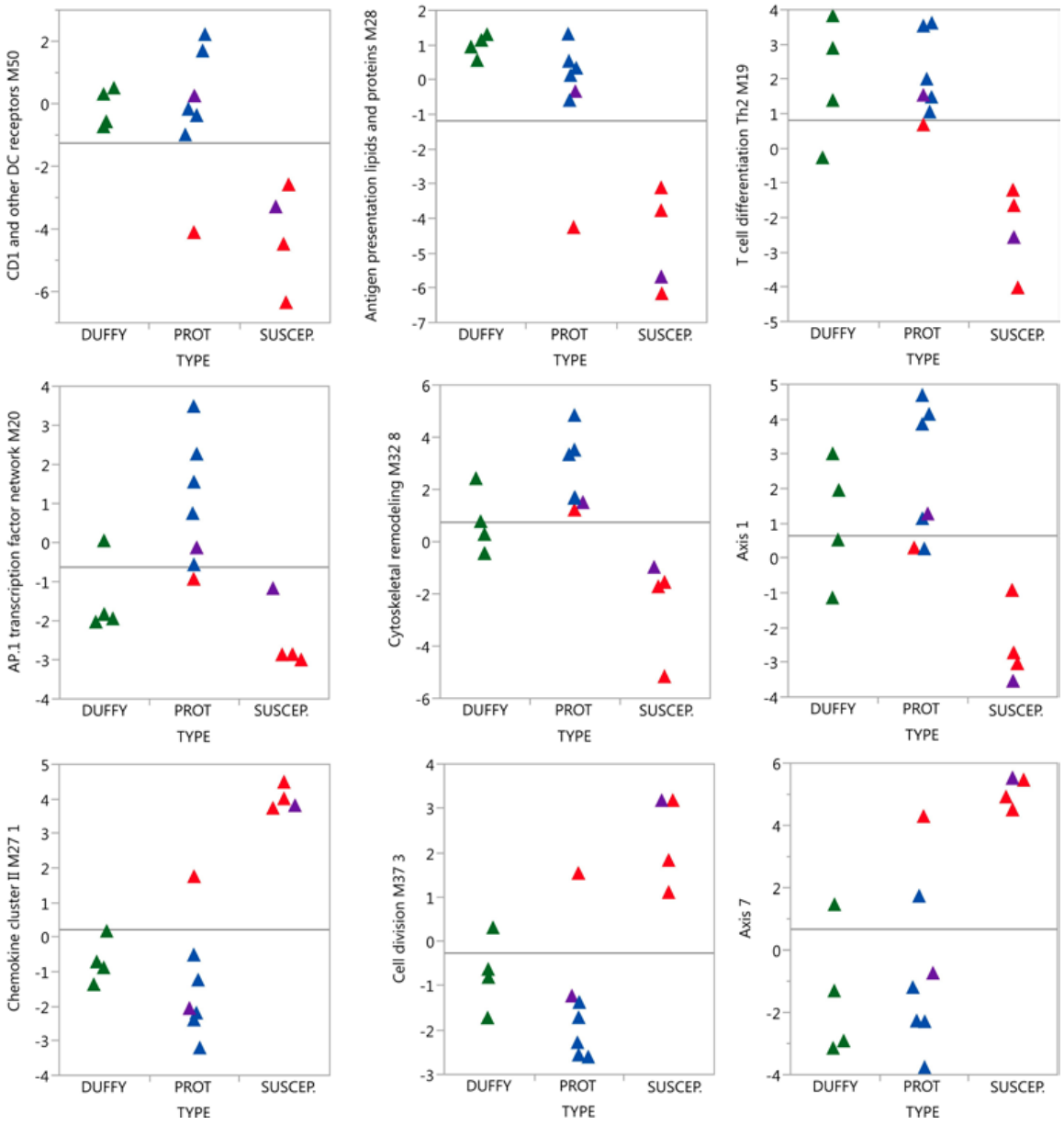

Figure 3.6. BTM and axes that distinguish Duffy Fy (-), protected and infected volunteers on diagnosis day. Each plot is an example of a modules or axis of variation, which differentiates volunteers who exhibit symptoms (A, B, G, H and I) from volunteers who did not become infected (C, D, E, F). Green dots represent Duffy negative volunteers, red represent non-protected RAS sick volunteers, blue represent RAS protected volunteers. The red dot in the

protected RAS column represents volunteer 026 who was semi-protected, while the purple dots are the controls, one of whom was unexpectedly protected.

A particularly important observation involves the expression of the plasma cell marker CD38, which is strongly elevated at diagnosis day in susceptible but not protected individuals. I identified 345 genes which are highly correlated with CD38, 63 negatively and 262 positively across the dataset, most of which are specifically elevated in the susceptible individuals. Figure 3.7 shows that these genes strongly tend to differentiate short-lived (SLPC) and long-lived plasma cells (LLPC) in human bone marrow (Halliley et al, 2015; S Garimalla, FE Lee, in preparation), with a clear bias toward up-regulation of genes enriched in short-lived CD19+CD38+ SLPC relative to CD19-CD38+ LLPC, and correspondingly down-regulation of the LLPC-enriched genes. A similar result was observed with suspected circulating peripheral blood cells expressing the same markers, but with less pronounced biases reflecting the less strong differentiation of the transcriptomes of the blood plasma cell types.

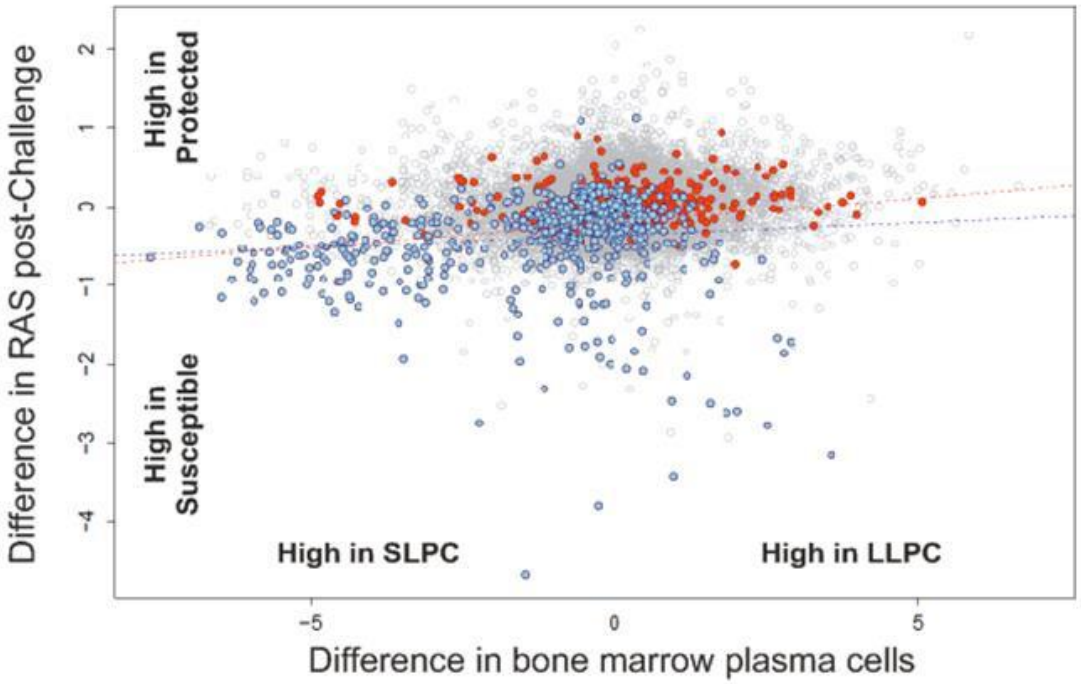

Figure 3.7. Apparent up-regulation of Short-Lived Plasma Cells in susceptible individuals. The plot shows all genes correlated with CD38 in this study at r>0.5, with genes positively correlated in red and negatively in blue, comparing susceptible and protected individuals at diagnosis (y-axis) and comparing SLPC and LLPC from the bone marrow of three human donors (x-axis) (Halliley et al, 2015).

3.4.2 Differential Expression Associated with Duffy negative status

Next, I asked whether there is any evidence for differential expression due simply to vaccination, as has repeatedly been reported for example in studies of influenza and yellow fever viruses. There is low power for this comparison given the small sample size after removing the Fy- individuals, who received non-attenuated parasite during the vaccination phase. Contrasting Pre-Challenge with Baseline in the RAS individuals revealed just a handful of genes at the stringent cut-off of $p<0.0001$ beyond which only one gene is expected by chance. Consequently, there is little evidence for an impact of vaccination alone. It should also be noted that the post-

vaccination sample was taken two weeks after the last of seven exposures to radiation-attenuated *P. vivax*, so will not capture early responses to the vaccine.

However, the contrast of Pre-Challenge with Baseline in the Fy- individuals did reveal substantial differential expression, which was qualitatively similar to the comparison of Diagnosis day with Baseline for these individuals (see Figure 3.5 for genes associated with Axis 1). This is as expected, since the challenge is essentially the same as their exposure to live parasite during the vaccination phase. Effectively, the blood gene expression is reporting a response to the presence of hypnozooites in the liver, as irradiated sporozoites are unable to progress to the blood stage. To establish this further, in Figure 3.8 I plot differential expression between Duffy pre-challenge and Baseline, contrasted with differential expression between individuals protected by RAS and Baseline. All of the genes significantly differentially expressed at Diagnosis day in the RAS individuals, even though not significant at Pre-challenge in the FY-, are trending in the same direction. This is particularly notable for the T-cell genes (red), but is also true of the neutrophil genes (blue), which are generally lower in individuals offered natural protection by the Duffy antigen as see in Figure 3.5.

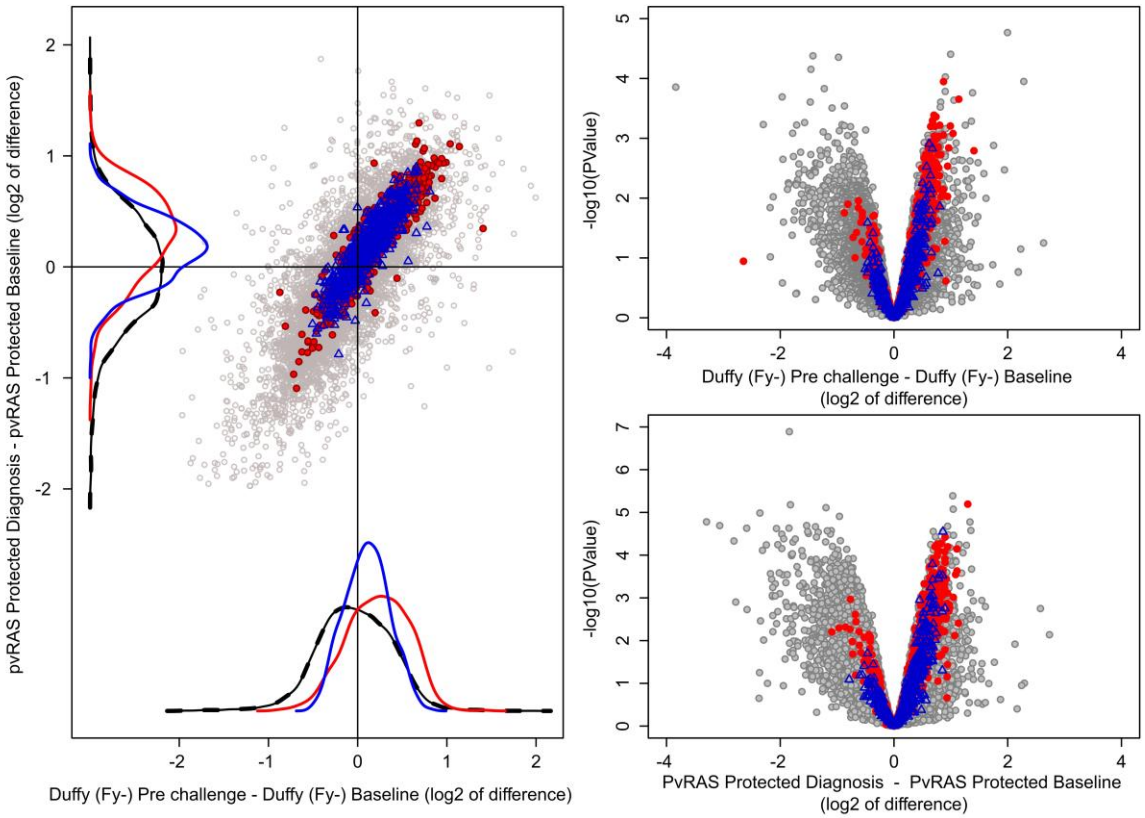

Figure 3.8. Similarities between natural and vaccine-induced protection. The scatter plot shows differential expression between *pv*RAS volunteers at the time of the diagnosis and baseline on the x-axis, and between Fy- volunteers pre-challenge and at baseline on the y-axis. Red circles represent 900 genes in the Axis 1, blue circles represent 1028 genes in the axis 5, and gray circles all remaining genes, with density distributions shown below and to the left.

Despite this similarity between the protected responses to live parasite, there are some very clear differences between the Duffy negative and positive samples overall. In order to ask whether there are differences among the sample types before challenge, we performed ANOVAs separately for each of the three response types, excluding the Controls. This analysis revealed no significant excess of differential expression between protected and susceptible individuals either at Baseline or Pre-challenge, but that Duffy individuals are somewhat differentiated even when I

combined Baseline and Pre-challenge to increase the power for the comparison of *pv*RAS and Fy- samples. This analysis showed that 273 genes are significantly up-regulated at FDR 5%, and 584 genes are down-regulated, confirming that Duffy status has a marked impact on the overall gene expression profile of adults.

The nature of this differential expression is explored further in Figure 3.9 which shows six examples of BTM that are differentially expressed in Duffy prior to the final malaria challenge. These include elevated gene expression related to subsets of chemokine signaling, B cell development, and integrin-mediated cell signaling, and down-regulation of neutrophil signaling, extracellular matrix and complement activation (suggesting reduced general inflammation at baseline).

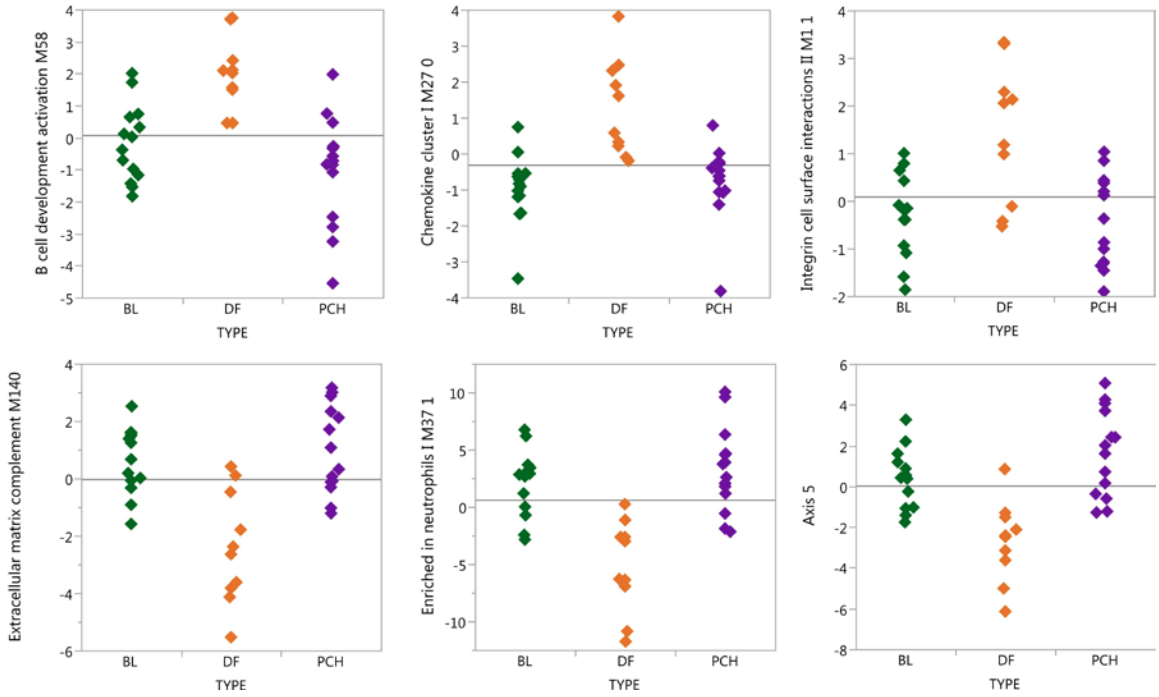

Figure 3.9. BTMs and axes that distinguish Duffy negative and positive before malaria challenge. Each plot represents a BTM or axis that differentiates Duffy negative volunteers from the other volunteers at Baseline or Pre-challenge. Green symbols represent baseline, purple pre-challenge and orange Duffy negative volunteers at either timepoint.

3.5 Discussion

This chapter describes a transcriptional profiling approach to characterize signatures of the impact of PvRAS immunization on the response to *P. vivax* malaria challenge. My results indicate that multiple innate and adaptive immune mechanisms are induced by *Pv*RAS immunization and that specific signatures seem to associate with parasite clearance and with clinical protection from malaria symptoms. Furthermore, I was able to identify specific transcriptional modules that differentiate Duffy negative-mediated natural protection from vaccine-induced protection.

Comprehensive BTM analysis suggests that both the B- and T-cell arms of the adaptive lymphocyte response are engaged in both natural and vaccine-induced protection. It is known that a B-cell response is required for regulation and elimination of infected red blood cells (Perez-Mazliah and Langhorne 2015) during the blood stage of infection, and correspondingly remarkable that 3 of 4 susceptible individuals profiled at diagnosis showed no sign of B-cell activation captured by Axis 3, which is heavily enriched for genes involved in B-cell functions, and is up-regulated in protected individuals after challenge. However, this deficit is offset by the strong up-regulation of CD38+ and hundreds of co-expressed genes that my fellow student Swetha Garimalla has recently observed to differentiate short- and long-lived plasma cells. I have not been able to identify another cell type with a similar profile, so the results strongly imply that the susceptible individuals are specifically engaged in immunoglobulin-based clearance of the parasite, while the protected individuals have effectively engaged memory B cells. We speculate that the protected individuals may have also been able to direct the precursor plasma cells toward a long-lived fate, since the LLPC cells express high levels of estrogen-response genes and are presumably more primed for the switch in females, possibly explaining the gender-bias in the effectiveness of the *pv*RAS vaccine.

Similarly, T cell activation and T cell differentiation were also up-regulated only in protected individuals, supporting previous findings reported with the RTS,S malaria vaccine, where proliferation of T cells was associated with RTS,S-immunized volunteers (Stoute et al. 1998, Dunachie et al. 2006). Other studies have also report an increase in T cell immunogenicity and a delayed prepatent period after malaria challenge (McConkey et al. 2003, Webster et al. 2005). A new algorithm for parsing the relative prevalence of immune cell types in whole blood transcript

profiles, CIBERSORT (Newman, et al. 2015) applied to this dataset also suggested that Treg cells are activated in protected individuals, whereas memory CD4+ T cells are activated in the susceptible ones. These results imply that the vaccine is facilitating the generation of specific B and T cell responses when it is effective. There is though some heterogeneity in the level of lymphocyte activation since two individuals, 007 and 012, had only a mild response. Since the same two individuals have specifically elevated activity of dozens of interferon response-related modules, it is possible that they have a delay in the timing rather than a dampening of the lymphocyte response.

In contrast to the protected volunteers, the non-protected individuals displayed up-regulation in modules related to type I interferon response, chemokine, pro-inflammatory cytokines, complement activation, cell migration, dendritic cells, and an extensive set of mitosis related modules. Previous studies have shown release of proinflammatory mediators like tumor necrosis factor (TNF) and interferon response to be a response to a malaria infection, which potentially contributes to organ damage (Miller et al. 2002). We postulate that the vaccination has facilitated an immediate recall of memory cells, which either rapidly clears the parasite or prevents it from transitioning from the hypnozoite to the merozoite stage. Consequently, there is no induction of the strong interferon response or induction of proliferation of immune cells.

One volunteer, 026, who displayed parasitemia but was asymptomatic, had a hybrid profile consisting of the strong interferon response as well as the inflammatory activation. This suggests that larger studies will enable us to distinguish between pathways that provide sterile protection mediated by *Pv*RAS, and ones that are more relevant to malaria symptoms. Differences were

observed in modules related to the cell cycle, which we also observed previously in semi-immune individuals in response to a malaria challenge (Rojas-Peña et al. 2015) and platelet activation which recently was identified to have protective properties against malaria infection, binding the infected erythrocytes and killing the parasite specifically by factor IV activity (Brendan et a. 2009, Love et al. 2012, McMorran et al. 2012, 2013)

Surprisingly, most of the RAS individuals that were infected after the challenge were male, suggesting that only women were protected by the *Pv*RAS immunization. This is particularly surprising since gender is only a minor contributor to variation in overall gene expression profiles observed in peripheral blood, and only a few dozen X-linked genes are typically highly divergent (Kukurba et al. 2016, Idaghdour et al. 2010). Only 11% of the variation was explained by gender in our dataset (Fig. 3.2), and most of this was because of the bias between responses at diagnosis: just 14% of the baseline and pre-challenge gene expression is between genders. Future studies should be directed at understanding whether and how hormonal differences influence the acquisition of immunity by RAS or other types of *P. vivax* vaccine, focusing to begin with on the LLPC compartments in the blood and bone marrow.

This study shows that differential expression is particularly strong in protected volunteers in comparison to non-protected individuals at the time of malaria diagnosis, providing a general picture of how the different arms of the immune system are engaged during the response to infection after immunization. Whole blood profiling does not generally have the resolution to define which immune cell sub-types are most important, and preferably should be supplemented with cellular immune profiling as has been used to highlight the roles of regulatory T cells and

dendritic cells in mediating the RAS response in the context of *P. falciparum*. However, it is apparent that signatures of these processes can be seen in the RNASeq data. Focused transcriptome profiling of sorted immune subsets in larger samples, combined with knowledge of differences between asymptomatic and complex malaria cases, should provide further insights that may be useful in enhancing vaccine development for malaria.

CHAPTER 4

4. LONGITUDINAL GENOMICS OF COMPLICATED MALARIA: A PILOT PROJECT IN COLOMBIA

4.1 Abstract

The majority of malaria infections are uncomplicated, and with appropriate treatment can be readily cured. However, with the increasing frequency of resistance to chloroquine, particularly of *Plasmodium falciparum*, there are concerns that we may see a resurgence not just of uncomplicated, but also complicated malaria cases. In recent years, transcriptomics has emerged as a helpful tool to explore and identify the mechanics of the roles of complex biological systems in disease, including malaria infection. To begin addressing the knowledge gap between cellular immunity and recovery from complicated malaria, I here describe the host transcriptomes of eight complicated malaria cases each over four or more days of recovery, and relate the gene expression changes to clinical features including parasitemia, disease severity, and rate of recovery. The first five principal components of transcriptional variation in the eight patients with complicated malaria captured 53% of the total variance in expression of 13,889 genes, while the effect of the other variables (*Plasmodium* species, gender, age, and day of hospitalization), explained 28% of the total variance, leaving 19% due to unidentified residual sources. Hierarchical clustering confirms this result. Patient CM02, a pregnant woman with symptoms of preeclampsia but no history of fever, showed the most gene expression differentiation over the period of her complicated malaria infection. Failure of a pregnant

woman to resolve infection corresponds with her variable transcriptome profile. In most patients, Axes 5 and 7 are clearly lower at time point 3 and generally time point 4 as well, with time point 1 always the highest. In contrast, Axis 2 tends to be highest at time point 3. These results are easily interpreted as signs of inflammation and elevated interferon signaling upon hospitalization, and of reticulocyte development associated with recovery from anemia after the peak of the complex malaria episode. Comparison of the Blood Transcript Modules (BTMs) at the day of diagnosis for samples from the irradiated sporozoite vaccine study in Chapter 3 with the day of hospitalization samples of this complicated malaria pilot project, indicates overall similarity of individuals based on infection versus sterile immunity. While I observe some commonalities in the responses across patients, there is wide individual variability and it is concluded that more detailed analyses of individual cell types linked to cellular and humoral data will be required to resolve the nature of personalized mechanisms of recovery.

4.2 Introduction

The majority of malaria infections are uncomplicated, and with appropriate treatment can be cured rapidly and reliably (Patel et al. 2003). However, with the increasing frequency of resistance to chloroquine, particularly of *Plasmodium falciparum*, there are concerns that we may see a resurgence not just of uncomplicated, but also complicated malaria cases (Oladipo et al. 2015; WHO 2015). These are the approximately one percent of infections where the patient shows extreme symptoms that generally require hospitalization for several days and can be lethal if untreated.

Complicated malaria is initiated in the erythrocytic stage, at which time it can manifest as a vascular obstruction due to sequestration of infected red blood cells. Subsequent inflammatory processes due to the presence of *Plasmoduim* parasites can cause cellular dysfunction, damage and death in different organs (Vasquez and Tobon 2012, Bassat and Alonso 2011). Complicated malaria during pregnancy can amplify morbidity in these patients, causing symptoms such as vaginal bleeding, abdominal pain, loss of perception of fetal movements, signs of preeclampsia and risk of early delivery (Tobon 2009), and it can also lead to death of the mother. Other consequences of complicated malaria include cerebral malaria, acute renal failure, hypoglycemia, respiratory distress syndrome, disseminated intravascular coagulation, hypotension and shock among others (Patel et al. 2003).

The incidence of complicated malaria is influenced by multiple factors, including the nature of the parasite (the species, multiplicity of infection, genetic variation, and parasitemia), host attributes (gender, age, nutrition level, level of immunity, and history of morbidity), as well as environmental (geography, weather) and socioeconomic conditions (access to healthcare and viable public health prevention strategies) (Chaparro-Narváez et al. 2016; Byakika-Kibwika, et al. 2009). Although complicated malaria can occur at any age, in some African regions complicated malaria is reported predominately in infants, while in other regions it affects both children and adults (Byakika-Kibwika, et al. 2009). In populations with high levels of malaria exposure (Bunn et al. 2004; Doolan et al. 2009) where complicated malaria is more prevalent, immunity developed through recurring exposure has been shown to control clinical

manifestations and limit the severity of infection. However, it is not known why some patients present complications, while others do not, or what physiological mechanisms are responsible for the diverse symptoms.

A limited understanding of human host biology, specifically the dynamics involved in pathogenesis and recovery, has delayed the development of prophylaxis, including vaccines. Most studies of complicated malaria have been performed in countries with high endemicity, notably in the African Sub-Continent. There has been very little published concerning complicated malaria in Latin America, a region with 21 countries at risk of *P. vivax* and *P. falciparum* infection (Guerra et al., 2008; Guerra et al., 2010). Approximately 90% of the population of Colombia live in areas prone to malaria transmission, and the country presents 14.2% of the malaria cases in Latin America, with half related to *P. vivax* and half to *P. falciparum* (WHO 2015). Immune surveillance has shown that individuals infected with malaria release inflammatory intermediaries, like tumor necrosis factor (TNF) and interferon gamma (IFN-γ), which can damage organs such as the brain, kidneys and lungs (Miller et al. 2002). One study has associated incidence of complicated cases with a specific sequence of the *P. falciparum* genome (Kirchgatter and Portillo 2002), and my collaborators have demonstrated elevated presence of multiple haplotypes of *P. vivax*, indicating multiple infections, in patients from Colombia (Pacheco et al. 2016).

In recent years, transcriptomics has emerged as a helpful tool to explore and identify the mechanics of the roles of complex biological systems in disease, including malaria

infection. For example, we have shown that naïve individuals exposed to *P. vivax* respond with divergent expression of genes with multiple roles in immunity, relative to semi-immune individuals (Rojas-Peña et al, 2015). Systems biology analyses seek to integrate profiles of gene expression with alterations in immune cell populations and function, as well as with cytokine profiles, metabolomic data, and antigen production (Li et al. 2014, 2016). To begin to address the knowledge gap between cellular immunity and recovery from complicated malaria, I here describe the host transcriptomes of eight complicated malaria cases each over four or more days of recovery, and relate the gene expression changes to clinical features including parasitemia, disease severity, and rate of recovery.

4.3 Materials and Methods

4.3.1 Study area:

Colombia is one of 21 countries in Latin America with high endemicity of malaria. The rate of incidence of malaria in Colombia fluctuates seasonally, with a tendency to increase during the wet seasons, and annual averages of between 80,000 and 120,000 cases have been reported by the Public Health Surveillance System (SIVIGILA). In Colombia, *P. vivax* is the predominant species, being responsible for 60% of cases. However, this pilot study was completed on some of the notably endemic areas of the Pacific coast of Colombia, in the states of Choco, Valle del Cauca and Nariño, where the primary malaria species is *P. falciparum* (Fig. 4.1) (Rodriguez et al. 2011; Chaparro-Narváez et al. 2016).

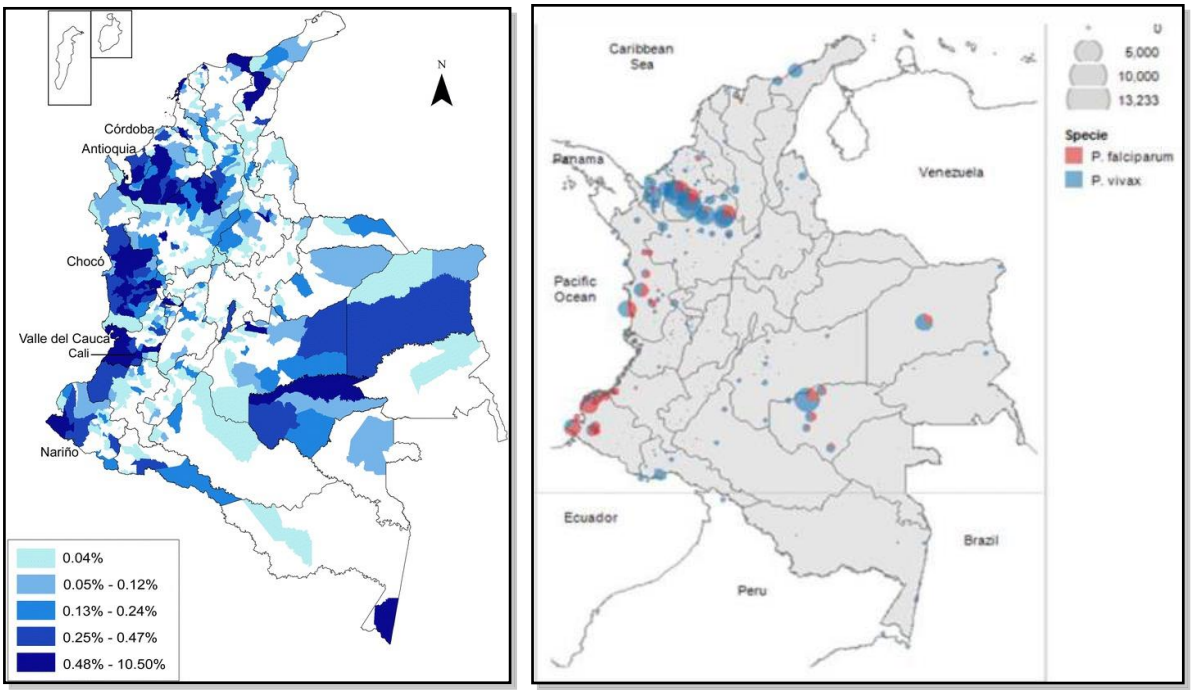

Figure 4.1. Complicated malaria cases distribution in Colombia between 2007 and 2013 (Chaparro-Narváez et al. 2016) (Left). Distribution of malaria cases per city and parasite 2013, Source: modify from SIVIGILA (2014) (Right).

4.3.2 Study design

This study, conducted over the course of 2015, was designed as a pilot project to assess the potential utility of longitudinal transcriptome profiling to resolve mechanisms of resolution of complicated malaria. A total of six *P. falciparum* and two *P. vivax* positive individuals were included. Individuals were enrolled in the study regardless of age, sex or ethnicity. Complicated malaria was defined based in the 2010 Colombian Ministry of Health (MoH) adaptation of the WHO guidelines (2000) to the Colombian population (MinSalud 2010) (Table 4.1).

Table 4.1. Complicated malaria classification guidelines: Ajusted from Chaparro-Narváez et al. (2016)

Criteria	Defined before 2010 (WHO 2000)	Defined after 2010 (MinSalud 2010)
Cerebral malaria	Impaired consciousness or coma (Blantyre score < 3 or Glasgow score < 9); unconsciousness with the possibility of waking up	Unchanged
Renal dysfunction	Serum creatinine > 3.0 mg/dL and/or urine vol < 400 mL in 24 h (adults) or <12 mL/kg of body weight in 24 h (children)	Serum creatinine > 1.5 mg/dL
Hepatic dysfunction	Serum bilirubin > 3 mg/dL and altered liver function tests	Serum bilirubin > 1.5 mg/dL or aminotransferases > 40 U/L
Respiratory distress	Increased respiratory rate at admission, presence of abnormal lung sounds or pulmonary oedema (X-rays)	Unchanged
Circulatory collapse or shock	Systolic blood pressure (SBP) < 70 mm Hg in adults or <50 mm Hg in children (3–5 years)	SBP < 80 mm Hg in adults
Hyperemesis	>5 episodes in 24 h	Not applicable
Hyperpyrexia	Axillary temperature >39.5 °C	Not applicable
Hypoglycaemia	Blood glucose level < 40 mg/dL.	Blood glucose level < 60 mg/dL
Severe anaemia	Haemoglobin < 5 g/dL or haematocrit < 15 %	Haemoglobin < 7 g/dL
Disseminated intravascular coagulation (DIC)	Abnormal bleeding in the presence of laboratory evidence of DIC	Unchanged
Acidaemia/acidosis and hyperlactemia	Acidaemia/acidosis (clinical signs)	Plasmatic bicarbonate < 15 mmol/L or base excess > −10; acidaemia pH <7.35; lactate acid > 5 mmol/L
Haemoglobinuria	Macroscopic haemoglobinuria	Macroscopic haemoglobinuria and positive urine dipstick
Hyperparasitaemia	>100,000 asexual parasites/μL of P. falciparum or in mixed infection with P. vivax and schizontaemia	>50,000 asexual parasites/μL

Patients were enrolled into the project after full informed consent. The experimental design protocol was approved by the Institutional Review Boards (IRB) at the Malaria Vaccine and Drug Development Center (CECIV, Comité de Ética Centro International de Vacunas, Cali), and Centro Médico Imbanaco (CMI; Comité de Ética en Investigacion, Cali), with subsequent approval for genomic analysis from the Georgia Institute of Technology IRB. The course of disease in patients with complicated malaria was monitored daily at a level 3 hospital until discharge, and clinical complications attributable to malaria infection were recorded.

4.3.3 Blood sample collection and RNA extraction

RNASeq analysis was performed for eight individuals at four time points (days), namely the first day of hospitalization, two further days of hospitalization generally at daily intervals, and at recovery (day of discharge). For sample CM02, two sets of samples (12 total) were collected due to a longer period of acute infection including a second malaria event. Approximately 1 mL of blood was taken for each sample, and was collected into a Tempus tube, which preserves whole blood RNA at 4°C indefinitely. Whole blood mRNA was extracted using Tempus Blood RNA Tube isolation kits and following the protocol provided by the manufacturer, Applied Biosystems/ThermoFisher Scientific. Sample quality was determined based on the Agilent Bioanalyzer 2100 RNA Integrity score (RIN).

4.3.4 Library construction and RNA sequencing

Library preparation for RNASeq was performed using the Illumina TruSeq Stranded mRNA Sample Low Throughput (LT) preparation protocol. Short read sequencing was performed in rapid run mode with ten samples per lane on an Illumina HiSeq 2100 at Georgia Tech, generating 100 bp single-end libraries with an average of ~30 million single reads per sample.

Quality control of the raw RNASeq reads (Fastq files) was analyzed using FastQC software, confirming for each sample the number of reads, GC content percent, median sequence length and median Phred score. The 100 bp single-end reads were then aligned to the human genome hg19/GRCh37 assembly using Bowtie as the short read aligner via Tophat2 (Trapnell et al. 2012). The human UCSC reference annotation was used to align and quantify transcript abundance, which was estimated as counts per million (cpm) using the tool htseq-count of the Python package HTSeq (Anders et al. 2015) 13,889 genes that had a cpm of 1 or more in at least 3 of the samples where used for subsequent analysis.

4.3.5 Data analysis

To simplify interpretation and to guarantee that the data were more normally distributed, cpm values were scaled by Trimmed Mean of the M-values (TMM) normalization and then log 2 transformed using code in edgeR (Robinson et al. 2010). All downstream analyses were performed on this TMM normalized data at the level of genes after collapsing of exon level estimates into a single value per gene. No effort was made to

evaluate transcript isoform abundance or alternative splicing. Principal component variance analysis (PCVA) was performed with the Basic Gene Expression routine in JMP Genomics version 8 (SAS Institute, NC), which was also used to perform hierarchical clustering using Ward's method to identify sub-types of expression profiles.

To identify enriched gene sets, functional annotation was performed using the ToppGene Suite (Chen et al. 2009), which uses a hypergeometric comparison of the proportion of genes differentially expressed in each class relative to their representation in the human genome. Enrichment was evaluated on various lists of genes down or up-regulated for sub-type comparisons (for example, Pregnant woman CM02 vs remaining 7 samples). Enrichment was evaluated in the gene annotation categories: molecular function, biological process, mouse phenotype, human phenotype, cellular component, pathway and disease-gene associations.

Blood informative transcript (BIT) analysis (Preininger et al, 2013) was used to define which of 8 major axes of variation are perturbed in each individual. Each axis is defined by the first principal component of 10 consistently highly co-expressed genes in blood gene expression datasets. Seven of the Axes were described in Preininger et al (2013), while the eighth one reported here (Axis 10) has not previously been described (Axes 8 and 9 reported in the initial study are too weak to resolve in this relatively small dataset). Gene set enrichment analysis of the Axes implies that they represent gene expression involved in particular immune functions, broadly speaking: T cell signaling (Axis 1), reticulocyte number (Axis 2), B cell signaling (Axis 3), inflammation/neutrophil number

(Axis 5), Interferon signaling (Axis 7), and regulation of the cell cycle/mitosis (Axis 10). Axis 4 involves generic housekeeping functions, while Axis 6 remains uncharacterized. Modulation of axis scores provides a high level overview of recovery of immune function, and is visualized for each individual on radar plots with 8 arms radiating from minimum to maximum scores, with polygons linking profiles from the day of hospitalization until the day of discharge.

Blood Transcript Modules (BTM) provide an alternate mode of tracking modulation of immune functions. Li et al (2014, 2016) reconstructed gene networks from over 30,000 transcriptomes downloaded from approximately 500 studies, and identified 334 gene modules via reverse engineering. The Blood Transcription Modules (BTM) are thought to summarize distinct transcriptional indications of molecular functions such as antibody responses to vaccination and cytokine production. As with the BIT, BTM are defined as the first PC of the genes in the module. We contrasted BTM across the eight individuals using hierarchical clustering with Ward's method in JMP Genomics version 8 (SAS Institute, NC), and also compared the complicated malaria samples on day one (hospitalization day) with the day of the diagnosis of a set of samples from individuals vaccinated with irradiated sporozoites and challenged with *P. vivax* .

4.4 Results

A total of eight patients were recruited for this pilot project, two infected with *P. vivax* and six with *P. falciparum*. The patients exhibited respiratory distress, severe thrombocytopenia and oral intolerance among other symptoms associated with

complicated malaria. Over the course of hospitalization, blood samples were taken from the patients daily to quantify parasitemia via PCR and thick blood smear. The parasitemia data showed that after prophylactic treatment, all patients were able to clear the parasite and were discharged after approximately one week, with the exception of patient CM02, a pregnant woman who continued to show levels of parasitemia in her blood after day 3 (Table 4.2). She remained hospitalized and relapsed into complicated malaria 15 days later, and a second series of samples were collected for gene expression profiling from her.

All of the RNA samples passed quality control filters (RIN > 6), few samples had RIN lower than 6, but these were not found to be outliers in the analysis. After normalization and filtering of raw RNASeq data, as described in the Methods, trimmed mean (TMM) log2 values for 13,889 genes were used for downstream analysis.

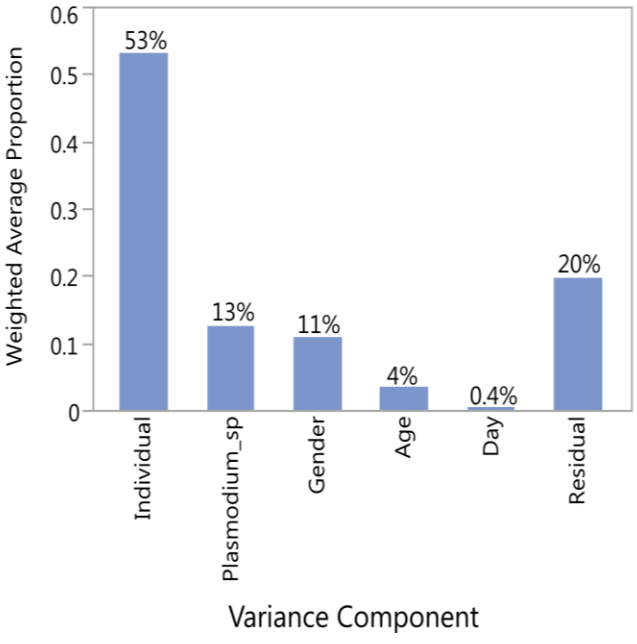

Figure 4.2. Variance component analysis of gene expression profiles. Bars show the weighted average of the variance captured by the first five principal components of the overall gene expression profiles that is explained by individual, *Plasmodium* species, gender, age or day of the sampling. The largest proportion of the variance is among the eight individuals.

4.4.1 Individual variation dominates gene expression in complicated malaria cases

The first five principal components of transcriptional variation in the eight patients with complicated malaria captured 53% of the total variance in expression of the 13,889 genes (Fig. 4.2), while the effect of the other variables (*Plasmodium* species, gender, age and day of hospitalization), explained 28% of the total variance, leaving 19% due to unidentified residual sources. This result implies that baseline differences in host peripheral blood transcriptomes, which we have shown known to be very consistent over a period of 12 months in healthy individuals (Tabassum et al, 2015), plays the dominant role in defining the gene expression profiles even of patients who are experiencing a malaria infection. Notably, day of hospitalization had the smallest detectable influence, but this may in part be because patients' profiles recover at different rates and the peak of disease relative to initial hospitalization was likely variable.

Hierarchical clustering (Figure 4.3) confirms this result, since the four samples for each patient typical cluster side-by-side, or at least in the same branch for patients with similar overall profiles. A notable exception was the clustering of the samples CM44 (D2) with CM41 (D1) and CM05 (D4), whereas samples CM09 (D3 and D4) clustered with sample CM44, as well as with the third sample of CM52.

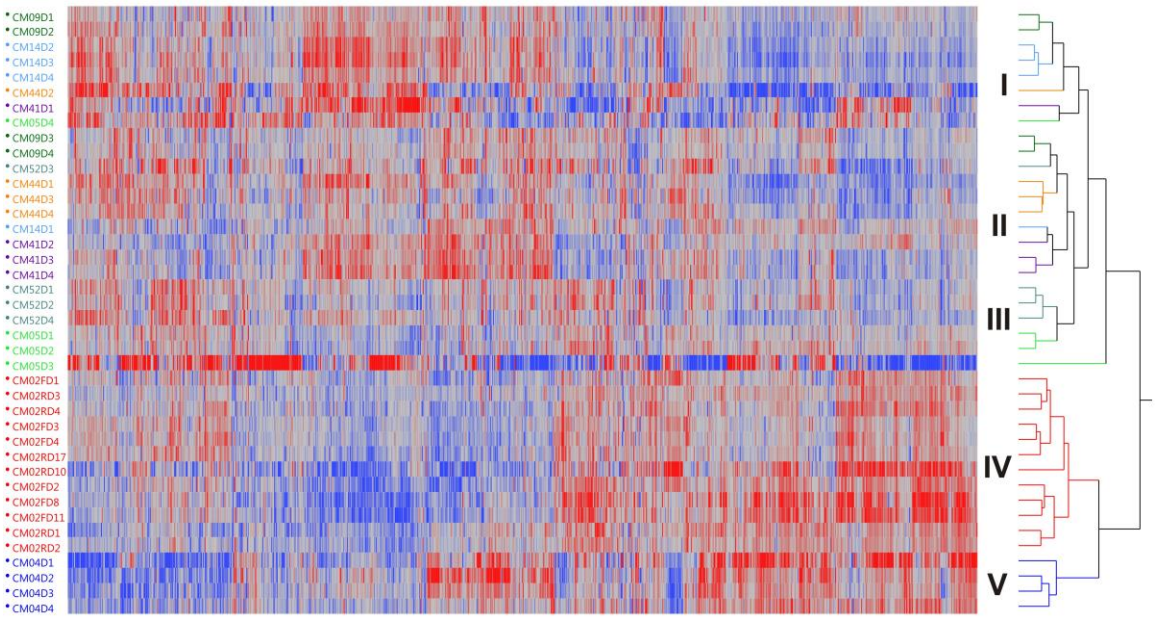

Figure 4.3. Two-way hierarchical clustering of complicated malaria samples. The heat map represents the overall profiles of eight patients with complicated malaria for 13,889 genes, standardized to z-scores across samples for each gene, such that blue indicates low transcript abundance and red high abundance. The heat map shows 5 clusters, cluster IV differentiates the pregnant patient, CM02, which clusters with sample CM04.

Table 4.2. Characteristics of study patients with complicated malaria *P. vivax* and *P. falciparum.*

Code	City	Gender	Age	Spp	Parasitemia	Clinical Criteria	Laboratory criteria	Hospital Stay (Days)
CM04	Barbacoas-Nariño	M	22	*Pf*	20800	Oral Intolerance	Renal and liver failure	4
CM05	Quibdó - Chocó	F	38	*Pf*	11640	Respiratory distress	Severe thrombocytopenia	4
CM02	Quibdó - Chocó	F	28	*Pf*	5200 4300	Respiratory distress (25 weeks pregnancy)	Severe Anemia Hb: 6,3	11 18
CM52	Alto Baudó-Chocó	M	3	*Pv*	5700	Respiratory distress	none	4
CM41	San José de Tadó-Chocó	F	49	*Pf*	142030	Oral intolerance jaundice	Severe Thrombocytopenia: 14000 hiper	4
CM44	Quibdó - Chocó	F	25	*Pv*	5742	Oral Intolerance macroscopic hemoglobinuria	Jaundice, Renal Injury, Severe Thrombocitopenya	5
CM14	Tumaco - Nariño	M	37	*Pf*	19609	Oral Intolerance	complication hepatica, Severe thrombocytopenia :29.000, hemoglobinuria.	5
CM09	Guacarí-Valle del Cauca	M	69	*Pf*	25119	Respiratory Distress	Severe Thrombocitopenia 15000, hyperparasitemia 25119	5

4.4.2 BIT expression reveals little change in gene expression for each patient at the four time points.

Despite the clustering of individual samples, I asked whether there are consistent changes in the gene expression profiles within individuals, by following the profiles of key signature genes from hospitalization to discharge. To do so, I computed Blood Informative Transcript (BIT) scores for ten major axes of blood gene expression, and drew radar plots, one for each individual infection (Fig. 4.4) for each individual. In each plot, the darkest line represents the earliest time point, and the lightest line the last time point, which should be close to the individual's healthy baseline. Each line links the eight summary axis scores for gene expression for each day, and each radar represents an individual patient. As explained in the methods, the Axes 1-7 can be regarded as measures of, respectively, T-cell, reticulocyte, B-cell, general cellular function, inflammation, an unknown immune function, and interferon response. Axis 10 is enriched for genes involved in cell cycle and mitosis.

Again consistent with Figure 4.3, the daily profiles are to a large extent superimposable and tend to have a patient-specific shape, indicating that they do not change dramatically from the day of recruitment until discharge. This is particularly notable for patients CM04 (Fig. 4.4A) and CM44 (Fig. 4.4B). Upon closer examination, however, there are some consistent trends. In each of patients CM05, CM09, CM14, CM41 and CM52 (Figs 4.4C-4.4G), Axes 5 and 7 are clearly lower at time point 3 and generally time point 4 as well, with time point one always the highest. In contrast, Axis 2 tends to be highest at time point 3. These results are easily interpreted as signs of inflammation and elevated

interferon signaling upon hospitalization, and of reticulocyte development associated with recovery from anemia after the peak of the complex malaria episode.

4.4.3 Failure of a pregnant woman to resolve infection corresponds with her variable transcriptome profile.

Sample CM02 (Fig. 4.4 H-I) does not show the same pattern, which may be related to her failure to resolve the complex malaria. Patient CM02 is a 28 year old Afro-Colombian pregnant woman from Quibdó-Chocó who was infected with *P. falciparum* malaria, and recruited on March 9, 2015. Her profiles cluster closest to those of CM04, a man infected also by *P. falciparum*, and are quite different from those of the other patients. They also showed more variability in the risk radars. The patient had 3 weeks of dyspnea (labored breathing), cough and lower extremity edema, with no history of fever. She is of normal weight, but showed severe hypertensive disorder possibly related to four previous pregnancies, with cardiopulmonary and hematologic compromise, and had experienced one other malaria episode in the preceding six months. After initial treatment, the patient had cleared the parasite successfully, but tested positive for *P. falciparum* malaria two weeks later with 4,300 parasites per ml of blood. Since this is considered a therapeutic failure, she further received a prophylactic treatment with Quinine plus Clindamycin (April 2 – April 10), which resolved the infection as assessed by thick blood smear and PCR (Table 4.3). She gave birth by induced delivery on May 2nd at 34 weeks and was discharged two days later.

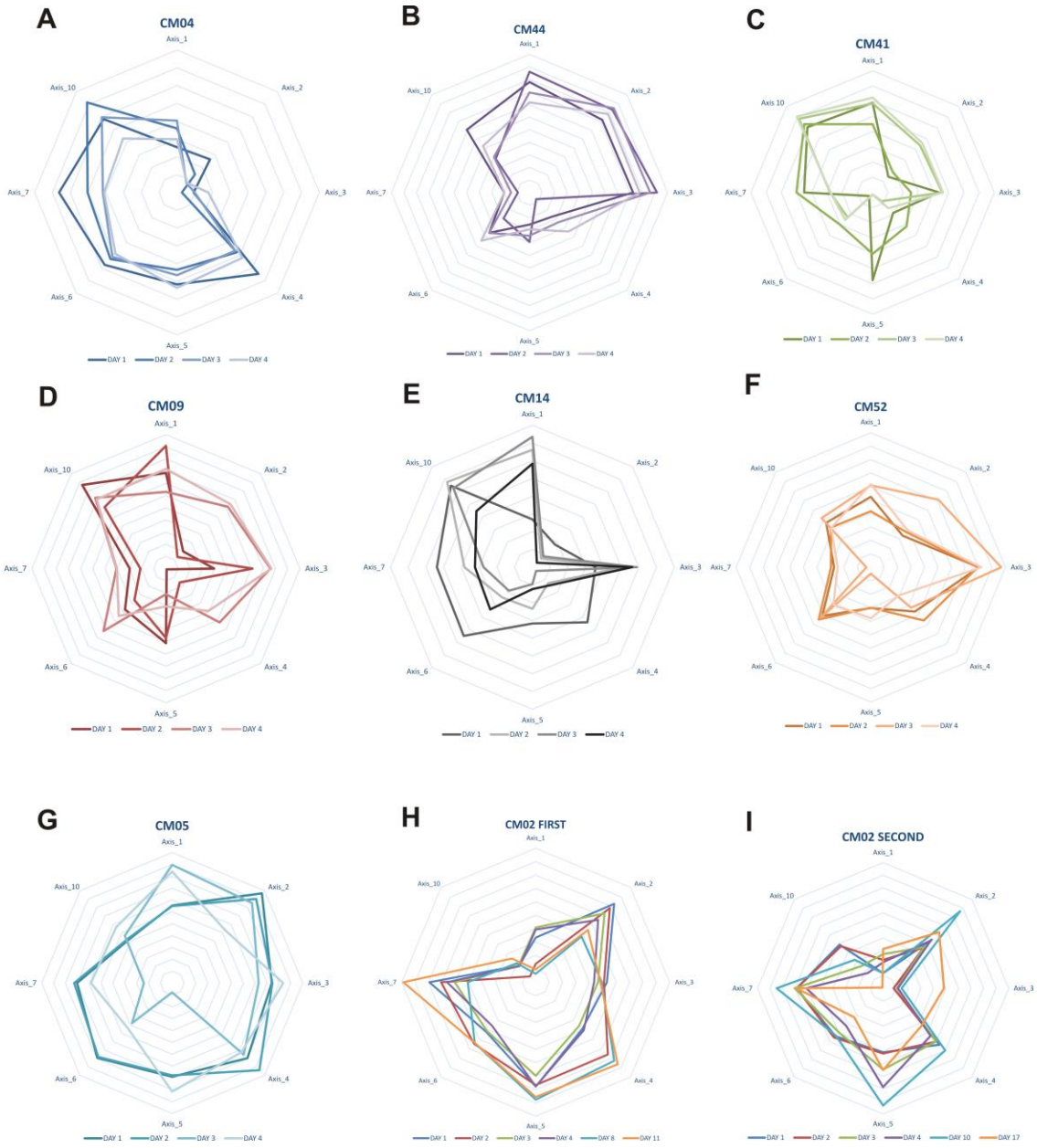

Figure 4.4. Complicated malaria transcriptional vectors. Each patient infection is represented as one radar plot, with CM02 having two successive infections shown in H, I. The lines on the radar plots join the Axis score (PC1 for 8 BIT across all samples in the 8 individuals) for consecutive time points, where earlier time points have darker colored lines.

Patient CM02's Axis 2 actually drops after several days in hospital, while her Axis 7 fluctuates, peaking at day 8, and her Axis 5 remains high throughout. She also has an increase in T-cell activity (Axis 1) at days 3 and 4, but this drops precipitously thereafter until the first malaria episode is resolved. Her fluctuating gene expression profiles continue in the second malaria episode, which is characterized by a consistently high interferon response (Axis 7), a bout of high mitotic activity (Axis 10, see below), and a precipitous decline in B-cell signaling (Axis 3) a few days into her relapse. It is not clear whether her profile at discharge represents her baseline state of gene expression, either, as it is actually unlike most of her other profiles. All of this indicates a deeply disturbed peripheral blood profile over a period of a month in this individual.

Table 4.3. Pregnant women CM02 parasite history and RNAseq sampling.

Day	Date	TBS	PCR	RNAseq
1	**3/8/2015**	Blood smear with parasites	-	YES
2	**3/9/2015**	5200	-	YES
3	**3/10/2015**	-	-	YES
4	**3/11/2015**	Negative	-	YES
5	**3/12/2015**	Negative	-	-
6	**3/13/2015**	Negative	-	-
7	**3/14/2015**	Negative	-	-
8	**3/15/2015**	Negative	-	YES
9	**3/16/2015**	Negative	-	-
10	**3/17/2015**	Negative	-	YES
	New Positive of malaria Infection			
1	**3/31/2015**	4300	-	YES
2	**4/1/2015**	10106	-	YES
3	**4/2/2015**	-	2100	YES
5	**4/3/2015**	680	300	YES
6	**4/4/2015**	600	4000	-
7	**4/5/2015**	400	2000	-
8	**4/6/2015**	120	-	-
9	**4/7/2015**	Negative	-	-
10	**4/8/2015**	Negative	-	-
11	**4/9/2015**	Negative	-	YES
12	**4/10/2015**	Negative	30	-
13	**4/11/2015**	Negative	-	-
14	**4/12/2015**	Negative	-	-
15	**4/13/2015**	Negative	-	-
16	**4/14/2015**	Negative	-	-
17	**4/15/2015**	Negative	-	-
18	**4/16/2015**	Negative	Negative	YES

4.4.4 Transcriptional changes associated with parasitemia

In order to assess whether the transcriptional changes associated with complication in malaria infection are related to those observed during standard malaria, I compared the first day of hospitalization complicated malaria samples with samples from the diagnosis day from the irradiated sporozoite project described in Chapter 3 (namely, vaccinated

individuals challenged with *P. vivax*). Degree of parasitemia captured one third of the variance for the first five principal components of transcriptional variation (Figure 4.5A), consistent with the findings of Idaghdour et al (2012) that parasite load is the major factor influencing transcript profiles in malarial infants in Benin, West Africa. Hierarchical clustering of the 250 Blood Transcript Module scores (BTMs; Li et al, 2015) confirms that much of this separation is between patients who were infected with the malaria parasite and individuals that were not infected (Fig. 4.5B). That is to say, the complicated malaria samples cluster with the volunteers who failed to mount an immune response and thus experienced parasitemia and showed mild malaria symptoms.

This analysis also confirms that the 7 Axes reported by Preininger et al (2013) cluster with BTM that have functions in the related components of immune function, and that Axis 10, which we only recently identified, is related to mitosis. Among the responses that differentiate infected and resistant (whether as a result of vaccination or natural Duffy negative protection) individuals are the interferon response, cell cycle/mitosis, and general metabolic processes, all of which are up-regulated in individuals with the infection, and B-cell and inflammation responses which are uniformly down-regulated in malarial subjects.

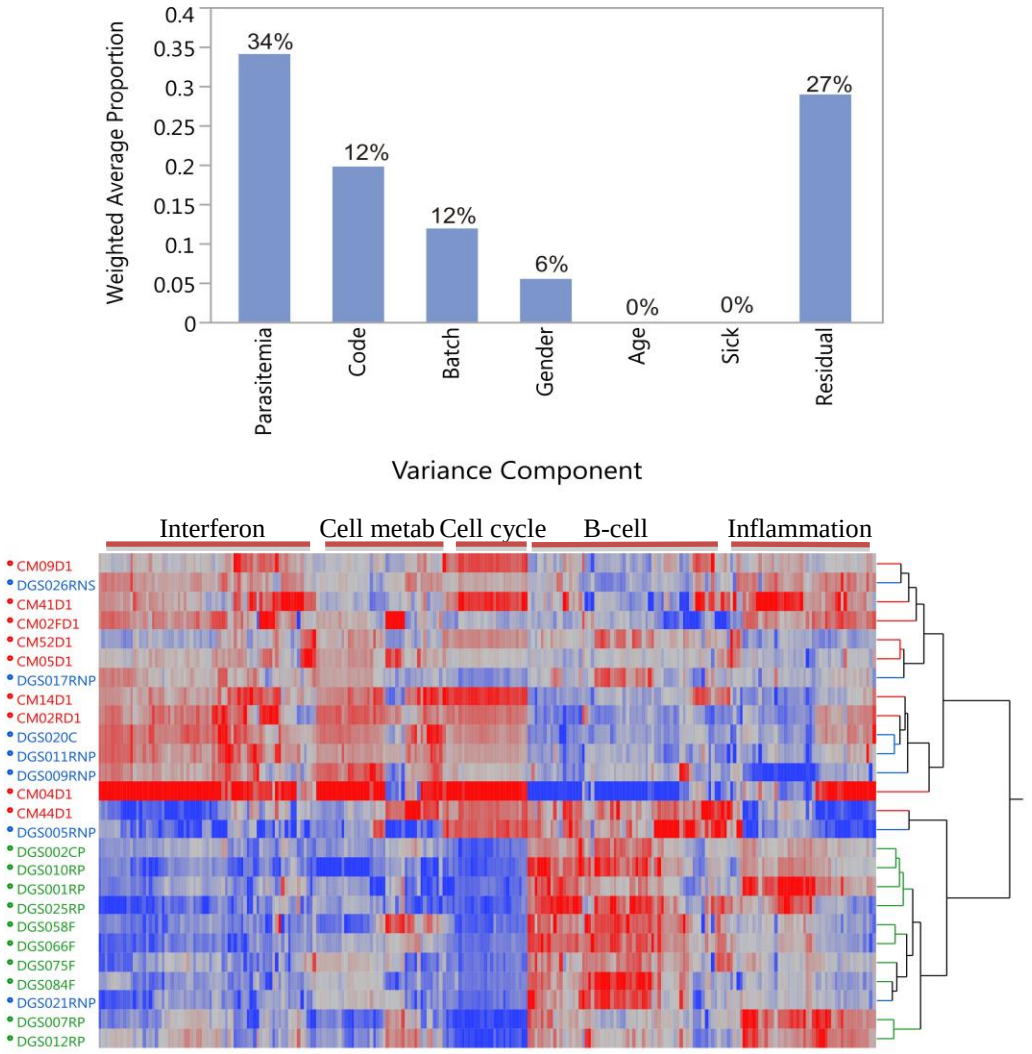

Figure 4.5. Gene profiling comparison between day one (hospitalization) of the complicated malaria study, and the diagnosis samples from irradiated sporozoite project described in Chapter 3. (A) Variance component analysis of gene expression profiles, showing the weighted average of the variance captured by the first five principal components. (B) Two-way hierarchical clustering of the 250 BTM and 8 Axis PC1 scores across all of the samples. The heat map represents the overall similarity whereby the profiles of the eight patients with complicated malaria (CM) cluster with the not protected (NP) vaccination study samples. Blue indicates low axis scores, and red high axis scores.

4.5 Discussion

In order to obtain a better understanding of the molecular events in patients who experience complicated malaria, a pilot study using RNASeq to longitudinally profile peripheral blood gene expression during recovery after initial hospitalization was conducted. Each of the eight patients clustered primarily as a group, indicating that any changes in gene expression over the time course of recovery are modest relative to baseline differences. We did not include control samples from several days after the patients were discharged from the hospital. Colborn et al. (2015) reported modulation of gene expression 7-10 days after recovery from a malaria infection, slightly longer than the period of recovery here, consistent with an earlier report suggesting that activation of a transcriptional response can continue after clearance of the infection (Schaecher et al. 2005).

Patient CM02, a pregnant woman with symptoms of preeclampsia but no history of fever, showed the most gene expression differentiation during the complicated malaria infection. Her delayed malaria diagnosis may have affected the observed responses, since progression of the infection before receiving treatment could be related to her inverted interferon and reticulocyte profiles relative to the other patients. However, her relapse infection was also unusual and it is further possible that the advanced pregnancy modulated her immune profile. Given the potential adverse impacts of malaria during pregnancy for both mother and child, further studies of more cases are warranted to document whether her situation is typical of pregnancy, or unique to her. It is probable

that such incidences of adverse trajectories of malaria resolution increase in communities with no access to appropriate medical services (Tobon et al 2006).

Comparison of the BTMs at the day of diagnosis for samples from the irradiated sporozoite project with the day one samples of the complicated malaria project, indicates overall similarity of individuals based on infection versus sterile immunity. A similar comparison with the discharge samples (typically day 4) failed to show clean separation of the complicated malaria patients with the malarial samples, indicating that the individuals had at least partially returned to normal. Concordant with previous analysis (Vasques and Tobon 2012), the interferon response was found up-regulated in individuals with malaria infection.

Two limitations of this study are (i) that we did not have healthy baseline samples, either before or after infection, against which the hospital profiles can be compared, and (ii) the sample size is too small to evaluate whether there is any association of particular modules of gene expression with specific symptoms. There does appear to be some variability among patients in terms of how much differential expression is observed across the 4-day resolution of complications. The BTM analysis indicates that hundreds of processes are disrupted, but also shows that these are embedded within the major axes of variation that dominate peripheral blood gene expression profiles (Preininger et al, 2013). This suggests that continuation of the project would be more informative if sampling could be specifically targeted to specific immune cell types, such as T-cells, B-cells, or neutrophils, each of which display patient-specific responses. Cell-type specific analysis

would be expected to overcome the contribution of variable abundance of the contributing cell types to the overall profiles obtained from RNA preserved from whole blood in Tempus tubes.

Our results provide the first evidence of host gene expression modulation in complicated malaria infections. There is some evidence gene expression differs subtly among samples infected with the two different species of *Plasmodium*, but there are also clearly commonalities to the profiles. Banchereau et al (2016) recently showed how longitudinal whole blood transcriptome profiling of 158 pediatric systemic lupus erythematosus patients can be used to classify eight subsets that they suggest might have different treatment responses. Lupus and malaria share many features of aberrant gene expression as well as prophylaxis, so my results can be seen as an initial exploration of the potential value of RNASeq for personalized medicine. Larger sample sizes will also be needed to support the bioinformatics that needs to be done before genomic profiling can be considered for integration as a component of personalized clinical diagnostics.

CHAPTER 5

5. CONCLUSIONS

In this thesis I have presented analyses of whole blood gene expression profiling in three different malaria settings with different outcomes of infection. The results provide useful information on the course of the immune response and pathological processes associated with infection with Plasmodium vivax in particular.

In Chapter II my experiments revealed that differential gene expression is particularly enhanced in naïve volunteers in comparison to semi-immune individuals at the time of malaria diagnosis. In the presence of chronic exposure, the host immune system moves toward an equilibrium where pathogen is tolerated by mounting a measured immune response, without requiring complete sterile immunity that would likely have a greater physiological impact on the infected individuals. This in turn implies that gene expression profiling of lymphocytes can be used to identify the type and duration of the immune signals that may be biomarkers for vaccine immunogenicity, and to establish how semi-immune exposure modifies their activation.

It would be interesting to evaluate purified CD4+T helper and Treg lymphocytes of semi-immune individuals and contrast them against naïve individuals. These cell types appear to mediate the quantitative modulation of memory responses, and are likely to be central to the clinical protection against malaria infection.

In chapter III I showed that differential expression is particularly prevalent in non-protected volunteers in comparison to protected individuals at the time of malaria diagnosis, providing a general picture of how the different arms of the immune system are engaged during the response to infection after immunization. Even though whole blood profiling does not generally have the resolution to define precisely which immune cell subtypes are most important, it provides an integrated picture of the combined immune response leading to protective immunity. Given sufficient funds and technical resources, it should preferably be supplemented with cellular immune profiling. Interestingly, my results implicate roles for regulatory T cells and dendritic cells in mediating the *P. vivax* irradiated sporozoite response, as has been observed by flow cytometry in the context of *P. falciparum*. Also noteworthy is the evidence for strongly enhanced short lived plasma cell gene expression in the susceptible individuals, since such a specific cell type would not normally be examined by flow cytometry. It will be interesting to learn whether these cells differentiate in the blood or are recruited from the bone marrow. Focused transcriptome profiling of sorted immune subsets in larger samples, combined with knowledge of differences between asymptomatic and complex malaria cases, should provide further insights that may be useful in enhancing vaccine development for malaria.

Blood Transcription Modules (BTM) emerged as a useful tool in blood analysis for evaluating host immune status and provided new insights into the immune repertoires

associated with malaria immunization. Combining multiple omics is critical to small “N”, human studies. Their integration can be driven by data mining or by knowledge models. Therefore describing in detail gene expression during the activation of the immune system against a malaria infection contributes greatly to the overall objective of developing a better malaria vaccine, since we can monitor the immune response in the blood rapidly and cost-effectively.

Finally, in chapter IV, my results provide the first evidence of host gene expression modulation in complicated malaria infections. There is some evidence that gene expression differs subtly among samples infected with the two different species of Plasmodium, but there are also clear commonalities to the profiles. These results demonstrate the potential value of RNASeq for studying the response of the host transcriptome of a malaria infection using leucocytes as markers of the severity and prognostics in malaria infections, reflecting various types of immune activity. All this can have implications for evaluation of new vaccines or treatments. These results can also lead to the identification of gene expression patterns that distinguish between mild and complicated malaria, which may be of use in the care provided following admission of complicated malaria patients to hospital.

To be able to continue the complicated malaria project I recommend (i) to also obtain healthy samples after a long period of the clearance of the infection, as these can be regarded as a patient’s baseline against which we can compare the aberrant hospital profiles and better identify specific patient-specific perturbations. This might offer insight

into the molecular source of the infection response that can be used to guide personalized health decisions, for example the means to suppress inflammation or promote the interferon response. It will also be essential (ii) to increase the sample size to be able to evaluate whether there is any association of particular modules of gene expression with specific symptoms, and also to differentiate responses to different species of Plasmodium, and different drug treatments.

From this dissertation I conclude that gene expression profiling is an extremely valuable tool with which to investigate the molecular features of the host's response to malaria immunization, malaria infection and clearance. It can offer insights to transcriptional mechanisms that impact both the pathogenesis of the infection and how individual people recover from infection. As a Colombian citizen, I am particularly hopeful that the new insights into the effectiveness of the irradiate sporozoite vaccine, particularly in women, might be used to help design adjuvants that may lead to further enhancements.

APPENDIX A

BTM THAT DIFERENTIATE PARASITE AND CLINICAL PROTECTION

BTMs -Absence of parasite (Protection)	***p*-value**
TBA M141	1.12E-04
complement activation I M112 0	2.96E-04
antigen presentation lipids and proteins M28	3.70E-04
activated dendritic cells M67	3.76E-04
proinflammatory dendritic cell. myeloid cell response M86 1	4.99E-04
RIG1 like receptor signaling M68	5.46E-04
viral sensing . immunity IRF2 targets network II M111 1	6.54E-04
viral sensing . immunity IRF2 targets network I M111 0	8.01E-04
chemokines and inflammatory molecules in myeloid cells M86 0	8.20E-04
TBA M121	1.00E-03
cell division M37 3	1.04E-03
enriched in activated dendritic cells I M119	1.20E-03
enriched in activated dendritic cells II M165	1.22E-03
putative targets of PAX3 M89 1	1.23E-03
type I interferon response M127	1.30E-03
antiviral IFN signature M75	1.38E-03
mismatch repair II M22 1	1.46E-03
mismatch repair I M22 0	1.51E-03
Axis 7	2.01E-03
proinflammatory cytokines and chemokines M29	2.01E-03
putative targets of PAX3 M89 0	2.18E-03
DNA repair M76	2.21E-03
cytoskeleton actin SRF transcription targets M145 0	2.38E-03
complement and other receptors in DCs M40	2.51E-03
innate antiviral response M150	2.52E-03
regulation of antigen presentation and immune response M5 0	3.55E-03
receptors. cell migration M109	3.56E-03
E2F1 targets Q4 M10 1	3.90E-03
cell cycle and transcription M4 0	5.42E-03
cell activation IL15 IL23 TNF M24	5.85E-03

Appendix A (Continued)

CD1 and other DC receptors M50	6.00E-03
mitotic cell cycle in stimulated CD4 T cells M4 9	6.23E-03
E2F1 targets Q3 M10 0	6.30E-03
cell division in stimulated CD4 T cells M4 6	7.55E-03
double positive thymocytes M126	7.89E-03
cytokines recepters cluster M115	8.97E-03
cell cycle III M103	1.15E-02
cell cycle II M4 10	1.17E-02
PLK1 signaling events M4 2	1.23E-02
mitotic cell cycle . DNA replication M4 4	1.34E-02
cell cycle I M4 1	1.36E-02
cell division stimulated CD4. T cells M46	1.45E-02
mitotic cell cycle in stimulated CD4 T cells M4 5	1.55E-02
chemokine cluster I M27 0	1.77E-02
mitotic cell division M6	2.08E-02
mitotic cell cycle M4 7	2.38E-02
Axis D	3.42E-02
cell division E2F transcription network M4 8	3.52E-02
C.MYC transcriptional network M4 12	4.21E-02
mitotic cell cycle in stimulated CD4 T cells M4 11	4.95E-02

BTMs - Clinical Protection	***p*-value**
plasma membrane cell junction M162 0	2.48E-05
chemokine cluster II M27 1	3.63E-05
TBA M70 1	1.64E-04
TBA M120	1.90E-04
targets of FOSL1.2 M0	1.95E-04
nuclear pore transport mRNA splicing. processing M143	3.19E-04
TBA M70 0	3.52E-04
T cell activation II M7 3	3.60E-04
nuclear pore complex M106 0	3.89E-04
enriched in T cells I M7 0	3.99E-04
leukocyte activation and migration M45	4.12E-04
inositol phosphate metabolism M129	4.24E-04
TBA M32 6	5.28E-04
CD28 costimulation M12	5.41E-04
TBA M148	5.53E-04
T cell differentiation Th2 M19	5.83E-04
TBA M161	6.39E-04

Appendix A (Continued)

viral sensing . immunity IRF2 targets network II M111 1	6.54E-04
TBA M32 5	6.64E-04
cytoskeletal remodeling M32 8	7.71E-04
platelet activation II M32 1	8.24E-04
CORO1A.DEF6 network II M32 4	8.30E-04
CORO1A.DEF6 network I M32 2	8.39E-04
platelet activation I M32 0	8.61E-04
Axis 6	8.93E-04
TBA M72 0	9.28E-04
TBA M125	9.70E-04
Axis 4	1.08E-03
KLF12 targets network M32 3	1.10E-03
TBA M137	1.10E-03
nuclear pore complex mitosis M106 1	1.54E-03
TBA M153	1.92E-03
phosphatidylinositol signaling system M101	2.02E-03
MAPK RAS signaling M100	2.12E-03
T cell activation and signaling M5 1	2.58E-03
T cell activation IV M52	2.60E-03
lymphocyte generic cluster M60	2.73E-03
Axis 1	2.75E-03
intracellular transport M147	3.02E-03
AP.1 transcription factor network M20	3.33E-03
TBA M151	3.44E-03
Axis-9	4.20E-03
enriched for ubiquitination M138	5.05E-03
TBA M72 1	5.34E-03
TBA M72 2	5.50E-03
TBA M128	6.43E-03
regulation of localization GO M63	8.52E-03
RA. WNT CSF receptors network monocyte M23	8.57E-03
Axis 8	1.00E-02
amino acid metabolishm and transport M154 0	1.20E-02
signal transduction plasma membrane M82	1.30E-02
TBA M174	1.44E-02
enriched in cell cycle M167	1.52E-02
cell cycle ATP binding M144	1.59E-02

REFERENCES

Akpogheneta OJ, Duah NO, Tetteh KK, Dunyo S, Lanar DE, et al. (2008). Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. Infect Immun. 76(4): 1748-1755.

Alexander DH, Novembre J, Lange K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 19:1655-1664.

Anders S, Pyl PT, Huber W. (2015). HTSeq a Python framework to work with high-throughput sequencing data. Bioinformatics. 31(2):166-169.

Anstey NM, Russell B, Yeo TW, Price RN (2009). The pathophysiology of vivax malaria. Trends Parasitol. 25:220-227.

Amino R, Thiberge S, Martin B, Celli S, Shorte S, et al. (2006). Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. Nat Med. 12(2): 220-224.

Ansong D, Asante KP, Vekemans J, Owusu SK, Owusu R, et al. (2011). T cell responses to the RTS,S/AS01(E) and RTS,S/AS02(D) malaria candidate vaccines administered according to different schedules to Ghanaian children. PLoS One 6(4): e18891.

Arévalo-Herrera M, Forero-Peña DA, Rubiano K, Gómez-Hincapie J, Martínez NL, et al. *Plasmodium vivax* Sporozoite Challenge in Malaria-Naïve and Semi-Immune Colombian Volunteers. PLoS One. 2014;9:e99754.

Ayers JS, Schnieder DS. (2012). Tolerance of infections. Annu Rev Immunol. 30:271-294.

Bassat Q, Alonso PL. (2011) Defying malaria: Fathoming severe *Plasmodium vivax* disease. Nature Medicine. 17: 48–49.

Banchereau R, Hong S, Cantarel B, Baldwin N, Baisch J. et al. (2016) Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. Cell. 165(6):1548-1550.

Byakika-Kibwika P, Ndeezi IG, Kamya MR. (2009) Health care related factors associated with severe malaria in children in Kampala, Uganda. Afr Health Sci. 9(3): 206–210.

Bojang KA, Olodude F, Pinder M, Ofori-Anyinam O, Vigneron L. et al. (2005). Safety and immunogenicty of RTS,S/AS02A candidate malaria vaccine in Gambian children.Vaccine. 23(32): 4148-4157.

Bouharoun-Tayoun HC, Oeuvray F, Lunel F, Druilhe P. (1995). Mechanisms underlying the monocyte-mediated antibody-dependent killing of Plasmodium falciparum asexual blood stages. J Exp Med. 182(2): 409-418.

Bunn AR, Escombe M, Armstrong, Whitty CJ, Doherty JF. (2004) Falciparum malaria in malaria-naive travellers and African visitors. Q J Med. 97:645-649.

Chaparro-Narváez PE, Lopez-Perez M, Rengifo LM, Padilla J, Herrera S. et al. (2016). Clinical and epidemiological aspects of complicated malaria in Colombia, 2007–2013. Malar J. 15: 269.

Chaussabel D, Quinn C, Shen J, Patel P, Glaser C. et al. (2008) A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity 29: 150–164.

Chen J, Bardes EE, Aronow BJ, Jegga AG. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37:305-311.

Clark IA, Budd AC, Alleva LM, Cowden WB. (2006). Human malarial disease: a consequence of inflammatory cytokine release. Malar J. 5: 85.

Cloonan N, Grimmond SM. (2008). Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol. 9:234.

Clyde DF, McCarthy VC, Miller RM, Hornick RB. (1973). Specificity of protection of man immunized against sporozoite-induced falciparum malaria. Am J Med Sci. 266:398-401.

Clyde DF, Most H, McCarthy VC, Vanderberg JP. (1973). Immunization of man against sporozite-induced falciparum malaria. Am J Med Sci. 266(3): 169-77.

Clyde DF, McCarthy VC, Miller RM, Woodward WE. (1975). Immunization of man against falciparum and vivax malaria by use of attenuated sporozoites. Am J Trop Med Hyg. 24:397-401.

Colborn JM, Ylöstalo JH, Koita OA, Cissé OH, Krogstad DJ. (2015) Human Gene Expression in Uncomplicated Plasmodium falciparum Malaria. J Immunol Res. 2015:162639.

Collins WE, Contacos PG. (1972). Immunization of monkeys against *Plasmodium cynomolgi* by X-irradiated sporozoites. Nat. New Biol. 236:176-177.

Cowman AF, Crabb BS. (2006) Invasion of red blood cells by malaria parasites. Cell. 124(4): 755-766.

Cox-Singh, J, Davis TM, Lee KS, Shamsul SS, Matusop A. et al. (2008). *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. Clin. Infect. Dis. 46:165-171.

Desai M, Kuile FO, Nosten F, McGready R, Asamoa K, et al. (2007). Epidemiology and burden of malaria in pregnancy. Lancet Infect Dis 7: 93-104.

Doolan D, Dobano C, Baird JK. (2009). Acquired immunity to malaria. Clin Microbiol Rev 22(1): 13-36.

Dunachie SJ, Walther M, Vuola JM, Webster DP, Keating SM. et al. (2006). A clinical trial of prime-boost immunisation with the candidate malaria vaccines RTS,S/AS02A and MVA-CS. Vaccine. 24: 2850-2859.

Egan JE, Hoffman SL, Haynes JD, Sadoff JC, Schneider I.et al. (1993). Humoral immune responses in volunteers immunized with irradiated *Plasmodium falciparum* sporozoites. Am. J. Trop. Med. Hyg. 49:166-173.

Epstein JE, Tewari K, Lyke KE, Sim BK, Billingsley PF. et al. (2011) Live attenuated malaria vaccine designed to protect through hepatic $CD8^+$ T cell immunity. Science. 334(6055):475-80.

Engwerda CR, Beattie L, Amante FH. (2005). The importance of the spleen in malaria. Trends Parasitol. 21(2): 75-80.

Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B et al. (2010) Shrinking the malaria map: progress and prospects. Lancet. 376: 1566-1578.

Finney OC, Riley EM, Walther M. (2010). Regulatory T cells in malaria--friend or foe?. Trends Immunol. 31(2): 63-70.

Frederich M, Dogne JM, Angenot L, De Mol P. (2002). New trends in anti-malarial agents. Curr Med Chem. 9(15): 1435-1456.

Frevert U, Nardin E. (2005). Arrest in the liver--a genetically defined malaria vaccine?. N Engl J Med 352(15): 1600-1602.

Garcon N, Heppner DG, Cohen J. (2003). Development of RTS,S/AS02: a purified subunit-based malaria vaccine candidate formulated with a novel adjuvant. Expert Rev Vaccines. 2(2): 231-238.

Gazzinelli RT, Kalantari P, Fitzgerald KA, Golenbock DT. (2014). Innate sensing of malaria parasites. Nat Rev Immunol. 14:744-757.

Genton B, D'Acremont V, Rare L, Baea K, Reeder JC. et al. (2008). Plasmodium vivax and Mixed Infections Are Associated with Severe Malaria in Children: A Prospective Cohort Study from Papua New Guinea. PLoS Medicine. 5(6): e127.

Gonçalves RM, Scopel KKG, Bastos MS, Ferreira MU. (2012). Cytokine balance in human malaria: does Plasmodium vivax elicit more inflammatory responses than Plasmodium falciparum? PLoS One. 7:e44394.

Gueirard PJ, Tavares S, Thiberge F, Bernex T, Ishino et al. (2010). Development of the malaria parasite in the skin of the mammalian host. Proc Natl Acad Sci USA. 107(43): 18640-18645.

Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL. (2008) The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide. PLoS Med. 5:e38.

Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP. et al. (2010) The international limits and population at risk of Plasmodium vivax transmission in 2009. PLoS Negl Trop Dis. 4:e774.

Greenwood BM, Fidock DA, Kyle DE, Kappe SH, Alonso PL. et al. (2008) Malaria: progress, perils, and prospects for eradication. J Clin Investig. 118: 1266–1276.

Grupo de vigilancia, Instituto Nacional de Salud SIVIGILA. (2014) Informe de evento Malaria 2014.

Gwadz RW, Cochrane AH, Nussenzweig V, Nussenzweig RS. (1979). Preliminary studies on vaccination of rhesus monkeys with irradiated sporozoites of *Plasmodium knowlesi* and characterization of surface antigens of these parasites. Bull. WHO 57(1):165–73.

Halliley JL, Tipton CM, Liesveld J, Rosenberg AF, Darce J. et al. (2015). Long-lived plasma cells are contained within the CD19. Immunity. 43:132-45.

Herrera S, Fernández O, Manzano MR, Murrain B, Vergara J. et al. (2009). Case Report: Successful Sporozoite Challenge Model in Human Volunteers with *Plasmodium vivax* Strain Derived from Human Donors. The American Journal of Tropical Medicine and Hygiene. 81(5), 740–746.

Herrera S, Fernandez OL, Vera O, Cardenas W, Ramirez O. et al. (2011). Phase I safety and immunogenicity trial of *Plasmodium vivax* CS derived long synthetic peptides adjuvanted with montanide ISA 720 or montanide ISA 51. Am J Trop Med Hyg. 84(2): 12-20.

Hoffman SL, Billingsley PF, James E, Richman A, Loyevsky M, et al. (2010). Development of a metabolically active, non-replicating sporozoite vaccine to prevent Plasmodium falciparum malaria. Hum Vaccin. 6(1): 97-106.

Idaghdour Y, Czika W, Shianna KV, Lee SH, Visscher PM, et al. (2010). Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat Genet 42: 62-7.

Idaghdour Y, Quinlan J, Goulet JP, Berghout J, Gbeha E. et al. (2012) Evidence for additive and interaction effects of host genotype and infection in malaria. Proc Natl Acad Sci. 109:16786-16793.

Jagannathan P, Nankya F, Stoyanov C, Eccles-James I, Sikyomu E. et al. (2015) IFNγ Responses to Pre-erythrocytic and Blood-stage Malaria Antigens Exhibit Differential Associations With Past Exposure and Subsequent Protection. J Infect Dis. 211(12):1987-1996.

Karunaweera ND, Grau GE, Gamage P, Carter R, Mendis KN. (1992) Dynamics of fever and serum levels of tumor necrosis factor are closely associated during clinical paroxysms in Plasmodium vivax malaria. Proc Natl Acad Sci USA. 89:3200-3203.

Kirchgatter K, Portillo Hdel A. (2002) Association of severe noncerebral Plasmodium falciparum malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. Mol Med. 8(1):16-23.

Kochar DK, Das A, Kochar SK, Saxena V, Sirohi P, et al. (2009). Severe *Plasmodium vivax* malaria: a report on serial cases from Bikaner in northwestern India. Am J Trop Med Hyg. 80: 194–198.

Komisar JL. (2007). Malaria vaccines. Front Biosci. 12: 3928-3955.

Kotepui M, Phunphuech B, Phiwklam N, Chupeerach C, Duangmano S. (2014). Effect of malarial infection on haematological parameters in population near Thailand-Myanmar border. Malar J. 13:218.

Krzych U, Zarling S, Pichugin A. (2014). Memory T cells maintain protracted protection against malaria. Immunol Lett. 161:189-195.

Kukurba KR, Parsana P, Balliu B, Smith KS, Zappala Z. et al. (2016). Impact of the X Chromosome and sex on regulatory variation. Genome Res. 26(6):gr.197897.115.

Laishram DD, Sutton PL, Nanda N, Sharma VL, Sobti RC, Carlton JM, et al. (2012). The complexities of malaria disease manifestations with a focus on asymptomatic malaria. Malar J.11:29.

Lamikanra AA, Brown D, Potocnik A, Casals-Pascual C, Langhorne J, et al. (2007) Malarial anemia: of mice and men. Blood.110:18-28.

Langhorne J, Ndungu FM, Sponaas AM, Marsh K. (2008). Immunity to malaria: more questions than answers. Nat Immunol 9(7): 725-732.

Landau I, Chabaud AG, Mora-Silvera E, Coquelin F, Boulard Y. et al. (1999). Survival of rodent malaria merozoites in the lymphatic network: potential role in chronicity of the infection. Parasite. 6(4): 311-322.

Li C, Seixas E, Langhorne J. (2001). Rodent malarias: the mouse as a model for understanding immune responses and pathology induced by the erythrocytic stages of the parasite. Med Microbiol Immunol.189(3): 115-126.

Lindner SE, Miller JL, Kappe SH. (2012). Malaria parasite pre-erythrocytic infection: preparation meets opportunity. Cell Microbiol. 14(3): 316-324.

Li S, Rouphael N, Duraisingham S, Romero-Steiner S, Presnell S. et al. (2014). Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. Nat Immunol. 15(2):195–204.

Li S, Todor A, Luo R. (2016). Blood transcriptomics and metabolomics for personalized medicine. Computational and Structural Biotechnology Journal. 14:1-7.

Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ et al. (2014) African origin of the malaria parasite *Plasmodium vivax*. Nat. Commun. 5, 3346.

Lou J, Lucas R, Grau GR. (2001). Pathogenesis of cerebral malaria: recent experimental data and possible applications for humans. Clin Microbiol Rev. 14(4): 810-820.

Love MS, Millholland MG, Mishra S, Kulkarni S, Freeman KB. et al. (2012). Platelet Factor 4 activity against P. falciparum and its translation to nonpeptidic mimics as antimalarials. Cell Host Microbe. 12:815-823.

Lyke KE, Fernández-Viňa MA, Cao K, Hollenbach J, Coulibaly D, et al. (2011). Association of HLA alleles with Plasmodium falciparum severity in Malian children. Tissue Antigens. 77(6):562-571.

Mackinnon MJ, Marsh K. (2010). The selection landscape of malaria parasites. Science. 328(5980): 866-871.
Marsh K, Forster D, Waruiru C, Mwangi I, Winstanley M. et al. (1995) Indicators of life-threatening malaria in African children. N Engl J Med 332:1399-1404.

Matuschewski, K. (2006). Getting infectious: formation and maturation of *Plasmodium* sporozoites in the Anopheles vector. Cell Microbiol. 8(10): 1547-1556.

Matuschewski K, Hafalla JC, Borrmann S, Friesen J. (2011). Arrested *Plasmodium* liver stages as experimental anti-malaria vaccines. Hum Vaccin. 7:16-21.

McCall MB, Sauerwein RW. (2010). Interferon-gamma--central mediator of protective immune responses against the pre-erythrocytic and blood stage of malaria. J Leukoc Biol. 88:1131–1143.

McConkey SJ, Reece WH, Moorthy VS, Webster D, Dunachie S. et al. (2003). Enhanced T-cell immunogenicity of plasmid DNA vaccines boosted by recombinant modified vaccinia virus Ankara in humans Nat Med. 9(6):729–735.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K. et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 20:1297-303.

McKenzie FE, Prudhomme WA, Magill AJ, Forney JR, Permpanich B. (2005). et al. White Blood Cell Counts and Malaria. J Infect Dis. 192:323–330.

McMorran BJ, Marshall VM, Drysdale GC, Shabbar M, Gordon K. et al. (2009). Platelets Kill Intraerythrocytic Malarial Parasites and Mediate Survival to Infection. Science. 323(5915):797-800.

McMorran BJ, Wieczorski L, Drysdale KE, Chan JA, Huang HM et al. (2012). Platelet factor 4 and Duffy antigen required for platelet killing of Plasmodium falciparum. Science. 338:1348–51.

Mecham BH, Nelson PS, Storey JD. (2010). Supervised normalization of micro- arrays. Bioinformatics. 26:1308-1315.

Mellouk S. (1990). Protection against malaria induced by irradiated sporozoites. Lancet. 335(8691):721.

de Mendonca VR, Goncalves MS, Barral-Netto M. (2012). The host genetic diversity in malaria infection. J Trop Med. 940616.

Miller LH, Baruch DI, Marsh K, Doumbo OK. (2002). The pathogenic basis of malaria. Nature. 415(6872): 673-679.

MinSalud. Ministerio de la Protección Social. Colombia. (2010). Guía de atención clínica de malaria Ministerio de la Protección Social, Bogotá. 132.

Moreno A, Joyner C. (2015) Malaria vaccine clinical trials: what's on the horizon. Curr Opin Immunol. 35:98-106.

MVI-PATH. (2013). Malaria Vaccine Technology Roadmap 2013. PATH Malaria Vaccine Initiative.

Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S et al. (2011) Systems biology of seasonal influenza vaccination in humans. Nat Immunol.12:786–787.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W. et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nature Methods 12, 453–457.

Oladipo OO, Wellington OA, Sutherland CJ. (2015) Persistence of chloroquine-resistant haplotypes of Plasmodium falciparum in children with uncomplicated Malaria in Lagos, Nigeria, four years after change of chloroquine as first-line antimalarial medicine. Diagnostic Pathology. 10:41.

Okell LC, Ghani AC, Lyons E, Drakeley CJ. (2009). Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis. J Infect Dis. 200(10): 1509-1517.

Pacheco MA, Lopez-Perez M, Vallejo AF, Herrera S, Arevalo-Herrera M. et al. (2016) Multiplicity of infection and disease severity in Plasmodium vivax. PLoS Negl Trop Dis. 10:e0004355.

Patel DN, Pardeep D, Surti MM, Agerwal SB. (2003) Clinical manifestations of complicated malaria: an overview. J Indian Acad Clin Med. 4: 323-31.

Perez-Mazliah D, Langhorne J. (2014). CD4 T-Cell Subsets in Malaria: TH1/TH2 Revisited. Frontiers in Immunology. 5:671.

Perkins DJ, Were T, Davenport GC, Kempaiah P, Hittner JB. et al. (2011). Severe malarial anemia: innate immunity and pathogenesis. Int J Biol Sci. 7(9): 1427-1442.

Portugal S, Moebius J, Skinner J, Doumbo S, Doumtabe D. et al. (2014). Exposure-Dependent Control of Malaria-Induced Inflammation in Children. PLoS Pathog. 10:e1004079.

Preininger M, Arafat D, Kim J, Nath AP, Idaghdour Y. et al. (2015) Blood-informative transcripts define nine common axes of peripheral blood gene expression. PLoS Genet. 9:e1003362.

Price RN, Tjitra E, Guerra CA, Yeung S, White NJ et al. (2007). Vivax malaria: neglected and not benign. Am J Trop Med Hyg 77: 79-87.

Pulendran B. (2014). Systems vaccinology: probing humanity's diverse immune systems with vaccines. Proc Natl Acad Sci USA. 111:12300-12306.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR. et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics. 81(3):559-75.

Regules JA, Cummings JF, Ockenhouse C. (2011). The RTS,S vaccine candidate for malaria. Expert Rev Vaccines. 10: 589–99.

Richie TL, Billingsley PF, Sim BK, Epstein JE, Lyke KE. et al. (2015) Progress with *Plasmodium falciparum* sporozoite (PfSPZ)-based malaria vaccines. Vaccine. 33(52):7452-61.

Rieckmann KH, Carson PE, Beaudoin RL, Cassells IS, Sell KW. (1974) Sporozoite induced immunity in man against an Ethiopian strain of Plasmodium falciparum. Trans R Soc Trop Med Hyg. 68:258-9.

Riley EM, Stewart VA. (2013). Immune mechanisms in malaria: new insights in vaccine development. Nat Med 19(2): 168-178.

Robinson MD, McCarthy DJ, Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26 (1): 139-140.

Rodriguez JC, Uribe GA, Araujo RM, Narvaez PC, Valencia SH. (2011). Epidemiology and control of malaria in Colombia. Mem Inst Oswaldo Cruz. 106(1):114-122.

Rogerson SJ, Hviid L, Duffy PE, Leke RF, Taylor DW. (2007). Malaria in pregnancy: pathogenesis and immunity. Lancet Infect. Dis. 7:105-117.

Rogerson SJ, Mwapasa V, Meshnick SR. (2007) Malaria in pregnancy: linking immunity and pathogenesis to prevention. Am J Trop Med Hyg. 77:14-22.

Rojas-Peña ML, Vallejo A, Herrera S, Gibson G, Arévalo-Herrera M. (2015) Transcription Profiling of Malaria-Naïve and Semi-immune Colombian Volunteers in a *Plasmodium vivax* Sporozoite Challenge. PLoS Negl Trop Dis. 9(8): e0003978.

Rosenberg R, Wirtz RA, Schneider I , Burge R. (1990). An estimation of the number of malaria sporozoites ejected by a feeding mosquito. Trans R Soc Trop Med Hyg. 84(2): 209-212.

RTS,S Clinical Trials Partnership. (2015). Lancet. 386 (9988): 31-45.

Schaecher K, Kumar S, Yadava A, Vahey M, Ockenhouse CF. (2005) Genome-wide expression profiling in malaria infection reveals transcriptional changes associated with lethal and nonlethal outcomes. Infect Immun. 73: 6091–6100.

Sachs J, Malaney P. (2002) The economic and social burden of malaria. Nature. 415: 680-685.

Schneider J, Dubovsky F, Tierney E, Williams J, G HD. et al. (2006). A clinical trial of prime-boost immunisation with the candidate malaria vaccines RTS,S/AS02A and MVA-CS. Vaccine. 24: 2850-2859.

Schellenberg D, Menendez C, Kahigwa E, Font F, Galindo C. et al. (1999) African children with malaria in an area of intense Plasmodium falciparum transmission: features on admission to the hospital and risk factors for death. Am J Trop Med Hyg. 61:431-438.

Schofield L, Grau GE. (2005). Immunological processes in malaria pathogenesis. Nat Rev Immunol 5(9): 722-735.

Serghides L, Smith TG, Patel SN, Kain KC. (2003). CD36 and malaria: friends or foes? Trends Parasitol 19(10): 461-469.

Spurgeon SL, Jones RC, Ramakrishnan R. (2008). High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. PLoS One. 3:e1662.

Stanisic DI, Javati S, Kiniboro B, Lin E, Jiang J. et al. (2013). Naturally acquired immune responses to P. vivax Merozoite Surface Protein 3α and Merozoite Surface Protein 9 are associated with reduced risk of *P. vivax* malaria in young Papua New Guinean children. PLoS Negl Trop Dis. 7:e2498.

Stoute JA, Kester KE, Krzych U, Wellde BT, Hall T. et al. (1998). Long-term efficacy and immune responses following immunization with the RTS,S malaria vaccine. J Infect Dis. 178:1139 -1144.

Stevenson MM, Riley EM. (2004). Innate immunity to malaria. Nature Reviews. Immunology 4(3): 169-180.

Tabassum R, Sivadas A, Agrawal V, Tian H, Arafat D et al. 2015. Omic personality: implications of stable transcript and methylation profiles for personalized medicine. Genome Medicine. 27(1):88.

The 1000 Genomes Project Consortium. (2015) A global reference for human genetic variation. Nature. 526, 68–74.

The International HapMap Consortium. (2003). The International HapMap Project. Nature. 18789 - 18796.

Tishkoff SA, Verrelli BC. (2003). Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet. 4:293-340.

Tobón AC, Giraldo CS, Pineros JG, Arboleda MN, Blair ST. et al. (2003). The epidemiology of complicated falciparum malaria: case and controls study in Tumaco and Turbo, Colombia, 2003. Rev. bras. epidemiol. 9(3): 1980-5497.

Tobón A. Signos de peligro en el paciente con malaria. Biomédica. 2009, 29(2): 320-329.

Tobón A, Piñeros J, Blair S, Carmona J. (2006). Clínica de la malaria complicada debida a P. falciparum. Estudio de casos y controles en Tumaco y Turbo (Colombia). Iatreia. 19:339-55.

Tjitra, E, Anstey, Sugiarto NM, Warikar P, Kenangalem N. et al. (2008). Multidrug-Resistant *Plasmodium vivax* Associated with Severe and Fatal Malaria: A Prospective Study in Papua, Indonesia. PLoS Medicine. 5(6), e128.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D. et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 7:562-578.

Tsang JS, Schwartzberg PL, Kotliarov Y, Biancotto A, Xie Z. et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. (2014) Cell.157:499–513.

Vásquez AM, Tobón A. (2012) Mecanismos de patogenia en la malaria por *Plasmodium falciparum*. Biomédica. 32(1):106-120.

Webster DP, Dunachie S, Vuola JM, Berthoud T, Keating S. et al. (2005). Enhanced T cell-mediated protection against malaria in human challenges by using the recombinant poxviruses FP9 and modified vaccinia virus Ankara. Proc Natl Acad Sci U S A. 102(13): 4836–4841.

Weiss WR. (1990). Host-parasite interactions and immunity to irradiated sporozoites. Immunol Lett. 25(1-3):39-42.

Weppelmann TA, Carter TE, Chen Z, Fricken ME, Victor YS et al. (2013). High frequency of the erythroid silent Duffy antigen genotype and lack of *Plasmodium vivax* infections in Haiti. Malar J. 12:30.

Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC. et al. (2003). Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci USA. 100:1896-1901.

WHO Severe falciparum malaria. (2000). World Health Organization, communicable diseases cluster. Trans R Soc Trop Med Hyg. 94(1):1-90.

WHO. (2015) World Malaria Report 2015. Geneva: World Health Organization.

Wright GJ, Rayner JC. (2014). Plasmodium falciparum Erythrocyte Invasion: Combining Function with Immune Evasion. PLoS Pathog. 10: e1003943.

Wykes MN, Kay JG, Manderson A, Liu XG, Brown DL. et al. (2011). Rodent blood-stage *Plasmodium* survive in dendritic cells that infect naive mice. Proc Natl Acad Sci U S A 108(27): 11205-11210.

Yilmaz B, Portugal S, Tran TM, Gozzelino R, Ramos S. et al. (2014). Gut microbiota elicits a protective immune response against malaria transmission. Cell. 6:1277-1289.

Zheng H, Tan Z, Xu W. (2014). Immune Evasion Strategies of Pre-Erythrocytic Malaria Parasites. Mediators of Inflammation. Mediators Inflamm. Article ID 362605.