In vivo Transcriptome Analysis of *Plasmodium falciparum* Clinical Isolates: Glimpses into Molecular Events in Complicated Malaria

THESIS

Submitted in partial fulfillment

Of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

AMIT KUMAR SUBUDHI

Under the Supervision of

Prof. ASHIS KUMAR DAS

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

PILANI (RAJASTHAN) INDIA

2014

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

PILANI, RAJASTHAN, INDIA

CERTIFICATE

This is to certify that the thesis entitled "***In vivo* Transcriptome Analysis of *Plasmodium falciparum* Clinical Isolates: Glimpses into Molecular Events in Complicated Malaria"** and submitted by Amit Kumar Subudhi, ID. No. 2007PHXF423P for award of Ph.D. Degree of the institute embodies original work done by him under my supervision.

Signature (Supervisor) :

Name (Supervisor) : Ashis Kumar Das, PhD

Designation : Professor
Department of Biological Sciences
BITS, Pilani, Pilani Campus

Date :

Dedicated to My Parents & to Those Who have Lost their Life Due to Malaria

Amit Kumar Subudhi

Acknowledgements

Dream and work hard to make it a reality. This is the mantra that has driven me to achieve many goals in my life. This thesis represents one such goal with hard work of near about seven years. My experience in BITS, Pilani, Pilani Campus has been nothing short of amazing and is an amalgamation of successes, failures, joys and sorrows. This phase of life during my thesis has taught me lessons that will allow me not only to be a complete scientist but also a complete human being. I am not in a planet where I am the only leaving creature moving in this journey. This thesis is in its present shape and through to completion as result of ideas, encouragement and support of many individuals who I wish to acknowledge.

First and foremost, I would like to express my deepest gratitude and special appreciation to my supervisor ***Prof. Ashis K. Das****, Department of Biological Sciences, a talented teacher and passionate scientist for his excellent guidance, support, encouragement and advice without which this work would not have been possible. I also thank Prof. Das for appreciating my research strengths and patiently encouraging to improve in my weaker areas. I am indebted and thankful for the opportunity he offered me by accepting me as one of his PhD student. I am proud to say my experience under his guidance was intellectually exciting and fun, and has energized me to continue in academic research. I sincerely hope I continue to have opportunities to interact with* ***Prof. Das*** *for the rest of my research career.*

I gratefully acknowledge ***Prof. Sanjeev Kumar*** *and* ***Dr. Vishal Saxena****, Members of my Doctoral Advisory Committee (DAC) for their valuable advice, constructive criticism and helpful suggestions and comments during my thesis progress.*

I take this opportunity to sincerely acknowledge ***Prof. B.N. Jain****, Vice-Chancellor, BITS, Pilani,* ***Prof. G. Raghurama,*** *Director, BITS, Pilani,* ***Prof. R.N. Saha****, Former Deputy Director,* ***Prof. Ravi Prakash*** *and* ***Prof. Ashis K Das****, Former Deans, Research and Consultancy Division,* ***Prof. S.K. Verma****, Dean, Academic Research Division,* ***Prof. Hemant R Jadav****, Associate Dean, Academic Research Division and Nucleus Members of BITS, Pilani, Pilani Campus for Academic, Administrative supports.*

This thesis would not have been completed without the close collaboration with Sardar Patel Medical College, Bikaner. I am extremely indebted to two physicians, ***Prof. Dhanpat K Kochar****, Former Head, Cerebral Malaria Research Center, S.P. Medical College, Bikaner and Visiting Professor-Medical Research, Rajasthan University Health Sciences, Jaipur and* ***Prof. Sanjay Kumar Kochar****, Associate Professor, Department of Medicine and Head, Cerebral Malaria Research center, S.P. Medical College, Bikaner for providing invaluable malaria infected blood samples. To mention specially,* ***Prof. Dhanpat K Kochar*** *has always been a supportive hand and his commitment towards malaria research is deeply appreciated. We (our group) have always enjoyed his warm hospitality and varieties of tasty foods he offered during our every visit to Bikaner for blood sample collection and discussion. I expand my thanks to* ***Dr. Sheetal Middha Bhasin*** *and* ***Mrs. Jyoti Acharya****. They played a critical role in my thesis and my warm appreciation goes to their hard work in collecting, processing and sending the blood samples neatly in accordance to the protocol.*

I would like to thank my laboratory seniors ***Dr. Vishal Saxena*** *(He was my DAC member too),* ***Dr. Shilpi Garg****,* ***Dr. Deepak Pakalpatti*** *and* ***Dr. Narayan Kumar*** *for guiding me, training me, encouraging me and supporting*

me morally and intellectually during the entire period of my thesis. They have always stand by me and laid their helping hand unconditionally during this tough journey. I would also like to extend huge warm thanks to my long standing friend, colleague ***Mr. P. A. Boopathi*** *for his affection, care and support. He has contributed to this thesis majorly by sharing his ideas, working with me in laboratory benches, writing research manuscripts and more importantly fighting with me to bring me on track. My special thanks go to laboratory juniors* ***Isha***, ***Ramandeep***, ***Gagan***, ***Zarna*** *for their timely help and supports in conducting some of my experiments.*

It's my fortune to gratefully acknowledge ***Dr. Jitendra Panwar****, HOD, Department of Biological Sciences and other faculty members of the Department for their care, understanding, suggestions and specially providing an environment which makes feel every research scholar of the Department like being in home. My thanks are due to all the research scholars of the Department for their cooperation and help.*

I am extremely indebted to ***Department of Biotechnology****, Govt. of India for providing Project Assistantship,* ***Council of Scientific and Industrial Research****, Govt. of India for Senior Research Fellowship and BITS, Pilani for Institute Research Fellowship. I also extend my thanks to* ***BITS, Pilani*** *for providing required infrastructural facilities to carry out my research work. This work would not have possible without the financial support by Department of Biotechnology in terms of research project to* ***Prof. Ashis Kumar Das****.*

My high regards goes to the non-teaching staff of the Department ***Mr. Kamlesh Soni***, ***Mr. Mukesh Saini***, ***Mr. Naresh Saini***, ***Mr. Anirudh***, ***Mr. Parmeswar*** *and* ***Mr.Subhash*** *for their timely help.*

Words are short to express my deep sense of gratitude towards my following childhood friends Mr. ***Subhransu Das***, ***Mr. Dileep Rao***, ***Mr. Balgopal Seth***, ***Mr. Manoj Meher*** *and some of my friends whom I met during my later part of education ,* ***Mr. Muthu Karupaiaah***, ***Dr. Jaskiran Kuljeet***, ***Dr. Madhvi Joshi***, ***Dr. Bhupendra Mishra***, *Mr.Swarna Kanchan,* ***Mr. Kuldeep*** *and* ***Dr. Vasanth Sekhar*** *for their constant encouragement and moral support.*

This list is incomplete without paying high regards to my parents ***Mr. Ashok Subudhi*** *and* ***Mrs. Januma Subudhi.*** *I bow down in front of them and owe everything to them. I know the amount of trouble and pain they have faced in bringing me to this stage of life. They remained in uncomfortable zone to provide me comfortable space. Their constant encouragement and inspiration has always provided me strengths much needed to achieve this tough goal. The moral support extended by my two sisters* ***Mrs. Aliva Patra*** *and* ***Mrs. Anjana Subudhi*** *and two brother-in laws* ***Mr. Pradeep Patra*** *and* ***Mr. Priyabrat Subudhi*** *is highly appreciated. I will miss something if I don't mention the name of my two nephews* ***Kunal*** *and* ***Krishh*** *who have delighted me always with their loving words and for constantly asking me the question: when is your studies going to complete?*

Last but not the least; I dedicate this thesis to all the patients who have agreed to donate blood for my research work and also to those who have lost their life due to malaria.

Amit Kumar Subudhi

Abstract

Malaria in humans is caused by six species of *Plasmodium,* of which *Plasmodium falciparum* is considered the most virulent and responsible for much of the mortality. The challenges in combating malaria can largely be attributed to the complex life cycle of this parasite coupled with its inherent ability to adapt to its host environment. Additional parasite traits that may contribute to its pathogenesis are virulence factors which interact directly with the host, use of multiple host cell invasion pathways, remodeling of host cell for survival and the ability to transmit from one host to another. Importantly, most of the aforementioned factors are under natural selection and may evolve in response to the changing host's genetic and physiological environment. Any phenotypic variation or adaptive changes of malaria parasites in response to changing environment of the host may be reflected through the differential regulation of gene expression. Although, gene expression studies have been carried out on this parasite and has given an extraordinary amount of information about the parasite biology, our understanding about the transcriptome of this parasite *in vivo* still remains elusive.

Pathogenesis studies in other systems have clearly demonstrated that organisms have distinct biology *in vivo* comparison to *in vitro.* Studies which have tried to partially mimic the host conditions in culture showed subsets of genes differentially regulated in response to these changes, suggesting that the host provides a substantially different environment compared to laboratory conditions with a distinct effect on regulation and expression of parasite genes. In temperate and sub-tropical regions of Asia and Latin America, complicated malaria manifests as hepatic dysfunction or renal dysfunction and is seen in all age groups. There has been a concerted focus on understanding the patho-physiological and molecular basis of complicated malaria in children, much less is known about it in adults. Central to this thesis is dissecting the *in vivo* transcriptome of the malaria parasite *P. falciparum* from adult patients with uncomplicated/complicated malaria and to differentiate between the transcriptome of complicated and uncomplicated parasite isolates.

In order to achieve the above mentioned goal, initial work involved validating a *P. falciparum* cross strain 15K array which has been designed by our group in collaboration with Genotypic Technology Pvt Ltd for the efficient detection of transcripts from field isolates. The array contains probes representing genome sequences of two distinct geographical isolates (i.e. 3D7 and HB3) and sub-telomeric *var* gene sequences of a third isolate (IT4) known to adhere in culture condition. Probes in the array have been selected based on their efficiency to detect transcripts through a 244 K array experimentation. A large percentage (91 %) of the represented transcripts was detected from Indian *P. falciparum* patient isolates. Replicated probes and multiple probes representing the same gene showed perfect correlation between them suggesting good probe performance. Additional transcripts could be detected due to inclusion of unique probes representing HB3 strain transcripts. Variant surface antigen (VSA) transcripts were detected by optimized probes representing the VSA genes of three geographically distinct

strains. The 15K cross strain *P. falciparum* array has shown good efficiency in detecting transcripts from *P. falciparum* parasite samples isolated from patients.

Another facet of this study dealt with exploring the diversity of natural antisense transcripts in *P. falciparum* clinical isolates from patients with diverse malaria related disease conditions. With the discovery of natural antisense transcripts (NATs) in this parasite and considering the various proposed mechanisms by which NATs might regulate gene expression, it has been speculated that these might be playing a critical role in gene regulation. The 244K array that has been used to select best probes for the custom cross strain 15K array design is actually a strand specific microarray and contains probes against both sense and antisense transcripts. Experimentation was carried out using isolates taken directly from patients with differing clinical symptoms caused by malaria infection. A total of 797 NATs targeted against annotated loci have been detected. Out of these, 545 NATs are unique to this study. The majority of NATs were positively correlated with the expression pattern of the sense transcript. However, 96 genes showed a change in sense/antisense ratio on comparison between uncomplicated and complicated disease conditions. The antisense transcripts map to a broad range of biochemical/ metabolic pathways, especially pathways pertaining to the central carbon metabolism and stress related pathways. Our data strongly suggests that a large group of NATs detected here are unannotated transcription units antisense to annotated gene models. The results reveal a previously unknown set of NATs that prevails in this parasite, their differential regulation in disease conditions and mapping to functionally well annotated genes.

In order to unravel the *in vivo* transcriptome of the parasite and to differentiate between the transcriptome of complicated and uncomplicated parasite isolates, microarray based gene expression data was generated by conducting experiment on the 15K custom array using parasite RNA material from adult patient samples, showing uncomplicated malaria, hepatic dysfunction or renal failure**.** The data has been analyzed with reference to variant surface antigens, encoded by the *var*, *rifin* and *stevor* gene families. The differential regulation profiles of key genes (comparison between *P. falciparum* complicated and uncomplicated isolates) have been observed. The exportome has been analyzed using similar parameters. Gene ontology term based functional enrichment of differentially regulated genes identified, up-regulated genes statically enriched ($p<0.05$) to critical biological processes. Systems network based functional enrichment of overall differentially regulated genes yielded a similar result. Up-regulation of *var* group B and C genes whose proteins are predicted to interact with CD36 receptor in the host as also the up-regulation of group A *rifins* and many of the *stevors* was observed. This is contrary to most other reports from pediatric patients, with cerebral malaria where the up-regulation of *var* A group genes have been seen.

The concluding part of the thesis dealt with constructing a weighted gene co-expression network from 20 parasite microarray expression profiles. Objectives of this part of the work were to identify complicated malaria disease specific modules and thereby parasite biological processes specifically perturbed during complicated disease conditions, and also to identify up-regulated

and stably expressed hub genes (highly connected genes) that might be potential candidate for intervention strategies. A total of 20 highly interacting modules were identified post network creation. Disease specific perturbed modules were identified by mapping the nodes with differentially regulated genes. Further, gene ontology based functional enrichment analysis was carried out to disclose the biological processes associated with these modules. Genes involved in biological processes like oxidation-reduction, electron transport chain, protein synthesis, ubiquitin dependent catabolic processes, RNA binding and purine nucleotide metabolic processes were found to be enriched in various modules populated with up-regulated genes. Additionally, for each module, highly connected hub genes were identified many of which were either up-regulated, stably expressed or down regulated genes. Detailed functional analysis of many hub genes with known annotated function underline their importance in parasite development and survival suggesting that other hub genes with unknown function might be playing crucial roles in parasite biology and are potential candidate for intervention strategies.

In conclusion, by exploring the transcriptome of field isolates from patients with different levels of disease severity, we have not only obtained unique information about the *in vivo* transcriptome that was not known earlier but also could show the virulence factors and biological processes that are specifically active during complicated disease conditions. We also able to show that there exists a unique diversity of NATs in field isolates and some of these were found to alter their expression in disease conditions. At the end, systems network based analysis has enabled identifying hub genes that are potential candidates for intervention strategies.

Contents

List of Tables

List of Figures

Abbreviations

ACD	Acid Citrate Dextrose
AMADID	Agilent microarray design identification number
AP2	Apitela 2
AS	Antisense transcripts
BEB	Back extraction buffer
CIDR	Cysteine rich inter-domain regions
CSA	Chndriotin sulfate A
Cy3	Cyanine 3
Cy5	Cyanine 5
DBL	Duffy binding like domain
DEPC	Diethyl Pyrocarbonate
DNA	Deoxy ribonucleic acid
dNTP's	Deoxy ribonucleoside triphosphates
EDTA	Ethylene Diamine tetra acetic acid
EST	Expressed sequence tags
HA	Haemaglutinin
HAS-LS	High antisense- low sense transcripts
HRP	Histidine rich protein
HS-LAS	High sense-low antisense transcripts
ICAM	Intracellular cell adhesion molecule
IDC	Intra-erythrocytic developmental cycle
iRBCs	Infected red blood cells
KHARP	Knob associated histidine rich protein
M	Molar
MOPS	3-[N-Morpholino]propanesulfonic acid
NATS	Natural antisense transcripts
NC-AS/S	No change in sense antisense transcripts ratio
P.	*Plasmodium*
PBMC	Peripheral blood mononuclear cells

PBS	Phosphate Buffered Saline
PECAM 1	Platelet/endothelial cell adhesion molecule 1
PEXEL	*Plasmodium* export element
Pf	*Plasmodium falciparum*
PfAQP	*Plasmodium falciparum* aquaglyceroporin
PFC	*Plasmodium falciparum* complicated cases
PfEMP1	*Plasmodium falciparum* erythrocyte membrane protein 1
PFU	*Plasmodium falciparum* uncomplicated cases
PHIST	Plasmodium helical interspersed sub-telomeric family
Pv	*Plasmodium vivax*
qPCR	Quantitative PCR
RDTs	Rapid diagnostic tests
RESA	Ring infected erythrocyte surface antigen
REX	Ring stage exported protein
RIFIN	Repetitive interspersed family genes
RMDT	Rapid malaria diagnostic tests
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RNA-Seq	RNA sequencing
S	Sense transcripts
SAGE	Serial Analysis of gene expression
S-AS	Sense and antisense transcripts
snRNA	Small Nuclear RNA
STEVOR	Sub-Telomeric variable open reading frame family genes
Tris	2-Hydroxy methylamine
tRNA	Transfer RNA
WGCNA	Weighted Gene Co-expression Network Analysis
VCAM	Vascular cell adhesion molecule

VSA	Variable surface antigens

1.1. Malaria- A brief historical perspective

Malaria parasites have been associated with mankind since time immemorial. Malaria as a disease has probably influenced to a great extent on human societies and human history. Many ancient documents like the Chinese medical writings "NeiChing" from 2700 BC, clay tablets from Mesopotamia from 2000BC, Egyptian papyri from 1570 BC and Hindu Sanskrit medical treatises like Charaka samhita and *Susruta* have mentioned about the symptoms of malaria (www.malariasite.com). Hippocrates, today regarded as the "Father of medicine" was probably the first to report the manifestations of the disease and also mentioned various forms of malaria fever such as quotidian, tertian and quartan, where temperature rises daily, every other day or every third day respectively. For a longer period of time malaria was attributed to stagnant waters and swamps and hence the word malaria is widely believed to have originated from the Italian word mal'aria which means "bad air". It was the discovery of bacteria by Antoni van Leewenhoek in 1676, and the postulation of the germ theory of infection by Louis Pasteur and Robert Koch in the year 1878-1879, when the search for the real cause of malaria intensified.

A series of discoveries covering a long period of time have shaped our knowledge about the malaria causing parasites and its biology. The discovery probably started with Louis Alphonse Laveran, a French army officer working in Algeria, who observed intra-erythrocytic organism including crescents, spherical motionless bodies with pigment and bodies with extruded flagella like structures in unstained blood of patients. These observations gave him confidence of discovering a parasitic protozoan which he later named it as *Oscillaria malariae* (Laveran, 1881). He was awarded the Nobel Prize in 1907. In between the years 1885-86, Golgi was to differentiate between tertian (48 hour periodicity) and quartan malaria (72 hour periodicity) (Golgi, 1886) and in 1889-1890, Golgi and Marchiafava mentioned the differences between benign tertian and malignant tertian malaria (Golgi, 1889). By 1890, it was proposed that there were three species of malaria parasites with specific periodicities of fever and other characteristics responsible for benign tertian, malignant tertian and quartan malaria caused by Heamaoeba vivax (*Plasmodium vivax*), Laverania malariae (*Plasmodium falciparum*) and Haemaoeba malariae (*Plasmodium malariae*) (Grassi, 1900). John Stephens, in the year 1922, described a fourth species, *Plasmodium ovale* which had characteristics similar to *P. vivax* (Stephens, et al., 1922).

Sir Ronald Ross, while working in India, had found developmental stages of human malaria parasites in *anopheline* mosquitoes and in his letters, he called August 20th 1897 as 'mosquito day'. In the same year, he had demonstrated the sexual cycle of *Plasmodium relictum* inside the culicine mosquitoes. Ross predicted correctly that human malaria was probably transmitted in the same way. In the year 1898, Bignami and Grassi were able to demonstrate that *Anopheles clavinger* mosquitoes when fed on infected patients could transmit the infection to uninfected patients. The discovery of erythrocytic and sexual stages of the parasite in human and mosquito still gave an incomplete picture of the life cycle of the malaria parasite. It remained incompletely understood as nobody knew where the parasites develop during the first few days (7-10 days) after infection during which they could not be seen in the blood. It was the discovery by Henry Shortt and Cyril Garnham working in London in 1947 that a phase of division in the liver preceded the development of parasites in the blood of primates, completed the full picture of the malaria life cycle stages (Shortt and Garnham, 1948a). Subsequently Shrott, Garnham and their co-workers found exoerythrocytic forms of *P. vivax* (Shortt and Garnham, 1948b), *P. falciparum* (Shortt, et al., 1949) and *P. ovale* in human volunteers (Shortt, et al., 1954).

1.2. Worldwide malaria prevalence

Malaria as a disease is caused by obligate intraerythrocytic protozoan parasites of the genus *Plasmodium*. For the past 80 years, four species of the genus *Plamsodium* are *Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae* and *Plasmodium ovale* have been known to infect humans of which *Plasmodium falciparum* is known to cause the most severe form of malaria. Although *P. vivax* is being increasingly associated with similar symptoms Recently, two more species of genus *Plasmodium* previously known to cause malaria in macaques has been identified to cause human malaria i.e. *Plasmodium knowlesi* and *Plasmodium cynomolgi* (Chin and Lach, 1965, Fong and Good, 1971, Singh, et al., 2004 and Ta, et al., 2014). As a disease, malaria continues to impose a great challenge to the health of most of the world's population and sustained economic burden on malaria endemic countries (Breman, 2001 and Gallup and Sachs, 2001). Malaria is the third most dangerous infectious disease along with HIV/AIDS and Tuberculosis that afflict mankind (Vitoria, et al., 2009).

In the year 2012, malaria was endemic in 99 countries and territories and 5 countries were in the prevention of reintroduction phase making a total of 104 countries endemic to malaria (WMR,

2013). According to WHO 2010 estimates, 215 million cases and 655,000 deaths were attributed to malaria. However, a report from Seattle Institute for Health with its detailed model of global deaths attributable to malaria in 2010 estimated 1,238,000 deaths (95% CI: 929,000-1,685,000), which is almost double the WHO's estimates underscoring the problems in assessing the disease at large (Murray, et al., 2012). More than half of the world population resides in malaria endemic areas in Africa, South and Southeast Asia and Central and South America. Globally 3.3 billion people were at risk of malaria in 2011, with population living in sub-Saharan Africa having the highest risk of acquiring malaria. Approximately 80% of cases and 90% of deaths are estimated to occur in the WHO African region, with children under five years and pregnant women most severely affected (**Figure 1.1**). Female mosquitoes of about 30 anopheline species were known to be the vector for malaria transmission in human.

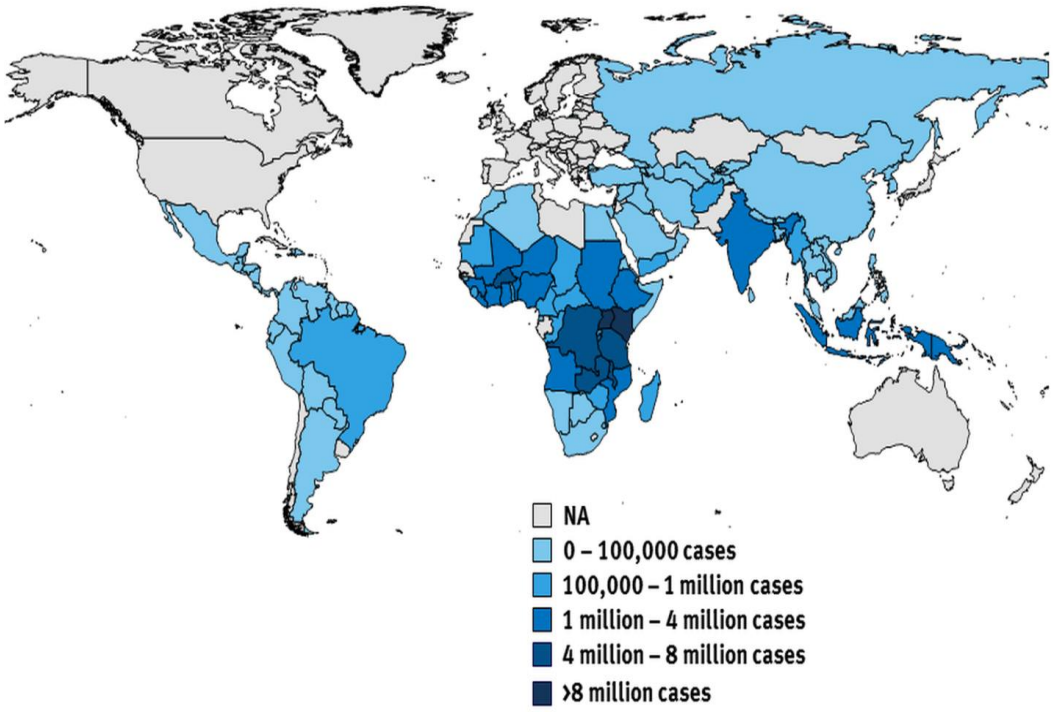


Figure 1.1. Worldwide reported malaria cases in 2011. Based on WHO, World Malaria Report, 2012; December, 2012. **Source**: Kiser Family Foundation, www.GlobalHealthFacts.org.

1.3. Prevalence of malaria in India

India is the seventh-largest country in the world in terms of geographical area inhabited by world's most culturally, linguistically and genetically diverse population after the African continent. It ranks as the second most populous country with an estimated population size of 1.2 billion. The Indian climate is heavily influenced by the Himalayas and the Thar Desert. Most parts of India receive rain fall between June and October and the total land can be divided into four major climatic groups: tropical wet, tropical dry, subtropical humid and mountainous. Part of the inhabited land which receives good amount of rain fall makes a good breeding ground for mosquitoes and is well suited for propagation of malaria parasite.

About half of the population (273 million) outside Africa at high risk of malaria transmission resides in India. In 2011, out of the 10 malaria endemic countries in South East Asia region (SEA) that reported 2.15 million of confirmed malaria cases, India contributed the most, 1.31 million (61%) cases followed by Myanmar 0.47 million (22%) cases and Indonesia 0.25 million (12%) cases (WMR, 2012). *P. falciparum* and *P. vivax* contribute almost equally to the total number of reported malaria cases in India (Singh, et al., 2009). Of the estimated 1.2 billion populations, 80.5% live in areas of high to low risk of malaria. States like Odisha, Chhatisgarh, West Bengal, Jharkhand and Karnataka contribute majorly to malaria cases in India. Although, there is an overall decrease in the incidence of malaria in recent years, the annual mortality rate

due to malaria largely remains unchanged. There were about 1018 and 753 deaths attributed due to malaria in 2010 and 2011 in India respectively (Annex 6E, World Malaria report [WMR] 2012) with estimated deaths of about 29,401 (Estimated based on fixed case fatality rate applied to case deaths, annex 6A, WMR 2012) in 2010 alone. **Figure 1.2** represents the region wise Annual Parasite Index (API) of India during 2010 as reported by National Vector Borne Disease Control Programme.

The absolute numbers of mortality due to malaria in India have been associated with significant uncertainties. Dhingra et al, in the year 2010 reported an estimated 205,000 (125000-277000) deaths per year in India that could be attributed to malaria during 2001-2003 based on verbal autopsies interviews (Dhingra, et al., 2010). This was supported again by report from Sharma et al (2011). These estimated numbers are quite high compared to the WHO recent estimate of

malaria deaths in India suggesting a complicated underestimation of deaths due to malaria in India. During the year 1997-1998, there were an estimated 144,000 deaths per year in India that could have happened due to malaria as determined based on certified death reports to the Government of India (Kumar, et al, 2011).

There are 9 species of Anopheline vectors transmitting human malaria parasites in India. Two species are widely distributed of which *A. culicifacies* is the major vector in rural areas where as *A. stephensi* is the major vector in urban areas (Nagpal and Sharma, 1994).

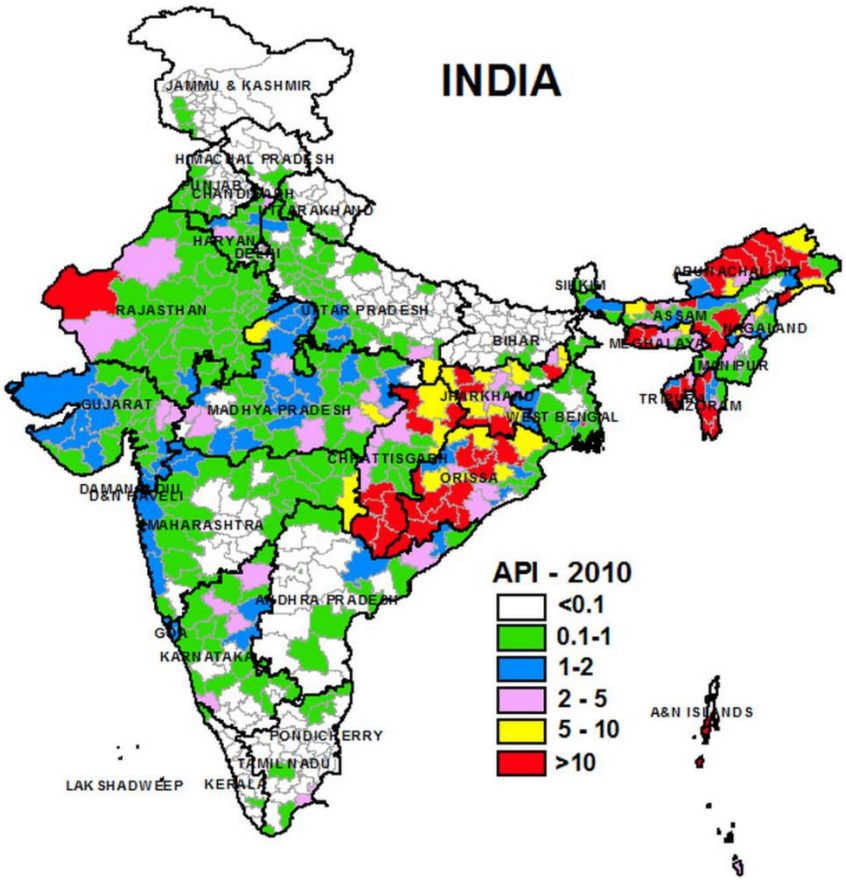


Figure 1.2. Annual parasite incidence of India during the year 2010. **Source**: NVBDCP www. nvbcdp.gov.in/maps

1.4. Malaria transmission and population at risk

Areas highly endemic to malaria, notably in sub-Saharan Africa, and young children with age ranging from 6 months to 5 years are considered to be particularly at the risk of developing complicated disease. This is because of high parasite inoculation rate and lack of pre-existing immunity. Pregnant women also constitute a high risk group because of pregnancy related immune suppression and affinity of *P. falciparum* to Chondriotin Sulfate A (CSA) in placenta (Hsiang, et al., 2009). Older children and adults belonging to stable malaria transmission areas, acquire semi-immunity which greatly reduces the risk of developing malaria related disease severity.

This is in contrast to a country like India and other countries of Asia and Latin America belonging to temperate and subtropical regions, where an altered pattern of disease presentation is seen. Populations of these countries suffer epidemics because of the seasonal transmission and insufficient ongoing exposure to induce or maintain immunity. Under these circumstances, residents of all age groups are at a high risk of developing complicated malaria. Moreover, it is the adult population which is at the highest risk of infection due to occupation and migration.

1.5. Life Cycle of malaria parasite

The life cycle of the malaria parasite is complex involving multiple stages of development and involves two hosts: Insect vector (the definitive host) and vertebrate hosts. Sexual phase of the parasite life cycle is completed in the invertebrate host while the vertebrate host is required for the asexual phase of its life cycle (**Figure 1.3**).

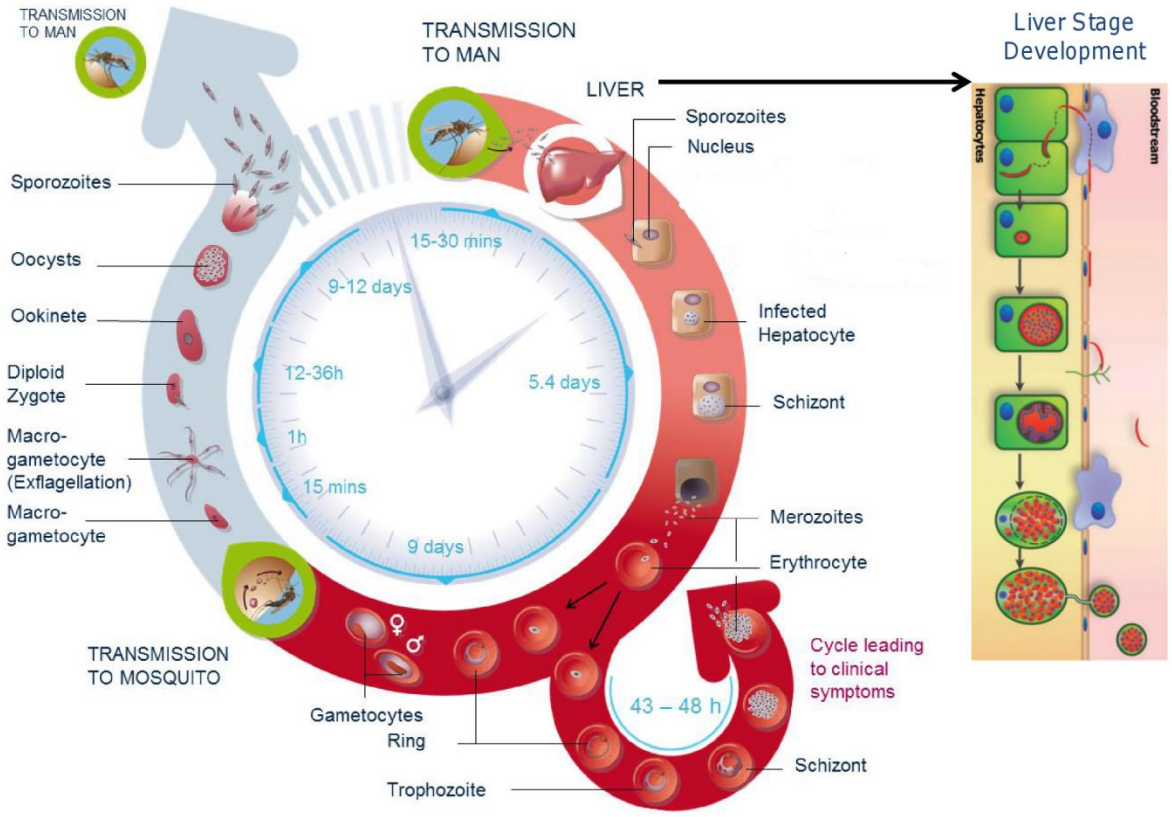


Figure 1.3. Life cycle of *Plasmodium*. A *Plasmodium* infected mosquito during a blood meal injects sporozoites into the host. From the release site, sporozoites make their way towards liver, infect hepatocyte and initiate the liver stage infection. Sporozoites replicate inside the hepatocytes, producing thousands of merozoites, which are released into the blood stream during hepatocyte egress. Released merozoites, invade the red blood cells, initiating erythrocyte schizogony, with parasites developing into a ring, followed by a mature trophozoite stage. After this nuclear division occurs, producing a schizont stage with 16-32 merozoites. Merozoite release occurs after the erythrocyte egress, where the erythrocyte schizogony commences again. A small proportion of released merozoites after infecting the red blood cells develop either into a male or a female gametocyte (sexual stages). Male and female gametocytes taken inside by mosquitoes during blood meals, where male gametocytes exflagellate and fertilize the female gametes to produce a zygotes that develop into ookinetes. Ookinetes, penetrate the midgut epithelium and develops into oocyst under the basal lamina. Oocysts filled with developing sporozoites, rupture releasing thousands of sporozoites that migrate to salivary glands where they are stored until the next bite of the mosquito. **Source**: Medicines for Malaria Venture, www.mmv.org/malaria-medicines/parasite-lifecycle.

1.5.1. Making way to the host liver: The pre-erythrocytic schizogony

The parasite's life cycle in the human host begins with the bite of an infected female *Anopheles* mosquito which inoculates 15-120 sporozoite forms (the infective stage of the parasite) into the host's skin along with saliva. Once injected into the skin, a portion (nearly 35%) of the motile sporozoites transmigrates through several cells until they encounter a blood vessel, move into the circulatory system and eventually reach the liver (Frischknecht, et al., 2004, Vanderberg and Frevert, 2004 and Amino, et al., 2006). The parasite binds to the highly sulfated heparansulfate proteoglycans (HSPGs) presented by hepatocytes (Coppi, et al., 2007). The binding with HSPGs signals the parasite to switch into its invasive mode. After crossing the space of Disse, sporozoites transmigrate through several hepatocytes before invading the target hepatocyte (Frevert, et al., 2005, Pradel and Frevert, 2001, Baer, et al., 2007a and Mota, et al., 2002). Following invasion of the hepatocyte by the sporozoite the latter is surrounded by a parasitophorous vacuole membrane (PVM), and undergoes rapid and extensive replication. Over 5 to 15 days (depending on *Plasmodium* species), each sporozoite inside the PVM differentiate and divide into thousands of merozoite forms. Thousands of merozoites in the host cytoplasm get surrounded by membrane to form merosomes which then bud off to be released into hepatic blood vessels (Sturm, et al., 2006 and Baer, et al., 2007b). This constitutes the first step of the transition from the liver to the blood stage.

Some sporozoites, in case of *P. vivax* and *P. ovale* infection convert to dormant forms called hyponozoites, which causes relapses after weeks, months or even years. However, these dormant stages do not appear in other *Plasmodium* parasites which infect humans (Hulden and Hulden, 2011).

1.5.2. Opening the door for disease pathology: The erythrocytic schizogony

The erythrocytic schizogony is the stage which leads to all clinical pathologies associated with the disease. This is the stage which has been investigated the most because of the development of continuous culture system. Upon the release of the merosomes into the blood stream, the merosome membrane breaks in an erythrocyte rich environment where the free merozoites rapidly invade the erythrocytes.

1.5.3. Erythrocyte invasion by parasite

Plasmodium merozoites invade erythrocytes by a complex multistep process which is probably similar for all *Plasmodium* species. The process of invasion appears to be very coordinated and depends on sequential release of proteins. Sub-cellular localization and sub-compartmentalization of proteins involved in invasion within secretory organelles also seem to play a critical role. The initial step includes reversible attachment of the merozoite and deformation of the erythrocyte surface, followed by quick apical orientation positioning of the merozoite apical end abutting the erythrocyte surface. This process involves molecules like merozoite surface proteins (MSP), serine repeat antigen protease like family proteins and MSP Duffy binding like proteins (Cowman, et al., 2012). The next step that follows is the invasion of the merozoites by forming a tight or moving junction, an area of electron density which facilitates the invasion and post invasion sealing of the parasite within the erythrocyte. This is facilitated by adhesin proteins like EBA-175, PfRH-4 and invasin protein like apical membrane antigen 1 (AMA 1) which resides in apical complex of secretory organelles (i.e. micronemes, rhoptries and dense granules) (Cowman, et al., 2012). The entry also initiates the formation of a parasitophorous vacuole (PV), where the parasite resides after invasion. Once the parasite enters the erythrocyte, sealing at the posterior end is followed by an event of echinocytosis of the erythrocyte which last for about 10 minutes (Gilson and Crabb, 2009). Inside the infected red blood cell, parasites undergo rapid changes in shape (Gruring, et al., 2011) and develop through ring, trophozoite and schizont stages before releasing merozoites by bursting the infected RBCs.

1.5.4. Exportome and host cell remodeling

During and after invasion, the parasite secrets proteins (exportome) into the PV which eventually traffick to the RBC cytoplasm and are required for host cell remodeling (**Figure 1.4**).These proteins modify the permeability and adhesive properties of the host cells making it a suitable place for the parasite's own survival. As the parasite starts developing inside the infected erythrocyte, structures like tubulovesicular network (TVN) derived from PV membrane and Maurer's cleft with distinct protein composition starts appearing in the erythrocyte cytoplasm (**Figure 1.5**) (Haldar and Mohandas, 2007 and Atkinson and Aikawa, 1990). These membrane

bound organelles appears to act as secretory organelles to deliver various exportome proteins to their destinations.

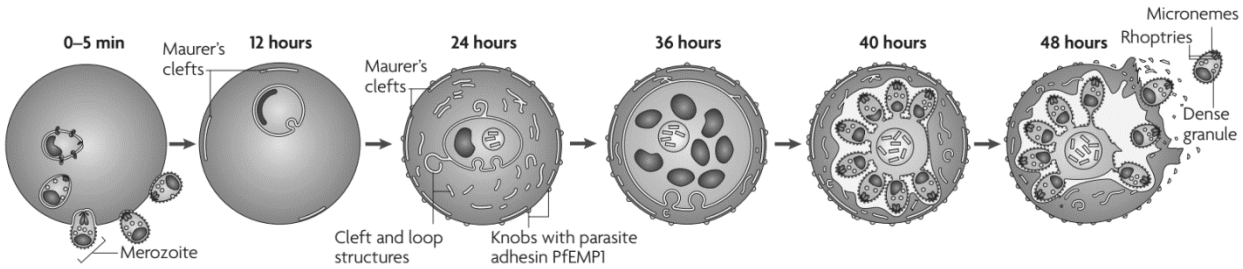


Figure 1.4. Erythrocyte remodeling during the intra-erythrocytic development of *Plasmodium falciparum*. The merozoite initially attaches to and then invades the RBCs. The parasite develops in a parasitophorous vacuole through ring, trophozoite and schizont stages. Membrane bound organelles like tubulovesicular network and Maurer's cleft appear in the host RBC cytoplasm as the parasite develops. These organelles originates and emanate from PV membrane. The release of the parasite cargo proteins through the newly formed membrane bound organelles produces knob like structures on the surface of the infected RBCs exposing of the adhesion proteins PfEMP1 on outer surface. **Source:** Maier, et al., 2009.

The adhesive properties of the infected RBCs are associated with the presentation of a membrane bound cytoadherence protein, *Plasmodium falciparum* erythrocyte membrane protein 1(PfEMP1). The PfEMP1 binds to various receptors expressed on the surface of endothelial cells. This is why the mature stage infected RBCs do not appear in the peripheral circulation as these are bound to the endothelial receptors of deep vasculature present in different organs. The appearance of TVN and Maurer's cleft represent a novel secretory system which is otherwise not present in an uninfected terminally differentiated RBC. Majority of exported proteins are characterized of having a novel pentameric amino acid sequence motif known as *Plasmodium* export element (PEXEL) / host targeting (HT) that directs the export of parasite proteins (Marti, et al., 2004 and Hiller, et al., 2004). Reports also suggest that some proteins that do not have a PEXEL/HT motif (for e.g. ring exported proteins 1 and 2 and skeletal binding protein 1) are also exported suggesting that either alternate signals can also be recognized by the machinery or the presence of additional protein export pathways (Spielman and Gilberger, 2010). The exported proteins interfere with the RBC membrane proteins which may play various roles like repair of

the RBC membrane after the invasion, uptake of lipid and nutrient, disposal of waste or maintenance of ion gradients (**Figure 1.5**, Maier, et al., 2009).

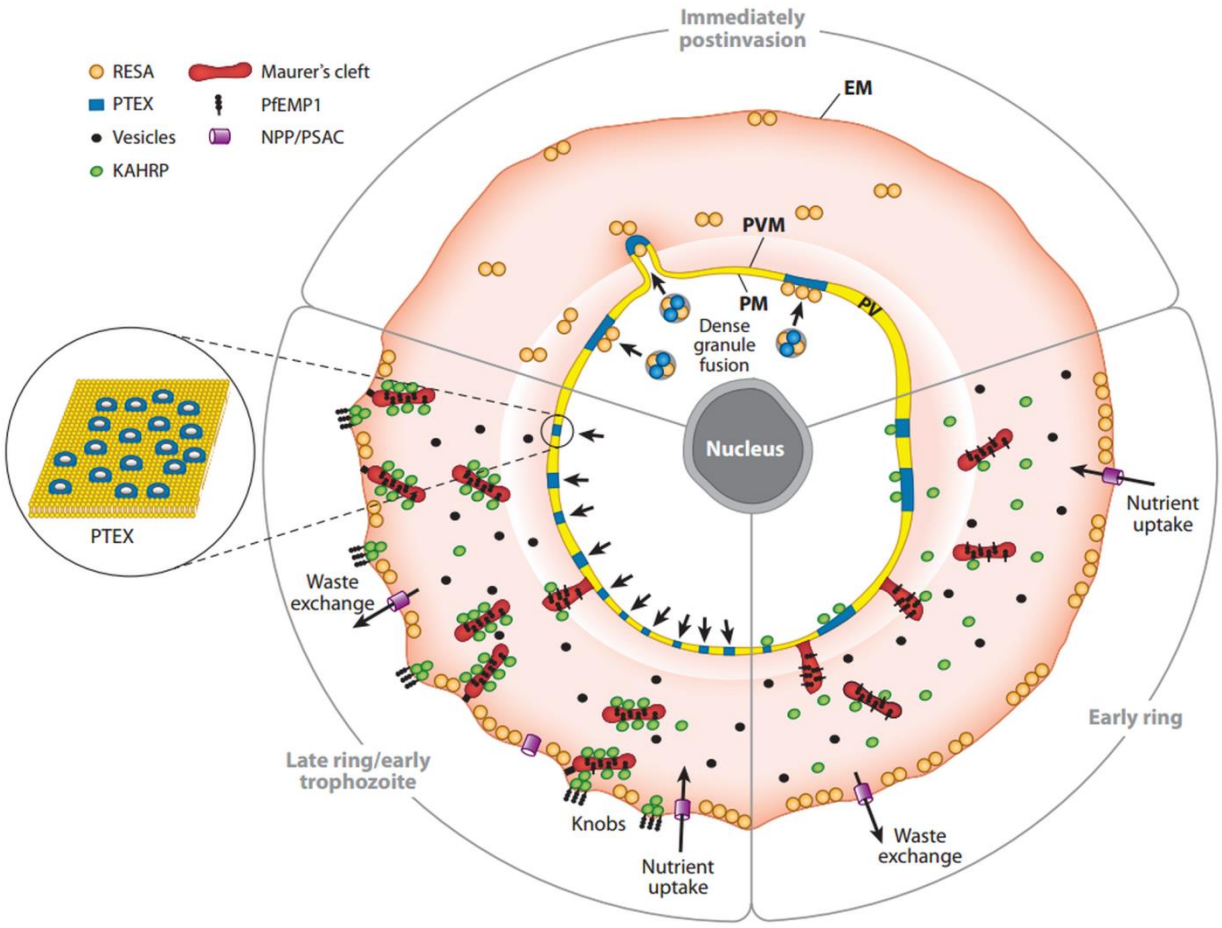


Figure 1.5. Remodeling of infected erythrocyte during parasite development: Immediately after parasite invasion dense granule fuse with the parasite membrane (PM). This fusion releases parasite proteins like *Plasmodium* translocon of exported proteins (PTEX) and ring-infected erythrocyte surface antigen (RESA) into the parasitophorous vacuole (PV). PTEX as a translocon assembles in the parasitophorous vacuole membrane, enabling export of proteins from PV. During the early ring stage, proteins like knob associated histidine rich protein (KAHRP) and PfEMP1 export through PVM. Following export, KAHRP and other proteins localize in the Maurer's cleft. The protein export continues throughout the ring and trophozoite stages. Further during the transport, KHARP reaches its final destination under the erythrocyte membrane, binds the erythrocyte protein spectrin and forms protrusions called as knobs. In this stage, translocation of PfEMP1 through the Maurer's cleft onto the infected erythrocyte surface begins and is presented on the cell surface by tethering with the knobs. New permeability pathways are established that is mediated by *Plasmodial* surface anion channel. **Source:** Boddey and Cowman, 2013.

1.5.5. Parasite intra-erythrocytic developmental cycle

The parasite after invading the RBC, initiates the intererythrocytic developmental cycle (IDC). The IDC depending on *Plasmodium* species completes in 24-72 h post merozoite invasion. The early stage is known as ring stage after which the parasite then enters a highly metabolic maturation phase, the trophozoite stage before its replication. In the trophozoite stage, it survives by ingesting host cell cytoplasm. It uses the host hemoglobin as a source of amino acids; however it's incapable of degrading heme, the byproduct of hemoglobin digestion. During the hemoglobin degradation, most of the heme is polymerized into hemozoin (malaria pigment) and stored within the food vacuoles. The conversion of heme to hemozoin is necessary as the free heme is toxic to the parasite (Weissbuch and Leiserowitz, 2008). In the next few hours, parasite grows in size and prepares itself for replication. In the schizont stage, the cell prepares to produce the merozoites by replicating and dividing to form 16-32 merozoites which are then released into the blood stream upon bursting of the infected RBCs. The released merozoites re-invade other RBCs immediately to continue the erythrocytic schizogony cycle.

1.5.6. Gametocytogenesis: The gateway to escape

Few merozoites which re-enter the RBC stop multiplying asexually and enter the sexual differentiation pathway. In the next 10 days gametocytes mature to develop into either male or female gametocytes. *P. falciparum* gametocytes mature through five different morphological stages (stage I- stage IV) (Field and Shute, 1956). It has been demonstrated in *P. falciparum* that all merozoites emerging from a single schizont either continue the IDC cycle or develop into gametocytes (Bruce, et al., 2000). This suggests that the fate of gametocyte development is believed to have been decided early in the late trophozoite/schizont stages of the preceding asexual generation. It has also been shown that all the gametocytes from a single schizont either develop into male or female gametocytes (Smith, et al., 2000 and Silvestrini, et al., 2000). However, exact factors that trigger the development of gametocytes *in vivo* have not been identified so far.

1.5.7. Mosquito stage development: Preparing for the infectious stage

When a female *Anopheles* mosquito takes a blood meal from an infected person, along with blood both the male and female gametocytes may get ingested into the mosquito gut where the sexual reproduction starts. The *Plasmodium* parasite sexual reproduction occurs once in its life cycle and that is inside the mosquito. During most of its life cycle *Plasmodium* remains haploid except for a brief period after fertilization. Formation of male and female gametes (Gametogenesis) completes right away in the mosquito midgut lumen where factors like mosquito derived molecule xanthurenic acid (Billker, et al., 1998), temperature shift and pH change (Billker, et al., 1997) are known triggers for exflaggelation (male gemetogenesis). This is followed by fertilization to form a diploid spherical zygote which then transforms into an elongated motile and invasive ookinete that penetrates the mosquito gut wall (Vinetz, et al., 1999). The ookinete rests below the basal lamina after crossing the midgut epithelium before transforming itself into an oocyst.

Oocyst development takes 10-12 days to complete and is the only developmental stage in the parasite life cycle which occurs outside a cell. The oocyst starts dividing mitotically to produce thousands of sporozoites. Fully developed sporozoites are eventually released into the hemocoel upon bursting of the oocyst. The sporozoites then find and invade the mosquito salivary gland to reach its duct where it remains ready for being injected into the new vertebrate host during a subsequent blood meal.

1.6. Complicated *Plasmodium falciparum* malaria

Most complicated malaria and much of related mortality are caused by *P. falciparum*, although in recent years *P. vivax* has also been implicated in causing complicated forms of malaria (Kochar, et al., 2005, Kochar, et al., 2009). Clinical symptoms of malaria, primarily starts with rupture of infected RBCs and the clinical presentation often resembles with those of common viral fevers which may sometimes delay in diagnosis (Murphy and Oldfield, 1996). The majority of patients with symptomatic uncomplicated malaria experience fever, chills, headaches and diaphoresis along with some other common symptoms like dizziness, malaise, myalgia, abdominal pain, nausea, mild diarrhea and dry cough. However, in absence of prompt

recognition of the disease, the uncomplicated malaria can lead to a complicated form of disease within a few hours and can eventually end with death of the patient. Complicated *falciparum* malaria can be defined as a patient with *P. falciparum* asexual parasitemia and no other confirmed cause for their symptoms and the presence of one or more clinical and laboratory features like prostration, impaired consciousness, respiratory distress, multiple convulsions, circulatory collapse/shock, pulmonary oedema, abnormal bleeding, jaundice, heamoglobinuria, and/or complicated anemia classify the patient as suffering from complicated malaria (WHO, 2000) (**Table 1.1**).

Complicated malaria as a disease is complex and depends on interaction between multiple parasite specific, host specific and environmental specific factors (Miller, et al., 2002). One of the parasite factors associated with complicated malaria is ability of the *P. falciparum* iRBCs to adhere to the host receptors on the endothelium lining of the small blood vessels. Adherence of the infected RBCs can obstruct the blood flow to a particular organ which may lead to organ failures and complicated malaria disease manifestations. The parasite adhesin protein which has extensively been studied and exported to the iRBCs surface is *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Baruch et al., 1995, Su, et al., 1995 and Smith et al., 1995).

Table 1.1. Clinical features associated with complicated malaria (WHO, 2000).

Clinical Symptoms	Features
Cerebral Malaria	Unrousable coma with a Glasgow Coma Scale score ≤ 9
Jaundice	Total serum bilirubin > 3 mg%
Renal Failure	Serum creatinine > 3 mg%
Severe anemia	Hemoglobin < 5 gm%
Abnormal bleeding/Thrombocytopenia	Platelet count < 100000/mm^3
Respiratory distress	The acute lung injury score is calculated on the basis of radiographic densities, severity of hypoxemia , and positive – expiratory pressure.
Repeated generalized convulsions	≥3 convulsions observed within 24 hours
Shock	Systolic blood pressure < 80 mmHg
Multiorgan dysfunction	≥ 2 complications

1.7. Variant *Plasmodium falciparum* surface antigens as virulence factors

PfEMP1 is a group of polymorphic proteins of high molecular weight antigens (200- 360 kDa) encoded by *var* gene family with approximately 60 members per haploid parasite genome (Gardner, et al., 2002). Not only surface location of PfEMP1 allows binding of the iRBCs to the host receptors but also makes it vulnerable to recognition by antibodies. In order to subvert the host immune response, this protein family undergoes clonal antigenic variation (Dzikowski and Deitsch, 2006). *var* genes have a two exon structure. Exon I encodes a variable extracellular part and a transmembrane region and exon II codes for the intracellular and acidic terminal segment. The variable extracellular part contains three domain types i.e. Duffy binding like domains (DBL domains), Cysteine-rich inter-domain regions (CIDR domains) and C2 of which the first two are adhesion domains and can bind to various host receptors simultaneously (Su, et al., 1995 and Smith, et al., 2000). Moreover, both adhesion domains have further been classified into different sequence types (α to ε) and sub-types on the basis of consensus sequence motifs (Smith, et al., 2000) and shown to have specific binding signature to the host receptors (Robinson, et al., 2003

and Springer, et al., 2004) (**Figure 1.6**). For example, CIDR:CD36 and DBLβ: intracellular adhesion molecule 1 (ICAM1) are best characterized pairs of binding interactors. Array of adhesion receptors expressed by endothelial cells which have been shown to serve as ligands for PfEMP1 binding are CD36, ICAM1, VCAM1(Vascular cell adhesion molecule 1), PECAM1 (Platelet/endothelial cell adhesion molecule 1), NCAM (Neuronal cell adhesion molecule), Chondriotin sulfate A (CSA), haemagglutinin (HA), E-selectin (CD62E), Heparan sulphate (Rowe, et al., 2009).

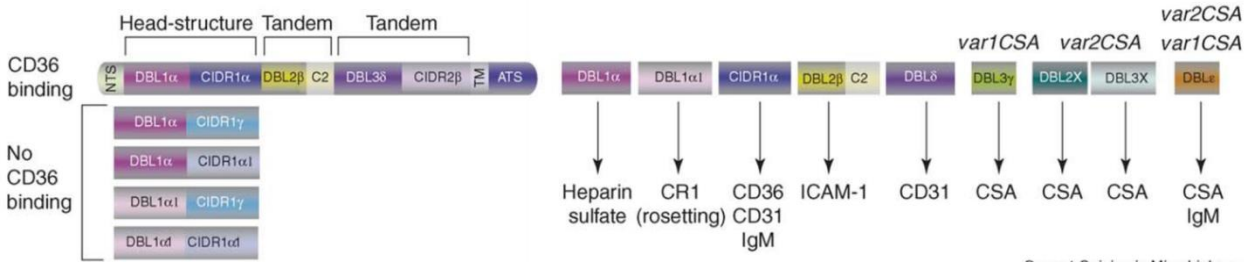


Figure 1.6. PfEMP1 domains and binding properties. Figure summarizes different domains of the PfEMP1 protein along with mapped adhesion properties to different host receptors. Binding is sequence sequence-context- dependent and not all domains of a particular type will bind to the same host receptor. Source: Kraemer and Smith, 2006.

Complete 3D7 genome sequencing project has revealed that most of the *var* genes are located in the highly polymorphic chromosome end regions and few in the central chromosome regions (Gardner, et al., 2002). Moreover, on the basis of 5' up-stream sequence and chromosomal location, *var* genes have been classified into three major (A, B and C) and two intermediate groups (B/A and B/C) (Lavstsen, et al., 2003). To elaborate, group A *var* genes are flanked by Ups A sequences present in sub-telomeric regions and transcribed towards telomere. In contrast, group B *var* genes are present in telomeric chromosomal regions, have Ups B flanking sequences and transcribed towards centromere. Group C *var* genes are centromeric in nature and are flanked by Ups C flanking sequences. Group B/A genes are similar in characteristics to group B genes but are sub-telomeric in nature like group A genes. Group B/C genes have Ups B flanking sequences, but are centromeric in nature like group C genes. In addition, the parasite has three unusual *var* gene sequences that have been classified into independent groups i.e. *var* D

(var1CSA), *var* E (var2CSA) and Type 3 *var* which are conserved across parasite isolates (Lavstsen, et al., 2003).

It has been proposed that aforementioned genomic organization might allow recombination of *var* genes from the same group and maintain separately recombining *var* groups. It is likely that group specific recombination might have functional implications. For example, member of *var* B, B/C and C effectively bind to host endothelial receptor CD36 whereas *var* group A members do not (Robinson, et al., 2003). Thus, recombination among members of the same group adds new *var* forms to the *var* repertoire which in turn generates enormous diversity but at the same time preserves the group specific functional and structural characteristics. To mention the host cell adhesion properties of *var* groups, members of group A genes are ICAM1, PECAM1 and CR1 binders, members of group B, C and B/C are CD 36 binders and single member group D (var1CSA) and group E (var2CSA) are CSA binders (Kraemer and Smith, 2006).

Distinct patterns of sequestration along with pathological consequences can be seen in the host with the expression of PfEMP1 proteins from a particular *var* group and expression of compatible host receptors in organs (**Figure 1.7**). This has clearly been seen in case of pregnancy associated malaria where over expression of var2CSA was observed (Rowe, et al., 2004 and Tuikue Ndam, et al, 2008). Infected RBCs expressing var2CSA bind to CSA on placental syncytiotrophoblasts which in turn lead to low birth weight and associated increase in neonatal mortality (Rogerson, et al., 2007). Additionally, expression of PfEMP1 from other *var* groups has also been associated with complicated malaria disease (cerebral malaria) with inconsistent results. Mostly, group A PfEMP1 have been linked to complicated childhood malaria (Bull, et al., 2000 and Nielson, et al., 2002). Studies from Tanzania suggest the involvement of *var* B genes along with *var* A genes in children with complicated malaria (Bull, et al., 2005). However, the same was not observed in the studies from Kenya (Rottman et al., 2006). Studies from Papua New Guinea suggest the involvement of *var* B genes alone in children with complicated malaria (Kaestli, et al., 2004). Although, it is evident from the above studies that parasite sequestration leading to microvascular obstruction is crucial in developing complicated malaria, the precise mechanism by which parasite exacerbate the disease process is under debate.

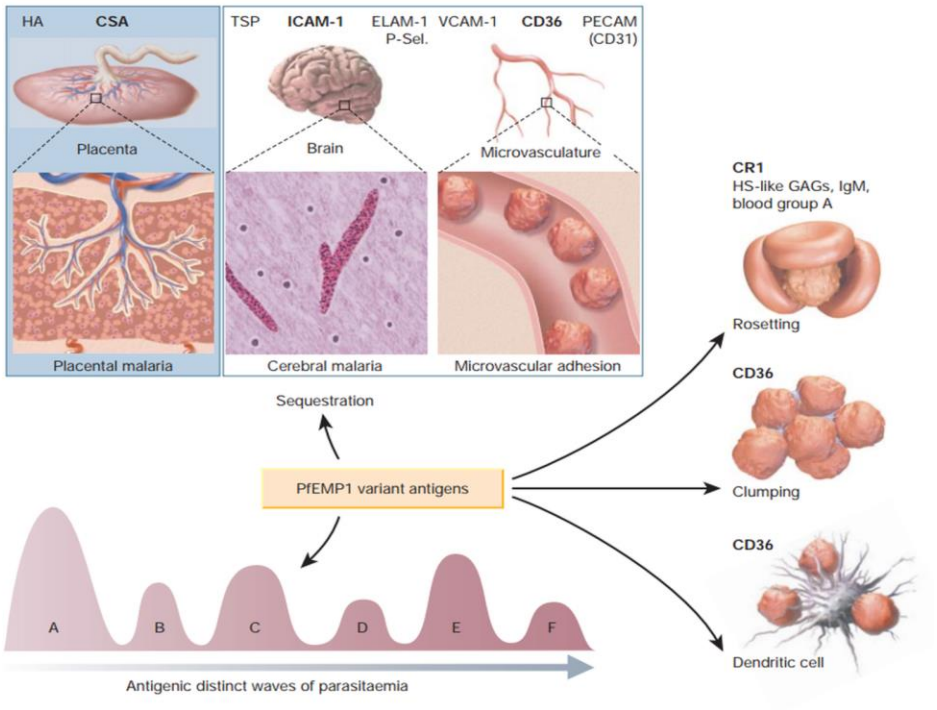


Figure 1.7. The PfEMP1 binding and pathogenesis. PfEMP1 with different binding signatures can bind to host receptors specifically or more abundantly expressed on the surface endothelium of particular organs. ICAM1 binding PfEMP1and CSA binding PfEMP1can sequester the infected RBCs in the brain and in placenta and may lead the patient to develop cerebral malaria and pregnancy associated malaria respectively. Simultaneous binding of PfEMP1 to several receptors, binding to uninfected RBCs (resetting), and clumping of infected RBCs thorough platelets are associated with the pathogenesis of malaria. The host immune response may get down regulated by binding of infected RBCs to the dendritic cells via CD36 receptors. Abbreviatiosn: HA, hyaluronic acid; TSP, thrombospondin; ELAM-1, endothelial/leukocyte adehesion molecule 1; P-Sel, P-selectin; VCAM -1, vascular cell adhesion molecule 1; PECAM (CD31), platelet endothelial cell adhesion molecule 1; CR1, complement receptor 1; HS-like GAGs, heparin sulfate-like glycosaminoglycans; IgM, immunoglobulin M. **Source**: Miller, et al., 2002

P. falciparum also expresses other sub-telomeric variant proteins like RIFINS and STEVORS that are linked to *var* genes but their exact biological role(s) has not been established yet. Both the proteins exhibit a similar two exon structure, a short exon I codes for a signal peptide and the larger exon II codes for a polypeptide with two transmembrane domains flanking a hypervariable region (Blythe, et al., 2004). RIFINs are encoded by *rif* gene family with around 150-500 copies in the genome (160 copies in 3D7 genome) of a single parasite. This is the largest multicopy gene family reported in *P. falciparum* yet. RIFINS have been shown to be expressed on the surface of iRBCs and to be clonally variable (Fernandez, et al., 1999 and Kyes et al., 1999).

Moreover, RIFINs are recognized by antibodies in hyper-immune sera from adult patients suggesting that they are antigenic as well as exposed to the immune system of the host (Abdel-Latif, et al., 2003 and Abdel-Latif, et al., 2002).

STEVORs are encoded by *stevor* genes with 28 copies in the genome of 3D7 reference strain (Gardner, et al., 2002). Recently, it has been demonstrated that STEVOR proteins are also exposed on the iRBC surface and their expression is clonally variant. More recent report suggested a dual role of STEVOR where its expression on iRBC facilitates rosette formation (binding of infected RBCs to uninfected RBCs) in schizont stage and in merozoite invasion (Niang, et al., 2014). Presence of these variant proteins on iRBC surface, their numbers, expression pattern, recognition by host antibodies and having hyper variable region strongly support the role of these proteins in antigenic variation and immune evasion.

Although complicated malaria (cerebral malaria, pregnancy associated malaria) can partly be explained by the process of antigenic variation and obstruction of the host vasculature due to adhesion of iRBCs, other parasite factors like differential multiplication rates (Chotivanich, et al., 2000) and use of multiple invasion pathways (Okoyeh, et al., 1999 and Dolan, et al., 1990) may also influence the outcome of the disease. Much of our understanding about the pathogenic basis of complicated malaria has come from studies involving *in vitro* culture of the parasite and animal models. However, the ultimate understating needs to come from an actual host context (human) for which data are limited.

1.8. *In vivo* conditions: How it differs from the *Plasmodium falciparum in vitro* culture condition?

In vitro culture conditions provide a highly suitable environment to organisms. It is important to note that establishment of continuous *P. falciparum* culture system has enormously helped the malaria research community for understanding the parasite developmental biology by conducting molecular biology and functional genomics studies, and testing antimalarial compounds. However, the *in vitro* culture condition substantially differs from the *in vivo* physiological conditions of the host where parasite resides (**Figure 1.8**, LeRoux, et al., 2009). Differences

between *in vitro* and *in vivo* conditions could reflect changes in parasite biology and hence the disease outcome of the patients too.

The *in vivo* environment differs in many ways compared to *in vitro* environment and includes difference in concentration of glucose (preferred carbon source) along with presence of other carbon sources *in vivo*, difference in concentration of hypoxanthine (preferred source for purine synthesis) along with presence of other purines (i.e. adenosine and inosine) in serum of patients that can be utilized by the parasite as a purine source and differences in the lipid milieu. Additional factors which are present exclusively in the *in vivo* system include inflammatory mediators, immune effector molecules, fluctuating temperature, hormones and metabolites. Parasites travel inside the blood through arterial and venous circulatory systems and in due course face changes in oxygen tension and pH of the blood. The different sequestration sites of the parasite inside the host (i.e. brain, liver, skin, placenta, gut and spleen) also present different microenvironments.

Several studies have indicated that change in the environment has detectable impact on parasite biology (reviewed by LeRoux, et al., 2009). A transcriptome study comparing the expression pattern of sporozoite stage, co-cultured with hepatocyte and without hepatocyte in media identified as many as 611 differentially regulated genes (Siau, et al., 2007). The other observation is that co-cultured sporozoites have an extended period of viability and hepatocyte infectivity. Parasites in culture conditions are grown in 5% hematocrit but an infected host may have 40-50% hematocrit or less in anemic condition; however 5% is relatively rare. Hematocrit can play a major role in determining the strength of infected RBCs adhesion to host receptors. Twelve fold increase in infected RBC adhesion was observed when hematocrit percentage in media was increased from 10 to 30 (Flatt, et al., 2005). Although the parasites *in vitro* are grown in an enriched glucose condition (11.1 mM), the host physiological glucose level is a bit lower and may vary from 4 to 8 mM. It has been reported that 560 genes of *P. falciparum* showed multifold change in their expression when cultivated under low glucose levels of 0.6 mM compared to standard *in vitro* conditions containing 11.1 mM glucose (Fang, et al., 2004).

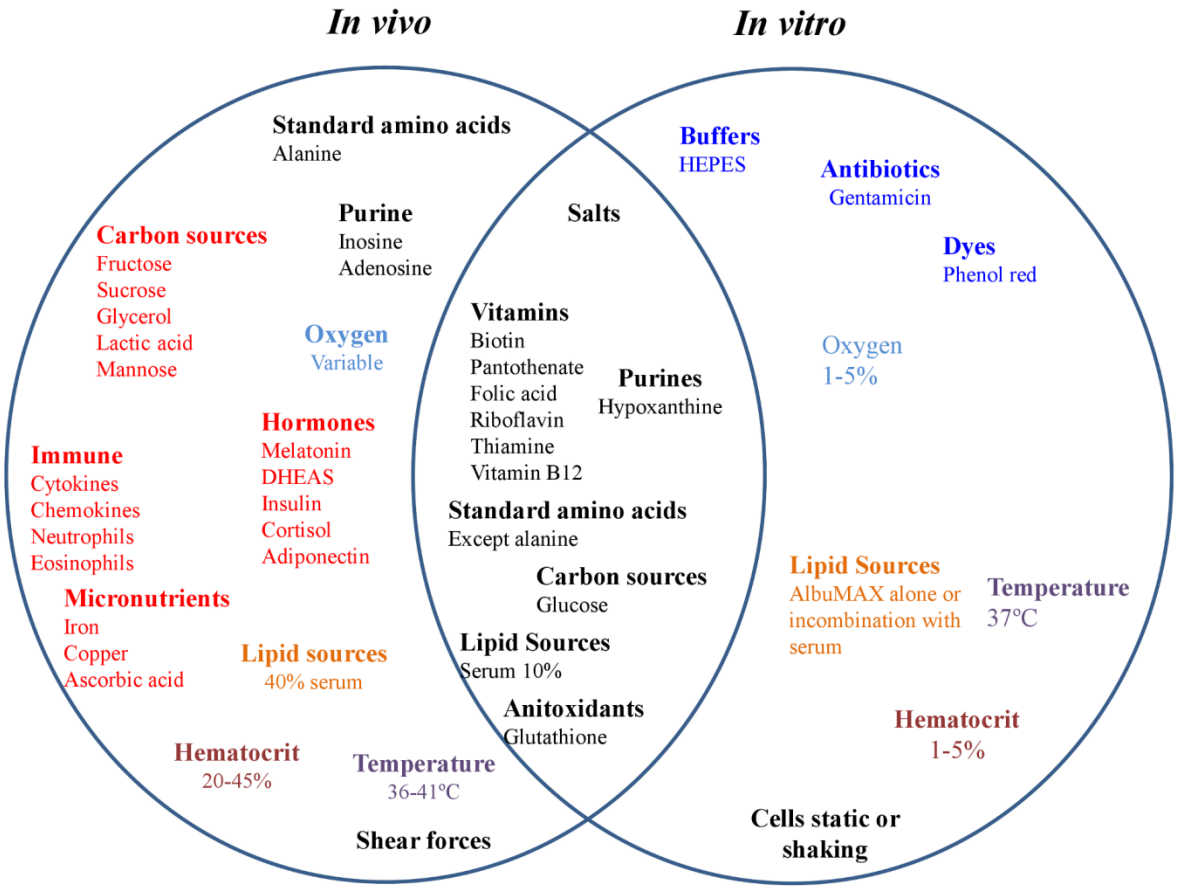


Figure 1.8. Comparison between *in vitro* and *in vivo* growth conditions of *Plasmodium falciparum*. **Source**: Le Roux, et al., 2009. Figure depicts the commonality and difference between *in vitro* and *in vivo* growth conditions of *P. falciparum*. There are additional factors and molecules (letters marked as red) that are not present in *in vitro* conditions. Although some of the factors and molecules like RBC, oxygen and glucose that are present common in both the conditions but fluctuates sometimes in certain *in vivo* host conditions.

A shift from asexual stage to sexual stage development has also been observed *in vitro* with low glucose concentrations indicating the effect of energy source on production of transmittable parasite stage (i.e. male and female gametocyte) (Fang, et al., 2004). Parasites *in vitro* are supplied with only glucose, but the host blood contains other carbon sources that might be utilized by the parasite as source of energy. The probability of utilizing these carbon sources other than glucose is supported by the presence of a putative galactose transporter (PF11_0141) and a putative mannose transporter (PFB0535w) in *P. falciparum* genome (Martin, et al., 2005). If this parasite is utilizing alternate carbon sources, it would be interesting to study and understand the complexities of energy metabolism in this parasite and its implications *in vivo*. Change in source of lipid as supplement from human serum to AlbuMAX *in vitro* had an impact on surface presentation of PfEMP1 and adherence to human endothelial cells (Frankland, et al., 2007).

Temperature is another factor which fluctuates in *in vivo* rising up to 40 ºC during fever in malaria infected patients whereas temperature *in vitro* remains constant maintained at 37 ºC. Febrile episodes of temperature allow synchronization of the parasites by eliminating non ring asexual stages (Gravenor and Kwiatkowski, 1998). It has been observed that parasites undergo transcriptional changes under febrile temperature, which significantly affects major parasite biological networks, most notably affecting genes involved in infected erythrocyte remodeling (Oakley, et al., 2007). A large transcriptome study by Daily, et al., (2007) identified parasites to be present in at least 3 physiological states *in vivo*. First state resembles to growth which depends on glycolytic metabolism, second state resembles to a starvation state further accompanied by utilization of alternate carbon sources and a third state resembles to an environmental stress response.

1.9. Genome of *Plasmodium falciparum*

Plasmodium falciparum has three types of genomes present in its nucleus (nuclear genome), apicoplast (apicoplast genome) and mitochondria (mitochondrial genome).

1.9.1. Nuclear genome

In 1996, collaborations between three international centers (Sanger Center, Stanford University and The Institute of Genomic Research) in genome sequencing temp started the *P. falciparum* genome sequence project to unearth unknown information. Of the 14 chromosomes that *P. falciparum* harbor, sequencing of chromosome 2 and 3 was completed first (Gardner, et al., 1998 and Bowman, et al., 1999) followed by sequencing of other 12 chromosomes (Gardner, et al., 2002a and 2002b and Hall, et al., 2002). The overall analysis of the whole genome was carried out to identify various features of the genome. The size of the genome is approximately 22.8 mb contained in 14 chromosomes. The *P. falciparum* genome is the most A+T rich genome sequenced till date. The overall A+T composition is 80.6% and increases to nearly 90 % in introns and intergenic regions. Gene prediction algorithms identified approximately 5300 protein coding genes with an average gene density of 1 gene per 4,338 base pairs. Introns are present in 54% of its genes and the average gene length is 2.3 kb that is substantially larger than the other sequenced organisms (1.3 – 1.6 kb). About 60% protein coding genes have little or no similarity to proteins in other sequenced organisms, annotating them as protein with unknown function (hypothetical proteins). All the 14 chromosomes are monocentric in nature with centromere size of approximately 3kb having extreme (A+T) richness. Pathologically, telomeres are thought to be the important part of the genome where many variant surface antigen gene families (i.e. *var*, *rifin* and *stevor*) are located. Members of these are known to involve in antigenic variation and immune evasion (Smith, et al., 2013). The presence of these variant genes in sub-telomeric regions gives an added advantage as these regions recombine more frequently than other regions to generate large population of variant gene products playing crucial role in host immune evasion (Scherf, et al., 2008) and are also implicated in disease pathology.

1.9.2. Mitochondrial genome

The mitochondrial genome is small in size (6 kb) and encodes only three proteins i.e. cytochrome oxidase I and III (cox I and III) and cytochrome b. The genome encodes small fragmented rRNAs and no tRNAs.

1.9.3. Apicoplast genome

Apicoplast, a relict four membrane plastid is homologous with the chloroplasts of plants and algae. It is an organelle essential for parasite survival in both of its erythrocytic (Goodman, et al., 2007) and liver stages (Vaughan, et al., 2009, and Yu, et al., 2008). The apicoplast harbors many plastid derived biochemical pathways including fatty acids II biosynthesis, isopernoid, haem and iron-sulfur clusters and is a major drug target. It has a 35 kb circular genome, one of the smallest genome known so far (Wilson, et al, 1996). The plastid genome is high in A+T content (86.9 %). The 35 kb circular genome encodes all the genes involved in gene expression: three subunits of eubacterial type RNA polymerase (rpoB, C1 and C2), 17 ribosomal proteins, the elongation factor Tu (EF-Tu), tRNAs (n =25), large and small subunit (LSU and SSU) rRNAs.

1.10. Functional genomics of *Plasmodium falciparum* – A transcriptomic view

Functional genomics, is a branch which utilizes wealth of genome sequence information to deduce the function of elements like genes and proteins. Functional genomics study evaluates the dynamic properties of functional elements like transcriptional, translational, post-transcriptional modification, protein-DNA and protein-protein interactions. It further attempts to reconstruct a system of biological functionalities in order to apprehend the intricate relationship between these functional elements. As this approach looks at a genome wide scale, technical approaches which can evaluate the entire system of a cell or unicellular organisms are being used.

Functional genomics approaches often depends on genome sequence information to design tools like microarrays and/or final data analysis approaches like next generation based RNA-Sequencing and mass spectrometry based proteomic analysis. In case of *P. falciparum*, completion of genome sequencing in the year 2002 (Gardner, et al., 2002) triggered the functional genomics study to be applied on this deadly parasite. Microarray based gene expression analysis of *P. falciparum* asexual stages were first conducted in the year 2000 (before

release of the first draft of *P. falciparum* genome sequence) using DNA microarrays generated using either sheared DNA fragments or cDNAs (Hayward, et al, 2000 and Ben Mamoun, et al., 2001). With the completion of *P. falciparum* 3D7 reference genome sequencing project, two DNA microarray platforms were created: A high density oligonucleotide microarray on AFFYMATRIX platform (Le Roch, et al, 2003) and a long oligonucleotide DNA microarray based on DNA spotting technology (Bozdech, et al., 2003). Utilizing these microarrays, transcriptional profiling of parasite's blood stage asexual development was carried out which indicated the stage specific expression pattern of many genes.

An original study by Bozdech, et al., 2003 captured transcriptional profile of 48 hours intra-erythrocytic developmental cycle (IDC) at single-hour resolution. This study demonstrated that at least 60% of the genome is transcriptionally active during this stage exhibiting single peak mRNA abundance profile. A parallel study by Le Roch, et al., 2003 using different developmental stages of the parasite identified higher number of genes to be expressed (88%) in at least one stage with expression values that may vary up to five orders of magnitude. Subsequent studies identified periodic transcriptional program of the IDC to be conserved across laboratory strains (Linas, et al., 2006) and freshly adapted field isolates (Mackinnon, et al., 2009). This suggests that the IDC based transcriptional cascade is driven by conserved transcriptional programs that does not vary amongst different field or laboratory isolates. Other microarray studies have focused on the effect of perturbagens on transcriptional variations of which some studies could not find any subtle differences (Ralph, et al., 2005, Le Roch, et al., 2008, and Ganesan, et al., 2008), whereas some studies could identify biologically relevant transcriptional changes that have been functionally associated to transcriptionally corregulated genes (Oakley, et al., 2007, Natalang, et al., 2008, Tamez, et al., 2008, Hu, et al., 2010 and Chaal, et al., 2010). Microarray based studies have also shown its ability to capture the transcriptional profiles that are unique to parasites *in vivo* compared to the parasites maintained *in vitro* (Daily, et al., 2005, Siau, et al., 2007, and Daily, et al., 2007).

The advent of next generation sequencing (NGS) technology has further contributed in understanding of *Plasmodium* transcriptome biology (Otto, et al., 2010 and Sorber, et al., 2011). NGS-based RNA-Seq has recently validated the microarray based IDC transcriptome data (Otto, et al., 2010 and Sorber, et al., 2011). The number of genes detected in IDC stage increased to 90

% of the genome that can be attributed to its higher sensitivity. Moreover, RNA-Seq studies have helped in detecting new open reading frames, corrected as many as 423 existing gene models and confirmed 75 % of mRNA splicing sites originally predicted by the *in silico* gene models.

It is important to note that, all of the above studies suggests *Plasmodium* transcriptome to be of highly dynamic in nature and indicates about the presence of mechanisms which precisely control the mRNA transcript levels. This could be mediated by transcriptional, post-transcriptional and translational regulatory mechanisms.

1.11. Genome regulation in *Plasmodium falciparum*

Chromatin modifications, an epigenetic process is the best characterized transcriptional regulatory mechanism in *P. falciparum*. Recent studies have illustrated the association of specific histone modifications with activation and repression of genes during IDC. Specifically, correlation of histone 3 lysine 9 tri-metylation with silencing of sub-telomeric *var* genes was reported (Lopez-Rubio, et al., 2009 and Salcedo-Amaya, et al., 2009). It has also been demonstrated that *P. falciparum* heterochromatin protein is strongly associated with histone 2 lysine 9 tri-methylation (H3K9me3) mark (Flueck, et al., 2009 and Perez-Toledo, et al., 2009). Recent data from genome-wide mapping of DNA methylation study identified asymmetric methylation pattern of malaria genome with a single *P. falciparum* DNA methyltransferase as the mediator. The data from the study suggests that DNA methylation could regulate *var* gene expression and transcription elongation (Ponts, et al, 2013). However, mechanisms driving expression of other variant surface antigen genes families like *rifin* and *stevor* are not well understood. These gene families shows temporal expression pattern during the IDC where *rifins* are transcribed during early trophozoite stage and *stevors* are transcribed during late trophozoite stage, however the precise mechanisms controlling such expression pattern is still a matter of investigation.

In *P. falciparum*, there are also many reports which indicate towards the role of post-transcriptional mode of gene regulation in development (Mair, et al., 2006 and Mair, et al., 2010). One such instance is translational repression of many proteins in female gametocytes by translational repression complex that includes the DEAD-box RNA helicase (PfDOZI) until a precise time of sexual development is attended (Mair, et al., 2006). Another mode of post-

transcriptional gene regulation could be due to differences in mRNA stability or decay during different developmental stages. The average mRNA half-life is approximately 9.5 minutes during the ring stage, whereas it reaches to an average half-life of 65 minutes during late schizont stage (Shock, et al., 2007). Until a few years ago, there was a significant gap in our understanding of transcription factor based gene regulation *in P. falciparum* due to identification of few specific annotated transcription factors and regulatory cis-acting elements. With the discovery of 27 Apicomplexan AP2 (ApiAP2) family of DNA binding proteins (Balaji, et al., 2005) as candidate transcription factors and work from several laboratories suggesting its role in transcriptional regulation and gene silencing, the gap has been narrowed down (Behnke, et al., 2010, Flueck, et al., 2010, Yuda, et al., 2009, Yuda, et al., 2010 and Balu, et al., 2009).

Other than the above known and postulated mechanisms behind *Plasmodium* gene regulation, Non-protein coding RNAs (npc-RNAs) in *P. falciparum* are also emerging as a new player that has the potential to regulate gene expression in many ways (Broadbent, et al., 2011, Raabe, et al., 2010, Lopez- Baraggan, et al., 2011 and Sorber, et al., 2011). As per the name, these RNAs do not code for proteins but rather perform various regulatory roles itself or in combination with RNA binding proteins (RBPs). npcRNAs can be divided into two broad classes based on their size: short npcRNAs (< 200 nucleotide in length) and long npcRNAs (> 200 nucleotide in length). Short npcRNAs include small interfering RNAs (siRNAs), microRNAs (miRNAs) and PIWI-interacting RNAs (piRNAs). Natural antisense transcripts (NATs) are npcRNAs that are transcribed from the opposite strand of that of sense transcript of either protein coding or non-protein coding genes (Pelechano, et al., 2013) Studies from yeast to humans suggest that npcRNAs participate in the regulation of diverse biochemical pathways including chromatin modifications, splicing, transcription, translation, developmental progression, differentiation, apoptosis and proliferation (Moazed, et al., 2009, Pelechano and Steinmetz, 2013, Prasanth and Spector, 2007, Faghini and Wahlestedt, 2009 and Lapidot and Pilpel, 2006). Expression of NATs is context dependent and can show distinct expression patterns during different process, such as cellular development and differentiation, in different environmental conditions, during disease progression or on different genetic backgrounds and sometimes linked to the activity of neighboring genes (Pelechano and Steinmetz, 2013).

1.12. Systems Network Biology of *Plasmodium falciparum* genome: a tool to interpret large scale data and opportunity for new discoveries

Systems biology can be defined as a science of integrative biology which combines biological information derived from multidisciplinary approaches of biochemical, bioinformatics, transcriptomic and proteomic technologies. This integrative approach enables gene functions to be predicted and networks to be elucidated. Through the use of genome sequence, automation and parallel technologies such as microarrays, large scale protein interaction studies and next generation sequencing, large scale data sets containing consistent and high quality data can be created, assembled and evaluated in a cost saving manner and informative compared to single gene studies. The collective data thus provides an understanding about how an organism operates, how different components of the organism interact (Kitano, 2002). The interactions between various components often visualized in the form of a network where genes or proteins are taken as unit of interaction. It is a prerequisite to have large amount of biological information for an organism before systems biology approaches can be applied. For model organisms like *Escherichia coli*, *Saccharomyces cerevisiae* and *Caenorhabditis elegans* which are well characterized, application of systems biology approaches has unraveled many hidden information. With accumulation of information, the malaria parasite *P. falciparum* stands as a good candidate for analysis by this approach.

After the completion of *P. falciparum* genome sequencing project, at least 60% of the genes identified were described as 'hypothetical' as they did not show significant homology to characterized genes from other species (Gardner, et al., 2002). This tremendous paucity in knowledge about the functions of large number of protein encoding genes urge for large scale functional genomics studies to be applied on this deadly parasite. High throughput transcriptomics (Bozdech, et al, 2003, Le Roch et al, 2003, Hu, et al., 2007), proteomics (Lasonder et al., 2002, Florens et al., 2002, Oehring, et al., 2012, Pease et al., 2013) and protein-protein interaction studies (LaCount, et al., 2005) along with computational approaches (Hu, et al., 2009) have generated huge amount of data which are now being used for systems level analysis and can be integrated together for systems level understanding of *P. falciparum* biology (**Figure 1.9**). Microarrays based gene expression analysis in this parasite provided a first step towards the goal of uncovering gene function in a global scale (Le Roch, et al., 2003). Further

contribution to gene function prediction came from a *P. falciparum* protein-protein interaction network based on 32,000 yeast two-hybrid screens with *P. falciparum* protein fragments (La Count, et al., 2005). Subsequently, other network based approaches utilizing computational modelling and integration of various high throughput datasets (primarily transcriptional profiling datasets along with sequence homology, domain-domain and yeast two-hybrid datasets) could enable prediction of the function of a large number of hypothetical proteins with more confidence (Date, et al., 2006, Wuchty and Ipsaro, 2007, Zhou, et al., 2008 and Hu, et al., 2009).

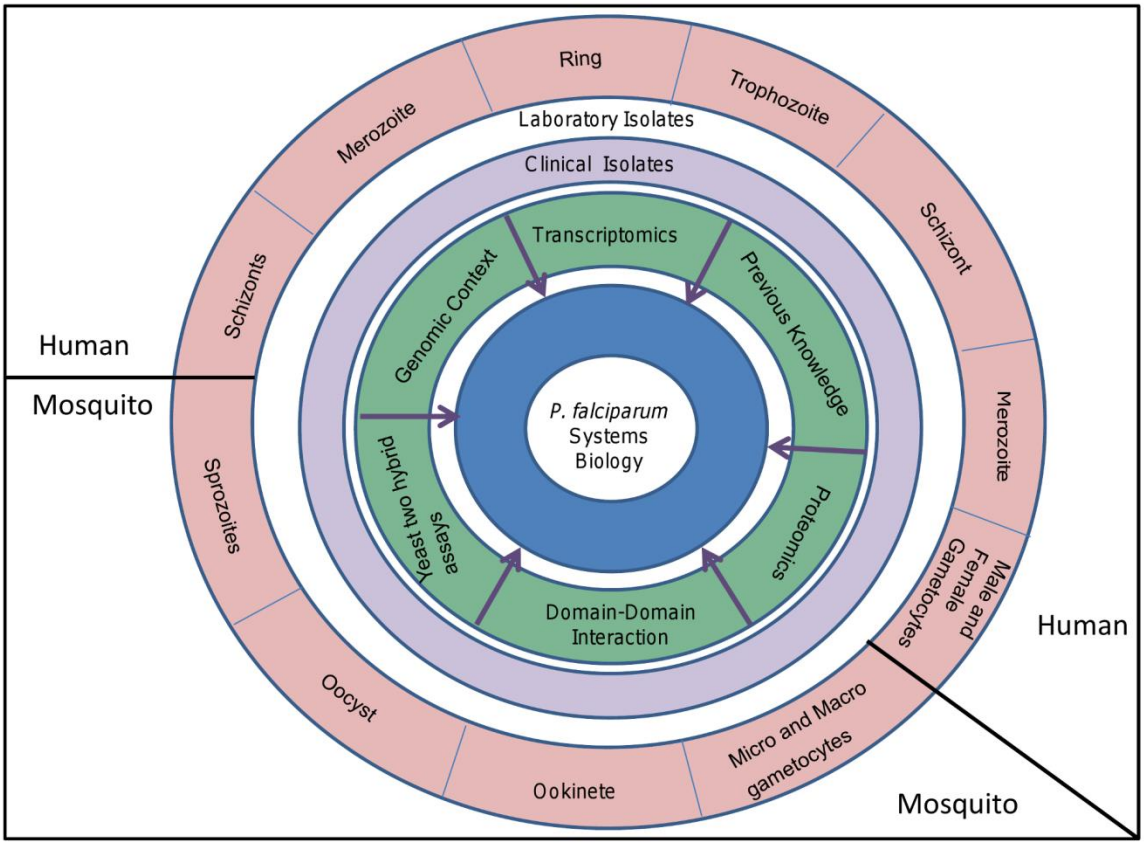


Figure 1.9. An overview of systems biology approach for understanding *Plasmodium falciparum* biology. An integrated approach incorporating data from the life cycle stages of the parasite in its vertebrate and invertebrate hosts, in culture condition with or without perturbagens would provide novel insights into the molecular biology of this parasite. **Source**: Adapted from Das, et al, 2008.

1.13. Research gaps

Understanding the transcriptome of this deadly parasite is one of the important aspects to decipher many unanswered questions in the field of malaria related disease pathogenesis. The major contribution for which has come from microarray studies. It has contributed in decoding many facets of parasite transcriptome biology including the stage specific gene expression pattern, response of genes to drug treatments, expression pattern of variant surface antigens and identifying multiple physiological states of the parasite in diseased hosts. However, this high throughput technique has a limitation as it depends on genomic sequences, utilizing which probes in the array are designed to capture transcripts. All microarrays designed till date for *P. falciparum* transcriptome studies were based on the single reference strain (3D7) genome sequence (Bozdech, 2003 et al., Le Roch et al., 2003, Kafsack, et al., 2012). Therefore, this limits the utilization of these microarrays in studying transcriptome of field/clinical isolates which have been reported to show sequence variations in their genomes (Pologe and Ravetch, 1988, Scherf and Mattei, 1992, Kidgell, et al., 2006, Volkman, et al., 2006 and Manske, et al., 2012). In order to capture the transcripts of *P. falciparum* field isolates efficiently, it would be prudent to include sequence information from multiple *P. falciparum* strains using which probes can be designed. Best probes can then further be selected based on their hybridization potential to transcripts of field isolates.

Malaria related disease severity in the host is orchestrated by interplay between factors that are of parasite, host and environmental origin. Through decades of research on this deadly parasite, it is now clear that disease progression inside the host is context dependent. Enormous efforts have been made to understand the life cycle and biology of this parasite through the use of molecular biology, biochemistry, genomics, transcriptomics and proteomics based approaches, which has taken the advantage of the established *P. falciparum* parasite culture system in laboratory settings. Almost all of these experiments were carried out by utilizing laboratory grown parasites. Studies in other systems have clearly shown that organisms *in vivo* have distinct biology compare to *in vitro* and some of these difference relate to pathogenesis/virulence (Mahan, et al., 1993).

In case of *P. falciparum*, little is known about its *in vivo* parasite biology. Few experiments that have utilized parasites, isolated directly from infected patients have shown that the parasite

inside the human host exists in at least three distinct physiological states (Daily, et al., 2007) and over express transcripts that encode for surface antigens (Daily, et al., 2005). To emphasize, these aforementioned studies have been conducted on parasites derived from *P. falciparum* infected pediatric patients with mild malaria related symptoms. However, a malaria infected patient can show diverse level of disease severity and it would be intriguing to see if the transcriptome biology of this parasite inside the host with two disease conditions (i.e. uncomplicated and complicated malaria) differs. Additionally, in South East Asia and Latin America, where malaria transmission is less intense, complicated malaria is shown by all age groups. However, an area with intense and stable malaria transmission like sub-Saharan Africa, complicated malaria is a disease of children under 5 years.

The complicated malaria related disease presentation by patients in different parts of the world also differs. Adult patients more frequently present hepatic and renal failure as main symptoms whereas pediatric patients more frequently present cerebral malaria, complicated anemia and respiratory distress as main symptoms. While much of the attention has been given to understand the malaria related disease biology in children, little is known about the parasite transcriptome in adult patients that differs immunologically from their pediatric counterpart. Moreover, expression pattern of virulent genes like *var* has extensively been studied in pediatric malaria patients and some of the studies have linked the over-expression of a particular type of *var* member to life threatening malaria conditions in children from areas of stable malaria transmission. No information is available about the expression status of these genes in adult patients with differing disease status. There are also lacunae of understanding about the expression pattern of other variant surface antigen genes like *rifins* and *stevors*, which have been shown to be clonally variant and expressed on the surface of the infected erythrocytes.

High throughput techniques like microarray, next generation sequencing including RNA-Seq, Mass spectrometry and yeast two-hybrid assay produce enormous amount of data pertaining to nature of genes and proteins in a cellular context (Reed, et al., 2006). Such huge data in combination with computational and systems biology approaches have started providing us a comprehensive map of gene and protein interactions networks (Kitano, 2000 and Nurse, 2003). These networks in turn have started providing us the functionalities of cellular and molecular pathways in organisms ranging from bacteria to humans. In case of *P. falciparum*, as mentioned

in the earlier section, employment of systems biology approaches have enabled the scientific community in predicting the functions of many hypothetical proteins, identifying new signaling pathways, surface proteins, transcription factors and identifying genes necessary for host red blood cell invasion (Ramprasad, et al., 2011). All of these studies have utilized transcriptomics, proteomics and yeast two-hybrid data derived from *P. falciparum* laboratory isolates or data generated from bioinformatics based approaches (Ramprasad, et al., 2011). These systems level studies have helped in educating a great deal about overall pathways that the parasite is capable of elaborating during growth. However, their significance in the context of the host (*in vivo*) is still not fully understood.

Questions that still need to be answered: are there specific pathways uniquely expressed in field isolates during severe disease states? If yes, do they link to processes that are critical for parasite survival? Are there specific gene/protein networks which can be linked to a particular disease state (disease specific signature)? Answer to these questions might identify candidates which could be drug targets or vaccine candidates. These questions can however be addressed by experimenting on parasite isolates from patients.

Natural antisense transcripts have recently been recognized as an important player in the field of genome regulation in lower organisms like bacteria to higher organisms like human (Pelechano and Steinmetz, 2013). NATs can regulate gene expression through multiple mechanisms which include DNA methylation, RNA stability, chromosome modifications, transcriptional interference, and translational repression (Pelechano and Steinmetz, 2013). In *P. falciparum*, there is a dearth of knowledge regarding the mechanisms through which this parasite regulates its gene expression. However, evidence suggests that this parasite might be employing chromatin modification, translational repression, transcription factors and non-coding RNAs based mechanisms to regulate its gene expression. The presence of alternate mechanism (non-coding RNA based) is evident from the discovery of non-coding RNAs including NATs in *P. falciparum* in culture condition (Broadbent, et al., 2011, Raabe et al., 2009, Lopez- Baraggan, et al., 2011 and Sorber, et al., 2011). Although NATs have been identified in multiple developmental stages of *P. falciparum* laboratory isolates, its presence and numbers in clinical isolates from patients with diverse disease conditions has not been documented yet. It would be

intriguing to dissect the diversity of NATs in clinical isolates and its further functional analysis might shed light on its possible role in gene regulation and disease progression.

Objectives

In order to address some of the aforementioned research gaps, the objectives identified for this study were

1. Validation of a custom cross strain *Plasmodium falciparum* microarray for the efficient detection of transcripts from clinical isolates.
2. Exploring the diversity of natural antisense transcript population in *Plasmodium falciparum* clinical isolates from adult patients with different malaria related disease symptoms.
3. Investigation and comparison between the *in vivo* transcriptome of clinical isolates from adult patients with uncomplicated or complicated malaria symptoms.
4. Creation of a gene co-expression based systems network by utilizing the *in vivo* parasite transcriptome data in order to identify disease specific modules thereby identify biological processes and probable candidate genes for intervention strategies.

2.1. Study site

Blood samples were collected from adult patients admitted in the classified malaria wards of Prince Bijay Singh Memorial (PBM) Hospital, S. P. Medical College, Bikaner, Rajasthan. Bikaner district is a part of the Thar Desert, Rajasthan, India having extremes of temperature. This region is regarded as a hypoendemic area for malaria and is basically an arid zone with recent changes in ecosystem due to increased rainfall and canal irrigation in the last three decades. This region also reported the changes in representation of clinical spectrum of severe *P. falciparum* malaria with a significant increase in jaundice and renal failure as important manifestations of severe malaria (Kochar, et al., 2006).

2.2. Patients, blood sample collection and processing

Patients who presented to the S.P. Medical College, Bikaner were triaged by the medical staffs to undergo diagnostic tests if they had symptoms suggestive of malaria. To check the presence of *P. falciparum* infection, a preliminary screening was done using slide microscopy and rapid diagnostic tests (RDTs) based on detection of specific *Plasmodium* spp. Lactate dehydrogenase (Optimal test; Diamed AG, Cressiersur Morat, Switzerland) and histidine rich protein 2 (HRP2) (Falcivax test; Zephyr Biomedical System, Goa, India) in the hospital. The patient samples were collected on informed consent by a team of clinicians at S. P. Medical College, Bikaner according to Hospital guidelines. Patient's samples showing positive results for *P. falciparum* infection were categorized as uncomplicated or complicated malaria based on guidelines for severe malaria (WHO, 2000). Patients with one or more manifestations of severe malaria, which include cerebral malaria (Glasgow coma score $\leq$ 9/14), jaundice (serum bilirubin > 3 mg%), renal failure (serum creatinine > 3 mg%), severe anemia (hemoglobin < 5 mg%), thrombocytopenia (platelet count < 100,000/mm^3), acute respiratory distress syndrome, circulatory collapse were managed according to the reported guidelines (WHO, 2010) and the rest patients were considered as uncomplicated malaria cases. Venous blood samples were collected (5ml and 1ml in separate tubes) from *P. falciparum* infected adult patients (n=44) during 4 transmission seasons (2007-2010).

5 ml blood was immediately (within 15 minutes of collection) subjected to Histopaque density gradient (Histopaque, Sigma Aldrich, USA) to separate the peripheral blood mononuclear cells

(PBMCs) from the RBCs following manufacturer's instructions. Both fractions (PBMCs and RBCs) were washed twice with Phosphate Buffered Saline (PBS) and lysed using four volumes of Tri-Reagent (Sigma Aldrich, USA) and stored at − 80 °C. Subsequently, these frozen samples were transported in cold chain to our lab for further processing.

Laboratory investigations to rule out possibility of any other cause for symptoms exhibited (complicated *P. falciparum* cases) were done as described by Kochar, et al (2005 and 2010). Briefly, appropriate blood test to rule out typhoid fever (widal test), leptospirosis (enzyme immunoassay for the differential detection of IgG and IgM antibodies), infectious mononucleosis (monospot test), and dengue infection (differential detection of IgG and IgM antibodies) were carried out.

2.3. Total RNA preparation from *P. falciparum* infected blood

List of reagents used:

1. Tri-Reagent (Sigma Aldrich, USA)
2. Chloroform
3. Isopropanol
4. 70 % Ethanol
5. Aboslute Ethanol (Molecular Biology Grade, Amresco)
6. DEPC treated MiliQ water
7. 3M Sodium Acetate (pH 5.2)

The frozen RBC samples were thawed at room temperature and transferred to 50 ml Oakridge tubes. To each sample, 0.2 ml of chloroform per ml of TRI Reagent used was added and shaken vigorously for 15 seconds. Samples were incubated for 10 minutes at room temperature and centrifuged at 12,000 × g for 10 minutes. This separates the mixture into three phases/layers: organic phase at the bottom, aqueous phase at the top and inter-phase at middle. Aqueous layer was transferred to a separate tube and 0.5 ml of isopropanol per ml of Tri-reagent used was added. Sample mixture was mixed by inversion and kept at room temperature for 20 minutes to precipitate RNA. The RNA was collected by centrifuging at 12,000 × g for 10 minutes, washed twice with 70% ethanol and air dried for about 10 minutes. The RNA pellet was dissolved in 90 µl of DEPC treated MiliQ water and heating at 65 °C for 10 minutes in a water bath. Five µl of RNA sample was kept separately in a DEPC treated PCR tube for denaturing agarose gel based analysis to check the integrity of total RNA. For long term storage, two volumes of absolute

ethanol and 10% v/v of 3M Sodium acetate (pH 5.2) was added to the rest of the RNA sample, mixed by inversion and kept in − 80 °C deep freezer.

2.4. DNA extraction using TRI reagent

List of reagents used:

1. Back Extraction Buffer (pH 9.5)
2. Isopropanol
3. 70 % Ethanol
4. Tris-EDTA (TE) buffer (pH 8.0)

After taking the aqueous layer from the sample mixture in the RNA extraction step, remaining small amount of aqueous layer was discarded. To the mixture 0.5 ml of back extraction buffer was added per ml of TRI Reagent used. The tubes were shaken vigorously for 15 seconds. Samples were incubated for 10 minutes at room temperature and centrifuged at 12,000 × g for 10 minutes. Aqueous layer was transferred to a separate tube and 0.5 ml of isopropanol per ml of Tri-reagent used was added. Sample mixture was mixed by inversion and kept at room temperature for 20 minutes for precipitation of DNA. The DNA was collected by centrifuging at 12,000 × g for 10 minutes, washed twice with 70% ethanol and air dried for about 30 minutes. The DNA pellet was dissolved by adding 90 µl of 1X TE buffer and kept at – 80 °C freezer till further use.

2.5. DNA extraction from whole blood

List of reagents used:

1. Lysis buffer A: 10 mMNaCl, 50 mM Tris, 10 mM EDTA
2. Lysis buffer B: 10mM NaCl, 50 mM Tris, 10 mM EDTA, 1% SDS
3. Proteinase K (20mg/ml stock solution in MiliQ water)
4. Phenol- Tris saturated solution (pH 8.0)
5. Chloroform ; Iso-Amyl Alcohol (24:1)
6. Absolute Ethanol (99.5%)
7. Sodium Acetate solution- 3M (pH 5.2)

8. 1X TE buffer (Ph 8.0)

One ml frozen blood samples were thawed at room temperature and split into two 0.5 ml volume and transferred into 2 ml microfuge tubes. Lysis of the cells was carried out using the lysis buffer A and B. Blood sample was treated with these lysis buffers at 37 °C for 1 hour, with thorough intermittent mixing by gentle tapping or inversions. Proteinase K (100 µg/ml) was added and the material was mixed by inversion and incubated for 1 hour at 50°C. Phenol: Chloroform: Iso-amyl alcohol (25:24:1) extraction was performed to obtain the aqueous phase. Two volumes of chilled absolute ethanol and 1/10 volume of Sodium acetate was added to the aqueous phase and the preparation was stored at − 20 °C overnight for DNA precipitation. The DNA was pelleted by centrifugation at 12,000 × g for 10 minutes. The pellet was washed twice with 70% ethanol, dried and dissolved in 1X TE buffer (pH 8.0). Diagnostic PCRs were performed to confirm *P. falciparum* infection as described elsewhere (Pakalpati, et al., 2013a &b, and Das, et al., 1995).

2.6. Total RNA detection using formaldehyde based denaturing agarose gel

List of reagents used:

1. Formaldehyde (40% v/v)
2. Formamide (distilled-deionized)
3. 10X Formaldehyde gel loading buffer
4. 10X MOPS electrophoresis buffer
5. Ethidium bromide (200 µg/ml): (Prepared in DEPC treated water)

For preparing 100 ml of 1.5% agarose gel containing 2.2M Formaldehyde, 1.5g of agarose was added to 72 ml of DEPC treated autoclaved water. Agarose was melted in a microwave oven. Solution was cooled to 55 °C and 10 ml of 10X MOPS electrophoresis buffer and 18 ml of deionized formaldehyde was added. Agarose gel was casted in a chemical fume hood and allowed to set for at least 1 hour at room temperature before running. RNA sample preparation was done as mentioned in Sambrook et. al., 2001. Briefly, in sterile microfuge tubes 2 µl of RNA, 2 µl of 10X MOPS electrophoresis buffer, 4µl of formaldehyde, 10µl of Formamide and 1 µl of Ethidium bromide was added. Sample mixture was incubated for 10 minutes at 65 °C and immediately placed on ice for 10 minutes. To each sample mixture, 2µl of 10X formaldehyde gel loading buffer was added and kept in ice bucket. Samples were loaded in the wells of

formaldehyde containing agarose gel placed in an electrophoresis tank containing sufficient 1X MOPS electrophoresis buffer. Gel was run at 5V/cm. RNAs were visualized and documented by placing the gel on a BIO-RAD Chemi-Doc XRS gel documentation system (**Figure 2.1**).

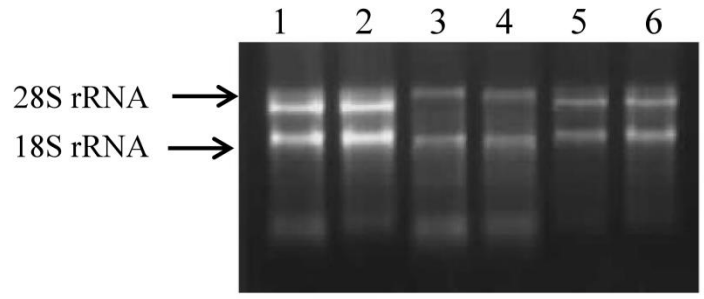


Figure 2.1. **Denaturing agarose gel for the detection of RNA**. Figure shows two sharp 28S and 18S rRNA bands in formaldehyde containing denaturing agarose gel. Lane 1-6 represents total RNA from 6 different *P. falciparum* clinical isolates.

2.7. RNA quantity and quality assessment

RNA samples quantity and quality were measured using automated electrophoresis stations (BIO-RAD Experion Automated electrophoresis station and Agilent's 2000 Bioanalyser). Both operate by using LabChip microfluidic technology. Chips and reagents that have been used to detect RNA can detect 25-500ng of RNA per µl of sample. The Experion automated electrophoresis station was used in the laboratory at BITS, Pilani where as the Agilent's 2000 Bioanalyser was used in Genomics Research facility at Genotypic Technology Pvt. Ltd., Bangalore. These two systems generate three main data (virtual gel image, electropherogram and RNA Quality Index (RQI) number in case of Experion system or RNA Integrity Number in case of Agilent's Bioanalyzer) to determine the RNA quality and quantity. Samples were run according to the manufacturer's protocol to determine the quality and quantity of the RNA samples (**Figure 2.2**). Total RNA purity was assessed by the NanoDrop® ND-1000 UV-Vis Spectrophotometer (Nanodrop technologies, Rockland, USA). RNA samples with RIN/RQI $\geq$ 7 have been used in this study. RNA from 36 samples passed the QC test, out of 44 samples from which RNA was isolated. **Table 2.1** represents clinical details of the 36 QC passed RNA samples.

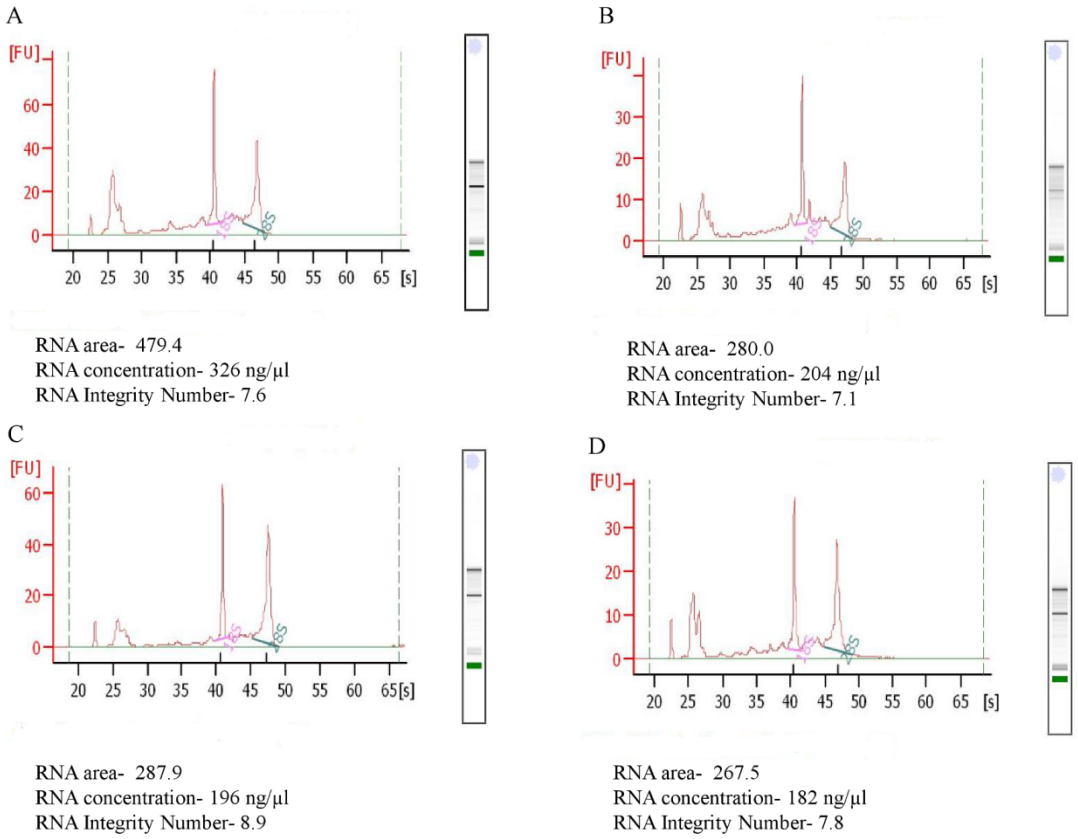


Figure 2.2. Electropherogram showing *Plasmodium falciparum* 28S and 18S rRNA peaks. **A-D**: Electropherogram of 4 representative samples. Electropherogram was generated by running total RNA sample in an Experion automated electrophoresis station (BioRad). Along with electropherogram, automated electrophoresis station provides the data about total RNA concentration and RNA quality.

Table 2.1. Clinical characteristics of patients with *Plasmodium falciparum* infection.

S. No.	Patient ID	Age(Year)/ sex	Clinical Presentation	Diagnostic tests for malaria		
				PBF	RMDT	PCR
			Uncomplicated malaria			
01	PFU-01	21, M	Unc	+	+	+
02	PFU-02	42, M	Unc	+	+	+
03	PFU-03	28, M	Unc	+	+	+
04	PFU-04	17, F	Unc	+	+	+
05	PFU-05	30, M	Unc	+	+	+
06	PFU-06	30, M	Unc	+	+	+
07	PFU-07	70, F	Unc	+	+	+
08	PFU-08	33, M	Unc	+	+	+
			Complicated malaria			
09	PFC-01	15, M	RF (ser.creat.-9), A (Hb- 5.9), T (platelet-51)	+	+	+
10	PFC-06	25, M	J (ser.bil.-3.1)	+	+	+
11	PFC-03	15, M	J(ser.bil.-6.5), RF (ser.creat.-3.5), T (platelet-12)	+	+	+
12	PFC-04	60, M	J (ser.bil.-17.6), RF (ser.creat.-4.8), T (platelet-70)	+	+	+
13	PFC-05	45, F	J (ser.bil.-9.81), RF (ser.creat.-4.7), T (platelet-22)	+	+	+
14	PFC-06	25, M	J (ser.bil.-8.3)	+	+	+
15	PFC-07	55, F	SA (Hb -5)	+	+	+
16	PFC-08	24, M	J (ser.bil.-4.9), T (Platelet-28)	+	+	+
17	PFC-09	29, M	J(ser.bil.-5.6), A (Hb- 6.2), T, (platelet-48)	+	+	+
18	PFC-10	42, F	J(ser.bil.-11), A (Hb -5.2)	+	+	+
19	PFC-11	40, F	J (ser.bil.-7), A (Hb -6)	+	+	+
20	PFC-12	17, F	J (ser.bil.-4.3), A (Hb -5)	+	+	+
21	PFC-13	23, M	J (ser.bil.-5.8)	+	+	+
22	PFC-14	32, M	CM, RF (ser.creat.-3.4), T (platelet-25)	+	+	+
23	PFC-15	23, M	J (ser.bil.- 3.3), T (platelet-38)	+	+	+
24	PFC-16	28, M	CM	+	+	+
25	PFC-17	25, M	J (ser.bil.-3), RF (ser.creat.-3.5)	+	+	+
26	PFC-18	60, M	J (ser.bil-8), RF(ser.creat.-3), T (platelet-71.2)	+	+	+
27	PFC-19	17, M	J (ser.bil.-8.2), RF (ser.creat.-3.2)	+	-	+
28	PFC-20	20, M	CM, J(ser.bil.-4), T (platelet-25)	+	+	+
29	PFC-21	68, M	RF (ser.creat.-3), T (platelet-26)	+	+	+
30	PFC-22	22, M	J (ser.bil.-3.9), RF (ser.creat.-3.5), T (platelet-32), A (Hb -4.1)	+	+	+

31	PFC-23	16, M	J (ser.bil.-3.5)	+	-	+
32	PFC-24	25, F	J (ser.bil.-3.1)	+	+	+
33	PFC-25	45, M	J (ser.bil.-4.2), T (platelet-65)	+	+	+
34	PFC-26	20, F	A (Hb-4.2), T (platelet-30)	+	-	+
35	PFC-28	28, M	T (platelet-45)	+	-	+
36	PFC-29	26, M	CM, A	+	+	+

Abbreviations: CM, cerebral malaria; J, jaundice; RF, renal failure; A, anemia; SA, severe anemia; T, thrombocytopenia; PBF, peripheral blood film, RMDT, rapid malaria diagnostic tests; ser.bil., serum bilirubin (mg/dL); ser.creat., serum creatinine (in mg/dL); Hb, hemoglobin (in gm/ dL); Platelet (in 10^3/ mm^3 of blood).

2.8. Designing of a custom 244K *Plasmodium falciparum* array

Plasmodium falciparum 244K gene expression array was designed on Agilent platform with probes having 60-mer oligonucleotides representing the 3D7 transcript sequences (PlasmoDB version 5.3), NCBI EST sequences of *Plasmodium falciparum* and apicoplast sequences of *Plasmodium falciparum* and *Plasmodium vivax.*

Out of 5532 *P.falciparum* transcript sequences available in the PlasmoDB (Release 5.3), probes were designed with an average of 8 probes per sequence in sense as well as in antisense direction (Strand specific). Out of 38413 *P. falciparum* EST sequences available in the NCBI EST database (2007), probes were designed with an average of 10 probes per sequence in sense as well as in antisense direction. Out of six *P. falciparum* apicoplast sequences and three *P. vivax* apicoplast sequences, probes were designed with an average of 10 probes per sequence in sense as well as in antisense direction. This array contains 2,41,399 *P. falciparum* transcript specific probes and additional 2105 Agilent control probes.

The 244K array was re-annotated with the PlasmoDB version 8.2. In the re-annotation process, 39,737 probes were assigned to the sense strand and 47,127 probes were assigned to antisense strand of 5,378 *P. falciparum* transcripts (PlasmoDB version 8.2). In the re-annotation process 8829 probes could not be assigned to any sequences in the database as the probes were not fulfilling the BLAST hit criteria for specificity (alignment of 30bp or more with greater than 84% identity). The probes were grouped into 8 probe groups and these are: ESTs antisense probe group, ESTs sense probe group, PlasmoDB antisense probe group, PlasmoDB sense probe group, *P. falciparum* apicoplast antisense probe group, *P. falciprum* apicoplast sense probe group, *P. vivax* apicoplast antisense probe group and *P. vivax* apicoplast sense probe group (**Table 2.2**).

Table 2.2. Overview of the 244K array probe distribution.

Total Number of features		243504			
Total number of probes designed		241399			
Agilent controls		2105			
	Probe groups				
	Probes	Transcripts		Probes	Transcripts
Sense probes- NCBI EST sequences	72631	19544	Antisense probes - NCBI EST sequences	72927	19725
Sense probes - PlasmoDB transcripts	39737	5369	Antisense probes - PlasmoDB transcripts	47127	5378
Sense probes - *P. falciparum* apicoplast sequences	46	6	Antisense probes - *P. falciparum* apicoplast sequences	56	6
Sense probes - *P. vivax* apicoplast sequences	23	3	Antisense probes - *P. vivax* apicoplast sequences	23	3

2.9. Designing of *Plasmodium falciparum* 15K cross-strain microarray

The probe screening was carried out from multiple probes per transcript present in the *P. falciparum* 244K array. The 244K array contains 2,41,399 probes and additional 2105 Agilent control probes. *P. falciparum* RNA samples were hybridized onto it in an Agilent two color format where pooled uncomplicated and complicated malaria RNA samples were labeled with Cy3 and Cy5, respectively. Best probes were chosen for 15K array based on the signal intensities (both Cy3 and Cy5 signal intensities were considered). Probes representing the same transcript and showing signal intensity greater than 96 (nearly twice the background intensity) were ranked and the top ranked probe/s were chosen. In some cases, probes showing intensity slightly lesser than 96 were not excluded and retained in the array. These were mostly towards the 3' end of the transcript. A total of 5629 probes representing the 3D7 transcript sequences were collected from the 244K array.

For the *P. falciparum* cross strain 15K custom array design, transcript sequences were collected from PlasmoDB v5.3(5595) (Aurrecoechea, et al., 2009) for 3D7 strain, from the Broad Institute for HB3 strain (5623) and from NCBI for IT4 strain (80 *var* transcript sequences) and a database was made. Probes collected from 244K array experiment were BLAST against the *P. falciparum* sequence database. Probes having single significant hit within any one of the strains transcript sequence were selected. For transcripts which were included for these probes, new specific probes were designed using Agilent eArray tool (http://earray.chem.agilent.com). In case of transcripts for which specific probes could not be designed, probes having minimum hits were selected. Criteria for significant hit were alignment of 30bp or more with greater than 84% identity with the transcript sequence. Probes were BLAST against Human transcript sequences and probes cross hybridizing with the human transcript sequences were removed from the list. Strain specific probes were determined by BLAST analysis and annotated accordingly as: probes specific to only one strain, common to two or all of the three represented strains, cross hybridizing to one strain but specific to the other strain or cross hybridizing probes. A total of 6362 probes were selected and included in the final array design. All the probes were replicated twice and randomly distributed across the array and empty features were filled with randomly selected duplicate probes.

2.10. cRNA preparation and labeling

2.10.1. 244K array experimentation

Complicated (n=9) and uncomplicated (n=2) malaria parasite RNA samples were pooled in equimolar amount separately. The two pooled RNA samples were labeled using Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies) as described by the manufacturer at Agilent certified service lab of Genotypic technology, Bangalore. Briefly, 500 ng each of the pooled uncomplicated (control) and complicated malaria (test) total RNA sample were incubated with reverse transcription mix at 40 °C, separately and converted to double stranded cDNA primed by oligodT with a T7 polymerase promoter. The double stranded cDNA was used as template for cRNA generation by *in vitro* transcription and during this Cy3 and Cy5 CTP dyes (Agilent) were incorporated for uncomplicated and complicated malaria pooled samples, respectively. The *in vitro* transcription step was carried out at 40 °C (**Figure 2.3**).

2.10.2. 15K array experimentation

Compared to 244K array experimentation for labeling and cRNA preparation which is based on two color microarray hybridization protocol, 15K array experimentation followed single color microarray hybridization protocol. Each total RNA sample for the 15K gene expression analysis was labelled using Agilent Quick-Amp labeling Kit (p/n5190-0442). Five hundred nanograms of each sample was incubated with reverse trancription mix at 40°C and converted to double stranded cDNA primed by oligodT with a T7 polymerase promoter. The cleaned up double stranded cDNA was used as template for cRNA generation. cRNA was generated by *in vitro* transcription and the dye Cy3 CTP (Agilent) was incorporated during this step. The cDNA synthesis and *in vitro* transcription steps were carried out at 40°C.

2.11. Hybridization and scanning

The labeled cRNA samples were hybridized to the custom arrays with Gene Expression Hybridization Kit (Agilent Technologies, Palo Alto, CA, part number 5188-5242). Hybridization was carried out in Agilent's Surehyb Chambers at 65 ºC for 16 hours. The hybridized slides were washed using Gene Expression Wash Buffer 1 (Agilent Technologies, Part No: 5188-5327), then washed in Gene Expression Wash Buffer 2 (Agilent Technologies, Part No: 5188-5328) at an elevated temperature. The processed microarray slides were scanned using the Agilent Microarray Scanner (Agilent Technologies, Palo Alto, CA, G Model G2565BA) at 5 micron resolution. Data extraction from Images was done using Feature Extraction software (Agilent Technologies, Palo Alto, CA).

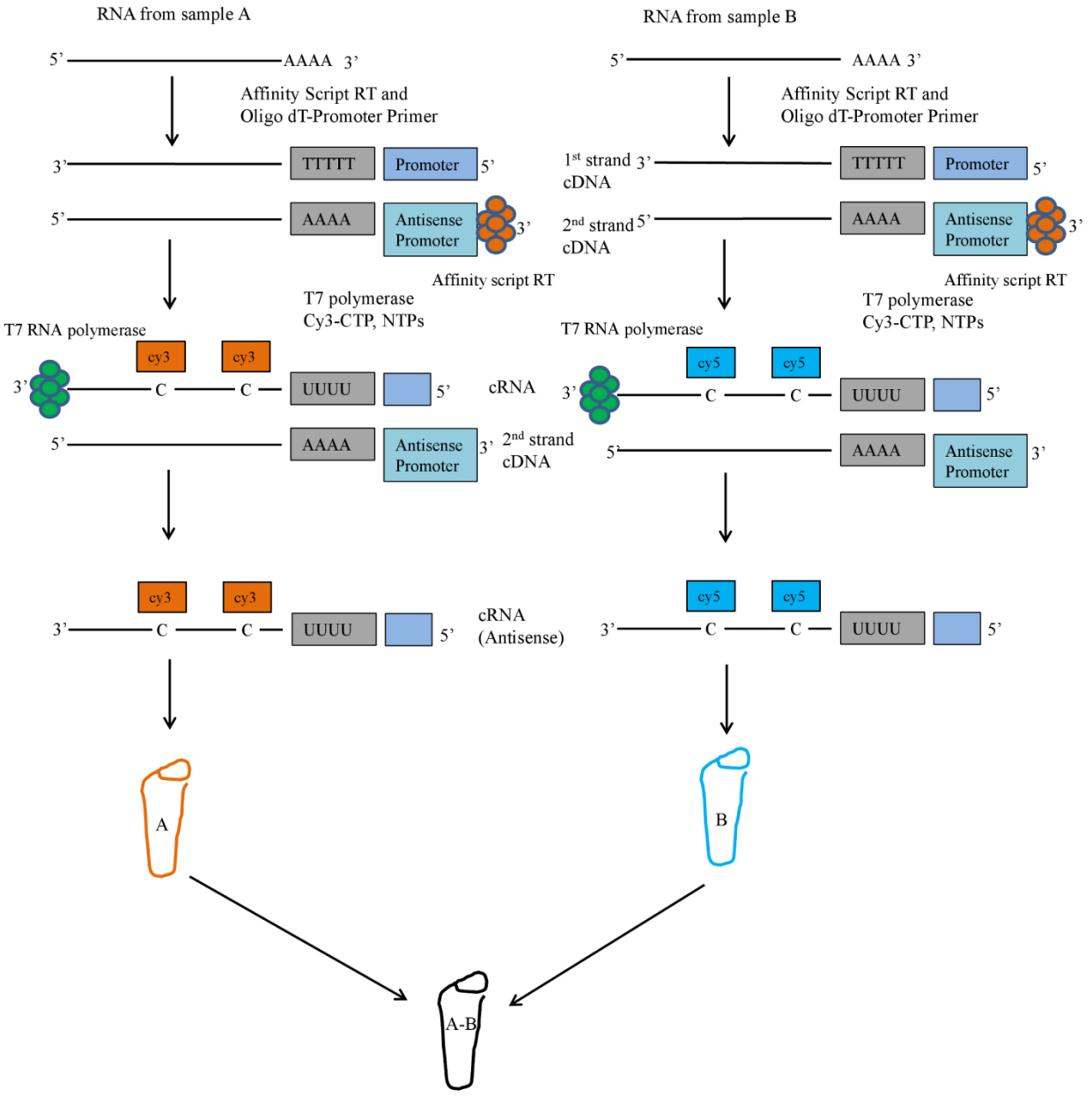


Figure 2.3. Schematic of cRNA preparation from total RNA. Generation of cRNA for a two-color microarray experiment is shown. To generate targets for one-color microarray experiments, only Cy3-lebeled samples are produced and hybridized.

2.12. Data analysis

2.12.1. 244K array experimentation

Data was analyzed separately for complicated (Red channel) and uncomplicated (Green Channel) samples using in-house perl program developed by Genotypic Technology Pvt. Ltd. The raw signal intensities and back ground intensities were obtained from the raw data files for each channel. Based on the re-annotation and probe orientation, the probes were classified as sense and antisense probes. Genes represented by both sense and antisense probes (5330 genes) were only considered for the analysis. Ratio was calculated by taking the raw signal intensity vs. background intensity for each of the probes. Probes representing the sense and antisense transcripts were filtered by considering signal intensity which was ≥ twice the background intensity for complicated cases and ≥ 1.5 times for uncomplicated cases. For genes represented by multiple probes, the probe data was correlated to gene data by taking the ratio of the median of raw signal intensity and background signal intensity.

Note: The number of total transcripts detected in uncomplicated samples was very less (less than 50%) compared to complicated cases, which is biologically not reasonable. To investigate this, we have analyzed our intermediate version 15K custom array (Genotypic AMADID ID: 024956) data, where the two uncomplicated pooled samples used in the 244K array were hybridized individually. We observed that one of the samples (PFU-1) is an outlier with a low number of transcripts (less than 50%) in comparison with other uncomplicated sample (PFU-2). The two samples were pooled in equimolar amount before hybridization. It is very probable that the transcripts present in the PFU-2 sample got diluted and there is insignificant contribution of transcripts from the sample PFU-1. To compensate partially for this, the signal intensity cutoff chosen less for the uncomplicated compared to complicated samples.

2.12.2. 15K array experimentation

Data extraction from Images, within array normalization, background subtracted signal intensity calculation and flagging of any outlier spots either due to saturation or non-uniformity was done using Feature Extraction software v10.5.1.1 of Agilent. Processed signals from the feature extraction software were used for the analysis. Further details can be obtained from the feature extraction user guide http://www.chem.agilent.com.

Feature extracted data was analyzed using GeneSpring GX v11.5 software from Agilent. For finding out differentially regulated probes, data was filtered by considering all those probes detected in at least 30% of the samples analyzed. Filtered data was subjected to between array normalization using "quantile" method and was baseline transformed using median signal intensity of all samples. GeneSpring GX uses the spot information in the data to flag the features. Spot information that can be used to flag a feature as present, marginal or as absent are features that are positive and significant, above background, not uniform, saturated and population outlier. Features which were positive and significant above background (IsPosAndSignif) established via a 2-sided t-test and well above background (IsWellAboveBG) were flagged as detected. IsWellAboveBG is a flagging option which first determines if the feature is IsPosAndSignif and additionally calculates if the background subtracted signal is greater than 2.6 times the background subtracted standard deviation for that feature.

For *var*, *rifin* and *stevor* (variant surface antigens) expression analysis, spots detected in at least one sample were considered. Filtered data were subjected to between array normalization using "quantile" method and were baseline transformed using median signal intensity of control samples (uncomplicated malaria).

2.13. Diagnostic PCR to confirm *Plasmodium falciparum* infection

The diagnostic PCR studies applied were targeted against the 18S RNA gene of the parasite and were based on conditions reported earlier (Das, et al., 1995, and Kochar, et al., 2005), utilizing one genus-specific forward primer and two species-specific reverse primers in the same reaction cocktail (**Table 2.3**). PCR amplifications were performed with 3 µl of purified DNA, 200 µM each of dNTP mix (dATP, dTTP, dGTP and dCTP)(Finzymes), 200 ng of each primer except the genus specific primer which was 300 ng, 1 U of *Taq* DNA Polymerse (Bangalore Genei) and 1X of *Taq* Buffer (Bangalore Genei) for a 50 µL reaction. The cycling parameter used was 1 cycle of 94 °C for 2 min, 35 cycles of 94 °C for 1.5 min, 52 °C for 2 min, and 72 °C for 2 min followed by 1 cycle of 72 °C for 10 min. All PCR products were visualized in a 2% agarose gel. The resulting amplicons were analyzed on a 1% agarose gel. Sample positive with *P. falciparum* amplified an amplicon size of ~ 1400bp.

Table 2.3. List of PCR primers used in the multiplex diagnostic PCR for the detection of *P. falciparum* and *P. vivax*.

S. No	Primer sequence (5'-3')	Detection	Orientation
1	ATCAGCTTTTGATGTTAGGGTATT	Genus	Forward
2	TAACAAGGACTTCCAAGC	*P. vivax*	Reverse
3	GCTCAAAGATACAAATATAAG C	*P. falciparum*	Reverse

Expected amplicon size from Genus specific forward and *P. falciparum* specific reverse primer is ~ 1400bp and expected amplicon size from Genus specific forward and *P. vivax* specific reverse primer is ~ 500bp.

2.14. Genotyping of *Plasmodium falciparum* clinical isolates

The *P. falciparum* clinical isolates were analyzed for molecular identification by amplification of two highly polymorphic regions of merozoite surface protein 1 (*Pfmsp1)* (Block 2) and merozoite surface protein 2 (*Pfmsp2)* (Block 3) using nested PCR as mentioned previously (Snounou, et al., 1999). Oligonucleotide primer sets, previously designed by Snounou et al. (1993) were used for detecting the different allelic variants of *Pfmsp1* (K1, MAD20 and RO33) and *Pfmsp2* (FC27 and 3D7/IC) (**Table 2.4**). The cycling conditions for primary and secondary PCR were as mentioned by Atroosh et al (2011). All the primary PCR amplifications were carried out in a final volume of 50µl containing 3 µl of template genomic DNA, 250 ng of

primers, 1U of Taq Polymerase, 1X buffer with 1.5 mM of $MgCl_2$ and 200 µM of dNTP mix. The primary PCR products were separated using agarose gel electrophoresis (using 2% agarose) and visualized using gel documentation system (BIO-RAD). Depending on the intensities of the amplicon, one, two or three µl of primary PCR products were used as a DNA template in the secondary PCR. The secondary/nested PCR had the similar concentrations of components to the primary PCR. The secondary/nested PCR products were also separated using agarose gel electrophoresis (using 2% agarose) and visualized using gel documentation system (BIO-RAD).

Table 2.4. List of primers used for genotyping *Plasmodium falciparum* in clinical isolates based on *Pfmsp* 1 and *Pfmsp* 2 genes (Atroosh, et al., 2011).

Amplification/Gene Primary PCR	Primer	Primer Sequence (5'-3')	
MSP-1	M1-OF	CTAGAAGCTTTAGAAGATGCAGTATTG	
	M1-OR	CTTAAATAGTATTCTAATTCAAGTGGATCA	
MSP-2	M2-OF	ATGAAGGTAATTAAAACATTGTCTATTATA	
	M2-OR	CTTTGTTACCATCGGTACATTCTT	
Secondary PCR			Expected amplicon size range (bp)
MSP-1	M1-KF	AAATGAAGAAGAAATTACTACAAAAGGTGC	160-225
	M1-KR	GCTTGCATCAGCTGGAGGGCTTGCACCAGA	
	M1-MF	AAATGAAGGAACAAGTGGAACAGCTGTTAC	130-220
	M1-MR	ATCTGAAGGATTTGTACGTCTTGAATTACC	
	M1-RF	TAAAGGATGGAGCAAATACTCAAGTTGTTG	160-210
	M1-RR	CATCTGAAGGATTTGCAGCACCTGGAGATC	
MSP-2	M2-FCF	AATACTAAGAGTGTAGGTGCARATGCTCCA	290-420
	M2-FCR	TTTTATTTGGTGCATTGCCAGAACTTGAAC	
	M2-ICF	AGAAGTATGGCAGAAAGTAACCTYCTACT	470-700
	M2-ICR	GATTGTAATTCGGGGGATTCAGTTTGTTCG	

2.15. Strand specific reverse transcriptase PCR

Strand specific reverse transcriptase PCR was performed to validate sense and antisense transcript candidates. The primers were designed using NCBI-Primer Blast (**Table 2.5**) (Ye, et al., 2012). Total RNA from each uncomplicated (n=5) and complicated (n=8) samples was quantified and treated with DNase I (Fermentas) for 30 min at 37°C. Inactivation of DNase I was performed by adding 1 ul of 50Mm EDTA per unit of DNase I used and incubated for 10 min at 65°C. The absence of DNA in RNA samples was confirmed by performing 40 cycles of PCR using Pfmahrp (MAL13P1.413) primers, which showed amplification in a positive reaction where genomic DNA was used as a template. DNase I treated total RNA samples from uncomplicated and complicated samples were pooled in equi-molar amount separately. Strand specificity was achieved by selective use of primers, where the forward primer (antisense reverse) was used to make first strand cDNA from antisense transcripts to detect the antisense transcripts, and the reverse primer (sense reverse) was used to do the same for the sense transcripts (Ho et al., 2010). For each reverse transcription reaction 250ng of total RNA was used. Reverse transcription was performed using M-MuLV reverse transcriptase (Thermo Scientific) in a total volume of 20µl following manufacturer's instructions. A separate reaction set was used where no primers were included to ascertain the level of self priming in further PCR reactions if any. Additional primers were designed near the 3′ end of the antisense (antisense forward) and sense (sense forward) transcripts to detect the respective transcripts following reverse transcription reaction. PCR reactions following reverse transcription were carried out in the presence of antisense reverse and antisense forward primers to detect the antisense transcripts and sense reverse and sense forward to detect the sense transcripts. The cycling parameter used was 1 cycle of 94 °C for 4 min, 55 °C for 2 min and 72 °C for 2.5 min, 35 cycles of 94 °C for 1.5 min, 55 °C for 1.5 min, and 72 °C for 2 min followed by 1 cycle of 72 °C for 10 min. All PCR products were visualized in a 2% agarose gel.

Table 2.5. List of PCR primers used in strand specific transcript detection experiment and strand specific quantitative real time PCR analysis

Amplification/ Gene	Primer	Sequence 5′-3′	Type	Amplicon size (bp)
PF07_0009	Pfcht-F	GGGCTCACTCACCATAGGTC	Antisense reverse/ sense Forward	193
	Pfcht-R	AACGTTTCCCTCCTGGGTTC	Antisense forward/ sense reverse	
MAL13P1.413	Pfmahrp-F1	AGAGCAAGCAGCAGTACAACC	Antisense reverse	269
	Pfmahrp-R2	TGTTCAAACCACGTTCTCTTGTG	Antisense forward	
	Pfmahrp-R1	CAAGTTCATGAGCGTGTGCAG	Sense Reverse	203
	Pfmahrp-F2	TCACGCCTTCTTATACCACTCA	Sense Forward	
MAL8b_28s	Pfunc-F1	TTTAGGAGGGCAAATCCGCT	Antisense reverse	313
	Pfuncf-R2	GCCTTTCACCCTCTTTGGGA	Antisense forward	
	Pfuncf-R1	ACGGTTGAACAATCCGACACT	Sense reverse	680
	Pfunc-F2	TGCGAAGGGCTTTTAGAATGT	Sense forward	
PF11_0175	Pfhsp-F1	TGTGTGCTCCCGATAATAAGCA	Antisense forward	251
	Pfhsp-R2	ATTATCCAATGGCGCTCCCG	Antisense reverse	
	Pfhsp-R1	AGCCTTTTCAGAAACAGATACTTGA	Sense reverse	235
	Pfhsp-F2	TGCCGACAATTCAGGTACTCC	Sense forward	
PFA0620c	Pfgarp-F1	TCTACTAAGTGCTTTTCCAATGGT	Antisense forward	94
	Pfgarp-R2	CGGTTTCGTTTAATAATCTTCCCGT	Antisense reverse	
	Pfgarp-R1	TGGGTACTACATTAACACGTCCT	Sense reverse	142
	Pfgarp-F2	AGACCACTAAGCCAACCACA	Sense forward	
PF10_0155	Pfeno-F1	TTCAGAGCTGCCGTACCATC	Antisense forward	293
	Pfeno-R2	GCTGCACCAGCTCTACATACA	Antisense reverse	
	Pfeno-R1	TTCTGCATGGTGCTCCTGTT	Sense reverse	286
	Pfeno-F2	ACAGCAGCTATTGGAAAGGATG	Sense forward	
PF14_0633	Pftap-F1	ATAAGCGTATGTGTGCGTGG	Antisense forward	118
	Pftapf-R2	TCTACTGCTGCTAGTCTTGCTT	Antisense reverse	
	Pftap-R1	TGTTGTTGTTCTAGGGGCATAC	Sense reverse	195
	Pftap-F2	TCAATTTAGTTATCAAAACCAGCCA	Sense forward	
PF14_0194	Pfsap-F1	TTGGATGCACAAGTTGACGA	Antisense forward	247
	Pfsap-R2	AATTTGCTCCCACATCGAAAG	Antisense reverse	
	Pfsap-R1	ATTAGGCGGTAGGTTGGGAG	Sense reverse	109
	Pfsap-F2	CTTTCCTCCTGGATTTCCACCAA	Sense forward	

2.16. Strand specific quantitative real-time PCR

Quantitative real time PCR measurements were performed on an IQ5 multicolor real-time PCR detection system (BIO-RAD). Reactions were prepared in volumes of 20 µl using IQ SYBR green PCR master mix and a primer concentration of 200nM. From each strand specific reaction cDNA template was taken separately in equal amounts to determine the antisense and sense transcript expression level of each of the eight genes in each condition under investigation. The PCR cycling conditions were as mentioned above in the strand specific reverse transcriptase PCR methodology section. The background expression levels of sense and antisense transcripts which might be due to false priming was determined by taking template from no primer reverse transcription reaction. Reactions were taken in duplicates and expression levels of each sense and antisense transcript was determined using Δct method with no control assigned. Formula used to calculate relative quantity was as follows

$$\boldsymbol{Relative\ quantity}_{\text{sample(Gene x)}} = \boldsymbol{E}_{\boldsymbol{Genex}}{}^{(C_{T(MIN)} - C_{T(SAMPLE)})}$$

Where E, efficiency of primer/(probe) set and efficiency was calculated as follows (% of Efficiency * 0.01 + 1) where 100% = 2 . $C_{T(MIN)}$ = Average C_T for the sample with the lowest average C_T for gene x. $C_{T(SAMPLE)}$ = Average C_T for the sample.

2.18. Quantitative real-time PCR

Total RNA from each sample was treated with DNaseI (Fermentas) for 30 min at 37 °C. Inactivation of DNase I was performed by adding 1 ul of 50 mM EDTA per unit of DNase I used and incubated for 10 min at 65 °C. The absence of DNA in RNA samples was confirmed by no detection of DNA band in 2% agarose gel after 40 cycles of PCR with seryl –tRNAsynthetase primers. DNase I treated total RNA samples from each comparison groups were pooled in equi-molar amount separately. First strand cDNA synthesis was carried out using iScript cDNA synthesis kit (BIO-RAD) in a total volume of 20µl according to the manufacturer's recommendations. Quantitative real-time PCR using IqSYBR green PCR supermix (BIO-RAD) was performed on an IQ5 real-time PCR cycler (BIO-RAD) using the seryl-tRNA synthetase gene as an endogenous control. Fold change expression was calculated using $2^{-\Delta\Delta Ct}$ method. The cycling parameter used was 1 cycle of 94 °C for 4 min, 55 °C for 2 min and 72 °C for 2.5

min, 35 cycles of 94 °C for 1.5 min, 55 °C for 1.5 min, and 72 °C for 2 min followed by an extension cycle of 72 °C for 10 min. All the primers except primer pairs A2, B1 and BC1 were designed using NCBI Primer Blast (Ye, et al., 2012) and are listed in **Table 2.6**. Sequences of primer pairs A2, B1 and BC1 have been reported elsewhere (Rottman, et al., 2006).

Table 2.6. List of PCR primers used in real-time qPCR validation experiments

Amplification/Gene ID	Primers	Primer sequence (5'-> 3')	Amplicon size (bp)
PF10_0155	PfENO-F	ACAGCAGCTATTGGAAAGGATG	286
	PfENO-R	TTCTGCATGGTGCTCCTGTT	
PF14_0683	PF14_0683-F	ATGTTGGAGGTGTGTGTGAGA	261
	PF14_0683-R	AACCATTCCTCTCCGTAACAACA	
PFF1280w	Pfsec14F	GCAGTGAAACAAAAGGACGTT	165
	Pfsec14R	GTGTCTCCTTATCACATGGGG	
MAL8P1.16	PfRHOPF	ACTCGGGGCAACTTATGGTC	255
	PfRHOPR	AACCCACTAGTTGATGCTCCT	
PF13_0269	PfGKF	TGATCCAAGTGAAGCTAGCGA	269
	PFGKR	TCATGCCTCCATCACATCGT	
PF07_0073	PfStRsF	TCAGGAGCTTTAAACAACGCA	283
	PfStRsR	GTGCAGCTACCATTGTTCCA	
PFC0710w-a	IPF	AGCGACATAGAAAAGTACTACCCA	363
	IPR	GAGATCTGGCTTGTAAGCACT	
PFL0795c	MDG1F	AGAAAGCACTCTCAAGCCTCT	135
	MDG1R	TCCGTTTCTTCATTAGCATTTCCG	
PF10_0013	HYP12F	ACGCAACAAAACAATTTAGGAGA	245
	HYP12R	TCTGGATTTTCCTTCGCCAT	
PFA0110w	RESAF	AGCTTTAAATGCCGCTGAACAA	252
	RESAR	GCATTATGTTGTACATGTTCGGGTA	
MAL7P1.146	DDSF	GGTACGAATATATTGACTGATCCGA	211
	DDSR	TAGGGGGAGGCAAATCACTT	
A2	A2F	AACCCATCTGTRRATGATATACCTATGGA	
	A2R	GTTCCAASGATCCATTRGATGTATTA	
B1	B1F	CATCCGCCATGCAAGTATAA	
	B1R	CGTGCACGATTTCGATTTTT	
BC1	BC1F	GACAAAACTTTCACCCAATAGA	
	BC1R	AATGATCGGTGTAACCACTATC	

2.19. Physical mapping of 3D7 genes with antisense transcripts

Annotation details of genes with antisense transcripts (AS) or both sense (S) and AS transcripts detected by at least 3 probes (for both complicated and uncomplicated cases) were extracted from the PlasmoDB v8.2 annotation file (Aurrecoechea, et al., 2009). A .BED file was prepared by taking all the information from the annotation file needed to create a BED detail file in a format specified by the UCSC Genome Bioinformatics team to visualize the gene positions across the 14 parasite chromosomes. The chromosomal locations of the genes were visualized using the UCSC *Plasmodium falciparum* genome browser gateway custom track (Kent, et al., 2002).

2.20. Gene ontology-based enrichment of genes

Biological process and molecular function ontology term which were highly represented by the list of genes expressing AS transcripts were noted. These genes were analyzed with the help of Bingo plug-in (Maere, et al., 2005) in Cytoscape v8.2 (Smoot et al., 2011). The biological process and/or molecular function ontology term were listed which showed a corrected *P* value ≤ 0.05 determined by the hypergeometric test after multiple testing corrections (Benjamini Hochberg's false discovery rate correction).

For functional enrichment analysis of gene list, Database for Annotation, Visualization, and Integrated Discovery (DAVID) Bioinformatics resource tool v6.7 was also used (Huang et al., 2009). Gene list enriched to any molecular functions (MF) and/or Biologicals processes (BP) with *P* value ≤ 0.05 were listed.

2.21. Pathway-based enrichment of genes with antisense transcripts

Genes with antisense transcripts in both complicated and uncomplicated cases were mapped to *P. falciparum* metabolic pathways from Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000). Information about the number of genes involved in each pathway was obtained and percentage of genes with antisense transcripts in each pathway was calculated and plotted against the respective pathways.

2.22. Comparison of *in vivo Plasmodium falciparum* antisense data with previously reported data from 3D7 laboratory strain.

Our microarray data was compared with previously published NATs data from different blood stages of *in vitro* cultured *P. falciparum* parasite using high throughput techniques like serial analysis of gene expression (SAGE), RNA-seq and cDNA library sequence analysis (Gunasekera, et al., 2004, Lopez-Barragan, et al., 2011 and Raabe, et al., 2010). For this, list of 688 genes yielding at least two antisense tag counts from the SAGE data was prepared by downloading from PlasmoDB (Gunasekera, et al., 2004), 312 genes with high levels of antisense transcripts reported from the RNA-seq experiment (Lopez-Barragan, et al., 2011) and 322 protein coding genes with complementary antisense RNA transcripts (non protein coding RNA) from cDNA library sequence experiment (Raabe, et al., 2010) were prepared.

2.23. Preparation of exportome gene list

List of genes predicted to be a member of exportome based on presence of a PEXEL motif in the protein product was prepared from three published reports (Marti, et al, 2004, Sargeant, et al, 2005 and Van Ooji, et al, 2008). Additional PEXEL negative genes were added to the above list, which are known or reported to be exported out from the parasitophorous vacuole and are members of exportome (Heiber et al, 2013).

2.24. Co-expression based systems network generation

2.24.1. Preprocessing of data

An initial expression matrix was created by taking expression profiles of 21 samples out of 33 samples hybridized on 15K array. Remaining 12 samples were not included for systems network generation and analysis study because either they had gametocyte stages in the blood sample as read from peripheral blood film or had very low overall intensity compared to the average intensity of all microarray samples. After performing quantile normalization on expression matrix using GeneSpring GX v 12.5, probes that were detected in at least 30% of the samples were retained. For gene wise analysis, only single probe per gene was kept: for genes with two or three corresponding probes, we chose the probe with highest expression. After filtering and removing duplicate probes, 4857 probes corresponding to 4857 gene were retained for further

analysis. Probe names were replaced with corresponding gene names in the expression matrix. Further analysis was carried out using WGCNA package in R (Langfelder and Hovarth, 2008).

2.24.2. Co-expression based network generation and module identification

Weighted gene co-expression network analysis assists in identifying network modules (clusters) containing highly co-expressed genes and also helps in defining intra-modular relationships and hub gene (genes with high degree of connection) identification. Weighted gene co-expression network differs from the un-weighted gene co-expression network by the fact that the latter one uses pairwise correlation of expression measures as the connection strengths between nodes (genes) whereas the previous one uses soft thresholding of pairwise correlation of expression measures to define connection strengths between two genes. Technical and theoretical aspects of a weighted gene co-expression network construction and analysis can be found elsewhere (Zhang and Horvath, 2005 and Langfelder and Horvath, 2008). One important aspect of weighted gene co-expression network is it preserves the continuous nature of underlying co-expression information and hence protects from information loss. Whereas, un-weighted gene co-expression network do not preserve the continuous nature of underlying co-expression information as it relies on hard thresholding. This mean if the hard thresholding parameter has been set to 0.85, any gene pairs with similarity ≤0.84 will not have connection between them.

Network construction was carried out following the weighted gene construction network analysis (WGCNA) framework (Zhang and Horvath, 2005) using WGCNA R package. An expression matrix was prepared by taking expression measures of 4857 genes across 21 samples. First step before constructing weighted gene co-expression network requires identification and removal of outlier samples if any. Clustering of 21 samples was performed using function flashClust to detect outlier samples. The flashClust provides a faster hierarchical clustering than hclust which is used normally as standard function in R for sample clustering. One outlier sample was detected and was removed from the analysis using cuttreeStatic function. Then, a similarity matrix was prepared by calculating pair wise Pearson co-expression coefficient between each gene with rest of the genes in the matrix. The similarity matrix was then converted into an adjacency matrix using a "soft" power adjacency function $a_{ij} = \left|cor(x_i, x_j)\right|^{\beta}$ which transforms the co-expression similarities into connection strengths. The power β was chosen based on scale free topology criterion. For this, pickSoftThreshold function in WGCNA package was used. This

function allows inspecting numbers of soft-thresholding powers that can provide a scale free network topology. The criterion states that the soft-thresholding power β, is the lowest integer using which the resulting network satisfies approximate scale free topology. In this case we have chosen soft-thresholding power (β) as 16.

The next step after choosing the soft thresholding power is to generate network and detect modules by using a one-step network construction and module detection function. The blockwiseModules function was used to construct network and detect modules. The default parameters were changed to construct a signed network. The soft thresholding power β = 16 was used, signed TOMType and netwrokType parameter was used to generate a signed network. To check whether the resulted network shows a scale free topology or not, scaleFreePlot function was used. This function generates a scatter plot between $\log_{10}$ p(k) and $\log_{10}$ (k), where k represents connectivity of nodes and p(k) represents frequency of connectivity. The linear model fitting index R^2 of the regression line was 0.78, which suggests that the resulted network approximately satisfies the scale free topology criterion. Modules were detected based on topological overlap measures. Topological overlap (TO) is a measure of node similarity (i.e. connection strengths between two genes in a network). Genes in the network were hierarchically clustered using 1- TO as the distance measure and modules were determined using the mergeCutHeight function (value used was 0.25). The one step network construction and module detection function identified 27 modules. Each module was subsequently assigned a color using function labels2colors. The network as well as identified modules was visualized in the form of a hierarchical clustering dendrogram (tree) by using the function plotDendroandColors. Modules information (nodes and edge information) were exported as a Cytoscape compatible file using export NetowrkToCytoscape function to visualize and analyze the network modules using Cytoscape.

2.25. Functional gene ontology based enrichment of modules

Genes in modules were checked for its enrichment to gene ontology terms present in each three gene ontology category (i.e. Biological Process, Molecular Function and Cellular Components) using functional annotation tool present in DAVID Bioinformatics Resources v6.7.

2.26. Materials used for various experiments

Composition of buffers and solutions

- Tris - Acetate (TA) buffer (50X): (1000 ml)

  Tris base 242 g
  Glacial Acetic Acid 57.1 ml
  0.5 M EDTA (pH 8.0) (100ml)

- Tris – Borate (TB) buffer (5X): (1000ml)

  Tris base 54 g
  Boric Acid 27.5 g
  0.5 M EDTA (pH 8.0) (20 ml)

- Tris EDTA (TE) buffer (pH 8.0):

  10 mM Tris-Cl (pH 8.0)
  1 mM EDTA (pH 8.0)

- Phosphate Buffered Saline (PBS) pH 7.4 (1X):

  137 mMNaCl
  2.7 mMKCl
  4.3 mM Na_2HPO_4
  1.4 mM KH_2PO_4

- MOPS electrophoresis buffer (10X):

  0.2 M MOPS (pH 7.0)
  20 mM sodium acetate
  10 mM EDTA (pH 8.0)

- Formaldehyde gel-loading buffer

  50 % glycerol (diluted in DEPC-treated water)
  10 mM EDTA (pH 8.0)
  0.25 % (w/v) bromophenol blue
  0.25 % (w/v) xylene cyanol

- Back extraction buffer (BEB): (250 ml)

  118.2 g Guanidine Thiocyanate
  3.68 g Sodium Citrate
  30. 29 g Tris

2.26. Various commercial kits used for this study

- Agilent Quick-Amp labeling kit (Cat#5190-0442) for cRNA labeling and amplification
- Low RNA Input Fluorescent Linear Amplification Kit for cRNA labeling and amplification
- QiagenRNeasyminikit(Cat#74104)
- Agilent *In situ* Hybridization kit (Cat#5188-5242)
- Agilent RNA 6000 Nano kit (Cat#5067-1511)
- BIO-RAD Experion RNA StdSens kit (Cat#700-7257)
- BIO-RAD iScript cDNA synthesis kit (Cat#170-8891)
- BIO-RAD iQSYBR Green Supermix (Cat#170-8882AP)

3.1. Introduction

Plasmodium falciparum has the inherent ability to adapt to its environment which in combination with its intricate biology can drive the interventions unsuccessful. The emergence and reemergence of drug resistant parasites and inability to develop an effective vaccine is a prominent example of this (Hay, et al., 2010 and Imwong, et al., 2010). Understanding the complex biology of this parasite has been partially achieved in recent years due to the advancement in techniques like microarray (DNA microarrays, Protein binding microarrays, SNP microarrays) (Le Rroch, et al., 2003 and Bozdech, et al., 2003), array-comparative genomic hybridization (aCGH) (Mackinnon, et al., 2009), RNA-seq (Otto, et al., 2010), ChIP-chip (Flueck et al., 2009), Chip-seq (Bartfai, et al., 2010), mass spectrometry techniques (Olszewski et al., 2009), transfection techniques (Caro, et al., 2012), mutagenesis methods (Balu, et al., 2005), gene knock out techniques (Duraisingh ,et al., 2005 and Maier, et al., 2008) and imaging techniques. All these techniques have helped in exploring some fundamental questions, which was not possible earlier. Interestingly, many basic aspects of its biology, while in the host, still needs to be explored.

Microarray, as a technique has the capability to analyze whole transcriptome of an organism, and has opened up the possibility of simultaneously exploring the expression and regulation of all the parasite's genes. The first versions of *P. falciparum* DNA microarrays were developed few years before the release of its first draft sequence and have used PCR amplified inserts from a *P. falciparum* DNA library (shotgun microarray) or cDNAs (Hayward, et al., 2000 and Ben Mamoun, et al., 2001). Many oligo-nucleotide based whole genome microarrays have been designed, following the release of the first draft of *P. falciparum* 3D7 genome sequence in the year 2002 from Malaria Genome Consortium (The Sanger center *Plasmodium falciparum* Genome Project, Stanford Genome Technology Malaria Genome Project, TIGR *Plasmodium falciparum* Genome Database). Importantly two custom arrays that have been extensively used are 1) high density short oligonucleotide (25mer) microarray on Affymetrix platform (Le Roch, et al., 2003) 2) long oligonucleotide (70mer) spotted microarray (Bozdech, et al., 2003). Upon release of subsequent updates of *P. falciparum* genome database annotation, many new versions of microarrays have evolved, and utilized various platforms.

DNA microarrays have been utilized to show the stage specific temporal expression pattern of the parasite during its intra-erythrocytic developmental cycle (IDC) (Bozdech, et al., 2003, Le Roch, et al., 2003 and Ben Mamoun, et al., 2001). Additionally, microarray based studies have also helped to explore the transcriptional status of genes in other life cycle stages like gametocyte and sporozoite stages, differences and commonality of transcriptional programming at a cross strain level (Mackinnon, et al., 2009), transcripts possibly involved in gametocytogenesis (Silvestrini, et al., 2005 and Young, et al., 2005), *in vivo* transcriptional states of the parasite in infected patients (Daily, et al., 2007), soft-wired transcriptional reprogramming in response to drug perturbations (Natalang, et al., 2008 and Tamez, et al., 2008) or environmental cues (Oakley, et al., 2007) and expression pattern of variant surface antigens particularly in clinical conditions (Tuikue Ndam, et al., 2008). Although the microarray technique has helped in addressing many fundamental questions, its use has been limited due to its cost and comparatively large amount of starting biological material required which is a limitation in the case of clinical isolates. Most of the arrays for *P. falciparum* have been designed based on a single reference strain genome (3D7 laboratory strain) and have not considered the possibilities of genomic variations among strains which could limit transcript detection.

Genome variation is a natural phenomenon observed among members of same species belonging to different geographical locations. This phenomenon can profoundly be seen in this parasite within and across geographical zones. This imposes challenge to gene expression studies (Mackinnon , et al. 2009) which are essential for understanding the biology of the parasite with a potential to discover intervention targets, diagnostic and prognostic markers. Prior to the genome screening using NGS and array comparative genomic hybridization (aCGH) based technologies, genes with several large and small deletions had been reported to be important for parasite virulence (Pologe and Ravetch, 1988, Scherf and Mattei, 1992, Cappain, et al., 1989 and Biggsn, et al., 1989). Subsequently, whole genome screening with the application of high throughput technologies (NGS and aCGH) have identified higher sequence diversity across the genome, in addition to the known sub-telomeric compartments, where key gene families like *var*, *rifin* and *stevor* are located (Bozdech, et al., 2003, Llinasn, et al., 2006, Kidgell, et al., 2006, Volkmannn, et al., 2006 and Manske, et al., 2012). Strategy which can take the genomic variations into

account during microarray designing could allow efficient detection of transcripts from regions of genomic variation.

This chapter deals with the strategies involved in designing of a cross strain *P. falciparum* custom 15K microarray on Agilent platform and its validation thereafter. Agilent provides in many ways the flexibility to design a custom array. This extends from ordering the number of slides, allowing updation of the microarray design as and when required. Further, due to the cRNA amplification strategy adopted, the amount of starting material required makes studies with field derived samples more feasible. The use of ink-jet based *in situ* synthesized 60 bp long oligo-nucleotide probes in the Agilent platform enhances its sensitivity and due to outstanding spot uniformity, it reduces the inter-array variability, has a broad dynamic range of signal intensity and shows low and uniform background signal (Kafsack, et al., 2012).

3.2. Results and discussion

Genome sequence information of Indian field isolates is limited. With this limited knowledge, one approach that can help to design a microarray for efficient detection of transcripts from the field isolates is to include probes representing the common and distinct sequences of geographically diverse *P. falciparum* laboratory strains for which genome sequence information is available. This increases the chance of representing the regions of variations in the array. The efficiency to detect transcripts from the field isolates can further be increased, if multiple probes representing a transcript/ gene are subjected to screening for transcript detection by using the RNA material from field isolates. This would screen the best probes based on hybridization intensities out of multiple probes to be used in final microarray design. A very similar strategy have been used in designing of the custom cross strain *P. falciparum* 15K array for the efficient transcript detection of Indian field isolates from North-Western India (**Figure 3.1**).

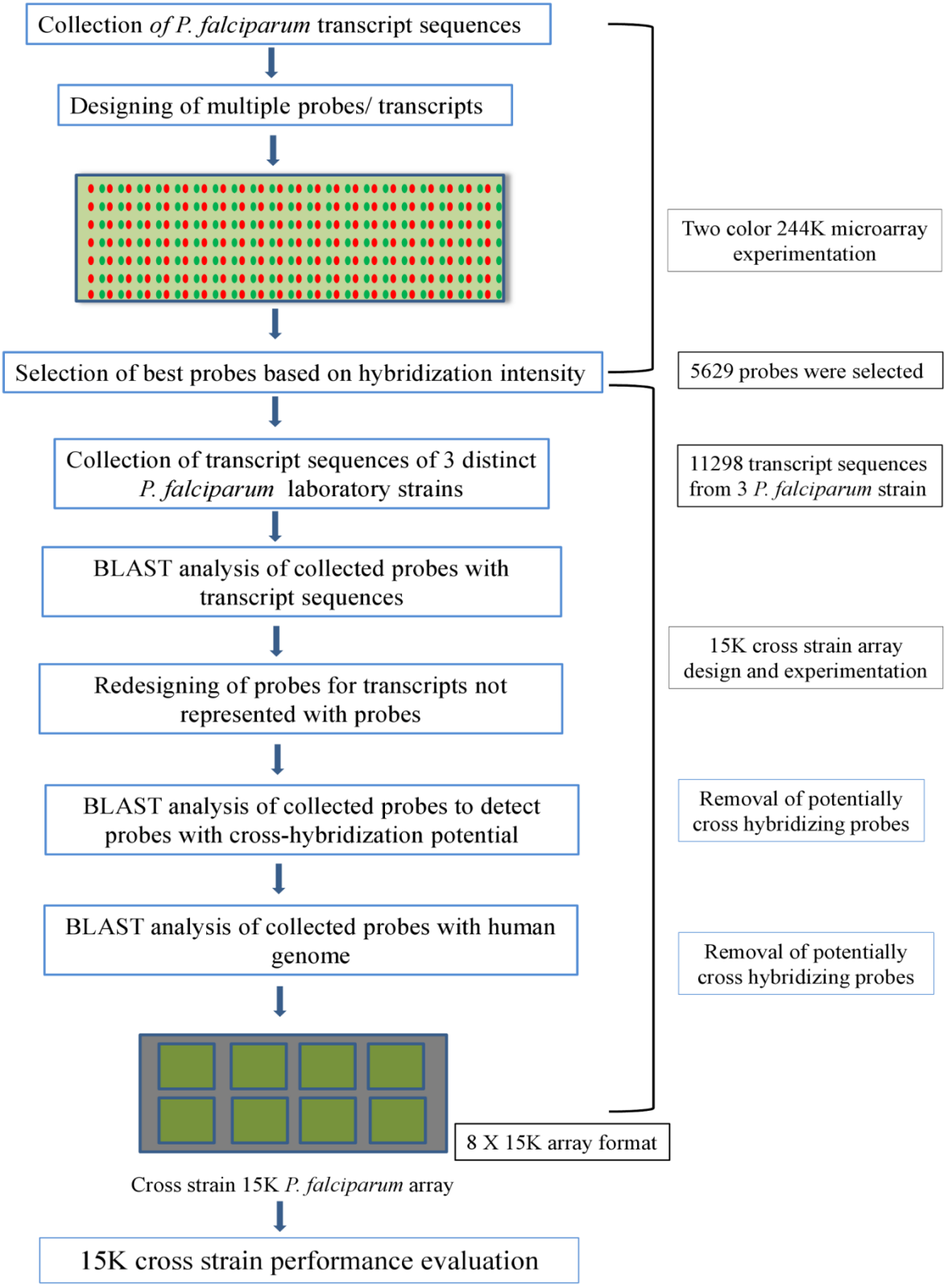


Figure 3.1. Schematic of 15K array design and its evaluation

3.2.1. Designing of 15K *Plasmodium falciparum* cross strain microarray

3.2.1.1. Selection of probes based on a custom designed 244K array experimentation

Streamlined procedures were adopted to screen best probes for transcriptome studies of *Plasmodium falciparum* Indian field isolates for which genome sequence information is limited. For this a 244K custom whole genome array was designed on an Agilent platform with 60mer oligonucleotide probes (Agilent Microarray Design Identity number (AMADID) 024956) representing *P. falciparum* 3D7 strain transcript sequences (PlasmoDB v5.3) (Aurrecoechea, et al., 2009) and NCBI EST sequences (NCBI 2007). The 244K array comprised of a total of 2,43,504 probes which included 2,41,399 probes representing the above mentioned sequences and 2105 Agilent controls. On an average 8 probes per transcript were designed in both sense and antisense direction representing 5532 *P. falciparum* transcript sequences (PlasmoDB v5.3, Aurrecoechea, et al., 2009) and 10 probes per transcript were designed in both sense and antisense direction representing 38,413 *P. falciparum* EST sequences (NCBI 2007).

244K gene expression array experiment was conducted according to Agilent Technologies' Two color microarray-based gene expression analysis protocol. Uncomplicated (n=2) and complicated (n=9) malaria parasite total RNA samples were pooled in equimolar amount separately, amplified and labeled using Low RNA Input Fluorescent Linear Amplification Kit, Two-color (Agilent Technologies). Labeled cRNA was prepared and hybridized to single 244K array.

Probe selection for 15K array was carried out by considering only sense probes representing PlasmoDB transcripts. GE gMedianSignal (green channel) and GE_rMedianSignal (red channel) intensity for each probe was collected. Maximum value of the GE green and red channel intensity of each probe representing a transcript was calculated, ranked and probe with maximum GE_Median signal intensity was selected for each transcript. Probes with signal intensity greater than 96 (~ twice the average background intensity) were t considered (3025 probes). Transcripts for which probes did not qualify the criteria of showing intensity greater than 96, probes were selected based on their proximity to the 3' end of the representing sequences. In this way, 5629 probes representing 3D7 transcript sequences were collected from the 244K array. Designing of multiple probes against a larger sequence region of genes/transcripts has helped in capturing transcripts more efficiently than if only single probes were used. Probe(s) falling in a sequence

region common/conserved between the field strain and reference strain will in all possibility, lead to efficient detection of such transcripts. Designing of multiple probes for selection of best probe(s) is crucial as there is limited genome sequence knowledge of the Indian field isolates. In this case best probe(s) were chosen out of multiple probes based on intensity.

3.2.1.2. Step-wise procedure implemented in selection of best probe(s)

Step 1: Collect the hybridization intensities of all the probes designed for the transcript

Probe name	Gene Name	GE_gMedianSignal	GE_rMedianSignal
X1	Y	57	88
X2	Y	51	51
X3	Y	46	90
X4	Y	62	297

X and Y represent hypothetical probe and gene name.

Step 2: Get the maximum value of GE_Green and Red channel intensity of each probe

Step 3: Rank the probes (of a target transcript) based on GE_Median intensity

Probe name	Gene Name	GE_MedianSignal	Rank
X1	Y	88	3
X2	Y	51	4
X3	Y	90	2
X4	Y	297	1

Step 4: Select the rank 1 probe if the intensity is higher than 96 (> twice the background intensity). In many cases, multiple top rank probes were also selected.

3.2.1.3. Final design of 15K *Plasmodium falciparum* cross strain whole genome GXP microarray

We also designed and included probes against sequences of two other geographically distinct sequenced *P. falciparum* strains i.e. HB3 and IT4 so as to represent diverse sequences of *P. falciparum* in the array. To design a cross strain *P. falciparum* array, transcript sequences were collected from PlasmoDB v6.3 for 3D7 strain (5595 transcript sequences), from Broad Institute for HB3 strain (5623 transcript sequences) and from NCBI for IT4 strain (80 *var* gene transcript sequences) and a unique database was made. Total number of sequences thus collected were 11298 out of which 9842 were unique sequences (although sequence similarity was observed between them) and 1427 transcripts sequences of HB3 were identical to 3D7 transcripts sequences (determined from BLAST results). Probes collected from the 244K array experiment were BLAST against the database and probes having only single significant hit with any of the strain were selected. New probes were designed for the transcripts for which no probes were found, using Agilent eArray tool. In the case of transcripts for which specific probes could not be designed, probes having minimum number of hits were selected (probes with a potential to cross hybridize) and these were flagged. Criterion for significant hit was alignment of 30bp or more with greater than 84% identity with the transcript sequence. All probes, collected and designed were BLAST against the human transcriptome to check for their cross hybridizing potential with the human transcripts. Potential cross hybridizing probes were removed from the array design.

A total of 6120 user defined long oligonucleotide (60mer) probes were designed against the transcript sequences present in the database. Based on the BLAST results, probes were annotated as specific to any one strain (Single significant hit within that strain-3D7specific/HB3 specific/IT4 specific), common in between any two or more strains (single significant hit within each strain-3D7 and HB3 or 3D7 and IT4 or HB3 and IT4 or 3D7, HB3 and IT4) and cross hybridizing probes (multiple significant hits within a strain). Array was re-annotated recently using the PlasmoDB v8.2. After the re-annotation, the final design of the *P. falciparum* cross strain whole genome GXP array (AMADID: 024956) is summarized in **Table 3.1** and **Supplementary Table S3.1**. Out of 6362 probes present in the array, 6120 probes can be used against the transcripts of any one of the strain. Briefly, 5791 probes representing 5276 3D7 transcripts, 5374 probes representing 5251 HB3 transcripts and 69 probes representing 43 IT4

var gene transcripts are present in the array. Further classification of the probes based on the specificity was performed and there are 691 3D7 specific probes, 224 HB3 specific probes and 41 IT4 specific probes. These probes were designated specific for these strains, as they either did not show alignment, or showed multiple transcript alignments, for the genome of the other strains.

Table 3.1 Probe distribution in the 15K cross strain array

Total probes in the array	6362		
Probes representing any of the strain	6120		
Cross hybridizing probes	194		
Human Probes (To check the Human RNA Contamination)	22		
Number of probes that can't be used for any strain after updated annotation	48		
Probe Classification	3D7 Strain	HB3 Strain	IT4 Strain
Probes specific to	691	224	41
Probe common to anyone	5164		
Probe common to 3D7 and HB3	5077		
Probe common to HB3 and IT4	6		
Probe common between 3D7, HB3, IT4	9		
Probe common to 3D7 and IT4	13		

The array also contains 22 probes representing 8 human housekeeping genes and 536 Agilent controls. The total number of features that were available in the 8x15K array after filling the 536 Agilent control features was 15,208. Each designed probe (6362 *P. falciparum* probes and 22 human probes) was replicated to fill the features. The remaining 2440 blank features were filled by randomly selected duplicate probes from the set of probes specific to any one strain. All the 60mer oligonucleotide probes were designed and synthesized *in situ* as per algorithms and methodologies used by Agilent technologies for 60mer in situ oligonucleotide DNA microarrays.

3.2.2. Performance of probes to detect transcripts of *Plasmodium falciparum* field isolates

To evaluate the cross strain 15K array, RNA material from 13 *P. falciparum* clinical isolates was used. Clinical isolates were obtained from patients with either uncomplicated or complicated malaria symptoms (**Table 3.2**).

Table 3.2. Samples used for cross strain 15K array evaluation

S.No	Uncomplicated samples[#]	S.No.	Complicated samples[#]
1	PFU-02	1	PFC-06
2	PFU-03	2	PFC-13
3	PFU-04	3	PFC-10
4	PFU-05	4	PFC-11
5	PFU-06	5	PFC-12
6	PFU-08	6	PFC-23
		7	PFC-24

[#]Clinical characteristics of these samples can be found in **Table 2.1**

Probe efficiency was evaluated for its ability to detect the transcripts of the Indian *P. falciparum* field isolates. Based on the normalized signal intensity, 5502 probes were able to detect 4846 3D7 and 4845 HB3 transcript sequences in at least one out of thirteen samples analyzed. Transcripts were not detected by 618 probes. These probes may represent genes which might express in parasite stages not present in the samples analyzed here or might represent variant surface antigen coding genes. This point has been analyzed in detail in the next section. Due to *in situ* probe creation process, probes were of high quality and the background signal was quite low. Probes performance can be evaluated in multiple ways. Probes that are replicated in the array (probes with same sequence), reproducibility in detecting a transcript can be measured by calculating the Pearson correlation coefficient of their hybridization intensities. Similarly, performance of multiple probes representing a transcript (probes with different sequences) can be determined. There are 6120 probes representing *P. falciparum* transcripts sequences which are present as duplicates in the array. All the duplicate probes showed almost perfect correlation ($r = 0.99$) between them (**Figure 3.2**).

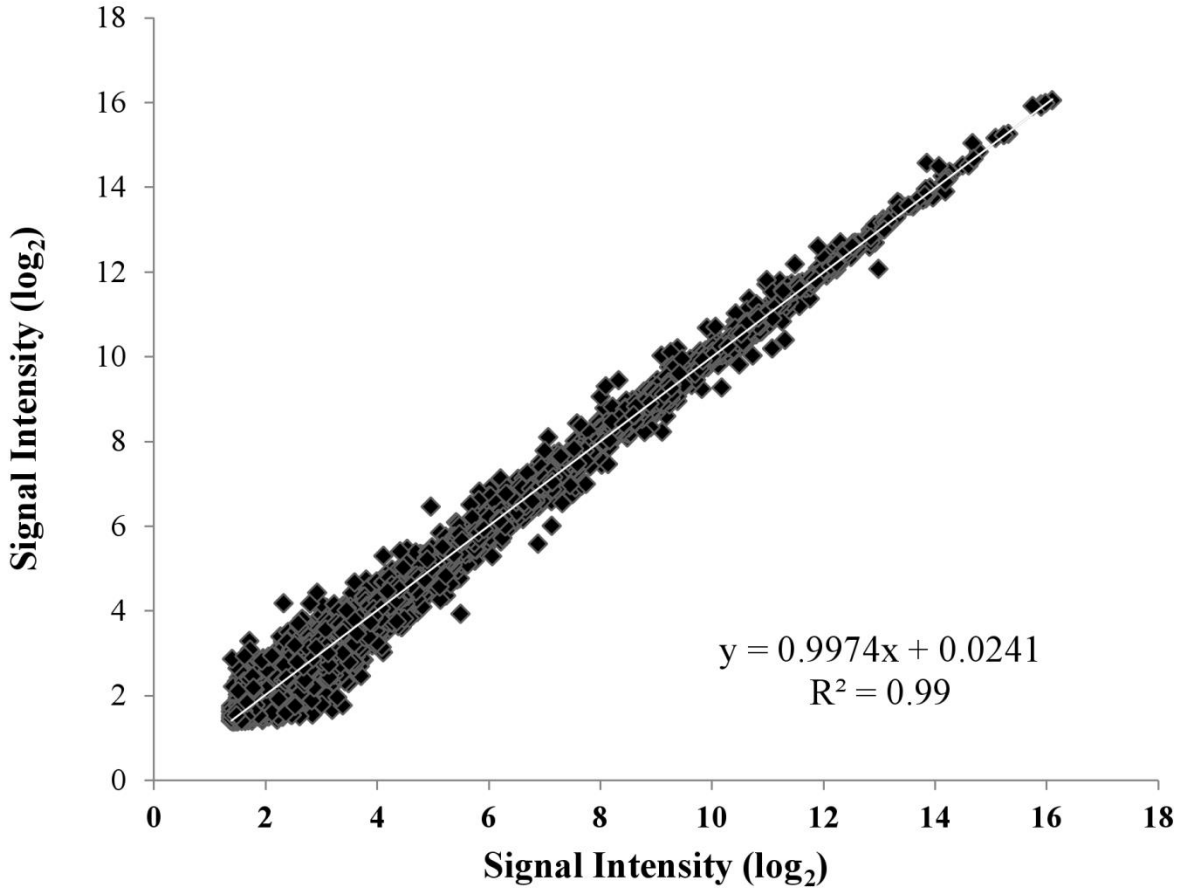


Figure 3.2. Correlation of replicated probes. Signal intensity (log_2) from the median of 13 RNA samples of replicated probe pairs (6120 pairs; $R^2 = 0.99$).

The array contains 1101 probes representing 522, 3D7 strain and accuracy of measurements of these individual probes was verified. Pair wise Pearson correlation of multiple probes representing the same gene was calculated. Seventy two percent of genes represented by multiple probes and detected in at least 4 out of 13 samples showed average Pearson correlation (r) above 0.8 between the probes representing the same gene. Percentage of genes with Pearson correlation between their representative probes >0.8 was found to increase, if probe detection status increased for number of samples, e.g. increased from 4 samples to 5 or 6 samples (**Figure 3.3**). This indicates that the correlation between probes representing the same transcript is high and multiple probes which could detect transcripts in the majority of samples can be taken for further analysis with high confidence level. One of the examples for genes with multiple probes is the gene encoding DBL like protein (PF11_0506). This gene is represented by 9 probes out of which

6 probes were detected in at least 4 samples and showed an average Pearson correlation of 0.986 (**Figure 3.4**). Probes representing a transcript with low Pearson correlation value may bind to sequence regions having variations, where one oligonucleotide may bind to the transcript efficiently in case of a test sample and the others may not. This would provide in-consistent signal intensity across the samples in comparison to the efficient binder. In case of such probes, those showing the highest consistent detection values across samples have been considered.

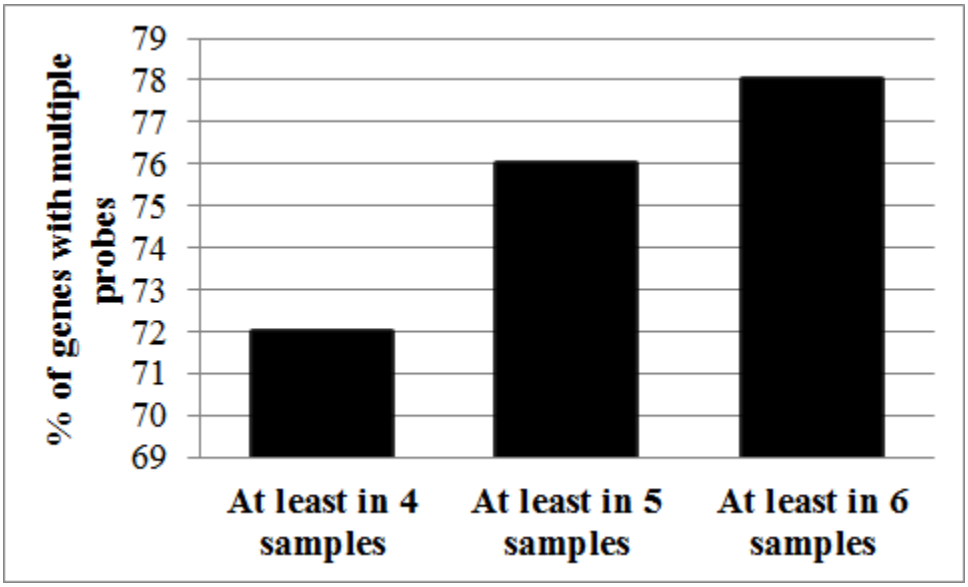


Figure 3.3. Relationship between correlation of multiple probes representing the same gene and probe detection status. The percentage of genes having average correlation of its multiple probes ≥ 0 .8 was determined for genes with its probes detected in at least 4, 5 and 6 samples out of 13 RNA samples.

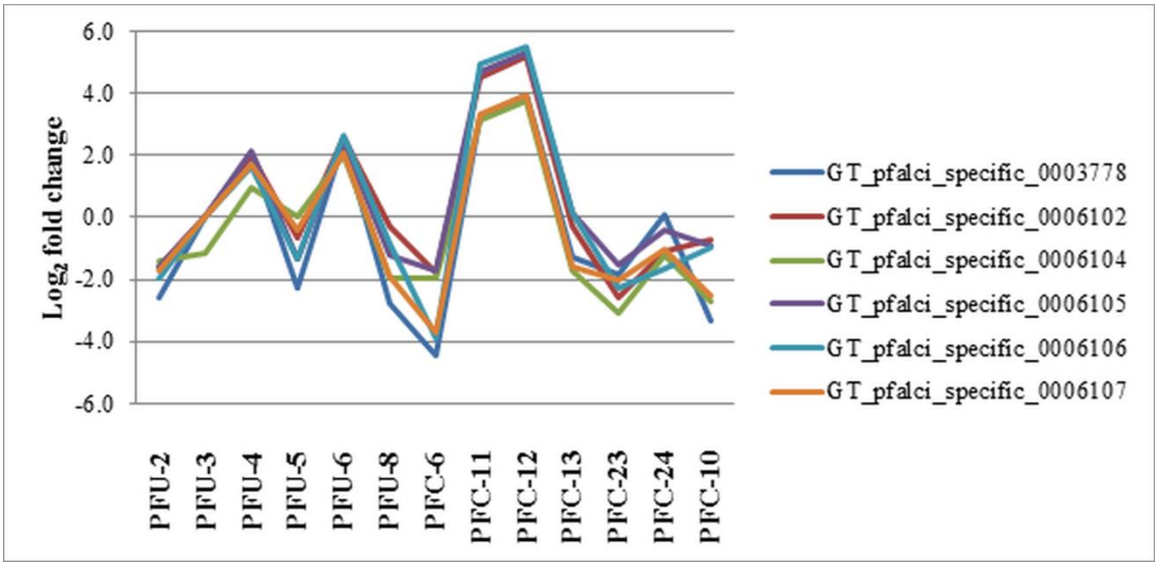


Figure 3.4. Expression profile of 6 probes representing antigen 332 encoding gene (PF11_0506). PF11_0506 is represented by 9 probes out of which 6 probes were detected in at least 4 samples. Pair wise Pearson correlation was calculated for detected probes. Average Pearson correlation of all the 6 detected probes was 0.986. Figure shows the expression patterns of all the 6 detected probes in 13 RNA samples.

The frequency of signal distribution of RNA/transcripts showed that many features were with low signal intensity, when probes irrespective of the filtering criteria (detected and not detected probe) were taken into consideration. After applying the filtering criteria and considering probes that were detected in at least one sample, many probes with low signal intensity were removed and the graph showed a bell shaped appearance adhering to the well accepted frequency distribution of signal intensities for whole genome expression profile (**Figure 3.5**). The low intensities for many features may be due to low expression of the genes in the blood stages of the parasite that are normally present in the peripheral blood of the infected humans from where the parasite material was isolated. The parasite exclusively expresses many of the genes that are stage specific (maximum expression at one point/stage of its lifecycle and express in low abundant or do not express in other stages) (Bozdech, et al., 2003 and Le Roch, et al., 2003).

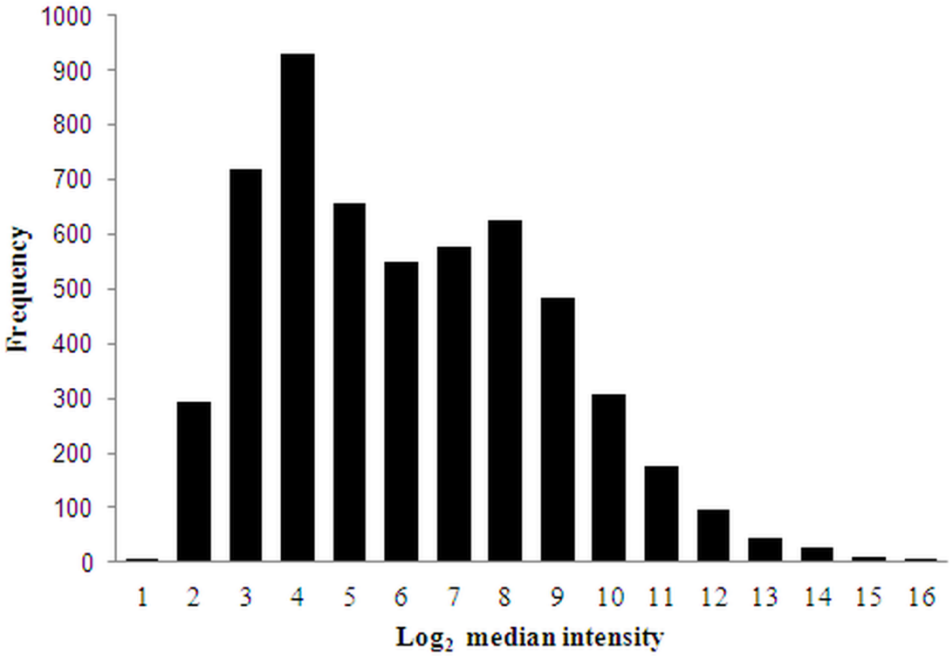


Figure 3.5. Probe intensity distribution: Frequency of intensity distribution of probes detected in at least 1 sample. Median intensity (Log_2) of probes in 13 RNA samples detected in at least 1 sample (n=5503) was taken and graph was plotted.

3.2.3. Stage specific expression pattern of detected and undetected transcripts

Developmental stage specific gene expression profiles of this parasite from culture condition have been reported and maximum and minimum expression stages for most of the genes have been determined (Aurrecoechea, et al., 2009, Bozdech, et al., 2003 and Le Roch, et al., 2003). Transcripts, which were not detected and detected in this study, were analyzed to find out the blood stages in which they express maximum and minimum (Aurrecoechea, et al., 2009, Bozdech, et al., 2003 and Le Roch, et al., 2003). In this study, 366 transcripts of 3D7 strain represented by probes were not detected (**Supplementary Table S3.2**). Investigation of their expression in the reported data revealed that 29 % of these (108 transcripts) expressed maximum in stages like early ring (ER)-5% (n=20), late ring (LR)-13% (49) and Gametocyte (G)-11% (39) which are normally found in peripheral blood of infected patients (**Figure 3.6**).

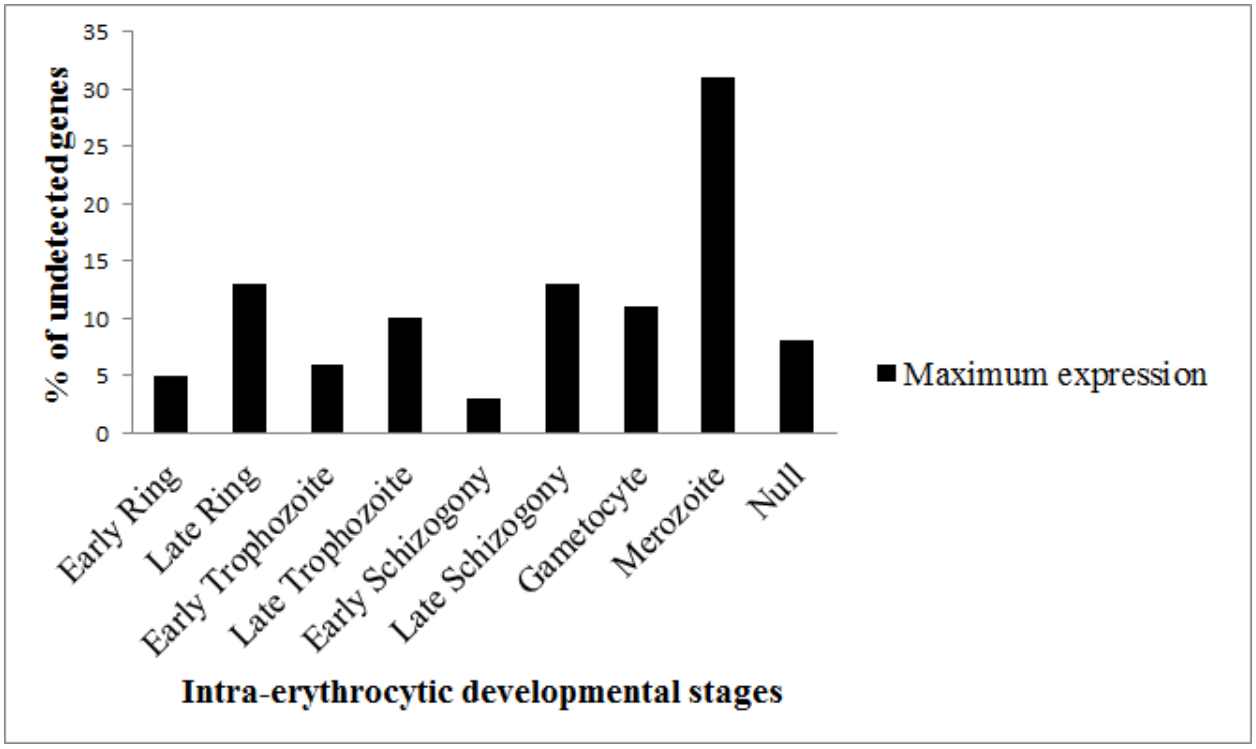


Figure 3.6. Maximum expression stage of undetected transcripts in different blood stages of *Plasmodium falciparum* (Aurrecoechea, et al., 2009 and Le Roch, et al., 2003). Blood stages at which each undetected transcripts were reported to be expressed maximum was retrieved from PlasmoDB v 8.2 (Aurrecoechea, et al., 2009**).**

We have investigated the probable cause for not detecting these transcripts although they were represented by specific probes. We could observe that many of these, which expressed maximum in early ring and late ring stages in culture condition constitutes variant parasite encoded erythrocyte surface antigens like *P. falciparum* erythrocyte membrane protein 1(PfEMP1) (n=5), repetitive interspersed family protein (RIFIN) (n=14), sub-telomeric variable open reading frame (STEVOR) (n=5) and *P. falciparum* Maurer's cleft- 2 trans membrane domain containing protein (PfMC-2TM) (n=4). There are 29 genes in the list for which expression data is not available in the PlasmoDB v9.2 (transcripts have not been detected in culture condition using microarray) (Aurrecoechea, et al., 2009 and Le Roch, et al., 2003). Many genes which express maximum in merozoite (M) stage might also express in ring stage due to possible involvement in linked processes like erythrocyte invasion and infected erythrocyte remodeling, so we have scanned the

list of genes which express maximum in this stage too. In the merozoite stage, 112 genes from the not detected list have been reported to express maximum, out of which again *rifin* (n=54) predominates in the list. Other genes that are present in the list are genes encoding for STEVOR (n=8), PfEMP1 (n=3), PHISTa (n=2) and hypothetical proteins (n=24).

Members of variant surface antigens (PfEMP1, RIFIN and STEVOR) are thought to be expressed in a mutually exclusive pattern where one member at a time expresses and others shut down. This has been well established in case of *var* genes encoding PfEMP1 antigens and might be the cause for not detecting these transcripts. This might also hold true for the other erythrocyte surface antigens (i.e. RIFINs and STEVORs). Members of *rifin* and *stevor* gene family were also reported to transcribe at a later stage of the intra erythrocytic developmental cycle. *Plasmodium* helical interspersed sub-telomeric (PHIST) family genes containing PHIST domains have been clustered into three groups (i.e. PHISTa, PHISTb and PHISTc). Microarray data from *P. falciparum* 3D7 strain has shown that PHISTa genes are generally not detectably transcribed (Sargeant, et al., 2006). Genes which expressed in other developmental stages (early trophozoite (ET), Late trophozoite (LT), early schizont (ES), late schizont (LS), liver and mosquito stages) and are not found in the peripheral blood of the infected patients may produce low amounts of transcripts under certain *in vivo* conditions, and are also found in the not detected transcripts list.

Transcripts were detected from 4,846 genes (93% of the genes represented in the array) in at least 1 sample out of the 13 samples analyzed (**Supplementary Table S3.3**). Majority of these have been reported to express maximum in G (21%), ER (14%), M (14%), LR (7%) and ET (13%) stages. Stages in which these genes expressed minimum were G (27%), ER (25%), M (12%) and LS (10%) stages (**Figure 3.7**). There were 218 genes for which expression data is not available from the previous studies using microarray (PlasmoDB v8.2) (Aurrecoechea, et al., 2009 and Le Roch, et al., 2003). This indicates that the probes present in the array were able to detect most of the represented transcripts if expressed in a detectable range.

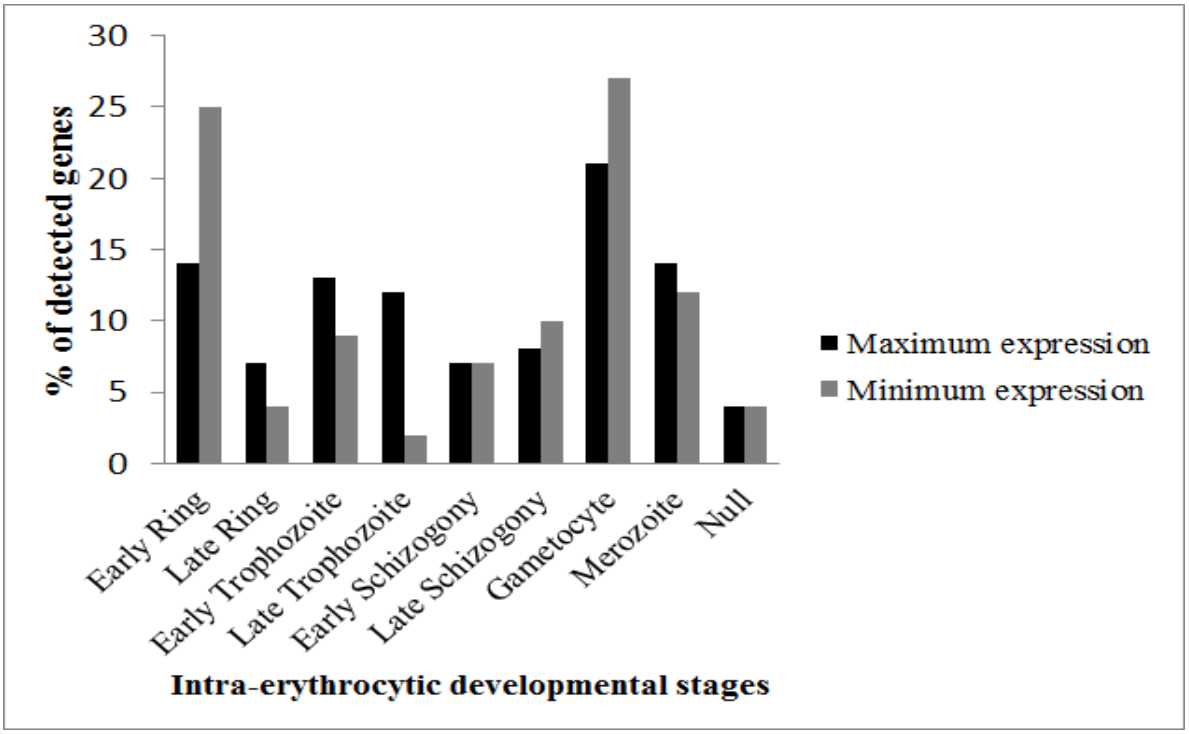


Figure 3.7. **Maximum and minimum expression stages of detected transcripts in different blood stages of *Plasmodium falciparum*** (Aurrecoechea, et al., 2009 Le Roch, et al., 2003). Blood stages at which each detected transcripts were reported to be expressed maximum and minimum was retrieved from PlasmoDB v8.2 (Aurrecoechea, et al., 2009).

3.2.4. Transcripts detection by probes specific to HB3 strain

The array contains 288 HB3 strain specific probes (Only can be used for HB3 strain) representing 282 HB3 transcripts. Out of these 288 probes, 147 probes are specific to HB3 strain because they did not show similarity with any of the 3D7 transcript sequences (based on BLAST results) and 141 probes were assigned as specific to HB3 strain because they showed multiple hits against 3D7 transcripts (cross hybridizing probes for 3D7 strain) but showed single hit to any one of the HB3 strain transcript (**Supplementary Table S3.4**). As many as 202 probes representing 196 transcripts were detected in at least one sample. Of the 147 probes, that are specific to HB3 (No hits found against the 3D7 transcripts), 106 probes representing 103 genes were able to detect transcripts (**Supplementary Table S3.4).** Most of these transcripts code for hypothetical proteins except for some PfEMP1 and RIFINs. This suggests that variations in the transcript sequences between strains and transcript of genes that are additional to a strain can be captured by the array. This would give information about the expression status of the transcripts that are expressed, but could not be earlier detected due to strain specific sequence variability, as

most of the arrays relied on the genome sequence of *P. falciparum* 3D7 reference strain. This could be true for studies investigating the *in vivo* transcriptome status such as by Daily, et al., 2007.

3.2.5. Expression pattern of variant surface antigens

This array is supplemented with probes representing the parasite genes encoding variant surface antigens (VSA) of three geographically distinct strains. Probes in the array represent the *var* genes of 3D7, HB3 and IT4 strains and *rifin* and *stevor* genes of 3D7 and HB3 strains. Transcript detection capability of these probes was ascertained.

Arrays used in this study contain specific probes representing *var* genes of the three strains (40 *var* genes of 3D7 strain, 30 of HB3 strain and 34 of IT4 strain) (**Supplementary Table S3.5**). Probes representing *var* genes were either specific for one strain or could be common between strains. All the full *var* genes were considered for this analysis and *var* pseudo genes, *var* gene fragments and *var* like genes were excluded from the analysis, although many of these are represented by specific probes. Many *var* genes that are not represented by specific probes are represented by cross hybridizing probes, where a single probe represents multiple *va*r genes of a strain. Usually these represent genes from the same *var* group. Because of the high sequence similarity between *var* genes, it is difficult to design 3' biased specific probes for all the members of the *var* gene family. Transcripts for 35, 3D7 *var* genes, 25 HB3 *var* genes and 23 IT4 *var* genes could be detected in at least 1 sample by representing probes (**Supplementary Table S3.5)**. Taking the 3D7 strain as an example, transcripts were detected from members of all the *var* group represented by probes but was dominated by *var* group B sub-family genes.

The other two variant surface antigens that express on the surface of the infected erythrocytes are RIFIN and STEVOR encoded by *rifin* and *stevor* gene families. Genome of a single parasite may have 150-200 *rifin* family genes, which is the largest multicopy gene family in *P. falciparum* (Petter, et al., 2007). RIFINs have been divided into two sub-groups based on the presence and absence of a 25 amino acid motif in its protein sequences. The two sub-groups have been named as A and B type RIFINs. The *P. falciparum* reference strain haploid genome contains 28 copies of the *stevor* gene and is closely related to RIFINs.

Joannin, et al., (2008) grouped 134, 3D7 and 59, HB3 RIFIN protein sequences into two groups (97 RIFINs in group A and 37 RIFINSs in group B for 3D7 and 50 RIFINs in group A and 9 RIFINs in group B for HB3 strain, respectively). Array contains specific probes for 80 group A and 19 group B *rifin* genes of 3D7 strain and 43 group A and 9 group B *rifin* genes of HB3 strain. Transcripts were detected from 40, 3D7 *rifin* genes (34 *rifin* group A and 6 *rifin* group B genes) and 27 HB3 *rifin* genes (22 *rifin* group A and 5 *rifin* group B genes) by probes in at least 1 sample (**Supplementary Table S3.6**).

Array contains probes for 26 *stevor* genes out of 28 genes present in the *P. falciparum* 3D7 reference strain (Gardner, et al., 2002). Transcripts were detected from 14 *stevor* genes (**Supplementary Table S3.7**). Out of 28 *stevor* genes reported to be present in the HB3 strain (Joannin et al., 2008), 23 are represented by specific probes. Transcripts were detected from 13 of these genes (**Supplementary Table S3.7**). The expression of *rifins* were reported to peak at 12-27 hours (ring and trophozoite stages) post invasion whereas *stevors* was reported to peak at 22-32 hours (trophozoite and schizont stages) post invasion of merozoites but expression of these genes was not only restricted to these stages and might express in other stages too (Scherf et al., 2008). Probes representing these VSA gene transcripts could efficiently detect the transcripts although some of them expressed at low levels. It is essential to detect these transcripts *in vivo* as their expression might determine possible interaction with the host molecules and henceforth the pathological outcome (Miller, et al., 2002).

3.2.6. Gene expression profiling of clinical isolates

Hierarchical clustering (similarity measure used Pearson centered, linkage rule used average) of the parasite whole genome expression profiles of 13 samples, segregated parasite samples into two broad clusters (i.e. Cluster 1 and Cluster 2). Cluster 1 contains 7 samples and cluster 2 contains 6 samples (**Figure 3.8**). Only probes representing the 3D7 strain were considered for this analysis. Comparison between cluster 1 and cluster 2 was performed to investigate genes, which differed between the two clusters. After allowing multiple testing corrections (Benjamini Hochberg False Discovery Rate) by applying a cut-off of ≥ 2.0 fold changes based on unpaired T test, the number of probes that differed between cluster 1 and cluster 2 were 1260 representing 1214 genes.

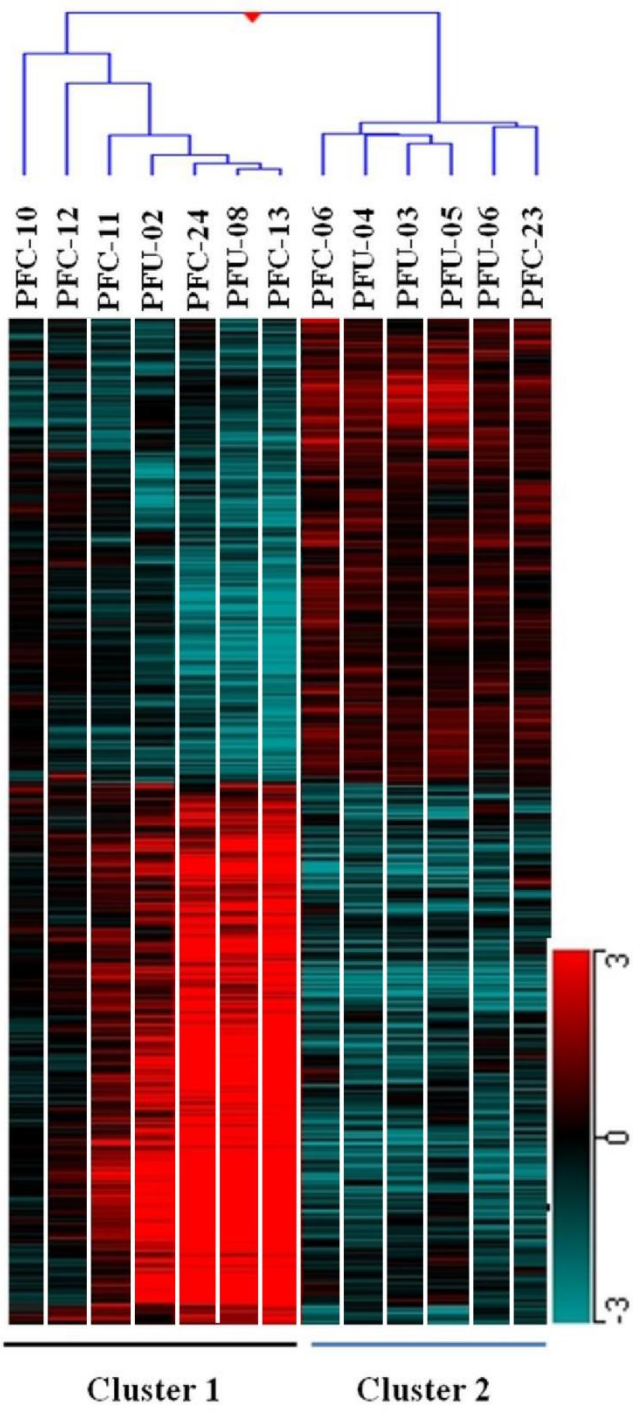


Figure 3.8. **Heatmap showing combined hierarchical clustering of differentially regulated probes in 13 samples**. Clustering applied on both samples and probes. Pearson Centered distance matrix and average linkage rule was used. 1260 differentially regulated probes (representing 1214 genes) were included to generate the tree. Hierarchical clustering demarcated total samples under investigation into two clusters.

Of the 1214 genes declared as significant, 53 % (649) were up-regulated and 47 % (565) were down-regulated in cluster 1 compared to cluster 2 (**Supplementary Table S3.8**). Genes that differed between the two clusters were compared with already reported data from laboratory strains to identify blood stages at which they expressed maximum and minimum (Aurrecoechea, et al., 2009 and Le Roch, et al., 2003). Of the 649 genes up-regulated, 63% (407) of them were reported to express maximum at gametocyte stage and of the 565 genes down regulated, 47% (268) of them were reported to express minimum at gametocyte stage (**Supplementary Table S3.8**). Up-regulation of a large proportion of genes in cluster 1 which reportedly express maximally at gametocyte stage(s), and similar down-regulation of a large proportion of genes in cluster 1, which express minimally at gametocyte stage(s), indicates the presence of gametocyte stages in samples present in cluster 1. This was confirmed by microscopic peripheral blood smear examination of samples (PFU-02, PFU-08, PFC-13 and PFC-24) present in cluster 1, where gametocyte stages along with ring stages were observed. Analysis of the sexual development stage transcriptome has identified genes (267 genes in the recent database but 246 genes reported in the original paper) possessing gametocyte-stage specific mRNA expression (Young, et al., 2005). Of these 267 genes, 171 genes (64%) were found to be up-regulated in cluster 1 (**Supplementary Table S3.9**). A total of 162 genes were reported to express maximum at the gametocyte stage, 7 at the merozoite stage, 1 at the late trophozoite stage and 1 at the late schizont stage compared to other blood stages (**Supplementary Table S3.9**). This again confirms microscopic observations showing presence of gametocyte stages in cluster 1 samples.

The remaining set of genes which differed between cluster 1 and cluster 2 and expressed maximum/minimum in stages (Early ring, late ring, early trophozoite, late trophozoite, early schizont and late schizont) other than gametocyte stage constitute 37% of the up-regulated gene group and 53% of the down-regulated gene group (**Supplementary Table S3.8**). Most of the time either one (ring stage) or two stages (ring and gametocyte stages) are found in the peripheral blood of infected patients, thus the over expression of genes in cluster 1 that were reported to be expressed maximum at stages like early and late trophozoite, and early and late schizont requires further investigation. The current understanding is that all the merozoites from a single schizont are committed to develop into gametocytes, suggesting the commitment is determined at earlier stages during the asexual cycle (Guttery, et al., 2012 and Silvestrini, et al., 2005). There are reports which have identified genes whose protein products contribute to the gametocyte stage

proteome and are specific to this stage (Florens, et al., 2002; Lasonder, et al., 2002). The reported gene lists were compared with the up-regulated gene list to determine in which stage they express maximum *in vitro*. We observed that 29 genes reported by Lasonder *et al* (2002) (**Supplementary Table S3.10**), and 28 genes reported by Florens *et al* (2002) (**Supplementary Table S3.10**), whose protein products are specific to gametocyte stages, expressed maximum at stages other than gametocyte stages (Aurrecoechea, et al., 2009 and Le Roch, et al., 2003). Most of them are from late trophozoite, early schizont and late schizont stages. This supports the earlier statement suggesting that commitment to sexual stage development is determined at asexual stages (Guttery, et al., 2012 and Silvestrini, et al., 2005). It is important to note that the up-regulated gene list from our experimentation contains many genes other than the above stated genes which reportedly expressed maximum at later stages of the asexual intra erythrocytic cycle. Investigation of these genes in the up-regulated gene list, other than housekeeping genes, might give us clues about their probable vital roles in sensing the external stimuli and inducing gametocytogenesis.

This study analysis aims at validating the custom cross strain microarray. The overall examination of the microarray data is partly presented as a preliminary analysis to validate the strength and accuracy of the array design and procedures.

3.2.7. Real-time PCR based validation of few differentially regulated genes

To confirm the microarray results, five genes were examined by real-time qPCR (**Table 3.3**). Real time qPCR results confirmed up-regulation of three genes out of four up-regulated genes from the microarray analysis (> 1 Log_2 fold change) and one gene showed Log_2 fold change expression of 0.13 (this did not qualify the criterion considered for designating up-regulated transcript expression) (**Figure 3.9**). In case of down-regulated gene, a Log_2 fold change expression of -0.70 was observed from qPCR analysis (**Figure 3.9**). Real time qPCR validated overall expression pattern of selected genes as seen in the microarray data.

Table 3.3. Comparison between expression status of selected genes in 15K array and real-time qPCR data.

Gene	Expression status in 15K array (Log_2 Fold Change)	Expression status in real-time qPCR (Log_2 Fold Change)
PF13_0269	4.99	3.59
MAL8P1.16	4.97	3.13
PFF1280w	3.78	0.13
PF14_0683	5.84	6.75
PF10_0155	-1.96	-0.70

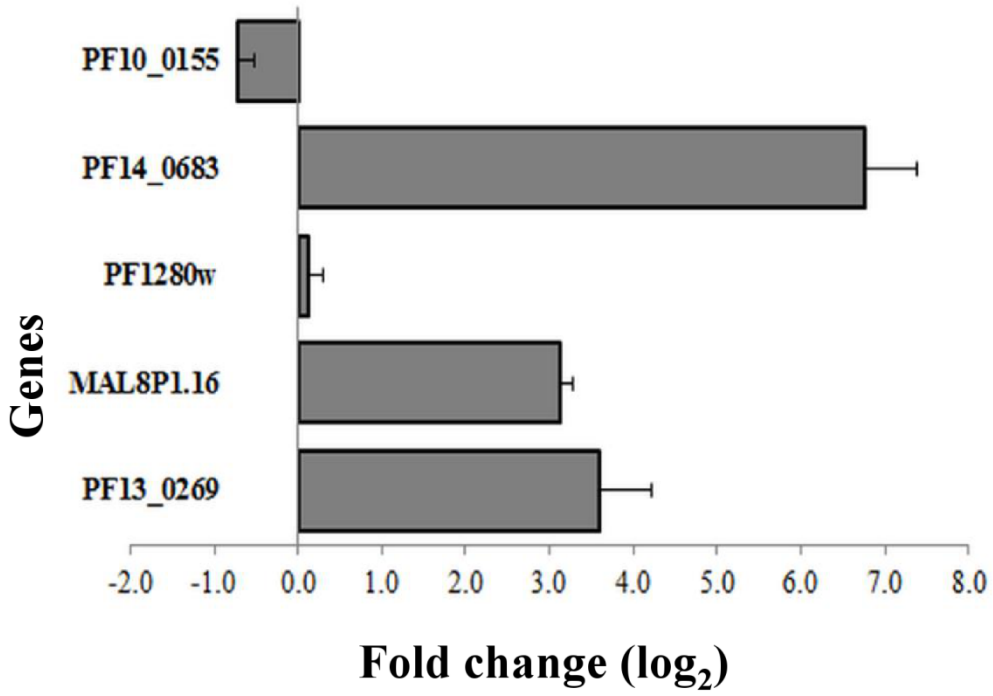


Figure 3.9. Quantitative real-time PCR based validation of 5 differentially expressed genes in cluster 1 compared to cluster 2. Log_2 fold change expression was calculated using 2^{-ddCt} method. Seryl tRNA synthetase was used as an endogenous control gene**.**

In summary, a cross-strain *P. falciparum* array has been designed for *in vivo* studies utilizing parasite material taken directly from patients. This array has been used for detection and analysis of the "sense" transcriptome of patient derived parasite material from North-West India. Analysis of the array signals has established the strength of the array design and procedures adopted.

As more sequence information of Indian *P. falciparum* strains emerge, additional oligonucleotide probes will be added to this flexible and evolving cross strain array. In addition to probes detecting "sense" transcriptome, probes detecting "antisense" transcriptome could also be included for the documentation of pervasive antisense transcription in this parasite. Emerging evidence also suggests the transcription of non-coding RNAs and probes against these may also be included as this platform provides flexibility in probes inclusion and updating array.

4.1. Introduction

Natural antisense transcripts (NATs) are transcribed from the strand opposite to the template DNA strand and may hybridize with the sense transcripts of same genomic loci (cis-NATS) or to the complementary transcripts of separate genomic loci (trans-NATs). These transcripts can derive from coding RNA and non-coding RNA including genic, intronic and intergenic regions of the genome. NATs have first been reported in prokaryotes and have shown to be encoded on plasmids, phage and transposons (. They are involved in diverse biological functions which include repression of tranasposase and toxin protein synthesis; regulation of transcription regulators levels; modulation of metabolic and virulence proteins, plasmid replication, incompatibility and conjugation and bacteriophage temporal control development (Thomason, et al., 2010). Subsequently, antisense transcripts were also described in eukaryotes. Antisense transcription was first reported from human and mouse mitochondria in 1981 (Anderson, et al., 1981 and Bibb, et al., 1981) followed by from *Drosophila* in 1986 (Spencer, et al., 1986).

They have also been implicated in regulating gene expression in eukaryotes through diverse postulated mechanisms (Faghihi and Wahlestedt, 2009). Gene expression regulation by NATs in different organisms includes genomic imprinting, transcriptional collision, X chromosome inactivation, alternative splicing and termination, RNA interference, translational regulation, and RNA editing (Faghihi and Wahlestedt, 2009 and Lavorgna, et al., 2004). In case of genomic imprinting, antisense transcripts have shown to guide chromatin and DNA modification of genomic loci that can expand to neighboring genes whereas X chromosome inactivation in mammals is executed by an antisense transcript called *TSIX* (X [inactive]-specific transcript, antisense). Transcriptional collision based gene expression control rely on the assumption that RNA polymerses in the convergent genes on opposite strands of DNA may collide during the transcription in the overlapping regions, blocking further transcription (Shearwin, et al., 2005). Antisense RNA in some cases may bind and mask the splice sites of the sense RNA thereby, changing the balance between splice variants (Hastings, et al., 1997). Same may also hold true for antisense transcripts that have the ability to cover the binding sites of micro-RNA which in turn can prevent the degradation of sense transcripts.

Microarray and proteomics based data have revealed significant regulation of transcripts and protein expression profiles associated with the morphological and developmental changes during the various stages of its life cycle (Bozdech, et al., 2003a&b and Le Roch, et al., 2003 & 2004) Remarkable changes in the transcript abundance of the *P. falciparum* transcriptome were observed when subjected to different perturbed conditions (Natalang, et al., 2008, Oakley, et al., 2007 and Tamez, et al., 2008). Transcriptional variation has also been documented between the culture adapted field isolates and long-term laboratory-adapted isolates (Mackinnon, et al., 2009). *In vivo* expression profiles of the *P. falciparum* parasite derived directly from blood samples of infected patients showed distinct transcriptional states with a probable role of epigenetic based mechanisms behind this has been suggested (Daily, et al., 2007).

The molecular mechanisms underlying this pattern of transcriptional regulation are not well understood (Painter, et al., 2011). The number of conserved transcription factors annotated in this parasite is extremely low. This has been partially compensated by the discovery of a large Apicomplexan AP2 (ApiAP2) protein family containing Apitela 2 (AP2) domains (Balaji, et al., 2005). There are many other possibilities of how gene regulation might happen in this parasite like controls at transcriptional (Cui and Miao, 2010 and Painter, et al., 2011), post transcriptional (Shock, et al., 2007) and/or at translational level (Mair, et al., 2006). Gene regulation could also be due to natural antisense transcripts (NATs), which may influence gene expression patterns (Militello et al., 2008).

Looking at the complex life cycle of the protozoan parasites which involves multiple developmental stages, hosts and changing environments, it is very much likely that NATs may have a similarly diverse regulatory role in these organisms. The identification of NATs from the transcriptomes of many protozoan parasites including *P. falciparum* (Gunasekera, et al., 2004 and Patankar, et al., 2001), *Toxoplasma gondii* (Radke, et al., 2005), *Theileria parvun* (Bishop et al., 2005), *Trypanosoma brucei* (Lininger, et al., 2001), *Leishmania Spp.* (Kapler and Beverley, 1989 and Dumas et al., 2006) and *Giardia lamibia* (Ullu, et al., 2005) supports the above mentioned statement. (Lavorgna, et al., 2004)

NATs in *P. falciparum* transcriptome were first reported from SAGE data (Gunasekera, et al., 2004 and Patankar, et al., 2001). In addition, NATs corresponding to the *P. falciparum var* genes

(Ralph, et al., 2005) and MSP2 gene (Kyes, et al., 2002) which are involved in immune evasion and merozoite invasion respectively have also been identified. Further evidence supporting the presence of antisense transcripts have come from different groups using various approaches like microarrays, nuclear run-on experiments, cDNA sequencing, and RNA-sequencing (Lopez-Barragan, et al., 2011, Lu et al., 2007, Militello, et al., 2005, Raabe, et al., 2010, Ralph, et al., 2005 and Sorber, et al., 2011). Many of these studies have reported the presence of NATs in various developmental stages of this parasite suggesting its important role in gene expression regulation. Moreover, NATs are confirmed to be synthesized by RNA polymerase II, evidence of came from nuclear run-on experiments using specific RNA polymerase inhibitors (Militello, et al., 2005). The diversity of NATs reported till date is from laboratory-adapted isolates. Dramatic transcriptional changes have been observed in *P. falciparum* in response to perturbations like febrile temperature and glucose starvation. As the evidence towards the role of NATs in genome regulation is increasing day by day, it would be logical to dissect the unexplored prevalence of NATs in *in vivo* disease conditions.

This chapter deals with the identification of natural antisense transcripts in *P. falciparum* clinical isolates from patients showing uncomplicated and complicated disease conditions on a genome wide scale using custom designed strand specific microarray. Their expression pattern, genome-wide distribution and differences in prevalence under differing clinical conditions have been explored. Functional analysis has been performed to try and understand the probable biological role. Finally, the data has been compared with the previously published data from *in vitro* culture conditions. To the best of our knowledge this is the first report of the presence of NATs in *P. falciparum* infections causing complicated disease conditions.

4.2. Results

To investigate the *in vivo* prevalence of antisense transcripts in clinical isolates, we have collected blood samples from 11 *P. falciparum* infected patients suffering from either uncomplicated (PFU) (n = 2) or complicated malaria (PFC) (n=9) (**Table 4.1**). Parasite RNA was isolated from un-complicated and complicated malaria samples and pooled in equimolar amount separately. Cy3 labeled cRNA was prepared from un-complicated malaria samples and Cy5 labeled cRNA was prepared from complicated malaria samples. Labeled cRNA samples

were hybridized to a custom 244K strand specific whole genome microarray and data acquisition was carried out.

The strand specific custom 244K *P. falciparum* microarray was originally designed for selecting best probe(s) for our custom cross strain *P. falciparum* 15K array. During the probe evaluation experiment, we could see that many antisense probes (antisense transcript specific probes) showed hybridization signal which was well above background. This indicated us that what we detecting are true antisense transcripts in this parasite which insisted us to perform detail analysis to understand its functional implications.

To validate the results obtained from array hybridization, strand specific RT PCR and strand specific quantitative real-time PCR (qPCR) were performed. The PFU pool for this experimental validation included five isolates (n=5). The PFC pool for the same consisted of nine (n=9) isolates (**Table 4.1**). The PFC and PFU samples which were pooled for strand specific RT and strand specific real-time qPCR included RNA from some samples which were not hybridized on the array.

Table 4.1. *P. falciparum* clinical isolates used in the strand specific 244K array experiment

S.No.	Uncomplicated samples[#]	S.No.	Complicated samples[#]
1	PFU-01	1	PFC-01
2	PFU-02	2	PFC-02
3	PFU-03	3	PFC-03
4	PFU-04	4	PFC-04
5	PFU-05	5	PFC-05
6	PFU-06	6	PFC-06
		7	PFC-07
		8	PFC-08
		9	PFC-09
		10	PFC-10

Samples with code PFU-01, PFU-02 and PFC-01 to PFC- 09 were used in microarray hybridizations. Samples with code PFU-01, PFU-03 to PFU-06, PFC-01 to PFC-08 and PFC-10 were used in strand specific RT and quantitative Real-Time PCR experiments. [#]Clinical details of these samples can be found in **Table 2.1.**

Out of 8 probe groups present in the array (**Table 4.2**), only data from sense and antisense probe groups representing the *P. falciparum* PlasmoDB transcripts were analyzed as these transcripts were well annotated. The pre-analysis includes transcripts detected by at least 1 representing probe. Initial analysis of the expression data revealed that 3,079 genes (58% of the total genes represented in the array) and 2493 genes (47%) expressed sense transcripts (S) while 1,803 genes (34%) and 1847 genes (35%) expressed antisense transcripts (AS) in complicated (PFC) and uncomplicated (PFU) samples respectively (**Supplementary Table S4.1**). When individual expression of S and AS transcripts was considered, 1532 and 1386 genes were with only S transcripts, 256 and 741 genes were with only AS transcripts and 1547 and 1106 genes were with both S and AS transcripts in PFC and PFU respectively (**Supplementary Table S4.1**). Chromosome wise comparison of genes showed the distribution of genes with S and AS transcript pairs across the genome (**Figure 4.1**). Genes in chromosome 13 and 5 (32.17 and 32.03%) showed the highest percentage of antisense transcription compared to the genome average (29%) whereas genes in chromosome 6 and 7 (25.46 and 26.4%) showed the lowest percentage of antisense transcription.

Table 4.2. Overview of the 244K array probe distribution.

Total Number of features			243504		
Total number of probes designed			241399		
Agilent controls			2105		
Probe groups					
	Probes	Transcripts		Probes	Transcripts
Sense probes- (NCBI EST sequences)	72631	19544	Antisense probes – (NCBI EST sequences)	72927	19725
Sense probes – (PlasmoDB transcripts)	**39737**	**5369**	**Antisense probes – (PlasmoDB transcripts)**	**47127**	**5378**
Sense probes – (*P. falciparum* apicoplast sequences)	46	6	Antisense probes – (*P. falciparum* apicoplast sequences)	56	6
Sense probes – (*P. vivax* apicoplast sequences)	23	3	Antisense probes – (*P. vivax* apicoplast sequences)	23	3

Probe groups considered in this analysis are marked in bold

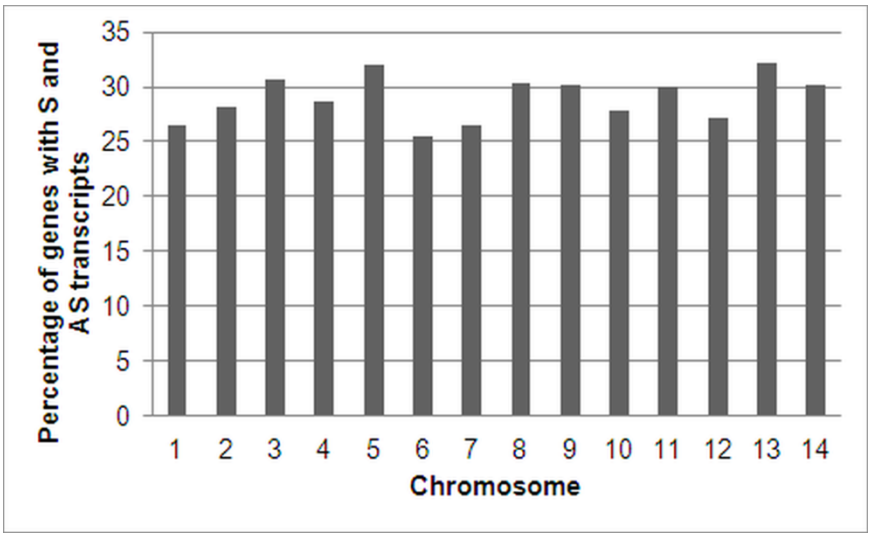


Figure 4.1. Overall chromosomal distribution of genes with both sense and antisense transcripts. S; sense transcripts and AS; antisense transcripts.

4.2.1. Prevalence of antisense transcripts

Stringent filtering criteria was applied on the gene based data and genes with S, AS or both S-AS transcripts detected by at least 3 probes were considered for further analysis. The genes were separated into three groups based on the type of transcripts they expressed; genes with only S transcripts, only AS transcripts and both sense and antisense (S-AS) transcripts. S-AS transcripts were detected from 736 and 267 genes and only AS transcripts were detected from 34 and 35 genes respectively from PFC and PFU (**Table 4.3**, **Supplementary Table S4.1**). In total, this study could identify 797 genes with AS transcripts with high confidence. Of these, 788 genes were protein coding, 7 genes were rRNA coding and 2 genes were non protein coding in nature. Both S and AS transcripts shown broad dynamic range in expression. The expression ratios of S transcripts were from as low as 1.5 up to 1465.01 and of the AS transcripts were from 1.5 to up to 240.58. Expression ratio of the AS and S transcripts of genes with both S-AS transcripts were found to be quite high (median expression ratio of AS transcripts 3.55 and 1.83 and of S transcripts 12.91 and 3.16 in PFC and PFU respectively) compared to the expression ratio of the AS and S transcripts of genes with only AS or S transcripts (median expression ratio of AS transcripts 2.89 and 1.73 and of S transcripts 3.93 and 1.85 in complicated and uncomplicated cases respectively).

Table 4.3. Summary of genes with antisense transcripts in complicated and uncomplicated isolates.

Antisense transcript detection status	PFC	PFU	PFC+PFU
Only antisense transcripts	34	35	
Both sense and antisense transcripts	736	267	
Total antisense transcripts	772	302	
Total number of antisense transcripts detected		797	
Protein coding			788
rRNA coding			7
Non protein coding			3

Note- Table shows list of genes with transcripts detected by at least 3 probes

No significant pattern of inter or intra chromosomal distribution could be found for genes with AS transcripts (**Figure 4.2**). Visual inspection identified 114 antisense bearing gene clusters where genes with AS transcripts were adjacent to each other (Physical cluster) (**Figure 4.3**, **Supplementary Table S4.2**). Genes in the physical cluster were found to be oriented in all the possible combinations: head to head and tail to tail for genes in opposite strands and head to tail for genes in the same strand. Most of the genes in the physical cluster were non-overlapping in nature. Physical clusters include 2-4 adjacent genes with antisense transcripts. One such example is operon like arrangement of four genes in head to tail orientation located in chromosome 14 (PF14_0075 to PF14_0078) (**Figure 4.4A**). Genes in many physical clusters also showed functional association between them. Examples include; four genes (PF14_0075, PF14_0076, PF14_0077 and PF14_0078) encoding aspartic-type endopetidase (Plasmepsin) and involved in hemoglobin catabolic process; four genes that code for *Plasmodium* exported protein (PFE0050w), heat shock protein 40 (hsp 40), type II (PFE0055c), Parasite-infected erythrocyte surface protein (PFE0060c) and skeleton binding protein 1 (PFE0065w) are members of the exportome and involved in infected erythrocyte remodeling (**Figure 4.4B**) and three genes (PF07_0029, PF07_0030 and PF07_0031) code for heat shock protein 90 and 86 family proteins (**Figure 4.4C**).

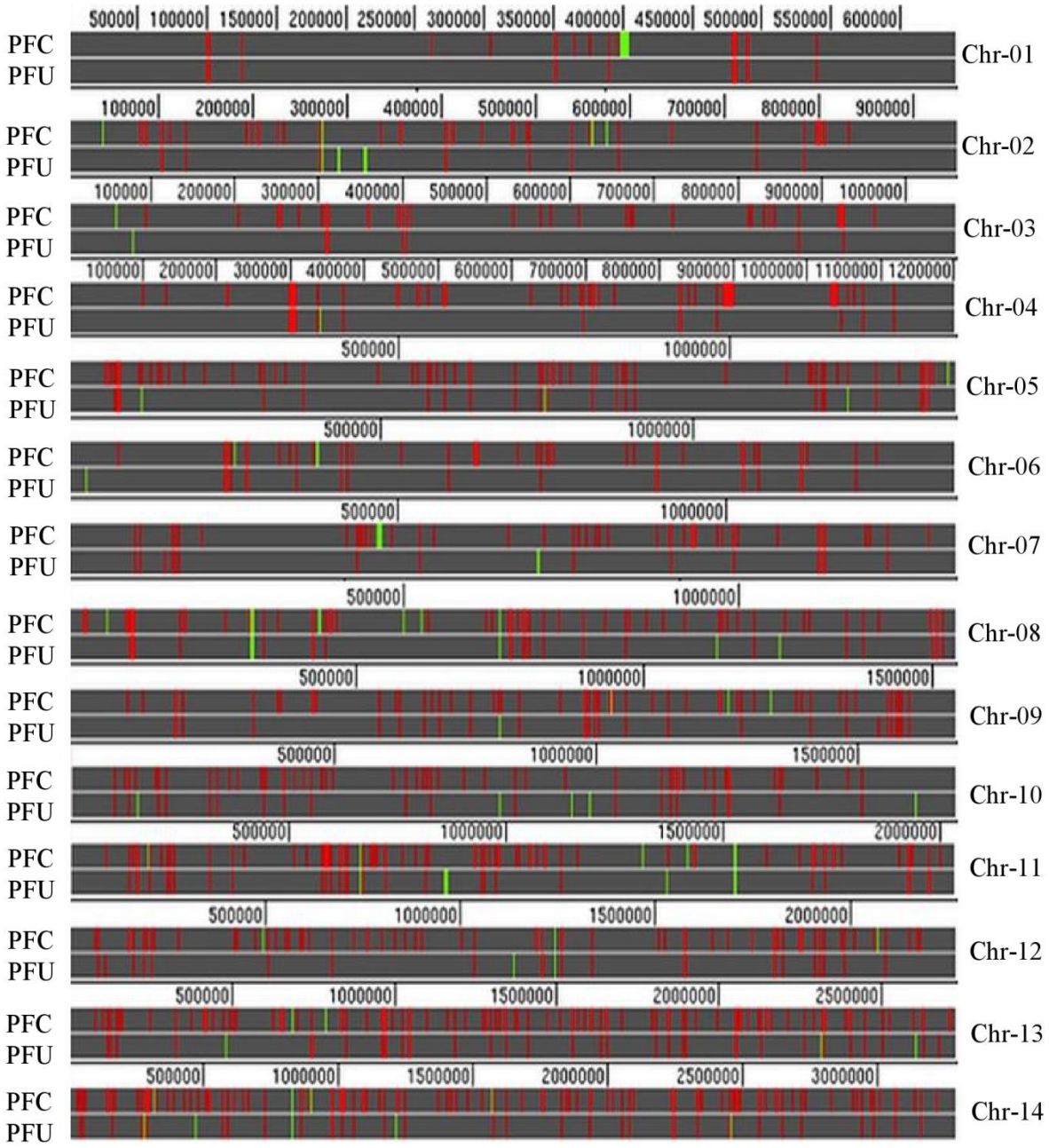


Figure 4.2. **Physical map of genes with antisense transcripts.** Colored vertical lines depicting the gene are plotted as a function of position across the parasite genome sequence. Chromosomes are depicted as grey horizontal bars. For each chromosome there are two grey horizontal bars, one representing uncomplicated malaria and the other representing complicated malaria samples. Fluorescent green vertical lines represent the genes with only antisense transcripts and red lines represent the genes with both antisense and sense transcripts. Abbreviations: Chr, chromosome; PFC, *P. falciparum* complicated isolates; PFU, *P. falciparum* uncomplicated isolates. Figure generated using University of California, Santa Cruz (UCSC) genome browser.

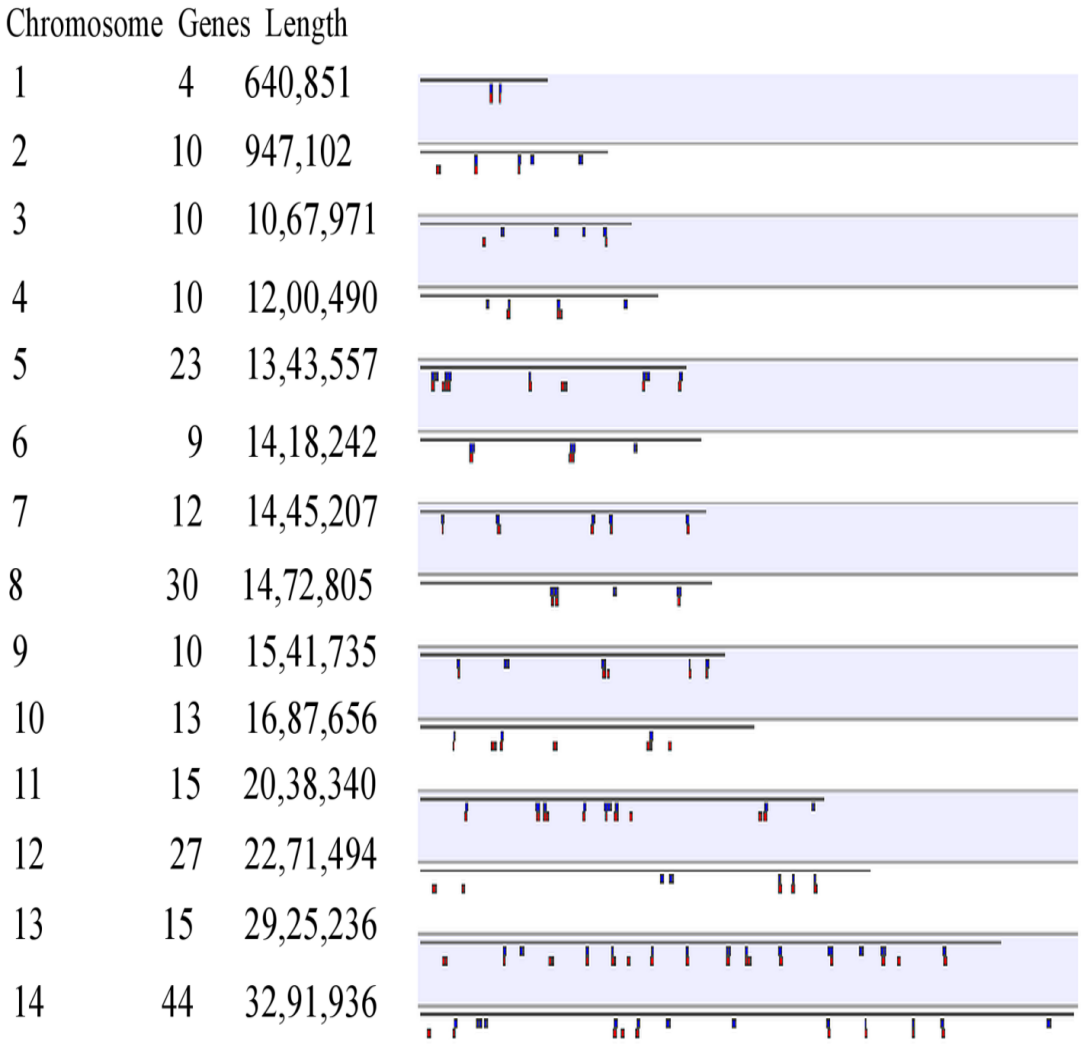
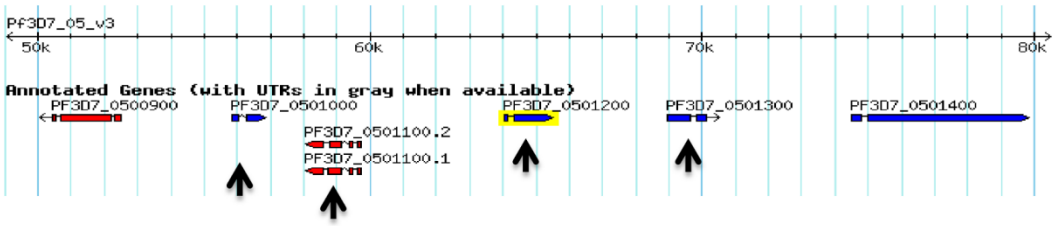


Figure 4.3. Genome-wide distribution of genes with antisense transcripts in physical clusters. Figure depicts two or more neighboring genes with antisense transcripts distributed across the 14 chromosomes of *P. falciparum*. Figure generated using PlasmoDB v8.2 genome browser.

PF14_0075- PF3D7_1407800, PF14_0076- PF3D7_1407900, PF14_0077- PF3D7_1408000, PF14_0078- PF3D7_1408100

PFE0050w-PF3D7_0501000, PFE0055c-PF3D7_050100.1, 050100.2, PFE0060w-PF3D7_0501200, PFE0065w-PF3D7_0501300, PFE0070w-PF3D7_0501400



PF07_0029 - PF3D7_0708400, PF07_0030 - PF3D7_0708500, PF07_0031 - PF3D7_0708600

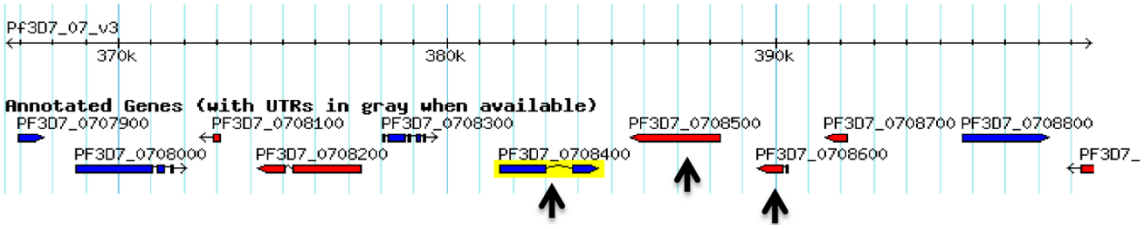


Figure 4.4. Cluster of genes with antisense transcripts. Figure shows few examples of genes with antisense transcripts physically adjacent to each other in the genome. Arrow marks show genes with antisense transcripts.

To find out the fold change expression difference between S and AS transcripts expressed from the same gene, median expression ratio of AS transcripts was divided by median expression ratio of S transcripts and was $\log_2$ transformed. Fold change values ≥ 1 was considered as genes with high antisense and low sense transcripts (HAS: LS) and values ≤ -1 was considered as genes with high sense and low antisense transcripts (HS-LAS). We observed that 70% of genes (n=515) and 40% of genes (n=104) expressed HS-LAS transcripts whereas only 1% of genes (n=10) and 2% of genes (n=6) showed the reverse phenomena and expressed HAS: LS transcripts (**Supplementary Table S4.3**) in complicated and uncomplicated cases respectively. Rest of the genes showed nearly equal level of S and AS transcripts abundance (NC-AS/S).

4.2.2. Confirmation of antisense transcripts by strand specific reverse transcriptase (RT) PCR

Eight genes were selected for validating the presence of sense/antisense transcripts detection as observed from the results of the hybridization (**Figure 4.5**). Antisense and sense transcripts were detected using strand specific primers to prime the first strand cDNA synthesis reaction and subsequent PCR amplification. As seen in the **figure 4.5** antisense and sense transcripts have been detected for six selected genes both from the pool of PFU (n=5) and PFC (n=9). The gene with ID PF14_0194 was detected with only antisense transcript in our microarray data and same was confirmed by the strand specific RT PCR which shows presence of antisense transcript and absence of sense transcript (**Figure 4.5**). We have also performed primer less cDNA synthesis, to detect the level of false priming, in spite of the fact that our array hybridization utilized cRNA. The data shows there is insignificant detection of antisense/sense transcript due to self-priming thus validating the true presence of antisense/sense transcript in the data presented (**Figure 4.6**).

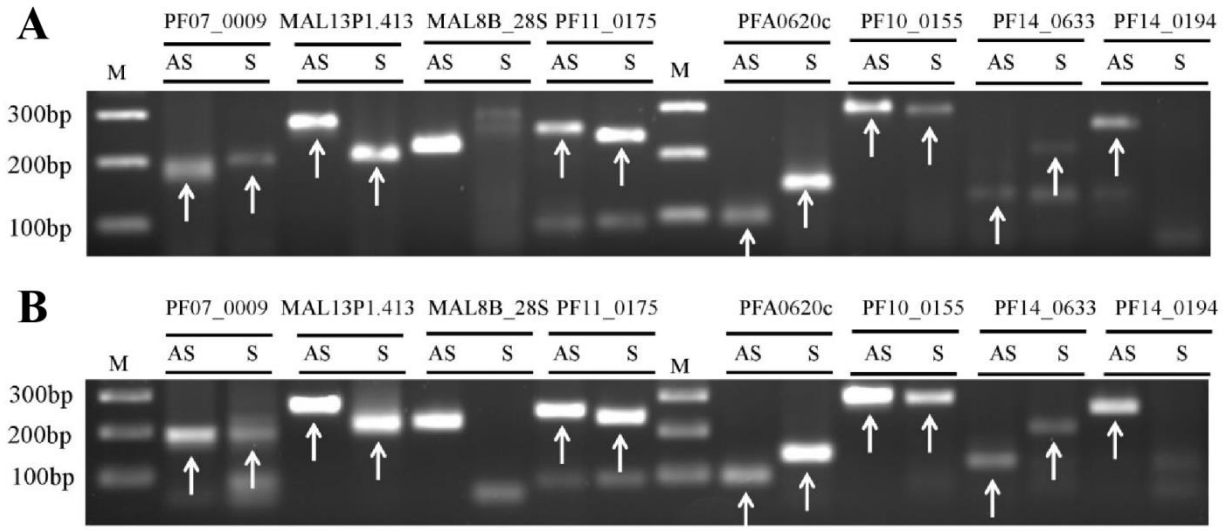


Figure 4.5. Validation of sense and antisense transcripts using strand specific RT-PCR. Antisense and sense transcripts were detected using strand specific primer to prime the first strand cDNA synthesis reaction. AS, antisense transcript; S, sense transcript; M, marker. Codes starting with letters PF/MAL represents the *P. falciparum* genes. gDNA and no template control as positive and negative PCR controls were included during gel running but not shown in the figure. (**A**) Represents the antisense and sense transcripts detection from uncomplicated pooled samples. (**B**) Represents the antisense and sense transcripts detection from complicated pooled samples. Arrows represent the desired bands.

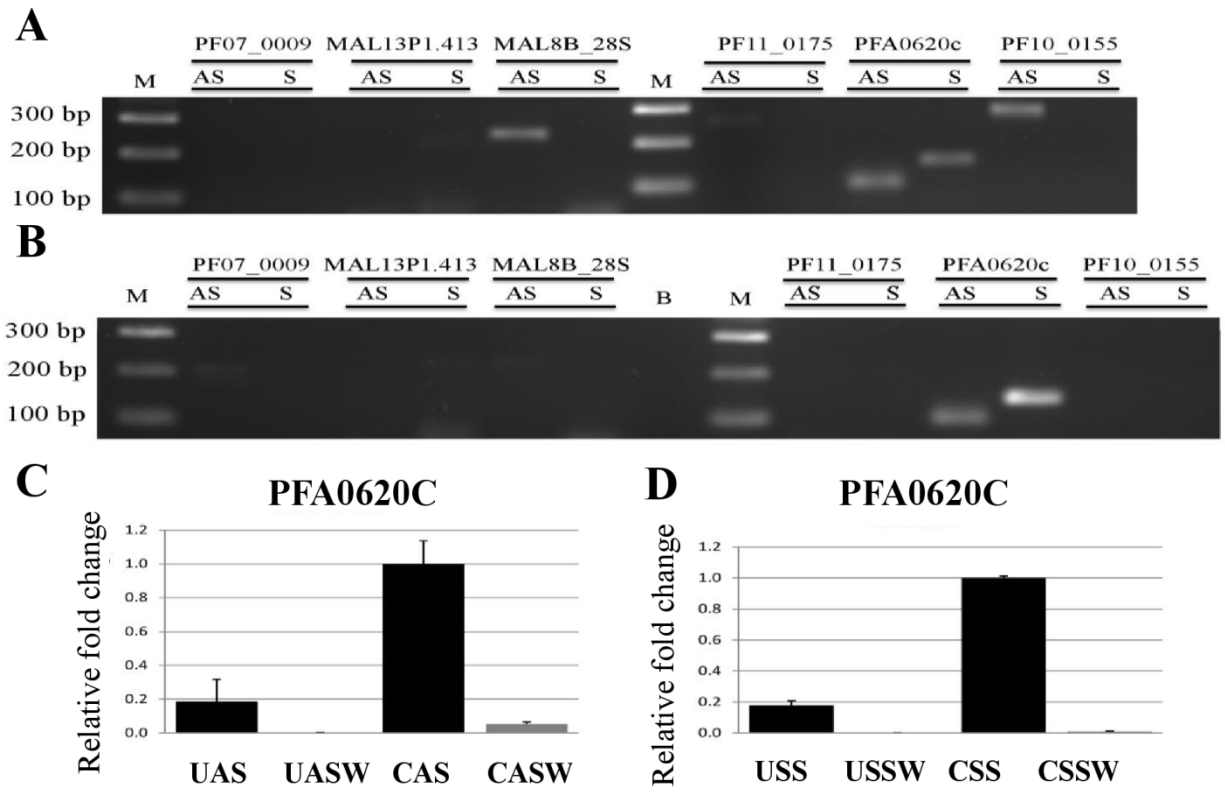


Figure 4.6. Sense and antisense transcript levels due to self-priming. Reverse transcriptase reactions were carried out where no primers were included during reactions to check the levels of antisense and sense transcripts detection due to self-priming if any. AS, antisense transcript; S, sense transcripts; M, marker; UAS: uncomplicated antisense transcript; UASW: uncomplicated antisense water (primer replaced with water in first strand synthesis reaction); CAS: complicated antisense; CASW: complicated antisense water (primer replaced with water in first strand synthesis reaction); USS: uncomplicated sense transcript; USSW: uncomplicated sense water (primer replaced with water in first strand synthesis reaction); CSS: complicated sense transcript; CSSW; complicated sense water (primer replaced with water in first strand synthesis reaction). Codes starting with letters PF/MAL represents the *Plasmodium falciparum* genes. gDNA and no template control as positive and negative PCR controls were included during gel running but not shown in the figure. (**A**) Represents the antisense and sense transcripts detection from uncomplicated pool samples due to self-priming. (**B**) Represents the antisense and sense transcripts detection from complicated pool samples due to self-priming. (**C**) Antisense transcripts expression level in reverse transcription reactions where specific primers were included compared to the reverse transcription reaction where no primer were used for the gene PFA0620c as determined by strand specific quantitative real-time PCR. (**D**) Sense transcripts expression level in reverse transcription reactions where specific primers were included compared to the reverse transcription reaction where no primer were used for the gene PFA0620c as determined by strand specific quantitative real-time PCR.

4.2.3. Comparison of antisense transcript expression pattern between complicated and uncomplicated cases

Comparison was done between complicated and uncomplicated cases to find out the common genes which expressed AS transcripts irrespective of its S transcripts expression status. 267 genes with AS transcripts were common between complicated and uncomplicated cases. Of the 267 genes, 258 genes expressed both S and AS transcripts and 7 genes expressed only AS transcripts in both the cases compared and 2 genes (PFE1430c and PFE0120c) expressed only AS transcripts in uncomplicated cases (**Supplementary Table S4.4**).

The 258 genes with S and AS transcripts in both clinical conditions were analyzed. 102 genes expressed HS-LAS transcripts, 2 genes expressed HAS-LS transcripts and 58 genes expressed NC-AS/S transcripts in both the cases (**Table 4.4**, **Supplementary Table S4.4**). Interestingly, 91 genes which expressed NC-AS/S in uncomplicated cases expressed HS-LAS transcripts in complicated cases. One gene (MAL8b_28s) with HS-LAS transcripts in uncomplicated cases expressed HAS-LS transcripts in complicated cases (**Supplementary Table S4.4**).

Table 4.4. Change in expression pattern of sense and antisense transcript pairs in two clinical conditions

Uncomplicated	Complicated	Genes
NC-AS/S	NC-AS/S	58
HS-LAS	HS-LAS	102
HAS-LS	HAS-LS	2
NC-AS/S	HS-LAS	91
NC-AS/S	HAS-LS	1
HAS-LS	NC-AS/S	3
HS-LAS	HAS-LS	1

Table shows the antisense and sense transcripts ratio status in both uncomplicated and complicated isolates. Abbreviations: NC-AS/S, nearly equal level of sense and antisense transcripts level; HS-LAS, high sense and low antisense transcripts and HAS-LS, high antisense and low sense transcripts.

It is important to note that many genes with HS-LAS transcripts in both PFC and PFU and genes with NC-AS/S transcripts in PFU that changed to HS-LAS transcripts in PFC belongs to exportome class and most genes of this class were transcribed either in early ring stages and/or in schizont stages (Maier, et al., 2008). Some of the genes which exhibited this phenomena are skeleton binding protein 1 (SBP1), early transcribed membrane protein 11.2, 2, 10.1 (ETRAMP), Membrane associated histidine rich protein (MAHRP), hyp class protein etc. (**Supplementary Table S4.4**). In a separate comparison, 175 genes which expressed only S transcripts in uncomplicated cases, expressed both S-AS transcripts in complicated cases (**Supplementary Table S4.5**).

Moderate to high positive correlation between sense and antisense transcript pairs was observed across the transcriptome for genes with both S and AS transcripts (**Figure 4.7**). In particular, when S transcript ratio and AS transcript ratio from 736 genes in complicated cases and from 267 genes in uncomplicated cases was plotted independently for each clinical condition, we observed that for majority of genes, increase in sense transcript ratio was accompanied with increase in their antisense transcript ratio (Positive correlation) for the same gene in both the conditions (**Figure 4.7**). For clearer picture, dataset from complicated and uncomplicated cases were segregated into three parts: genes with HS-LAS transcripts, genes with HAS-LS transcripts and genes with NC-AS/S transcripts and the same analysis was performed. Interestingly, Genes with HAS-LS transcripts showed the highest positive correlation (r=0.94 and 0.96 for PFC and PFU cases respectively) followed by genes with NC-AS/S transcripts (r=0.88 and 0 .76) and genes with HS-LAS transcripts (r=0.74 and 0.65) (**Figure 4.7**). These results clearly suggest that the majority of genes with S and AS transcripts are positively regulated.

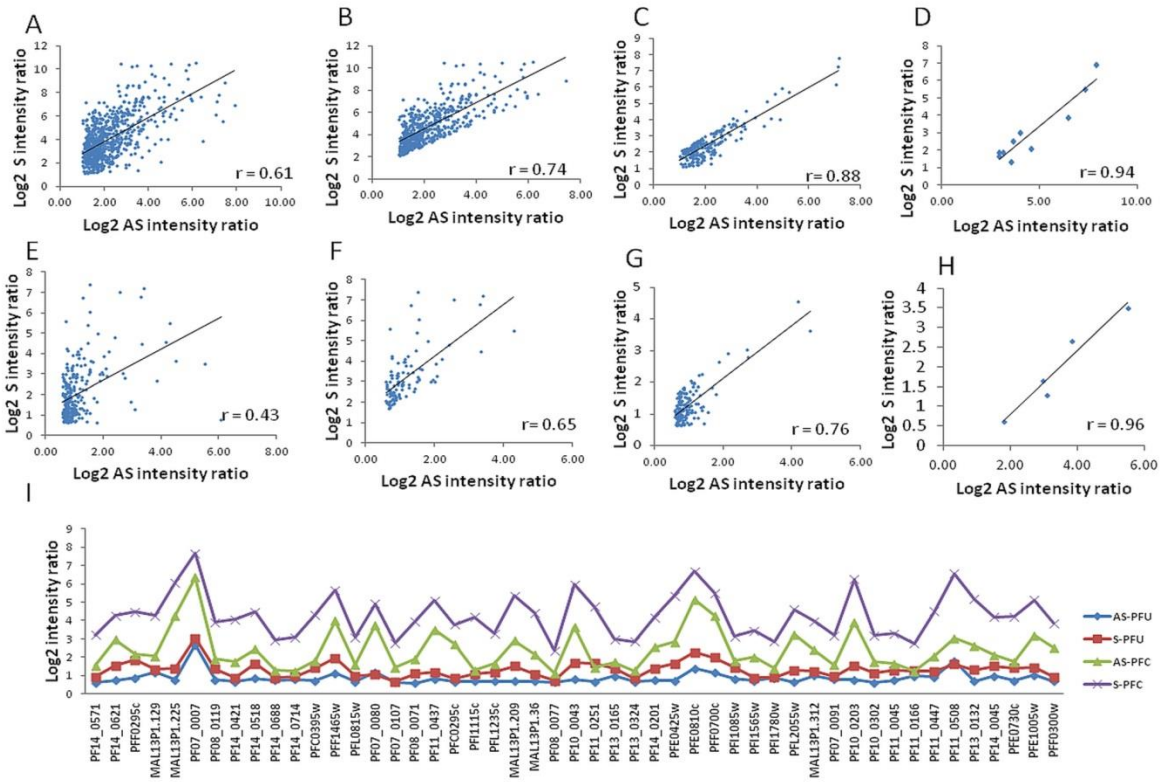


Figure 4.7. Correlation between sense and antisense transcript expression. Sense and antisense transcript expression ratio was plotted to find out the type of correlation (Positive/negative) that exist between them. r, Pearson correlation, x axis, Log_2 transformed AS transcript ratio, y axis, Log_2 transformed S transcript ratio. (**A** to **D**) S and AS transcript pairs in complicated cases. (E to H) S and AS transcripts in uncomplicated cases. (**A** and **E**) All the genes with S and Antisense transcripts. (**B** and **F**) Genes with HS-LAS transcripts. (**C** and **G**) Genes with NC-AS/S transcripts. (**D** and **H**) Genes with HAS-LS transcripts. I) Comparison of S and AS transcripts expression pattern in both complicated and uncomplicated cases. Abbreviations: AS, antisense, S, sense, PFU, *P. falciparum* uncomplicated samples, PFC, *P. falciparum* complicated samples.

4.2.4. Strand specific real-time PCR confirms the change in expression ratios

Quantitative real-time PCR reactions were performed by utilizing the primers used for the genes selected for the strand specific RT PCR (**Figure 4.8**). The RNA pool from samples for PFU and PFC remained the same as used in strand specific reverse transcriptase PCR experimentation. This experiment was performed to validate the overall ratio between AS/S transcripts for six genes, ascertained from the array hybridization data. One gene (PF07_0009) showed similar ratio change as we had defined in the array. In case of gene PF14_0633, The overall ratios in both uncomplicated and complicated disease states remain similar (**Figure 4.8** and **Table 4.5**). Two genes (MAL13P1.413 and PF11_0175), which showed no change in AS/S ratios between PFU and PFC in microarray data, showed change in AS/S ratios in strand specific real time quantitative PCR data. The remaining two genes (PFA0620c and PF10_0155), which showed change in AS/S ratios between PFU and PFC in microarray data, showed no change in AS/S ratios in strand specific real time quantitative PCR data. The variation obtained on comparison of the microarray hybridization data with the strand specific quantitative real time PCR data is due to inherent differences in the techniques and experimentation. Our data however validates that there is change in AS/S ratios in differing disease conditions in the case of some genes.

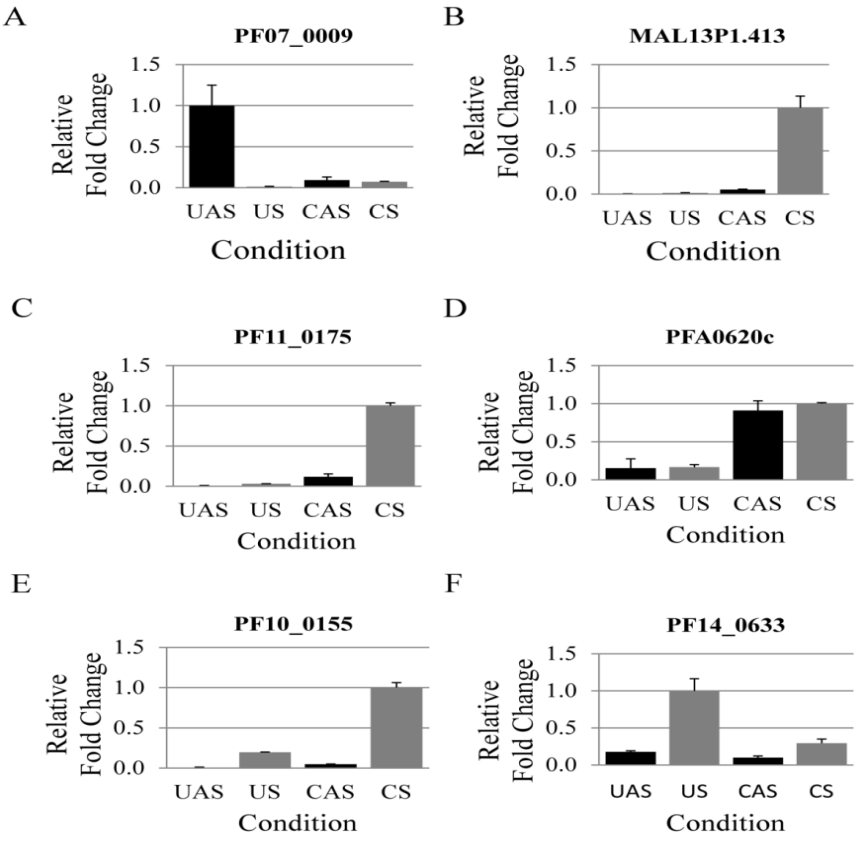


Figure 4.8. Antisense and sense transcript levels of selected genes in uncomplicated and complicated cases measured by strand specific quantitative real-time PCR (mean ± corrected S.D.; n=2). Relative fold change expression determined by ΔCt method with no control assigned. UAS, antisense transcript in uncomplicated cases; US, sense transcript in uncomplicated cases; CAS, antisense transcript in complicated cases; CS, sense transcript in complicated cases. Antisense specific primers and sense specific primers were used to detect antisense and sense transcripts level respectively. The real time data from gene MAL8b_28s and PF14_0633 has not been shown.

Table 4.5. Expression status of sense and antisense transcripts for six genes in the microarray and real-time qPCR data.

Gene	Microarray data		Real time q PCR data	
	Uncomplicated	Complicated	Uncomplicated	Complicated
PF07_0009	HAS-LS	NC-AS/S	HAS-LS	NC-AS/S
MAL13P1.413	HS-LAS	HS-LAS	NC-AS/S	HS-LAS
PF11_0175	NC-AS/S	NC-AS/S	NC-AS/S	HS-LAS
PFA0620c	NC-AS/S	HS-LAS	NC-AS/S	NC-AS/S
PF10_0155	NC-AS/S	HS-LAS	HS-LAS	HS-LAS
PF14_0633	NC-AS/S	HAS-LS	HS-LAS	NC-AS/S

Abbreviations: NC-AS/S, nearly equal level of sense and antisense transcripts level; HS-LAS, high sense and low antisense transcripts and HAS-LS, high antisense and low sense transcripts.

4.2.5. Possible mechanisms behind antisense transcription in *Plasmodium falciparum*

AS transcripts could derive from overlapping transcriptional units present in opposite strands, where the transcript pairs might overlap by their 3′ ends (convergent) or by their 5′ ends (divergent). Of the 797 AS transcripts reported here, 10 AS transcripts could be mapped to 7 pairs of overlapping protein coding transcriptional units (**Table 4.6**). Out of 7 pairs of protein coding overlapping transcriptional units, one gene pair (PF14_0688/PF14_0689) is annotated as overlapping, 5 pairs had evidence of overlap from EST where single EST overlapped the genomic region of two nearby gene models and one gene pair (PFL2055w/PFL2060c) had evidence of overlap from RNA-seq data reported previously (Sorber, et al., 2011).

In the absence of overlapping gene pairs which could be the reason for antisense transcript detection, three additional mechanisms have been proposed: Transcriptional run-through based, if the genes are oriented in tail-tail orientation; bi-directional promoter based, if the genes are oriented in head-head orientation and independent promoter based. To identify possible mechanisms of antisense transcription, the data obtained from the strand specific quantitative real time PCR (**Figure 4.9**) were used to identify genes on a particular strand. Subsequently, neighboring genes on either strand were mapped. The hybridization signals obtained for these neighboring genes for sense/antisense or both sense and antisense transcripts were then determined. The data suggests that independent promoters and bidirectional promoters as well as transcriptional run through are involved in the mechanism of antisense generation.

Table 4.6. Overlapping gene pairs with antisense transcript

Gene ID- Gene 1 Gene 2	Product description	Orientation	EST evidence, if any	Genomic Location Gene 1 Gene 2	Comment	Gene (s) with AS transcripts
PF14_0688 PF14_0689	pre-mRNA-splicing factor ISY1 homolog, putative Yip1 protein, putative	Tail to tail	-	2,940,415-2,941,540 2,941,537-2,943,196	Annotated	PF14_0688
PFB0300c PFB0295w	merozoite surface protein 2 (MSP2) adenylosuccinate lyase (ASL)	Tail to tail	PFACA05TF	Pf3D7_02_v3: 273,689 - 274,507 Pf3D7_02_v3: 271,594 - 273,009	Evidence from EST	PFB0300c PFB0295w
PF10_0096 PF10_0097	conserved Plasmodium protein, unknown function mitochondrial ribosomal protein L22/L43, putative	Tail to tail	BI814915	Pf3D7_10_v3: 399,207 - 401,778 (+) Pf3D7_10_v3: 401,876 - 402,711 (-)	Evidence from EST	PF10_0097
PFL0815w PFL0820c	DNA-binding chaperone, putative conserved Plasmodium protein, unknown function	Tail to tail	PFMC384TR	Pf3D7_12_v3: 670,447 - 673,266 (+) Pf3D7_12_v3: 674,349 - 674,777 (-)	Evidence from EST	PFL0815w
PFL1235c PFL1230w	conserved Plasmodium protein, unknown function small subunit rRNA processing factor, putative	Tail to tail	PFOCC08TR	Pf3D7_12_v3: 1,035,261 - 1,038,155 (-) Pf3D7_12_v3: 1,033,551 - 1,035,112 (+)	Evidence from EST	PFL1235c
PF13_0128 MAL13P1.129	beta-hydroxyacyl-ACP dehydratase (FabZ) conserved Plasmodium protein, unknown function	Tail to tail	BU496018	Pf3D7_13_v3: 963,661 - 964,690 (-) Pf3D7_13_v3: 961,213 - 963,561 (+)	Evidence from EST	PF13_0128 MAL13P1.129
PFL2055w PFL2060c	40S ribosomal protein S17 rabGDI protein	Tail to tail	-	-	Reported by Sorber *et al.* (29)	PFL2055w PFL2060c

Table shows list of overlapping gene pairs with one or both the members expressing antisense transcripts. Overlapping gene pair list was prepared by gathering information from gene annotation, evidence from EST and/or scientific reports.

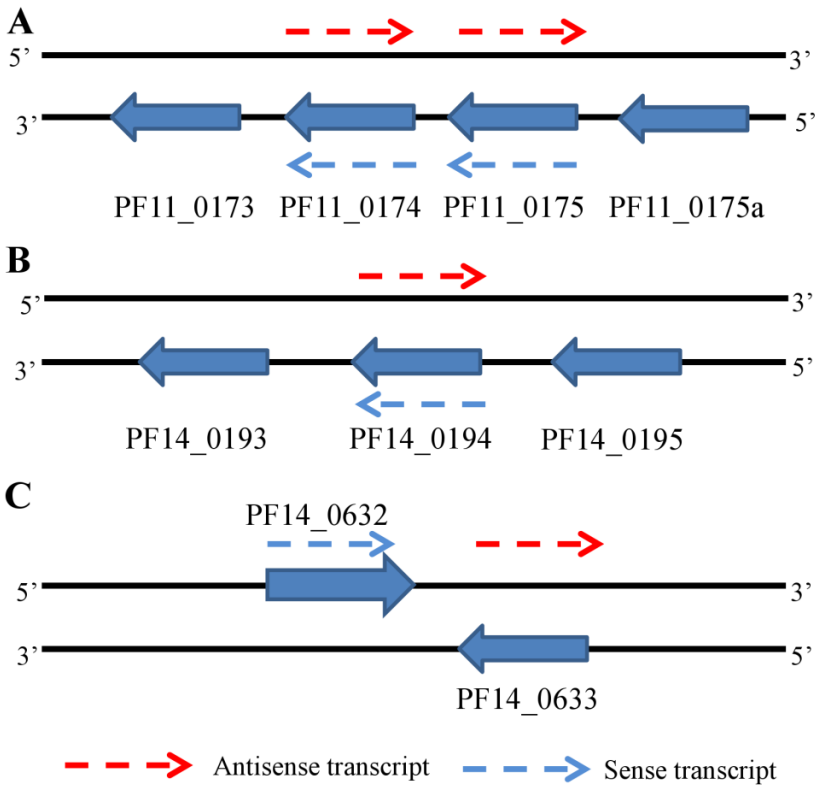


Figure 4.9. Possible mechanisms behind antisense transcript expression. Blue color dashed arrow depicts the sense transcript and red color dashed arrow depicts the antisense transcript detected in microarray data. (A)The four genes PF11_0173, PF11_0174, PF11_0175 and PF11_0175a are encoded on the negative strand of the chromosome 11. Sense and antisense transcripts were detected for the gene PF11_0174 and PF11_0175 and no transcripts were detected for the gene PF11_0173 and PF11_017a. Presence of sense and antisense transcripts for the gene PF11_0175 was also validated using strand specific RT and Real-Time PCR. The detection of antisense transcripts for the gene PF11_0175 might be due to bidirectional promoter, simultaneously driving expression of sense transcript of downstream gene PF11_0174 and the antisense transcript expression of upstream gene PF11_0175. An independent promoter might be driving the expression of antisense transcript of the gene PF11_0174 as no transcripts were detected for the downstream gene PF11_0173. (B)The three genes PF14_0193, PF14_0194 and PF14_0195 are encoded on the negative strand of chromosome 14. Sense and antisense transcript were detected for the gene PF14_0194 and no transcripts could be detected for the gene PF14_0193 and PF14_0195. Presence of sense and antisense transcripts for the gene PF14_0194 was also validated using strand specific RT and real-time PCR. This suggests that an independent promoter might be driving the expression of antisense transcript of the gene PF14_0194. (C) The two genes PF14_0632 and PF14_0633 are encoded on opposite strands of chromosome 14 and are in tail to tail orientation. Sense and antisense transcripts were detected for PF14_0633 whereas only sense transcript was detected for PF14_0632. This suggests that the antisense transcript detection for PF14_0633 might be due to transcriptional read-through of PF14_0632 sense transcript.

4.2.6. Functional analysis of genes with antisense transcripts

Gene ontology enrichment analysis on biological process was performed for genes with AS transcripts in complicated and uncomplicated cases separately. In complicated cases, genes were found to be enriched to biological processes like primary macromolecule and cellular metabolic processes, gene expression, RNA splicing and regulation of translation, carbohydrate biosynthetic and metabolic processes which mainly involve gluconeogenesis, pentose phosphate pathway and nitrogen compound metabolism (**Figure 4.10**, **Supplementary Table S4.6**). Interestingly, the number of genes with AS transcripts in uncomplicated cases was less and not enriched to any biological processes.

Genes with AS transcripts in both cases were mapped to a diverse variety *P. falciparum* biochemical/metabolic pathways present in KEGG database (Kanehisa and Goto, 2000) (**Supplementary Table S4.7**). Percentage of genes involved in each pathway with AS transcripts was calculated. Pathways in which at least 40% of genes were with AS transcripts are shown in **figure 4.11**. Most importantly, genes with AS transcripts were mapped to pathways that are crucial components of the central carbon metabolism in the *P. falciparum* parasite (**Figure 4.12**, Olszewski and Llinas, 2011). Specifically, the non-oxidative arm of pentose phosphate pathway which converts glycolytic intermediates to ribose-5-phosphate through a series of reversible reactions, part of the oxidative branch which converts ribulose-5-phosphate to either ribose-5-phosphate or xylulose-5-phosphate, the mannose metabolism pathway that contributes to the production of glycophosphatidylinositol (GPI) anchors, the glycolytic link to mannose-6-phosphate production (Note- MAL8P1.156-mannose-6-phosphate isomerase detected by two probes), pathways for the sn-glycerol-3-phosphate production from glycolytic intermediates and/or from extracellular glycerol, aspargine and aspartate metabolism pathways and pathways for 2-oxoglutarate production via ammonia and glutamine were mapped. Genes encoding enzymes for tricarboxylic acid cycle (TCA) were devoid of AS transcripts except for one gene (PF14_0295) which codes for ATP-specific succynyl-CoA synthetase beta subunit for which only AS transcript was found in uncomplicated cases. The central carbon metabolism may allow this parasite to utilize wide variety of nutrients (glucose, glycerol, glutamine, lactic acid, lipids,) depending upon their availability in the patient's blood for energy production (Le Roux, et al., 2009).

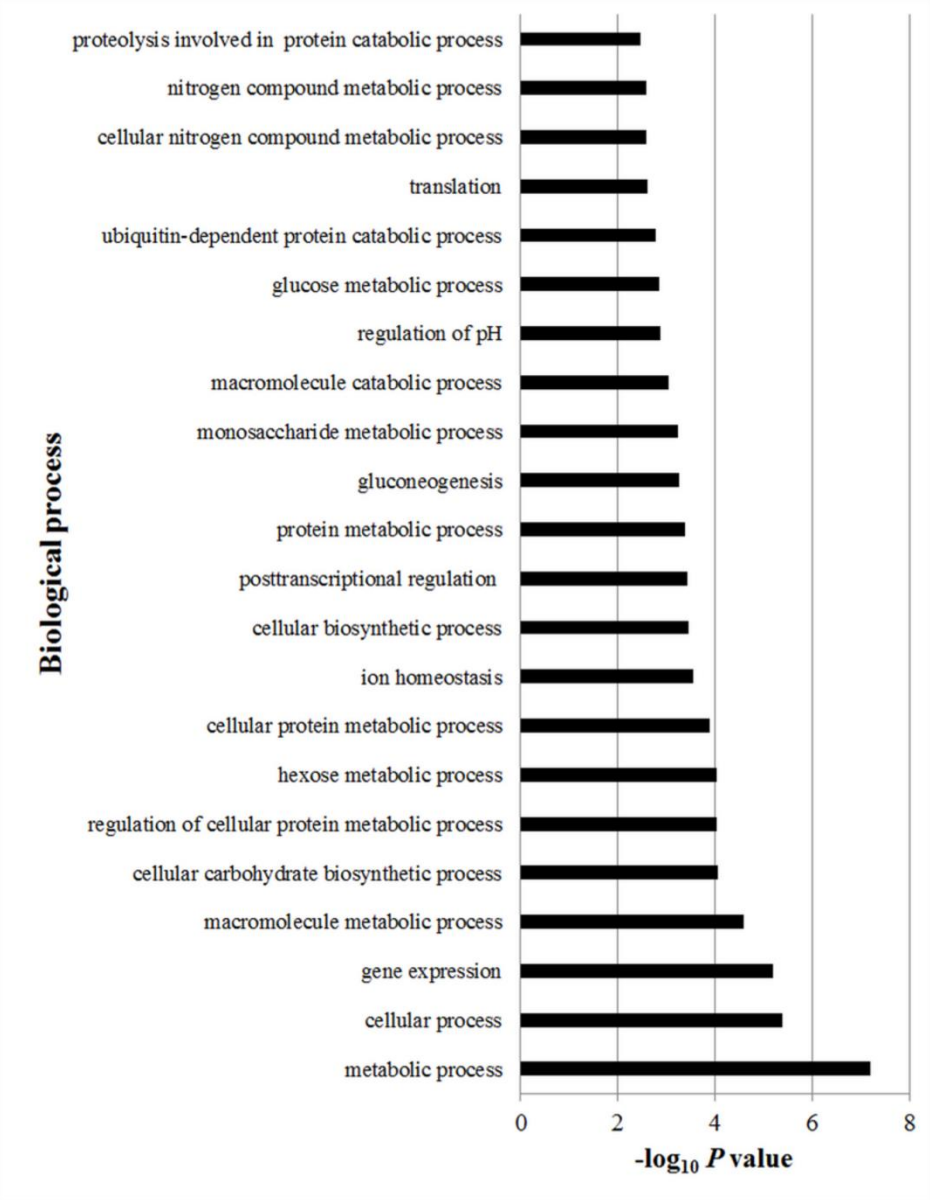


Figure 4.10. Enrichment of genes with antisense transcripts to biological process gene ontology (GO) terms. Over-represented biological processes were determined by taking list of genes with antisense transcripts in complicated cases using the Bingo plug-in (Maere, et al., 2005) in Cytoscape2.8(Smoot, et al., 2011). Overrepresented biological processes term were included in the list with P values ≤ .05 determined by hypergeometric test after multiple testing corrections (Benjamini Hochberg's false discovery rate correction). Y axis denotes the –Log10 P value and X axis denotes the enriched biological process terms.

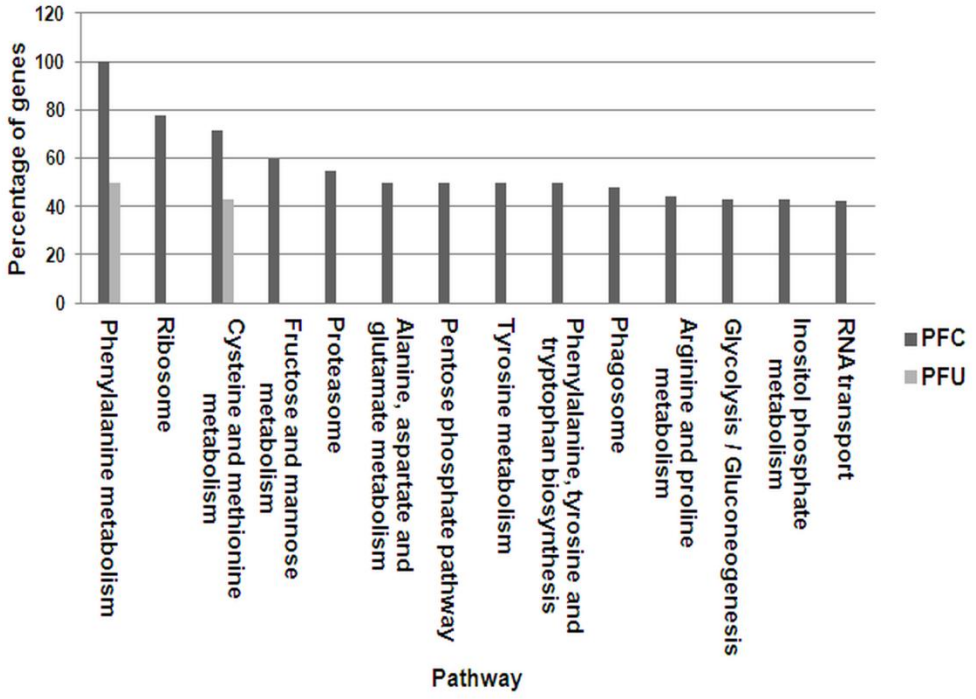


Figure 4.11. **Mapping of genes with antisense transcripts to various biochemical pathways.** Genes with antisense transcripts in complicated cases were mapped to *P. falciparum* metabolic pathways present in KEGG (Kanehisa and Goto, 2000). Information about the number of genes involved in each pathway was obtained. Percentage of genes with antisense transcripts in each pathway was calculated and was plotted against the respective pathways. Figure shows the pathways in which at least 40% of their genes expressed antisense transcripts. Abbreviations: PFC, *P. falciparum* complicated isolates, PFU, *P. falciparum* uncomplicated isolates.

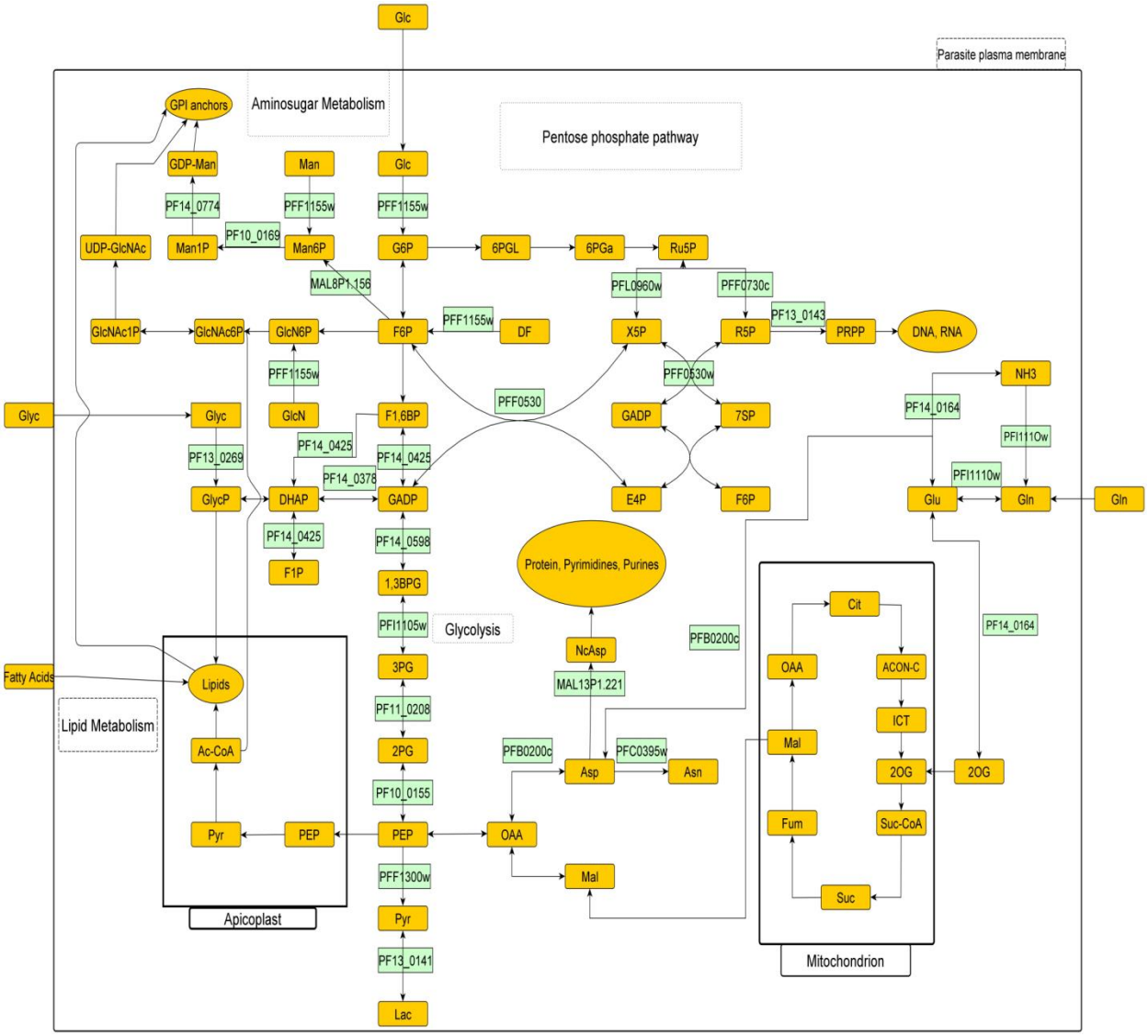


Figure 4.12. Mapping of genes with antisense transcript to an integrated map of central carbon metabolism of *P. falciparum* (Adapted and modified from Olszewski *et al,* 2011). Modifications were done by referring to the KEGG metabolic pathway database (Kanehisa and Goto, 2000) and Malaria parasite metabolic pathways site maintained by Hagai Ginsburg (Malaria Parasite Metabolic Pathways. http://priweb.cc.huji.ac.il.malaria/. This figure depicts the location of genes with antisense transcripts coding for the enzymes involved in respective pathway/s. Gene IDs starting with letters PF/MAL represents genes with antisense transcripts detected in this study. Gene with ID MAL8P1.156 was detected by 2 representative probes. Yellow box represents substrate and product of each enzymatic reaction.

Abbreviations : Glc, glucose; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; DF, D-fructose; F1P, fructose-1-phosphate; F1,6BP, fructose-1,6-bisphosphate; DHAP, dihydroxy-acetone-phospahte; GADP, glyceraldehydes-3-phosphate; 1,3 BPG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglucerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; Lac, lactate; Ac-CoA, acetyl-CoA; GlycP, glycerol-3-phosphate; Glyc, glycerol; Man, mannose; Man6P, mannose-6-phosphate; Man1P,

mannose-1-phosphate; GDP-Man, GDP-mannose; GlcN, glucosamine; GlcN6P, glucosamine-6-phosphate; GlcNAc6P, N-acetyl-glucosamime-6-phosphate; GlcNAc1P, N-acetyl-glucosamine-1-phosphate; UDP-GlcNAc, UDP-N-acetyl-glucosamine; 6PGL, 6-phosphoglucono-δ-lactone; 6PGa, 6-phosphogluconate; Ru5P, ribulose-5-phosphate; R5P, ribose-5-phosphate; X5P, xylulose-5-phosphate; S7P, sedoheptulose-7-phosphate; E4P, erythrose-4-phosphate; PRPP, phosphoribosylpyrophosphate; Asp, aspartate; Asn, aspargine; NcASP, N-carbamoyl-L-aspartate; Gln, glutamine; Glu, glutamate; 2OG, 2-oxoglutarate; ICT, isocitrate; Cit, citrate; ACON-C, cis-aconitate; OAA, oxaloacetate; Mal, malate; Suc-CoA, succinyl CoA; Suc; succinate, Fum, fumarate; GPI, glycophosphatidylinositol . Gene coding for protein: PFF1155w, hexokinase; PF14_0425, fructose-bisphosphate aldolase; PF14_0598, glyceraldehydes-3-phosphate dehydrogenase; PFI1105w, phosphoglycerate kinase; PF11_0208, phosphoglycerate mutase; PF10_0155, enolase; PFF1300w, pyruvate kinase; PF13_0141, L-lactate dehydrogenase; PFF0530, transketolase; PFL0960w, D-ribulose-5-phosphate-3-epimerase; PFF0730c, enoyl-acyl carrier reductase; PF13_0143, phosphoribosyl pyrophosphate synthetase; PF14_0164, NADP-specific glutamate dehydrogenase; PFI1110w, glutamine synthetase; PF10_0169, phosphomannomutase; PF14_0774, mannose-1-phosphate guanyltransferase; PF14_0378, triosephosphate isomerase; PF13_0269, glycerol kinase; PFC0395w, aspargine synthetase; PFB0200c, aspartate aminotransferase; MAL13P1.221, aspartate carbamoyl transferase; MAL8P1.156, mannose-6-phosphate isomerase.

4.2.6. Comparison with the previously reported *in vitro* data

We compared our microarray data with previously published data from SAGE, RNA-seq and cDNA library sequence analysis, obtained from different blood stages of *in vitro* cultured parasites for genes with AS transcripts (Gunasekera, et al., 2004, Lopez-Barragan, et al., 2011 and Raabe, et al., 2010). These publications provide the main body of information available about NATs from *P. falciparum* in literature. 146 (19%), 84 (11%), and 32 (4%) genes were found to be common between our data when compared with SAGE library, RNA-seq and cDNA library sequencing data respectively (**Figure 4.13**). Our data overlaps maximum with the SAGE library data and minimum with cDNA library data. We also compared our data with the RNA-seq data from Sorber *et al* (Sorber, et al., 2011), which reports a small number of AS transcripts (149) from this parasite.

The comparisons revealed 545 genes with AS transcripts unique to our dataset (**Supplementary Table S4.8**). 142 of these 545 genes encode for conserved hypothetical proteins with unknown function, 47 of these represent genes encoding proteins of the exportome. The rest of the genes encode for variant surface antigen proteins (PfEMP1, Rifin and Stevors) heat shock proteins (like HSP 86, HSP 101, HSP40 type II), proteins with hydrolase activity (like Vacuolar ATP synthase, NUDIX hydrolase), various ATP and GTP binding proteins, DNA binding proteins (like polymerases), Histone and Zinc ion binding protein and proteins involved in diverse biochemical

pathways etc. Apart from these, three genes (PF14_0633, PF11_0163 and PF07_0126) encoding Apicomplexan AP2 family of transcriptional regulators with both S-AS transcripts were also identified.

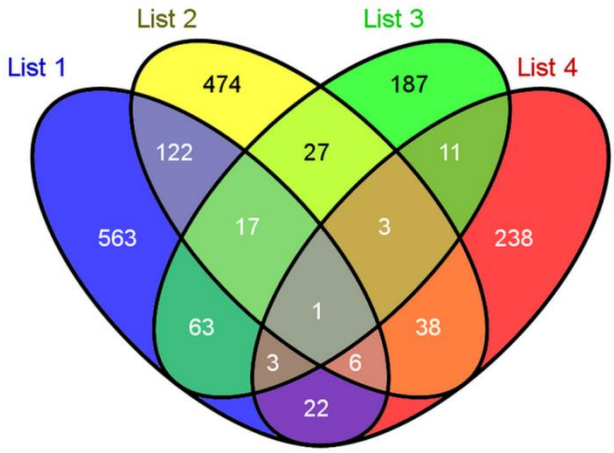


Figure 4.13. Comparison of detected NATs in this study with the previously reported data. The 4 group Venn diagram depicts the shared and unique number of genes with antisense transcripts reported in this study and from previous studies (Gunasekera, et al., 2004, Lopez-Barragan, et al., 2011 and Raabe, et al., 2010). List 1, genes with AS transcripts from our data, List 2, genes with AS transcripts from SAGE data (Gunasekera, et al., 2004), List 3, genes with AS transcripts from RNA-seq data (Lopez-Barragan, et al., 2011) and List 4, genes with AS transcripts from cDNA library sequencing data (Raabe, et al., 2010)

Gene ontology enrichment analysis for biological processes was carried out separately for genes with AS transcripts reported earlier, and similar genes unique to this study. There was a clear difference between the biological processes to which these two groups of genes were enriched. Genes reported in the previous studies (Gunasekera, et al., 2004, Lopez-Barragan, et al., 2011 and Raabe, et al., 2010) were mostly seen to be involved in biological processes like antigenic variation, cytoadherence to microvasculature, cell-cell adhesion, stress response and regulation of cellular process. Genes with AS transcripts unique to this study enriched to biological processes as mentioned in the earlier section. In this study, two major pathways (Mannose metabolism and non-oxidative branch of pentose phosphate pathway) of central carbon metabolism, and the nitrogen metabolism pathway were mapped with genes expressing AS transcripts and identified additionally compared to the earlier reports (Gunasekera, et al., 2004, Lopez-Barragan, et al., 2011 and Raabe, et al., 2010).

A total of 1834 genes which expressed AS transcripts were identified based on all the studies including ours (**Supplementary Table S4.9**). This constitutes 29 percent of total nuclear encoded genes (6275 genes, PlasmoDB version 8.2) predicted in *P. falciparum.*

4.2.7. Genotyping of Clinical Isolates

Pfmsp1 and *Pfmsp2* based genotyping of all the isolates used in this study was carried out using nested PCR based method (Snounou, et al., 1999). On the basis of the MSP1 and MSP2 nested PCR data obtained, it appears that all the allelic variants are represented in these samples. In a few cases isolates show the presence of multiple parasite populations as per the data for these two genes (**Figure 4.14** and **4.15**). The genotyping information has been included to provide an idea of the diversity in the parasite populations being used for the study.

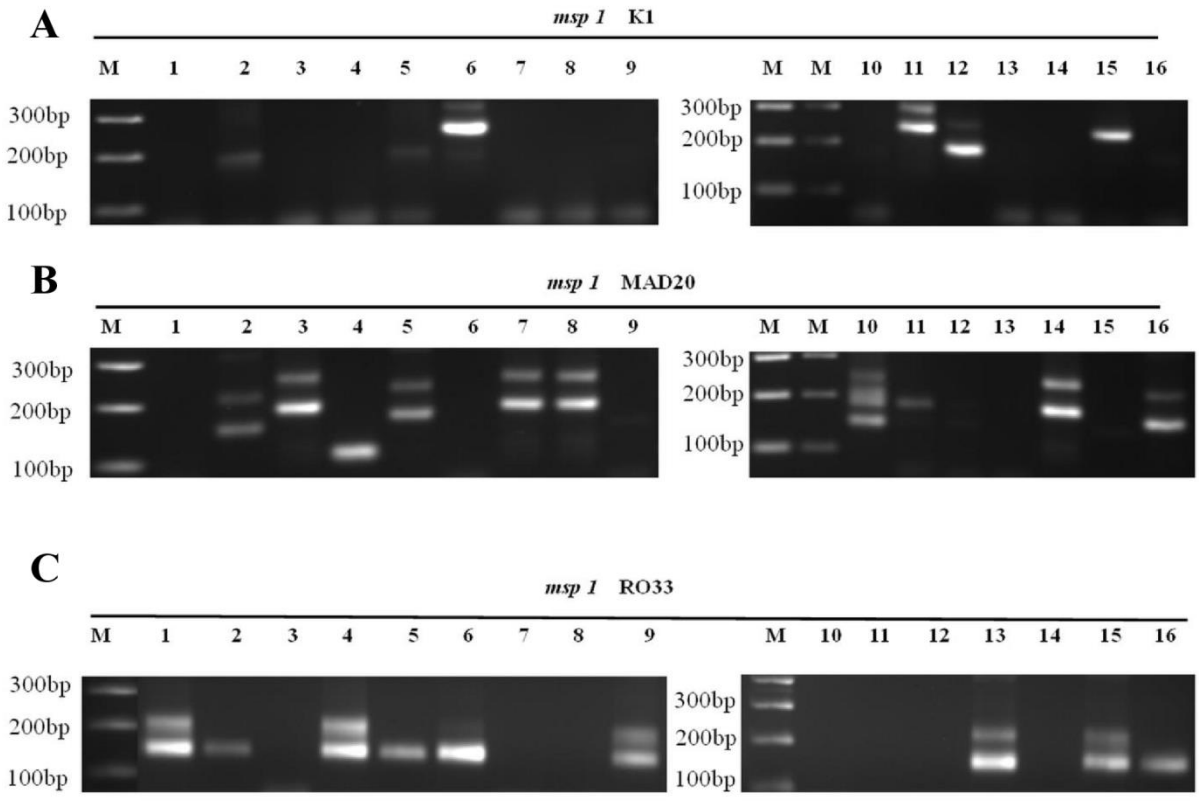


Figure 4.14. Genotyping of *P. falciparum* parasite isolates using *P. falciparum* merozoite surface protein 1 (*msp1*) gene. Nested PCR amplification: **A**, Primer pair used to detect the K1 allelic variant of *msp1* **B**, Primer pair used to detect the MAD20 allelic variant of *msp1* **C**, Primer pair used to detect the RO33 allelic variant of *msp1*. Lane 1 to lane 16 in all panels represents PFU-1, PFU-3 to PFU-6, PFC-1 to PFC-10 and PFU-2.

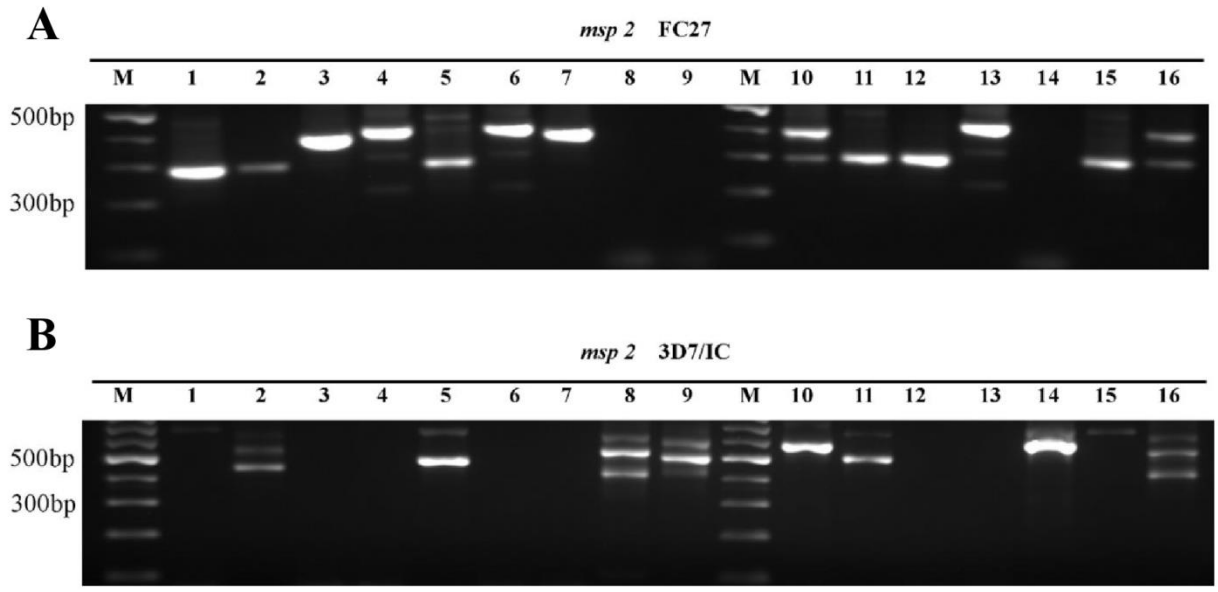


Figure 4.15. **Genotyping of *P. falciparum* parasite isolates using *P. falciparum* merozoite surface protein 2 (*msp2*) gene**. Nested PCR amplification: A, Primer pair used to detect the FC27 allelic variant of *msp2* B, Primer pair used to detect the 3D7/IC allelic variant of *msp1*. Lane 1 to lane 16 in all panels represents PFU-1, PFU-3 to PFU-6, PFC-1 to PFC-10 and PFU-2.

4.3. Discussion

Evidence of wide spread occurrence of NATs in eukaryotic transcriptomes added an extra layer of transcriptome complexity with its postulated involvement in transcriptional and post transcriptional gene regulation. With the discovery of natural antisense transcripts in *P. falciparum*, a protozoan parasite where the mechanisms of gene expression regulation has remained less characterized, the role of NATs in regulating gene expression has been speculated. *In vivo* clinical conditions provide a substantially different environment compared to *in vitro* conditions. Variations in these could determine the parasite biology and might affect disease manifestations in humans. The diversity of NATs reported from *P. falciparum* till date has focused on laboratory cultured isolates. In the present work, we have investigated the prevalence of antisense transcripts in *P. falciparum* clinical isolates collected directly from patient blood.

In this study, we have identified 797 genes (combined number from PFU and PFC) which expressed AS transcripts. 545 of these 797 genes were unique to this study. The detection of antisense transcripts has been validated for a subset of seven genes by strands specific RT PCR. Although, there are overlaps between our and previously reported datasets, it should be noted

that the percentage of overlaps between the previously reported datasets amongst themselves also vary. There could be multiple reasons for identification of an additional set of antisense transcripts in this study: the different techniques used in all these studies, which have their own limitations and advantages to detect antisense transcripts, differences in *in vivo* and *in vitro* conditions, different parasite strains (Indian clinical isolates and 3D7 strain) and the parasite stages used. Although, the overlap between the published dataset and our dataset is limited, this comparison allowed us to identify the additional set of NATs that prevails in this parasite. The combined data has enabled the identification of 29 percent of the predicted transcripts coding genes in *P. falciparum* with antisense transcripts. We have selected stringent criteria for transcripts to qualify as detected or else the total number of NATS might have been more. In comprehensively analyzed transcriptomes, the occurrence of NATs ranges from 16.7-85.2% of the transcription units (Donaldson and Saville, 2012). It is important to emphasize the prevalence of an additional set of NATs which exist in *P. falciparum* when isolated from diverse clinical conditions.

The comparison of expression data between complicated and uncomplicated cases showed that genes with only S or AS transcripts in one clinical condition produced both S and AS transcripts in the other clinical condition. The level of S and AS transcripts (NC-AS/S to HS-LAS, HAS-LS to NC-AS/S etc) expressed from the same gene was also found to be altered between the two clinical conditions suggesting condition specific differential expression of S and AS transcript pairs. It has been established in other systems that expression of antisense transcripts is condition specific (Ni, et al., 2010). In yeast, antisense transcripts and their sense targets were found to be differentially regulated in response to conditions ranging from nutrition limitations to stress (Yassour, et al., 2010). It is likely that the expression pattern of antisense transcript change during developmental progression and also in response to different *in vivo* clinical conditions in *P. falciparum*. Thus, it may be possible to detect more antisense transcripts if the parasite could be isolated from a wider range of clinical conditions or the liver and mosquito stages of the life cycle. Interestingly, RNA-seq experiment using strand specific library from various IDC stages of *P. falciparum* has revealed the differential expression of the AS transcripts in a stage specific manner (Lopez-Barragan, et al., 2011).

Parasites mainly utilize glucose as a carbon source which is well maintained in the homeostatic blood environment. The level of glucose might fluctuate (reduce) in a malaria infected host and in such a situation parasite may utilize other readily available carbon sources (Mannose, fructose and monosaccharides) as an alternate substrate (LeRoux, et al., 2009). Parasite might also metabolize other carbon compounds like amino acids, fatty acids, pyruvate, glycerol phosphate, lipds etc, either for its growth or for various other functions (like glycosylation) (Olszewski and Llinas, 2011). *In vivo* transcriptome study on this parasite has shown distinct transcriptional states which have been related to the differences in substrate levels like glucose, amino acids and lipids and differences in environmental stresses like heat shock, and oxidative/osmotic stress. "These differences might influence the parasite to express genes which are involved in alternative pathways of energy production and substrate utilization in conditions when the main substrates are in limit in the host" (Daily, et al., 2007). This might be achieved by utilizing the interconnected pathways of carbon metabolism and enzymes that have the potential to reverse the reactions steps (Olszewski and Llinas, 2011). Daily *et al* (2007) postulated that "epigenetic mechanisms might have a role in establishing these observed transcriptional shifts" (Daily, et al., 2007).

Gene ontology enrichment and pathway mapping analysis in this study has demonstrated that genes with antisense transcripts enriched to a wide variety of biological processes/biochemical pathways. Importantly, major components of the carbon metabolism were mapped with genes expressing AS transcripts. Genes involved in environmental stress response like heat shock response and ion homeostasis/ regulation of pH were also had AS transcripts. Concentrating at various proposed mechanisms by which AS transcripts might regulate its S transcripts counterpart. This suggests a possible regulatory role of transcriptional changes allowing the parasite to adapt to different environmental perturbations.

Of the various functionally annotated genes with antisense transcripts, the exportome is of particular importance in parasite pathogenesis. This class of genes code for protein products which interact with the infected RBC membrane and induce changes in the morphology, physiology and adhesive characteristics of the RBC by displaying variant cytoadhesive molecules (Hiller, et al., 2004, Maier, et al., 2008, Sargeant, et al., 2006 and van Ooij, et al., 2008). Correct timing of expression of these genes after the merozoite invasion appears to be

essential in infected RBC remodeling (establishing cytoadhesive characteristics) which helps the parasite to avoid non-specific clearance by spleen. It has also been established that parasites from field isolates express more of this class of transcripts compared to long term laboratory isolates which is essential for parasite in *in vivo* conditions (Mackinnon, et al., 2009). Presence of antisense transcripts for these genes suggests regulatory roles of antisense transcripts in expression of these genes in a stage specific manner.

Majority of genes with both S-AS transcript expressed HS-LAS transcripts and few genes expressed HAS-LS transcripts. Antisense RNA transcripts have been shown to express in lower abundance compared to the sense transcripts in human cells (He, et al., 2008). The comparison of the complicated and uncomplicated disease manifestations has shown 91 genes whose antisense and sense ratios have changed from insignificant change, to high sense and low antisense. Strand specific quantitative real time PCR data for a subset of 6 genes using pooled PFU (n=5) and PFC (n=9) total RNA has confirmed that there is a change in AS/S ratios in certain genes amongst differing disease complications. The isolates which were chosen for array hybridization were not identical to those which were used for quantitative real time q PCR analysis. This was due to the fact that we used total RNA isolated from parasites directly taken from patients. Inspite of the fact the RNA pools for PFU and PFC are not identical for the different experimentations, a good correlation is seen for detection of strand specific transcripts. However, the 6 genes analyzed by strand specific real time qPCR though showing antisense/sense ratio change between PFU and PFC, differ in the extent of change as detected by array hybridization. This suggests modulation of parasite gene expression which could be leading to adaptation against environmental stresses, or additionally could be involved in host parasite interactions leading to the severity of disease symptoms. Interestingly, only seven genes were shared between uncomplicated and complicated samples which showed only AS transcripts. Of these five are conserved *P. falciparum* proteins with unknown function but the remaining two include MSP-2 and spliceosome-associated proteins. Although, it would be imprudent to comment on systemic implications of these, there could be indications of modulation of protein maturation and invasion pathways.

It has previously been reported that presence of cis-NATs might negatively affect gene expression due to transcriptional collision (Lavorgna, et al., 2004). However, increasing evidence

has shown positive correlation between S and AS transcript pairs to be more frequent in extensively studied genomes (Conley and Jordan, 2012, Katayama, et al., 2005 and Luo, et al., 2012). *P. falciparum* developmental expression profiles of many sense and antisense pairs tested, exhibited positive co-expression profiles (Raabe, et al., 2010). Consistent with these reports, our data also suggests positive correlation between majorities of the detected S and AS transcript pairs as determined from global Pearson correlation coefficient of both the S and AS transcripts expression levels. Genes with both S and AS transcript showed higher individual S and AS expression ratio compared to genes with only S or AS transcripts which again supports the positive effect of AS transcripts on S transcripts expression. Here, we have shown the global trend of the data showing moderate to high positive correlation between S and AS transcript pairs. It is quite possible that there might still be S and AS transcript pairs with an inverse correlation or no obvious correlation (positive or negative correlation) between their expression. One example of negative correlation from our data is in case of MAL8b_28s gene which showed negative correlation between its S and AS transcript pairs.

In extensively studied transcriptomes, the majority of cis-NATs reported were from overlapping transcriptional units present in opposite strands. In *P. falciparum*, the number of overlapping transcriptional units present throughout the genome is very low whereas the number of genes with AS transcripts detected in this study is very high. Many genes with AS transcripts are neighbors on either side present in the same strand. This suggests a possibility of these AS transcripts being unannotated transcription unit antisense to the annotated gene models and not overlapping transcription units. This supports the argument of Sorber *et al* (2011) with RNA-Seq data from *P. falciparum* cultures. The cis-NATs might have their own promoter for transcription or might express through other mechanisms. Some of them may have originated because of bidirectional promoter activity of promoters present in between genes in tandem or due to transcriptional read through of non-overlapping tail to tail oriented genes in very close proximity. Comparison of the strand specific quantitative real-time PCR data with array hybridization data for neighboring gene pairs, suggests presence of different types of transcription machinery. Antisense specific promoters, bidirectional promoters and transcriptional run through mechanisms all appear to be present in this parasite.

The concordant expression of S-AS transcript pairs might be due to sharing of common or similar regulatory elements (promoter region) that can bind to or compete for same transcription regulators. This might also be influenced by the chromatin environment surrounding the regulatory elements (chromatin profiles and protein binding properties at promoter level), which favors the co-expression (Conley and Jordan, 2012 and Lapidot and Pilpel, 2006). We found that genes in many physical clusters with AS transcripts were functionally associated. Genes involved in same/similar function(s) may need to be expressed synchronously under similar inducing conditions. This co-expression could be achieved by sharing of common or similar regulatory elements as mentioned and explained above. Although, we do not have additional experimental evidence to confirm the exact mechanism by which the antisense transcripts detected here play a role in regulating gene expression. The presence of antisense transcripts expressed from these functionally associated physical gene clusters hints about its probable regulatory role.

NATs exert gene regulation by various mechanisms: transcriptional collision, RNA masking, genomic imprinting, alternative splicing and termination and alteration of DNA and chromatin (Faghihi and Wahlestedt, 2009 and Lapidot and Pilpel, 2006). The association between S-AS transcripts expression patterns may provide the information about the type of mechanism that might takes place (Lapidot and Pilpel, 2006). Gene regulation by RNA masking might require concordant regulation of S-AS pairs, where S-AS transcript pairs might interfere with splicing, translation and degradation. RNA editing by double stranded duplex formation might also require concordant expression of S-AS pairs (Chen, et al., 2005). Transcriptional interference /collision is another regulatory mechanism which might require S and AS transcript expression to be inversely correlated. Possible regulatory interactions (positive regulation or negative regulation) between sense and antisense transcript pairs could be assessed by monitoring the expression during the intra-erythrocytic development of the parasite or by biochemical and genetic studies, including gene knockdown from *in vitro* studies.

In summary, a total of 797 NATs targeted against annotated loci have been detected. Out of these, 545 NATs are unique to this study. The majority of NATs were positively correlated with the expression pattern of the sense transcript. However, 96 genes showed a change in sense/antisense ratio on comparison between uncomplicated and complicated disease conditions.

The antisense transcripts map to a broad range of biochemical/ metabolic pathways, especially pathways pertaining to the central carbon metabolism and stress related pathways. Our data strongly suggests that a large group of NATs detected here are unannotated transcription units antisense to annotated gene models. The results reveal a previously unknown set of NATs that prevails in this parasite, their differential regulation in disease conditions and mapping to functionally well annotated genes. The results detailed here call for studies to deduce the possible mechanism of action of NATs, which would further help in understanding the *in vivo* pathological adaptations of these parasites.

5.1. Introduction

The severity of malaria and population at risk differs in different geographical locations, and partly depends on endemicity (stable or seasonal transmission). In parts of the world (Tropical areas like Sub-Saharan Africa), where malaria transmission is stable, complicated malaria is predominantly seen in children (≤5 years). However, areas like temperate and sub-tropical regions of Asia and Latin America with seasonal/unstable malaria transmission, all age groups can show complicated malaria due to insufficient exposure to induce or maintain immunity (Trampuz, et al., 2003; World Health Organization, 2000). The frequency of the type of clinical presentation of severe malaria also differs in adults compared to children (World Health, 2000). Frequency of jaundice is higher and renal failure is almost exclusively seen in adults compare to young children (World Health Organization, 2000). Children in hyperendemic areas like sub-Saharan Africa contribute more than 90% of the total mortality due to malaria worldwide (WMR, 2013). Therefore, much effort has focused on understanding the pathophysiological and molecular basis of complicated malaria in children (mostly in the case of cerebral malaria). However, the phenomenon of complicated malaria in adults has been studied in much lesser detail.

One of the parasite factors that have been recognized to contribute to the disease severity is expression of clonally variant adhesion proteins on the surface of the infected RBCs (iRBCs). The infected erythrocyte sequestration/adhesion to receptors present in deep micro vasculature is mediated by the variant surface antigen *P. falciparum* erythrocyte membrane protein-1 (PfEMP1). PfEMP1 antigens are encoded by the *var* gene family with approximately 50-60 different var genes in the genome (Kyes, et al., 2001). *var* genes have been grouped into 5 sub-groups: three major groups (group A, B and C) and two intermediate groups (B/A and B/C) based on their chromosomal location and 5' upstream region (Lavstsen, et al., 2003). *var* gene groups have different functional properties relating to binding to a particular host receptor and determined by type of domain they encode for. Group B, B/C and C binds to host receptors like CD36 (Robinson, et al., 2003), whereas some members of Group A bind to ICAM-1. Over expression of selected members of a particular group of PfEMP1 has been associated with certain malaria associated disease severity. Particularly, *var* A group in case of severe pediatric malaria and *var* E group in case of pregnancy associated malaria (PAM) (Bull, et al., 2005,

Jensen, et al., 2004, Kaestli, et al., 2006 Kyriacou, et al., 2006, Rottmann, et al., 2006, Salanti, et al., 2004 and Salanti, et al., 2003). All of these studies have focused either on pediatric malaria cases or on pregnant women with malaria. Variant surface antigens (VSAs) other than PfEMP1, repetitive interspersed family proteins (RIFINs) and sub-telomeric variable open reading frame proteins (STEVORs) are also seen to contribute to the antigenic variation process allowing the parasite to establish prolonged chronic infection (Abdel-Latif, et al., 2003, Abdel-Latif, et al., 2002, Fernandez, et al., 1999, Kyes, et al., 1999, Niang, et al., 2009).

Extensive knowledge about the parasite biology has been gathered through experiments which were possible due to the advent of a continuous parasite culture system (Trager and Jensen, 1976). However, culture systems provide a static environment which is substantially different from the *in vivo* conditions (LeRoux, et al., 2009). The host environment may vary in different hosts and during the disease progression. As a result, parasite biology might also get altered in response to changes in one or many of the host factors. The knowledge about the impact of altered parasite biology on disease outcome is limited and any such information would be important for malaria intervention and elimination strategies. As the adult population presents jaundice and/or renal failure as one of the main and most frequent malaria associated severe symptoms (World Health, 2000), it is important to study the parasite biology in these patients in comparison to patients with uncomplicated malaria.

This chapter includes studies which aimed at investigating the transcriptome of *P. falciparum* clinical isolates in adult patients with diverse disease symptoms. *P. falciparum* clinical isolates were directly collected from adult patients with uncomplicated or complicated malaria (jaundice and/or renal failure as main symptoms with or without other associated symptoms) and subjected to microarray based transcriptome analysis. Parasite transcriptomes from patients with uncomplicated malaria were compared with that of complicated malaria to decipher differences in parasite biology in the two clinical conditions with special reference to genes encoding exportome proteins and variant surface antigens. This is the first study, to our knowledge, which investigates the *in vivo* parasite transcriptome and identifies differences in gene expression between parasites from patients with uncomplicated (PFU) malaria as compared to those with complicated malaria (PFC).

5.2. Results

Parasite RNA was isolated from 12 *P. falciparum* infected patients showing either uncomplicated (n=5) or complicated malaria (n=7) symptoms (**Table 5.1**). Parasite RNA was hybridized onto a 60mer custom cross strain *P. falciparum* 15K array containing 6180 probes which represent genomes of two *P. falciparum* (3D7 and HB3) reference strains and *var* genes of IT4 strain for transcriptome analysis (Details of the array design mentioned in chapter 4). The objective of this study was to identify differential expression of genes with a special reference to variant surface antigens between adult patients with uncomplicated and complicated malaria (uncomplicated vs. complicated). All the samples included in this study showed only ring stage parasites in the peripheral blood smear. Genotyping of clinical isolates was carried out by assessing merozoite surface protein 1 (*Pfmsp*1) and merozoite surface protein 2 (*Pfmsp*2) allelic variants to determine the number of clones per sample (Snounou et al., 1999). Data suggests monoclonal or polyclonal isolates with reference to *Pfmsp*1 and *Pfmsp*2 (**Table 5.2**).

Table 5.1. *Plasmodium falciparum* clinical isolates used in this study

S.No.	Uncomplicated isolates[#]	S.No.	Complicated isolates[#]
1	PFU-03	1	PFC-03
2	PFU-04	2	PFC-05
3	PFU-05	3	PFC-10
4	PFU-06	4	PFC-18
5	PFU-07	5	PFC-19
		6	PFC-21
		7	PFC-22

[#]Clinical details of these samples can be found in **Table 2.1**.

Table 5.2. ***Pfmsp1* and *Pfmsp 2* based genotyping of clinical isolates**

Variants	PFU03	PFU04	PFU05	PFU06	PFU07	PFC18	PFC03	PFC05	PFC19	PFC10	PFC21	PFC22
MSP1												
FC27	+	+	+	+	+	+	-	+	+	+	+	-
3D7/IC	-	+	-	-	+	+	+	+	+	+	+	+
K1	-	+	-	-	+	+	-	-	-	+	-	+
MSP2												
MAD20	-	+	+	+	+	+	+	+	+	-	-	+
R033	+	+	-	+	+	-	-	-	+	+	+	-

(+) indicates present of a particular variant and (-) indicates absence of a particular variant in parasite isolates collected from patients.

5.2.1. Differentially regulated genes in complicated malaria isolates

For whole genome expression analysis, only probes representing the 3D7 transcripts were considered. However, for *var* gene expression analysis, probes representing all the three strains were considered. In the case of *rifins* and *stevors*, probes representing 3D7 and HB3 strains were considered.

Of the 5276 genes (3D7 strain) represented by specific probes, 4727 genes were detected in at least one sample. This constitutes around 90% of the total genes represented in the array. A more stringent criterion was applied for further analysis to identify differentially regulated genes, where probes detected in at least 6 out of 12 samples representing 3847 genes (74% of the total genes represented in the array) were considered. Hierarchical clustering of whole genome expression profiles of all samples clearly differentiates the uncomplicated and complicated samples except for sample PFU-04 (**Figure 5.1**).

Based on a volcano plot after applying a cut-off of ≥ 2 fold , total number of genes that differentially expressed in PFC compared to PFU at $P < 0.05$ (moderated t test) were 380 (9.9%). Of the 380 genes declared as significantly differentially expressed, 194 (5%) genes were up-regulated and 186 (4.9%) genes were down-regulated in complicated samples compared to uncomplicated samples (**Figure 5.2A**, **Supplementary Table 5.1**). The differentially regulated gene list contains members of the exportome class of proteins, ATP, GTP, DNA and RNA binding proteins, phosphatases and kinases, sexual stage specific genes, genes involved in host cell invasion and in many other biological processes (**Supplementary Table 5.1**). **Figure 5.2B** displays the expression profiles of top 20 up-regulated and top 20 down-regulated genes in complicated samples compared to uncomplicated samples.

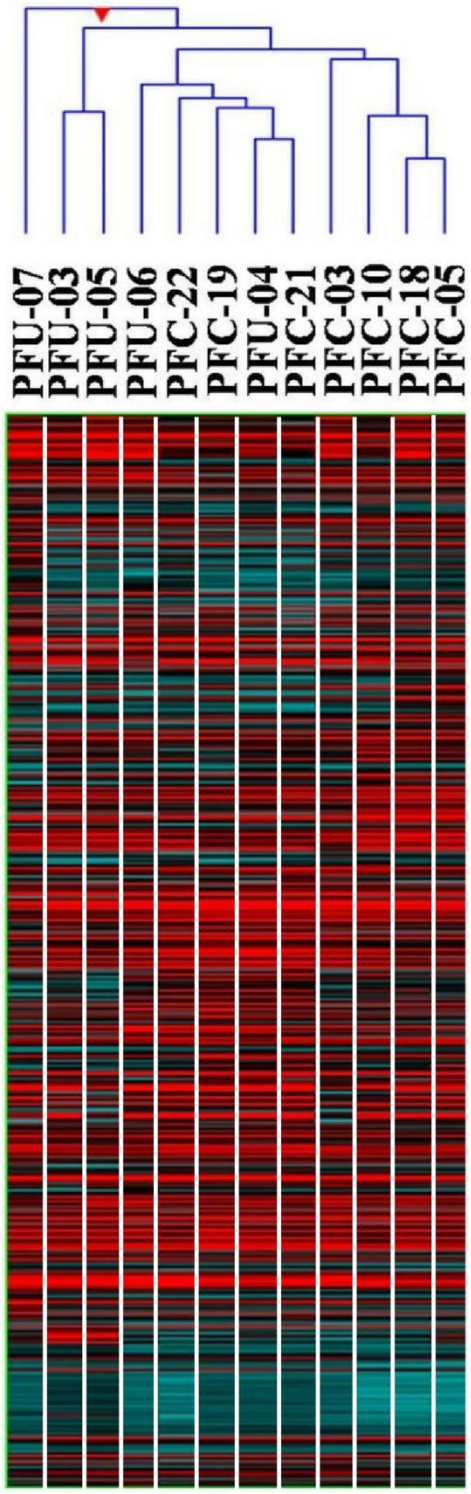


Figure 5.1. Clustering of samples based on expression values. Heat map showing hierarchical clustering of all the samples (column) based on expression values of all the probes. The Pearson centered distance matrix along with average linkage rule were used for clustering the samples.

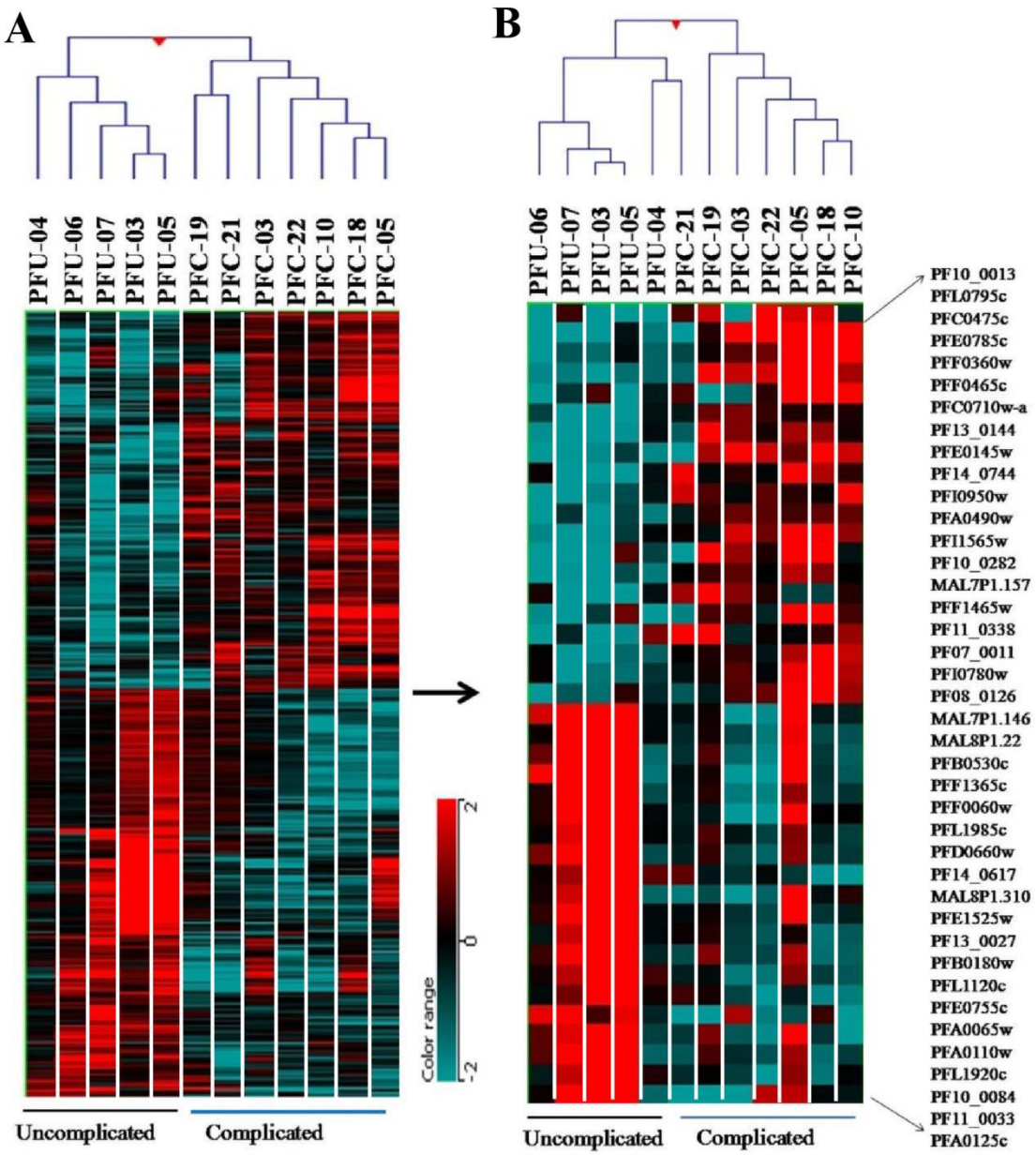


Figure 5.2. Hierarchical clustering of differentially regulated genes in complicated isolates. Pearson distance matrix and average linkage rule were used on samples for cluster generation. Clustering was applied both on samples and genes **A**, Heat map showing hierarchical clustering of all the differentially regulated genes; **B**, Heat map showing hierarchical clustering of top 20 up and down-regulated genes.

The top 20 up-regulated genes includes a metabolite/drug transporter gene (PFE0785c), known to over express in parasite strains having mutations in their *Plasmodium falciparum* chloroquine resistance transfer (*pfcrt)* gene (Jiang et al., 2008); two genes (PF10_0013, PF14_0744) encoding exportome class proteins, of which the former has been reported to localize in parasite derived structures called Maurer's clefts (Vincensini et al., 2005). Two sexual stage specific genes (secreted ookinete protein, putative (PFF1465w) and a male development gene 1 (PFL0795c)), the latter has been found to be over expressed in field strains compared to laboratory strains (Mackinnon et al., 2009); a gene encoding protein disulfide isomerase (PFI0950w), involved in oxidative protein folding, an apicoplast targeted uroporphyrinogen III decarboxylase protein (PFF0360w), a key enzyme in heme biosynthetic pathway (Nagaraj et al., 2009), Oxidoreductase (PF13_0144); phosphatidylinsitol 3- and 4- kinase (PFC0475c) which is involved in lipid metabolism; inorganic pyrophosphatase (PFC0710w), reported to be over expressed in pumilio/FBF (PUF) RNA binding protein (PfPUF2) deleted gametocytes compared to wild type gametocytes (Miao et al., 2013) also fall in the same catrogory. Further genes encoding organelle ribosomal protein L28 precursor, putative protein (PFE0145w); a profilin protein (PFI1565w) which showed delayed translation both in asexual and sexual stage proteome (Le Roch, et al., 2004), an aquaglycoporin protein (PF11_0338), a multifunctional membrane channel protein that facilitates the passage of both water (Chen, 2013) and glycerol and a major drug target (Hansen, et al., 2002) were also found to be up-regulated. Five genes (PFA0490w, PF10_0282, MAL7P1.157, PFL0700w and PF07_0011) representing conserved proteins with unknown function are also part of the top 20 up-regulated gene list.

Likewise, the top 20 down-regulated genes include two genes [rhomboid protease 9 (PFE0755c) and erythrocyte binding antigen 181 (PFA0125c)], of which the latter is believed to be involved in erythrocytic invasion (Cowman et al., 2012); phosphoglycerate mutase (PFD0660w), a serine-threonine protein phosphatase involved in amino sugar & nucleotide sugar metabolism and galactose metabolism; a ring infected erythrocyte surface antigen (RESA, PFA0110w), member of exportome class of proteins that might prevent reentry of the parasite to already parasitized RBCs and could enhance resistance of the infected RBCs to both mechanical and thermal stress, considered to be a potential vaccine candidate (Maier, et al., 2009 and Genton, et al., 2003). Genes encoding Hydroxyethylthiazole kinase involved in thiamine biosynthetic process; a HECT domain (ubiquitin-transferase) encoding protein with amino acid ligase activity involved

in cellular protein modification process; DNA gyrase subunit A protein (PFL1120c); dehydrodolichyl diphosphate synthetase (MAL8P1.22), part of terpenoid back bone biosynthesis process; a probable protein (PF11_0033); a tubulin beta chain protein (PF10_0084) and eight genes encoding conserved *Plasmodium* protein with unknown function were also part of the top 20 down-regulated gene list.

The lists mentioned above indicate that these genes are involved in processes important for parasite invasion, parasite survival in host intra-cellular environment (infected erythrocyte remodeling by exportome class of proteins), parasite transmission to the new host (sexual stage specific genes), nutrient utilization and cellular homeostasis. Functional analysis of all genes has been briefly presented below.

5.2.2. Functional enrichment of differentially regulated genes in complicated *Plasmodium falciparum* clinical isolates.

5.2.2.1 Gene Ontology term based functional enrichment of differentially regulated genes

In order to understand the biological role(s) of differentially regulated genes, gene ontology (GO) based term enrichment analysis was performed using gene functional classification tool from DAVID Bioinformatics resources v6.7 (Huang da et al., 2009, Huang da et al., 2009). This analysis identified up-regulated genes statistically enriched ($P < 0.05$) to biological processes (BP) category like generation of precursor metabolite and energy, DNA packaging, chromosome organization and electron transport chain (**Table 5.3**). Two other BP categories i.e. cell cycle process, cellular macromolecular complex assembly were also enriched (6.8 and 4.4 fold enrichment respectively) however, the values were just above our cut off value ($P = 0.065$ and $P = 0.056$) (**Table 5.3**). The enriched molecular function categories were monovalent inorganic cation transmembrane transporter activity ($P = 0.001$), cytochrome c oxidase activity ($P = 0.01$) and DNA binding (5.7 fold enrichment) (**Table 5.4**). None of the down-regulated genes enriched to any GO category.

Table 5.3. Gene Ontology biological process term enrichment of differentially regulated genes

Term	*P* Value	Fold Enrichment
GO:0006091-generation of precursor metabolites and energy	0.022	3.56
GO:0006323-DNA packaging	0.035	9.60
GO:0051276-chromosome organization	0.038	5.12
GO:0022900-electron transport chain	0.050	8.00
GO:0034622-cellular macromolecular complex assembly	0.056	4.41
GO:0007049-cell cycle	0.066	6.86
GO:0034621-cellular macromolecular complex subunit organization	0.088	3.66

Acronym: BP; biological process. Gene ontology biological process term enrichment was carried out using DAVID functional annotation tool (Franceschini, et al., 2013).

Table 5.4. Gene Ontology molecular function term enrichment of differentially regulated genes

Term	*P* Value	Fold Enrichment
GO:0015077-monovalent inorganic cation transmembrane transporter activity	0.001028	7.18
GO:0004129-cytochrome-c oxidase activity	0.016014	14.37
GO:0003677-DNA binding	0.085687	1.90

Abbreviation: MF; Molecular Function. Gene ontology biological process term enrichment was carried out using DAVID functional annotation tool (Franceschini, et al., 2013).

5.2.2.2. Systems network based functional enrichment

Network-based analysis allows understanding of the processes which might be affected due to linked gene perturbation during a particular condition. Biological cellular networks are modular in nature, and contain inter-connected proteins responsible for specific cellular functions (Hartwell, et al., 1999). In order to identify biological processes which are linked to perturbed genes and which might get affected, all the differentially regulated genes were mapped to a *P. falciparum* network derived from StringDB v9.1(Franceschini, et al., 2013). A sub-network was generated by considering only the differentially regulated genes which connected to at least one or more differentially regulated genes (**Figure 5.3A**). Further, the generated sub-network was subjected to MCL clustering algorithm (inflation value = 2.5) for identification of densely connected interaction clusters or modules. MCL clustering identified 53 clusters with maximum cluster size of 28 nodes (node represents gene) and minimum cluster size of 2 nodes (**Figure 5.3B**). Clusters having at least 6 nodes (total 6 clusters with ≥ 6 nodes) were further analyzed. To investigate the biological processes under the influence of differentially regulated gene clusters, each cluster was grown by adding 10 first interacting partners and subsequently mining for their enrichment to a given gene ontology based biological process term, using the advanced analysis option present in StringDB v 9.1 (**Figure 5.4**) (Franceschini et al., 2013). This analysis identified biological processes which might be perturbed in complicated malaria cases.

The largest and most dense inter-connected module (**Figure 5.4A**) enriched to the host cell entry BP gene ontology (GO) category. This module contains the majority of down-regulated genes with few up-regulated genes and invasion related genes like EBA-181, MSP 6, PfRh5, PfRH2b and RhopH2. The second largest cluster (**Figure 5.4B**) enriched to regulation of intracellular pH BP GO term. This cluster contains four genes (PF13_0034, PF11_0412, PF14_0615 and PFD0305c) involve in regulation of intracellular pH but were not differentially regulated and linked to the original cluster through an up-regulated gene (PF11_0412). Majority of the members of this module were up-regulated. The important member of this module encodes for a *Plasmodium falciparum* aquaglycoporin protein (PfAQP). The members of the third module enriched to three inter-related BP GO terms like RNA metabolic process, nitrogen compound metabolic process and gene expression, and contains majority of up-regulated genes as its member (**Figure 5.4C**). The fourth module has all the members up-regulated and enriched to the

RNA splicing BP gene ontology term (**Figure 5.4D**). The fifth module did not enrich to any particular BP class whereas the sixth module (**Figure 5.4E**) enriched to vitamin biosynthetic process BP GO term.

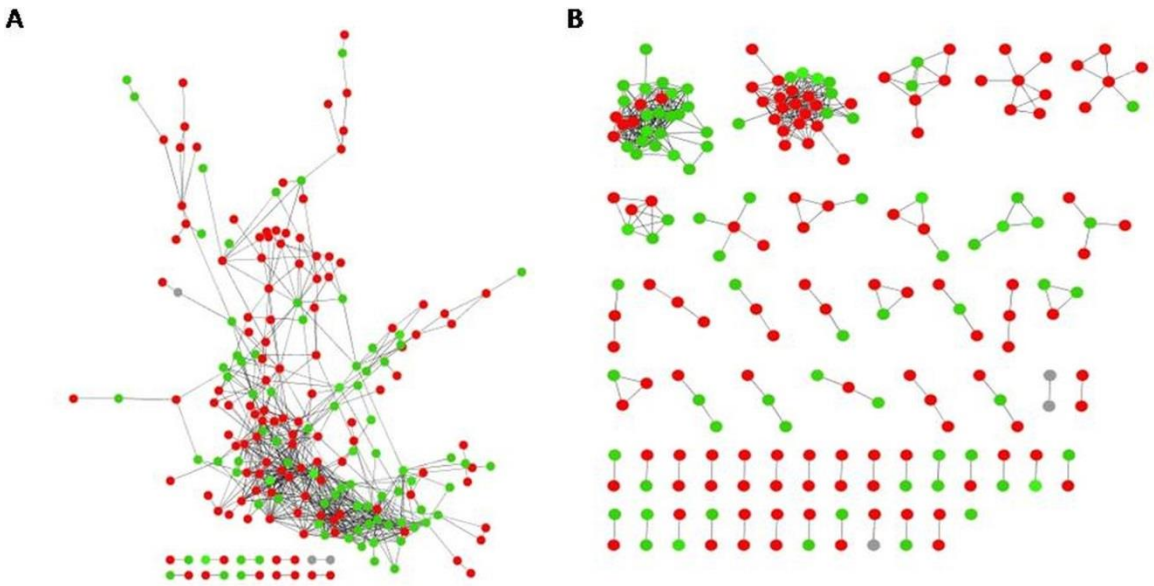


Figure 5.3. **Sub-networks of differentially regulated genes**. A, whole sub-network extracted from *P. falciparum* network present in StringDB v9.1 (Franceschini, et al., 2013) after mapping the differentially regulated genes; B, Sub-networks after applying MCL clustering. Red node indicates up-regulated genes and green node indicates down-regulated genes.

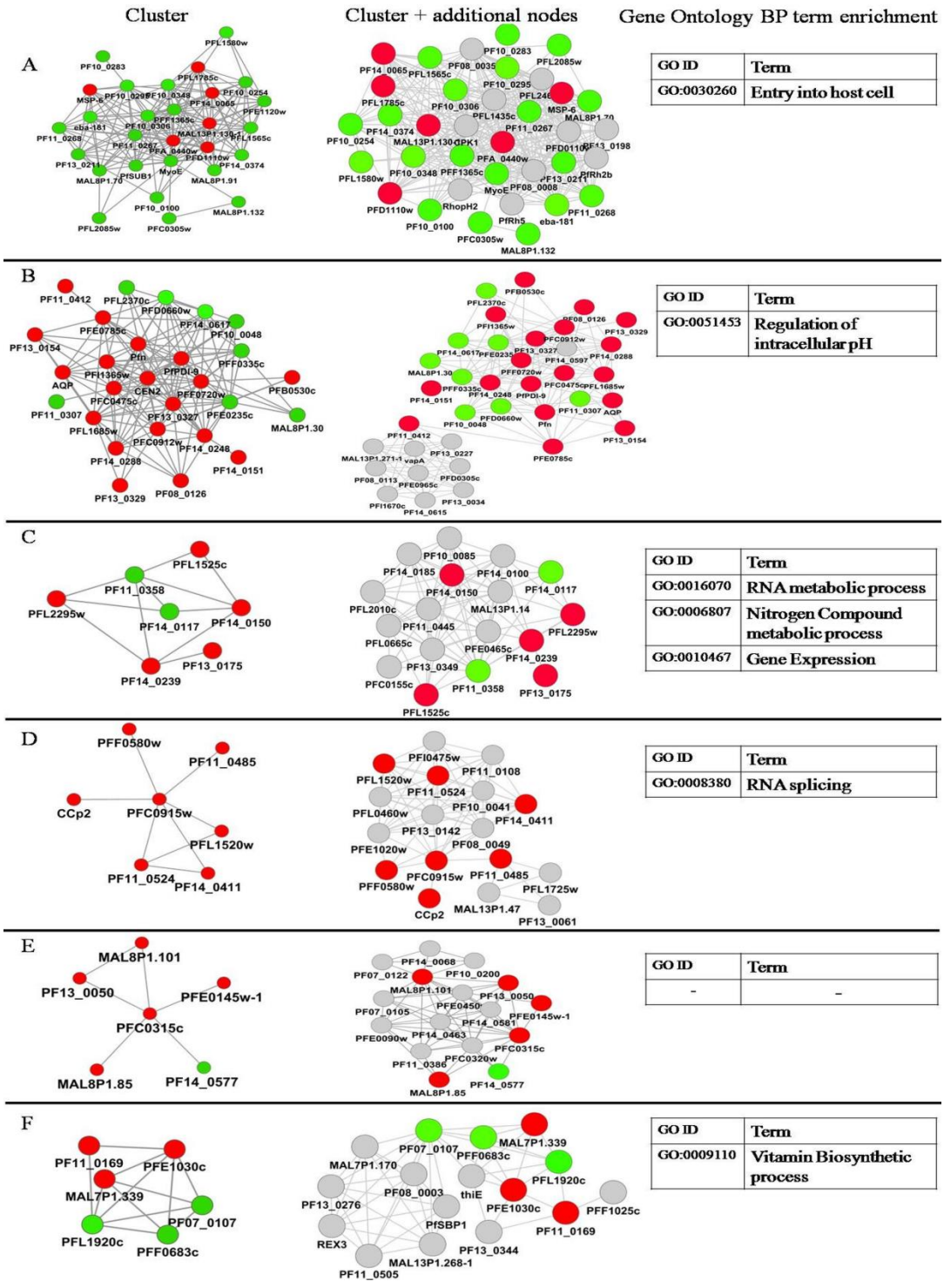


Figure 5.4. Clusters identified from sub-network of differentially regulated genes. **A-F**, 6 different clusters with at least 6 interacting nodes identified for analysis. Cluster, original cluster from sub-network of differentially regulated genes; cluster + additional nodes, 10 additional nodes added to the original cluster using advanced analysis option of StringDB v9.1 (Franceschini, et al., 2013); gene ontology BP term enrichment; enrichment of each cluster to a particular Biological Process gene ontology category. Red nodes indicate up-regulated genes and green nodes indicate down-regulated genes.

5.2.3. **Differential expression pattern of genes encoding exportome class of proteins**

Proteins in the exportome class are predicted to be exported from the parasite, crossing the parasitophorus vacuole membrane (PVM) to reach specific locations in the host cells by establishing *de novo* trafficking machineries which is lacking in a normal cell (Cooke et al., 2004 Lanzer et al., 2006, Marti et al., 2005, Przyborski et al., 2003). These proteins are involved in infected erythrocyte remodeling, converting the terminally differentiated RBCs to one with altered permeability, structural, mechanical and cytoadherence properties (Deitsch and Wellems 1996, Kyes et al., 2001, Rowe and Kyes, 2004). The parasite induced erythrocyte remodeling creates an environment inside and outside the host cell which facilitates parasite survival (Maier et al., 2008). Understanding the expression of these genes is therefore very important in, any attempt to analyze the disease conditions at the molecular level. The exportome class contains > 400 proteins predicted to be exported through the parasitophorous vacuole (PV) membrane with a majority of members coding for hypothetical proteins (Hiller et al., 2004, Marti et al., 2005, Sargeant et al., 2006, Van Ooij et al., 2008). Most of the members of exportome class are characterized of having a pentameric signal motif (RXLXE) known as protein export element/ host targeting signal (PEXEL/HT) located about 35 amino acids downstream from the hydrophobic signal sequence. There are also some members in the exportome class which lack PEXEL/HT motif but are known to be exported through the PV membrane (Pexel Negative Exported Elements (PNEPs).

It was observed that 23 genes encoding proteins of the exportome class were differentially expressed in PFC isolates (**Supplementary Table 5.1**). Of the 23 differentially regulated genes, 9 genes were up-regulated and 14 genes were down-regulated (**Supplementary Table 5.1**). Of the 9 up-regulated genes, two exported genes (PF10_0013 and PF14_0744) are already mentioned in the top 20 up-regulated gene list analysis. Remaining members of the up-regulated exportome class include one member encoding *Plasmodium* helical interspersed subtelomeric family c (PHISTc) domain containing protein (PFE1595c), a member of PHISTc protein subfamily with its ortholog found only in *P. vivax* and 6 gene encoding *Plasmodium* exported protein with unknown function. Members of the PHISTc subfamily have orthologs found in

many *Plasmodium* species and are believed to perform a function (unknown) common to the genus *Plasmodium* (Sargeant et al., 2006).

Some selected down regulated exportome genes include three PHISTb domain encoding proteins (MAL7P1.7 and PFD1180w, PFA0110w-RESA), one of which (PFA0110w-RESA) also contains an additional DnaJ domain mentioned earlier in the top 20 down-regulated gene list. Another DnaJ domain containing protein (PF14_0013) which is a member of type IV Hsp40 protein subfamily (Botha et al., 2007), two *Plasmodium* exported proteins (MAL7P1.177 and MAL8P1.160) belonging to hyp 9 and hyp 7 subfamily of proteins, glycosylphosphatidylinsositol anchor attachment 1 protein, putative (MAL13P1. 348) and two Pfmc-2TM Maurer's cleft two transmembrane proteins (PFA0065w and PFFOO60w) were also found to be down-regulated.

5.2.4. **Expression pattern of variant surface antigens**

5.2.4.1. **Differential expression of *var* genes**

The differences in expression patterns of genes between adult patients with uncomplicated and complicated malaria were investigated. The arrays utilized here contain probes representing *var* transcript sequences of three different geographical reference strains i.e. 3D7, HB3 and IT4. Array contains specific probes for 40, 30 and 34 *var* genes of 3D7, HB3 and IT4 strains respectively, representing all the *var* groups. Probes representing 3D7, HB3 and IT4 transcript sequences detected 35, 23 and 22 *var* gene transcripts respectively. Although, a few probes are common between the referral strains, there are also probes which are specific to *var* genes from each strain. Transcripts were detected from members of *var* gene family representing each of the five groups in either PFU or PFC. This may be due to the presence of polyclonal populations of parasites in the host as determined through the genotyping of the isolates (**Table 5.2**).

Differential expression analysis of probes representing 3D7 *var* genes identified over expression of members belonging to *var* B and C groups in PFC as compared to PFU (**Figure 5.5**). Probes considered for 3D7 analysis also cover some of the genes from HB3 and/or IT4 strains. Interestingly, the expression pattern showed that in each of the PFC sample at least one of the members of the two *var* groups, namely group B and C were found to be up-regulated. In the samples analyzed, only two of the *var* A genes were singly detected and up-regulated in just two

of the seven PFC cases analyzed (**Figure 5.5**). *var* E group which includes a single member (PFL0030-*var2csa*) and has been associated with pregnancy associated malaria (Salanti et al., 2004, Salanti et al., 2003), One of the five probes representing a single member of the *var* E group was detected in all the PFU and in five PFC cases but did not show up-regulation in any of the PFC cases (**Figure 5.6** and **Figure 5.5**).

Differential expression analysis of probes representing only HB3 *var* genes (Probe specific to only HB3 strain) showed an overall similar expression pattern for all *var* group members as seen in case of 3D7 (**Figure 5.7**). In case of HB3 as compared to 3D7, only one member of *var* group C in one PFC case, and one or more than one member of *var* group B in five PFC cases was found to be up-regulated. This difference could be due to the presence of few HB3 specific probes (only specific to HB3) that could be designed. However, many other *var* B group genes have been represented by probes which are common to both 3D7 and HB3 strains and have shown transcript detection as mentioned earlier. Out of five genes of the *var* A group represented by specific probes, only two genes showed up-regulation in two individual PFC cases (**Figure 5.7**). However, the transcripts for *var* A group genes were detected in all the five PFU isolates (**Figure 5.8**).

In the case of probes representing only IT4 *var* genes (specific to IT4 strain only), members of *var* B group showed similar expression pattern as seen in case of probes representing the 3D7 strain (**Figure 5.9**). Probes specific to IT4 could not be designed for any of the *var* B/C group members. There is only a single *var* C member of the IT4 strain (IT4_var47) represented by specific probe. This was not detected in any of the PFC isolates but detected in three PFU isolates (**Figure 5.10**). In case of *var* group A, one member (IT4_var7) showed over expression in four PFC isolates but in three of the PFC isolates (PFC-18, PFC-5 and PFC-22) where IT4_var7 was up-regulated, their counterpart from *var* B group showed higher fold expression in the same samples (**Figure 5.9**). This confirms the trend in expression seen in case of probes representing the 3D7 and HB3 *var* genes where *var* group B and *var* group C genes were found to be up-regulated.

PFC-18	PFC-5	PFC-3	PFC-19	PFC-10	PFC-21	PFC-22	3D7 *var*	HB3 *var*	IT4 *var*
							***var* A**		
					1.1		PF08_0141		
2.3					4.9		PFE1640w	PFHT_03840	
							Var B/A		
-1.9	-0.9	-0.8		-1.3	-0.7	0.9	PF08_0140		
					4.6		PFL0020w		
				1.9			PFL0020w		
							***var* B**		
		4.0			4.8		PFA0005w		
-0.4	1.8	1.4	-4.4	2.9	1.6	2.2	PFA0765c		
	3.0	9.4		3.6			PFB1055c	PFHT_03476	
1.0	-1.1	-2.7	1.0	1.5	-2.5	-0.7	PFC0005w		
					3.9		PFC1120c	PFHT_05200	
		3.6					PFC1120c		
		1.7					PFE0005w		
		0.1				0.1	PFE0005w		
1.7	-0.3	-3.5	-1.0	-0.3	-1.8	-1.3	PFF1595c	PFHT_04769	
		0.0	0.0		0.7	-0.7	MAL7P1.212		
-2.6	-0.1	-6.1	-3.2	-3.8	-4.5	-5.4	PF08_0142		
				0.6	1.2	1.8	MAL8P1.220		
3.1			2.6	9.7		4.1	PF10_0406		
	-2.2	-2.3	-1.5	-2.3	4.5	7.6	MAL13P1.1		
	-2.2			-0.9	-1.7	-2.2	MAL13P1.356		EF158089.1
-5.0	-2.9	-2.1	-3.6	-3.3	-2.4	-4.8	PFI1830c		AY578326.1
2.8					10.6		MAL7P1.187		EF158086.1
							***var* B/C**		
-2.2	-1.9	-3.3	-1.6		-3.3	-3.8	MAL7P1.50		AY372122.1
0.0	-0.3	-3.3	-2.3	-2.7	0.0	-3.6	MAL7P1.55		
				-0.1	2.8		PF08_0103	PFHT_02429	
	0.2						PF07_0050	PFHT_03234	
							***var* C**		
		5.2	0.7				PFD1015c	PFHT_03478	
						0.4	PF07_0048		
5.9					9.3		PFL1960w		
	6.9		9.0	3.0			PF07_0051		
-0.5		-0.7				-0.8	PF07_0051		
3.2				1.9			PFD0995c		EF158088.1
		6.7	0.2		1.7	3.8	PFF0845c		
-2.1	-1.1	-1.8	2.3	-2.5	0.1	-1.7	MAL7P1.56	PFHT_04081	
1.1		0.8			5.6		MAL7P1.56		
							***var* E**		
	-0.4						PFL0030c		AY372123.1
-1.1	-1.2	-1.0		-1.1		-2.7	PFL0030c		AY372123.1

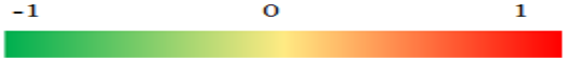


Figure 5.5. Heatmap showing differential expression pattern of 3D7 *var* genes in complicated *P. falciparum* isolates. Grey color cell indicates no transcript was detected for the particular gene in sample(s).

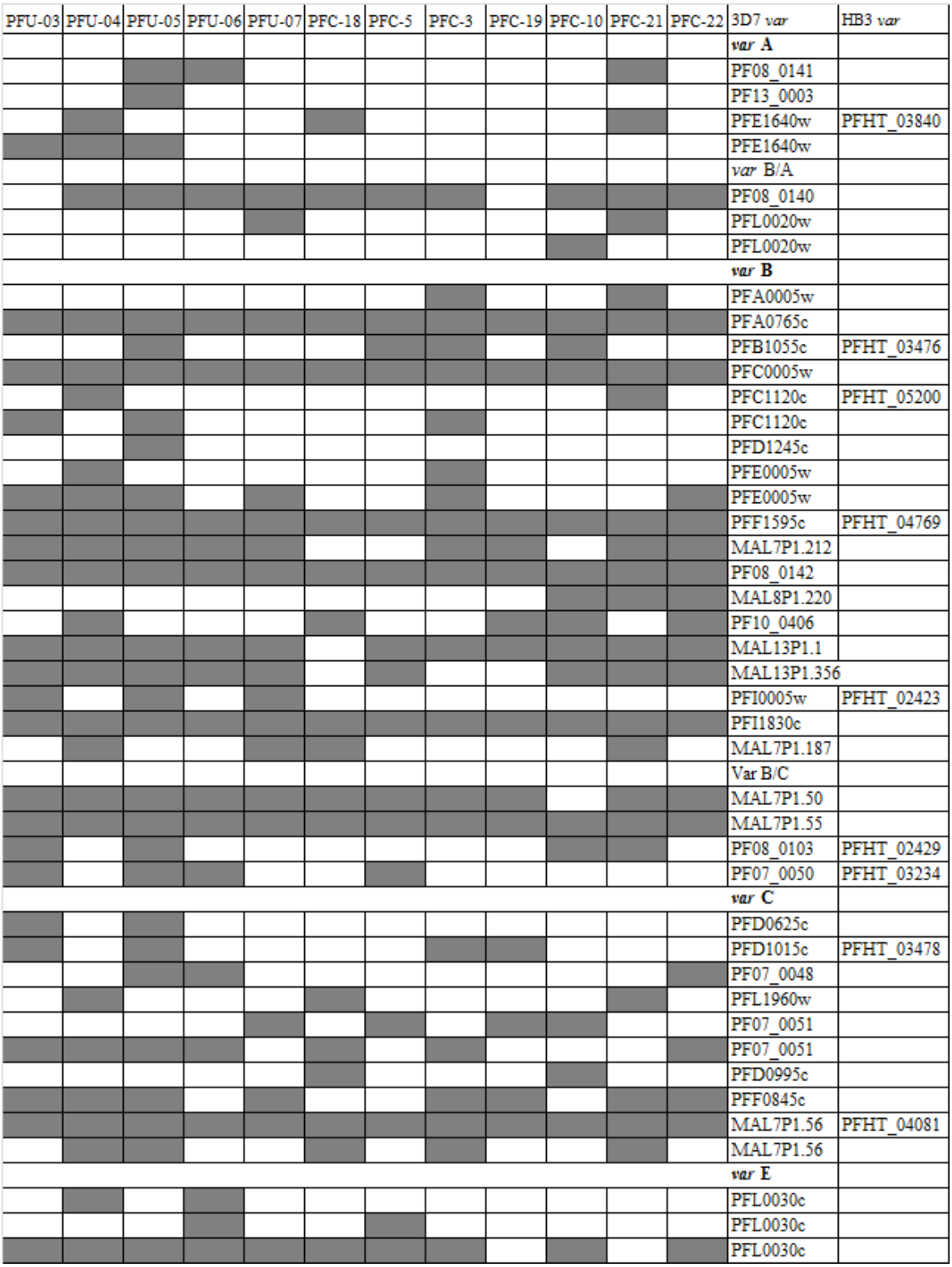


Figure 5.6. Heatmap showing detection status of 3D7 *var* genes in all *P. falciparum* clinical isolates. Grey color cells represent detection of transcript and white color cell represents unexpressed transcripts.

PFC-18	PFC-5	PFC-3	PFC-19	PFC-10	PFC-21	PFC-22	HB3 *var*
							Var A
		-0.3	0.6		0.7	-0.5	PFHT_03234
							PFHT_03234
		-1.0	-0.3		1.1	-0.8	PFHT_05052
		-0.8	-0.4		-0.7	2.3	PFHT_04861
		-1.0	-0.3		1.1	-0.8	PFHT_05052
							Var B/A
						0.4	PFHT_04769
							Var B
-0.4	3.1		-1.2	0.8	2.3	-1.1	PFHT_04593
-1.4	5.1	0.9	3.2	1.9		-0.9	PFHT_04491
				0.6	-0.5		PFHT_04491
						0.4	PFHT_04769
							Var B/C
-2.6	0.1	-2.0	-2.1	-3.0	-3.2	-4.7	PFHT_04620
	-1.3	-0.6	0.0		-0.5	-0.9	PFHT_04910
							Var C
		-0.5	3.8		0.7		PFHT_02277
		-0.5	-0.2		0.0	-1.0	PFHT_00592
	-0.2			-1.6			PFHT_04012
							Var Group E
	-0.5	-0.7	-0.1	1.5	-0.4	-1.3	PFHT_05155

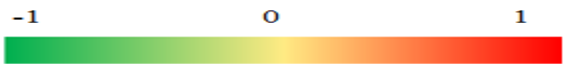


Figure 5.7. **Heatmap showing differential expression pattern of HB3 *var* genes in complicated *P. falciparum* isolates**. Grey color cell indicates no transcript was detected for the particular gene in sample(s).

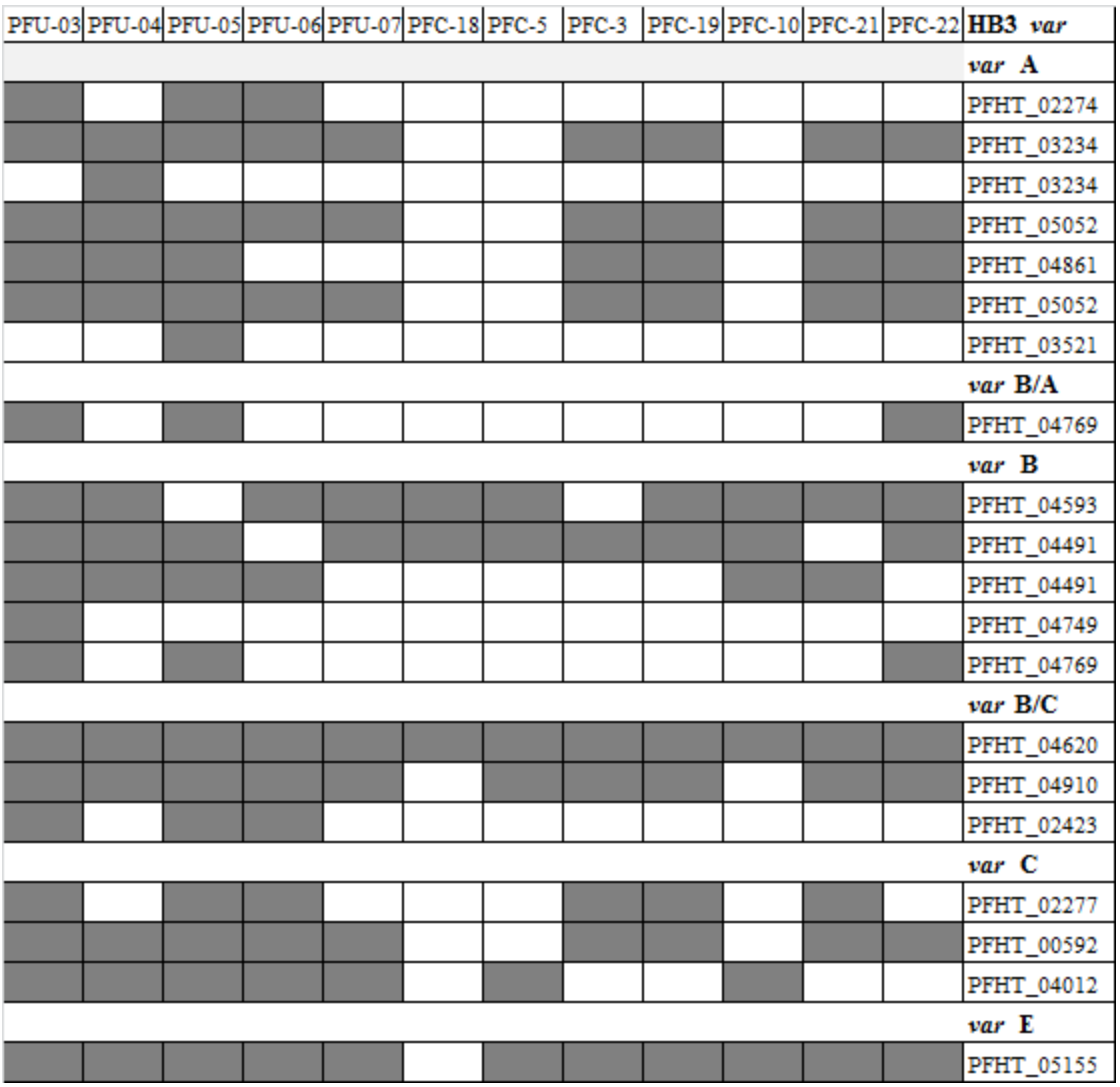


Figure 5.8. Heatmap showing detection status of HB3 *var* genes in all *P. falciparum* clinical isolates. Grey color cells represent detection of transcript and white color cell represents unexpressed transcripts.

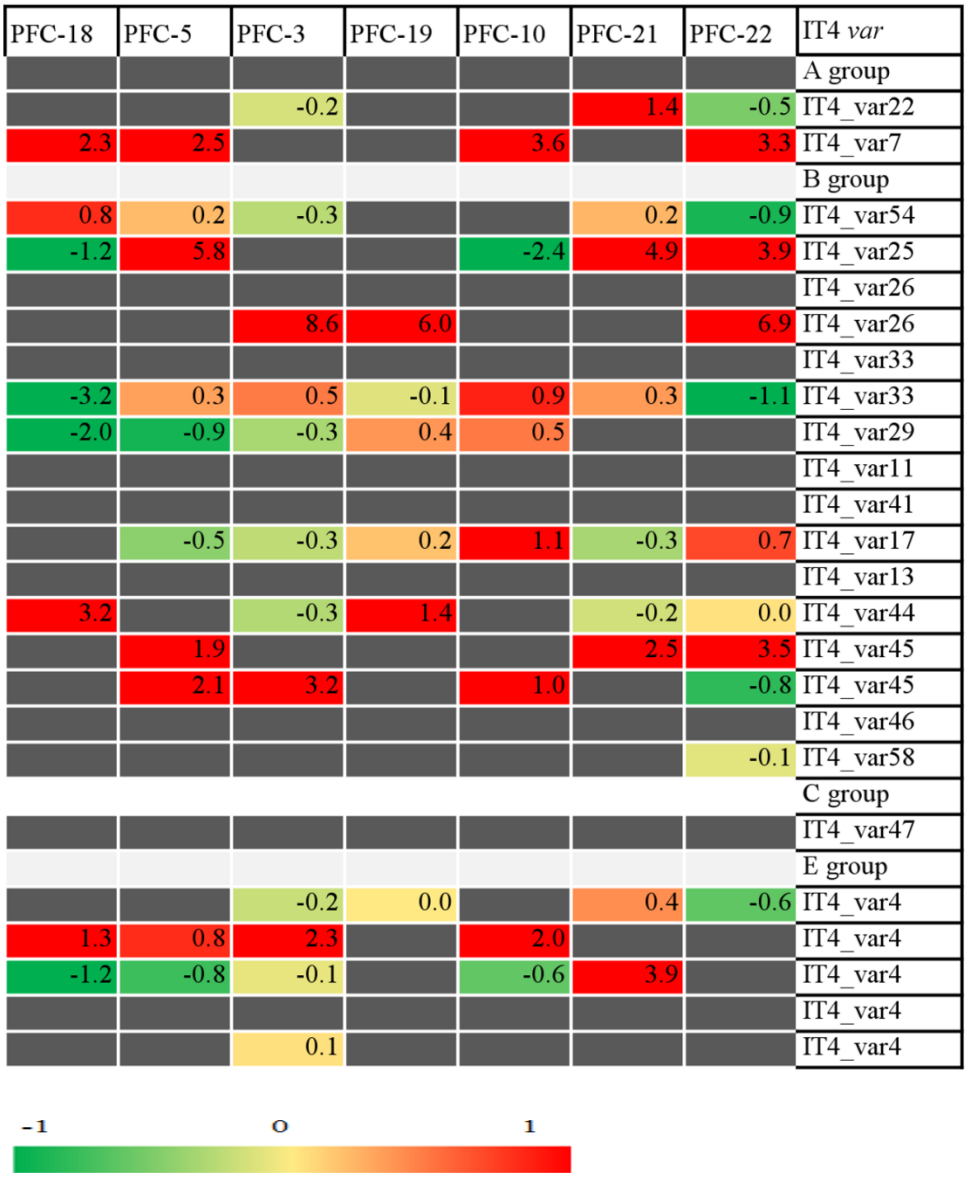


Figure 5.9. **Heatmap showing differential expression pattern of IT4 *var* genes in complicated *P. falciparum* isolates**. Grey color cell indicates no transcript was detected for the particular gene in sample(s).

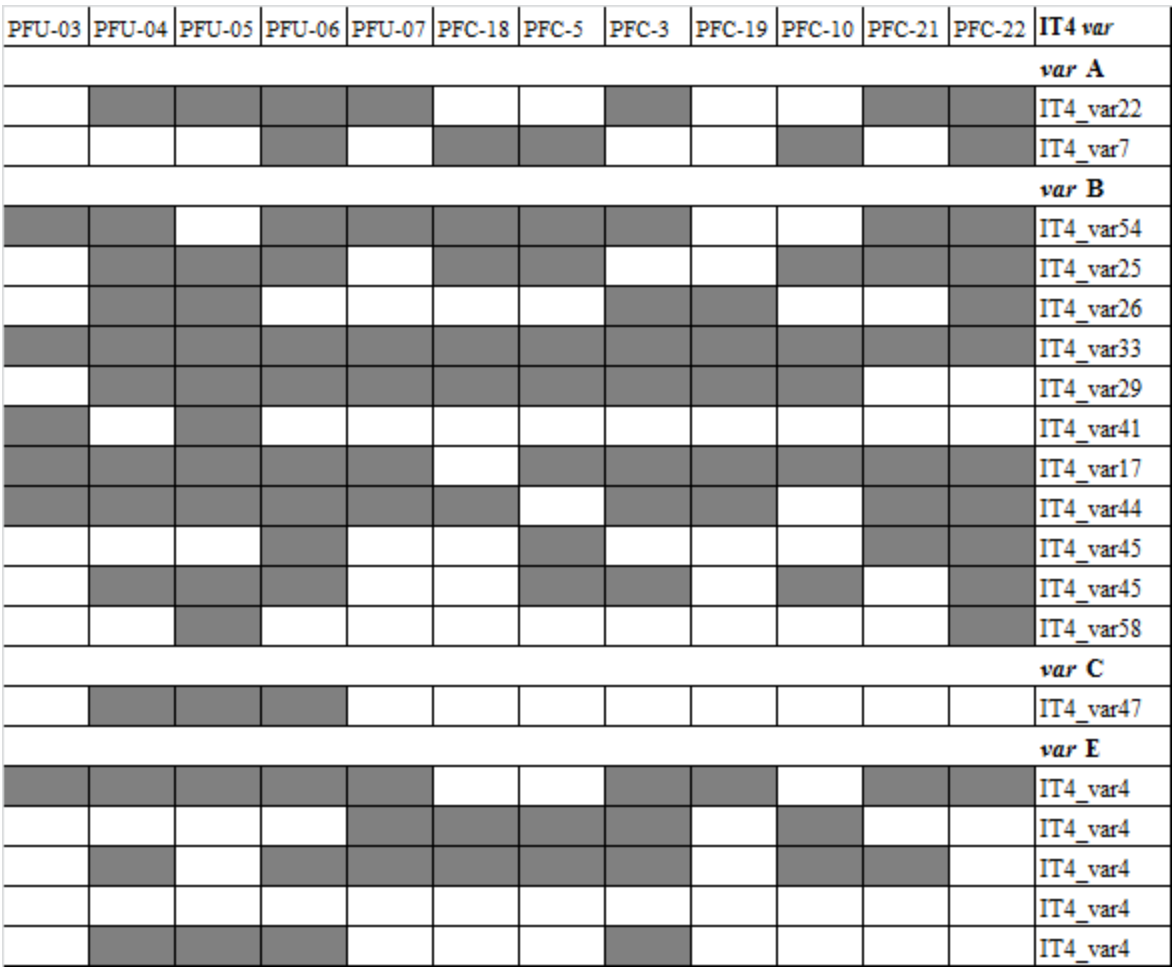


Figure 5.10. **Heatmap showing detection status of IT4 *var* genes in all *P. falciparum* clinical isolates**. Grey color cells represent detection of transcript and white color cell represents unexpressed transcripts.

5.2.4.2. Differential expression pattern of other variant surface antigen genes

Array used in this study contains specific probes for 80 type-A and 19 type-B *rifin* genes of 3D7 strain and 43 type-A and 9 type-B *rifin* genes of HB3 strain (detailed in Chapter 3). Transcripts were detected from 43, 3D7 *rifin* genes (34 type-A and 9 type-B *rifins*) (**Figure 5.11**) and from 30, HB3 *rifin* genes (25 type-A and 5 type-B rifins) (**Figure 5.12**). Expression analysis was carried out to find differentially expressed *rifins* between PFC and PFU isolates which identified up-regulation of type-A *rifins* in PFC cases. Each of the PFC isolate was represented by at least one up-regulated *rif* A group member (**Figure 5.13**). Except for six genes, most of the probes representing the members of type-B *rifins* were not detected in PFC isolates (**Figure 5.13**). Similar expression pattern was observed for probes representing type-A and type-B HB3 *rifins* as seen for type-A and type-B 3D7 *rifins* (**Figure 5.14**). These observations confirmed the difference in the expression pattern of group-A and group-B *rifins* in this parasite.

Microarray used in this study contains specific probes for 26 *stevor* genes out of 28 genes reported in the 3D7 strain. Of the 26 genes, transcripts were detected from 14 *stevor* genes in at least one sample. *stevor* transcripts were detected in at least one to as many as seven isolates (**Figure 5.15**). *stevor* genes were found to be up-regulated in two PFC isolates (PFC-03 and PFC-19) with 2-4 up-regulated members per sample (**Figure 5.16**).

The Array also contains specific probes for 23 *stevor* genes out of 26 genes reported to be present in the HB3 strain (Joannin et al., 2008). Transcripts were detected from 12 *stevor* genes of the HB3 strain (**Figure 5.17**). Probes specific for only HB3 *stevor* genes identified 2 more PFC isolates (PFC 21 and PFC 22) having *stevor* transcripts in comparison to 3D7 (**Figure 5.18**). The advantage of having probes from different strains of *Plasmodium falciparum* is the possibility of detecting more transcripts than with probes designed against only one of the type strains. In the case of this study which investigates field isolates with uncharacterized genomes this provides a definite advantage to the analysis.

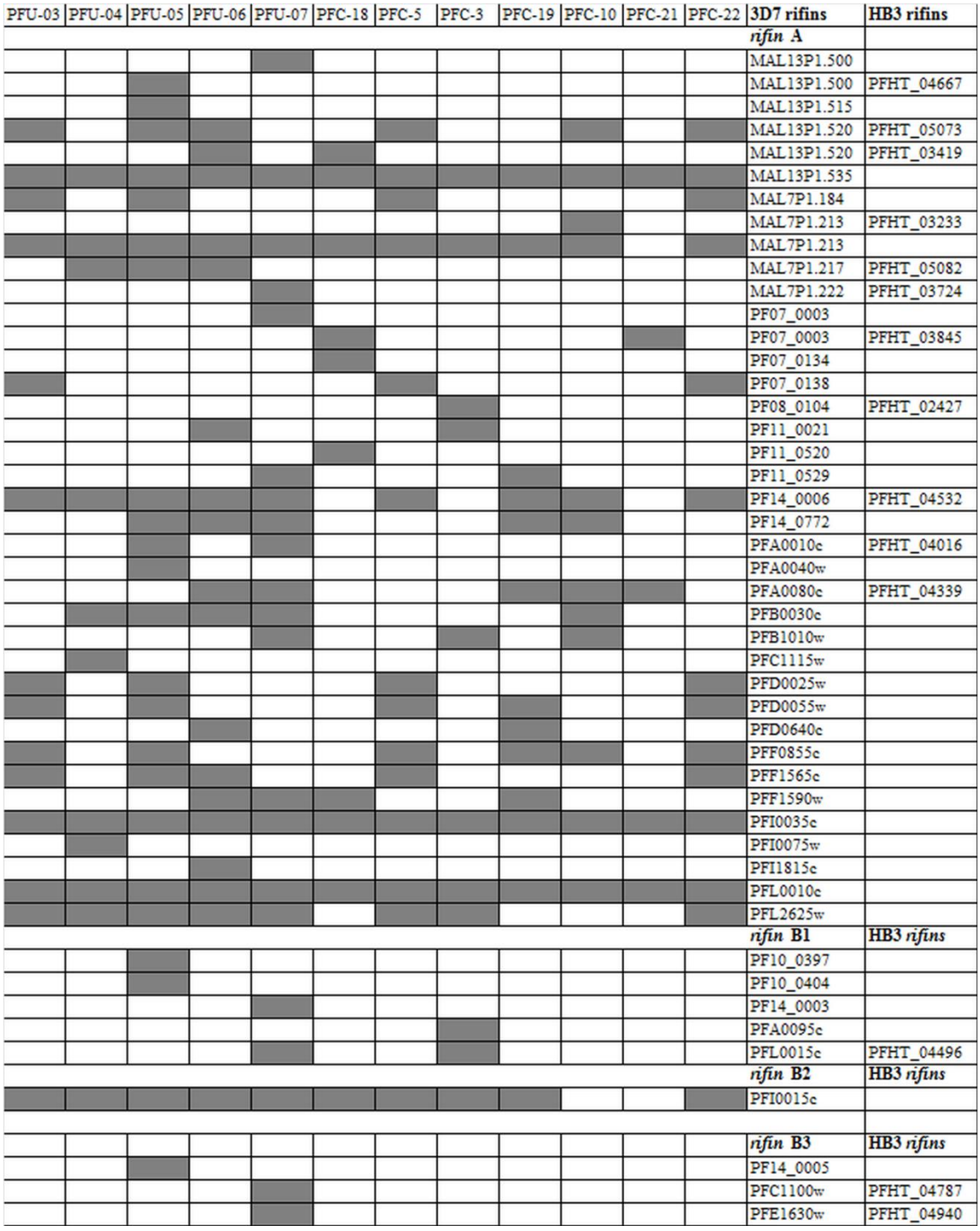


Figure 5.11. Heatmap showing detection status of 3D7 *rifins* in all *P. falciparum* clinical isolates. Grey color cells represent detection of transcript and white color cell represents unexpressed transcripts.

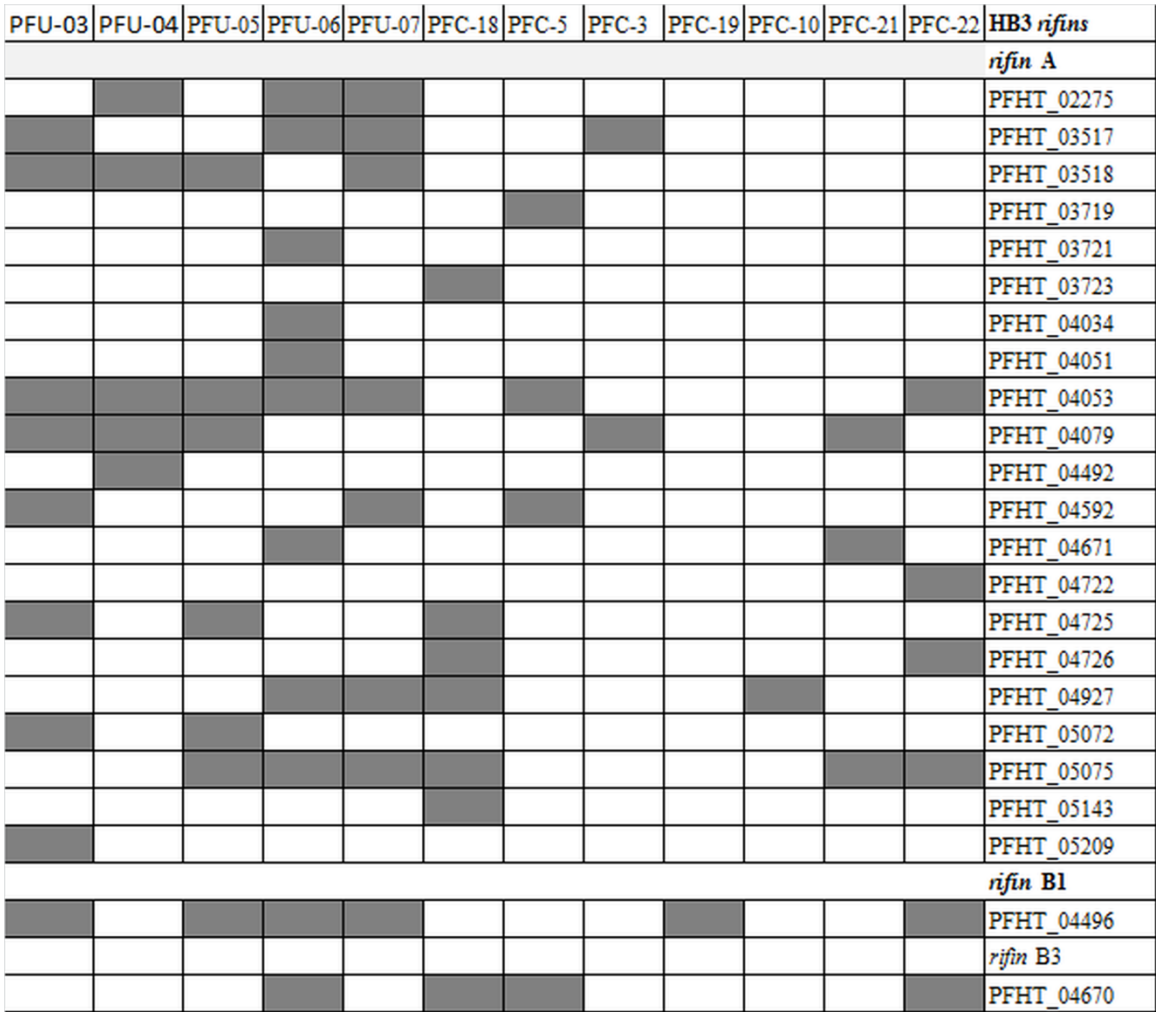


Figure 5.12. **Heatmap showing status of HB3 *rifins* in all *P. falciparum* clinical isolates**. Grey color cells represent detection of transcript and white color cell represents unexpressed transcripts.

PFC-18	PFC-5	PFC-3	PFC-19	PFC-10	PFC-21	PFC-22	3D7 *rifin*	HB3 *rifin*
							rifin A	
	0.2			0.2		0.2	MAL13P1.520	PFHT_05073
1.9							MAL13P1.520	PFHT_03419
-2.9	4.7	-0.1	-2.4	-1.3	-2.1	4.7	MAL13P1.535	
	-0.3					-0.3	MAL7P1.184	
				4.9			MAL7P1.213	PFHT_03233
-5.0	-4.2	-4.1	-5.0	0.7		-4.2	MAL7P1.213	
0.9					0.7		PF07_0003	PFHT_03845
3.8							PF07_0134	
	0.8					0.8	PF07_0138	
		2.7					PF08_0104	PFHT_02427
		1.6					PF11_0021	
3.0							PF11_0520	
			4.0				PF11_0529	
	0.2		-0.6	-0.2		0.2	PF14_0006	PFHT_04532
			-0.1	0.6			PF14_0772	
			3.4	5.8	1.8		PFA0080c	PFHT_04339
				1.1			PFB0030c	
		1.7		7.4			PFB1010w	
	2.5					2.5	PFD0025w	
	1.0		0.5			1.0	PFD0055w	
			0.6				PFD0640c	
	0.6		0.2	0.6		0.6	PFF0855c	
	0.2					0.2	PFF1565c	
2.1			1.5				PFF1590w	
-3.6	3.6	-0.6	-2.8	-1.5	-3.3	3.6	PFI0035c	
-3.9	4.2	-0.7	-3.8	-2.4	-4.5	4.2	PFL0010c	
	3.0	-0.5				3.0	PFL2625w	
							rifin B1	HB3 *rifin*
		2.6					PFA0095c	
		0.5					PFL0015c	PFHT_04496
							rifin B2	HB3 *rifin*
-3.4	0.2	-2.1	-3.4			0.2	PFI0015c	

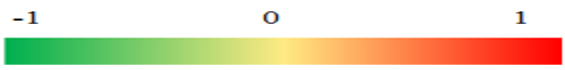


Figure 5.13. Heatmap showing differential expression of 3D7 *rifins* in complicated *P. falciparum* clinical isolates. Grey color cell indicates, no transcript was detected for the particular gene in sample(s).

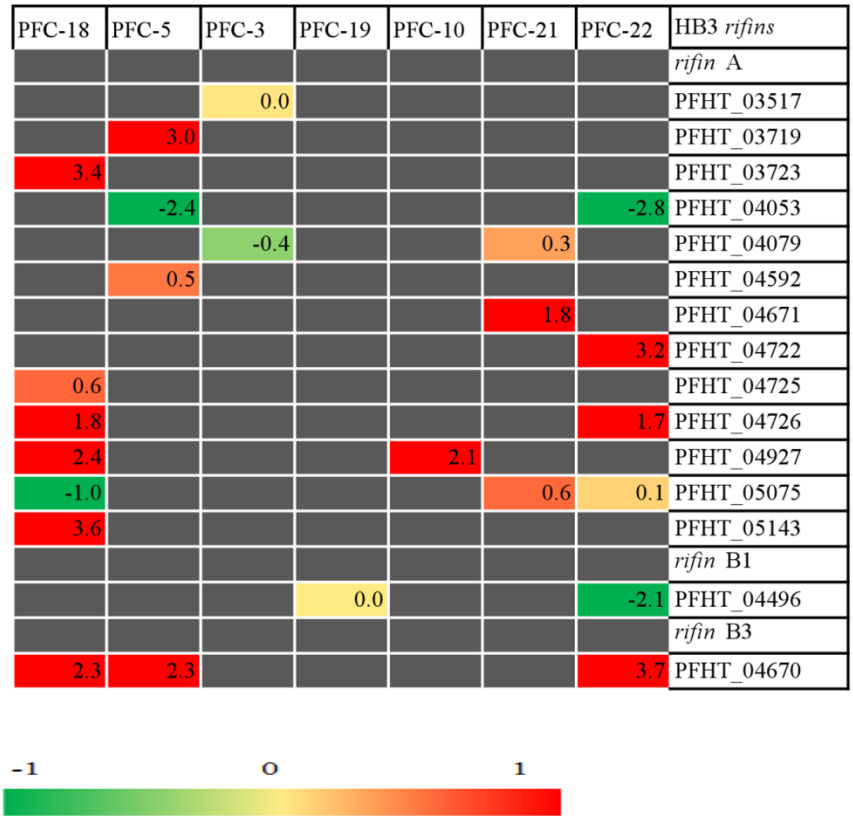


Figure 5.14. Heatmap showing differential expression of HB3 *rifins* in complicated *P. falciparum* clinical isolates. Grey color cell indicates, no transcript was detected for the particular gene in sample(s).

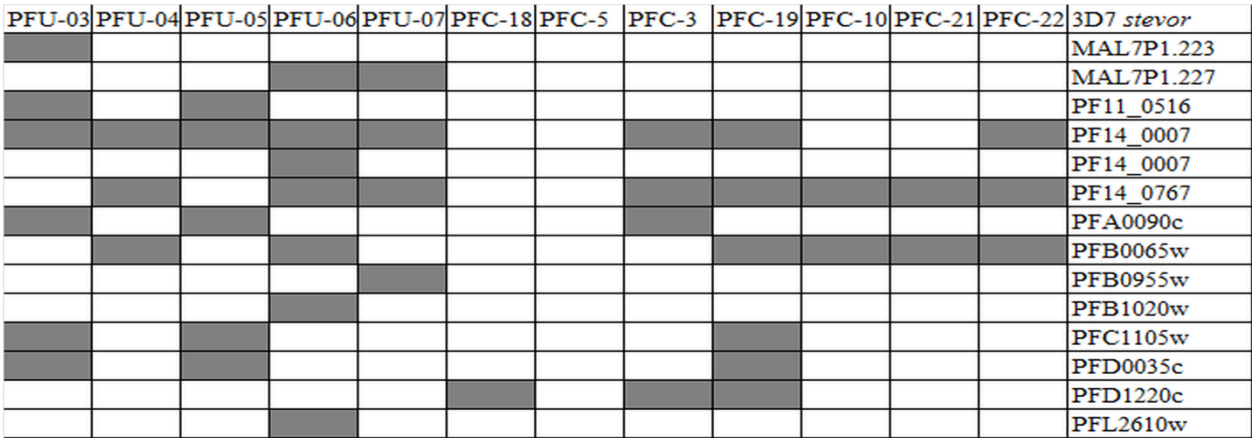


Figure 5.15. Heatmap showing detection status of 3D7 *stevors* in all *P. falciparum* clinical isolates. Grey color cells represent detection of transcript and white color cell represents unexpressed transcripts.

PFC-18	PFC-5	PFC-3	PFC-19	PFC-10	PFC-21	PFC-22	3D7 *stevor*
		-0.6	0.2			-0.9	PF14_0007
		1.7	2.8	0.4	0.9	0.0	PF14_0767
		-0.1					PFA0090c
			1.2	0.9	0.8	0.0	PFB0065w
			1.0				PFC1105w
			0.7				PFD0035c
0.3		1.4	1.6				PFD1220c

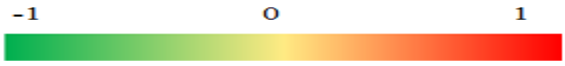


Figure 5.16. Heatmap showing differential expression of 3D7 *stevors* in complicated *P. falciparum* clinical isolates. Grey color cell indicates, no transcript was detected for the particular gene in sample(s).

PFU-03	PFU-04	PFU-05	PFU-06	PFU-07	PFC-18	PFC-5	PFC-3	PFC-19	PFC-10	PFC-21	PFC-22	**HB3 *stevor***
												PFHT_03712
												PFHT_04052
												PFHT_04531
												PFHT_04535
												PFHT_04591
												PFHT_04727
												PFHT_05035
												PFHT_05217

Figure 5.17. Heatmap showing detection status of HB3 *stevors* in all *P. falciparum* clinical isolates. Grey color cells represent detection of transcript and white color cell represents unexpressed transcripts.

PFC-18	PFC-5	PFC-3	PFC-19	PFC-10	PFC-21	PFC-22	HB3 *stevor*
			1.2	0.6	0.5	0.1	PFHT_03712
					2.7		PFHT_04531
-0.7	-2.4	0.4	-1.2		-2.3	-2.5	PFHT_04535
-0.3			2.2		0.7	1.4	PFHT_05035
-1.0	-2.6	0.2	3.3	-2.0	-1.4	1.9	PFHT_05217

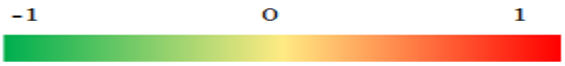


Figure 5.18. Heatmap showing differential expression of HB3 *stevors* in complicated *P. falciparum* clinical isolates. Grey color cell indicates, no transcript was detected for the particular gene in sample(s).

5.2.5. **Network based connectivity of exportome and variant surface antigen genes**

A protein-protein interaction (PPI) network was created by considering reported exportome and variant surface antigens. The entire network consists of 251 interconnected nodes and shows two dense regions of connectivity (**Figure 5.19**). One of these regions has mostly members from the exportome class of proteins, whereas, the other has genes from the variant surface antigens, as well as, some of the DnaJ domain containing proteins.

The two dense clusters are linked by mostly members of PHISTA, PHISTB, and DNAJ domain containing proteins along with a few FIKK kinases and hypothetical proteins. To analyze the relationships within the differentially regulated exportome genes and between the exportome genes and the variant surface antigens, the expression data from the relevant genes was mapped on to the PPI network.Four up-regulated exportome genes (PF10_0013, PF14_0239, PFL0600w and PF14_0744) were mapped in the network. The first three exportome genes just mentioned (PF10_0013, PF14_0239, PFL0600w) were seen to interact with multiple exportome proteins.

Hyp 12 (PF10_0013) interacts with as many as 50 other exportome proteins and is a hub gene. Its interacting partners include proteins containing PHISTa, PHISTb, and PHISTc domains, ring exported protein (REX 3 and 4) and parasite infected erythrocyte surface protein (PIESP). The PHIST domain containing proteins are specific to genus *Plasmodium* whereas the PHIST A proteins are specific to *P. falciparum*. REX 3 is a soluble exported protein with unknown function but are localized to locus associated with virulence (Spielmann, et al., 2006)

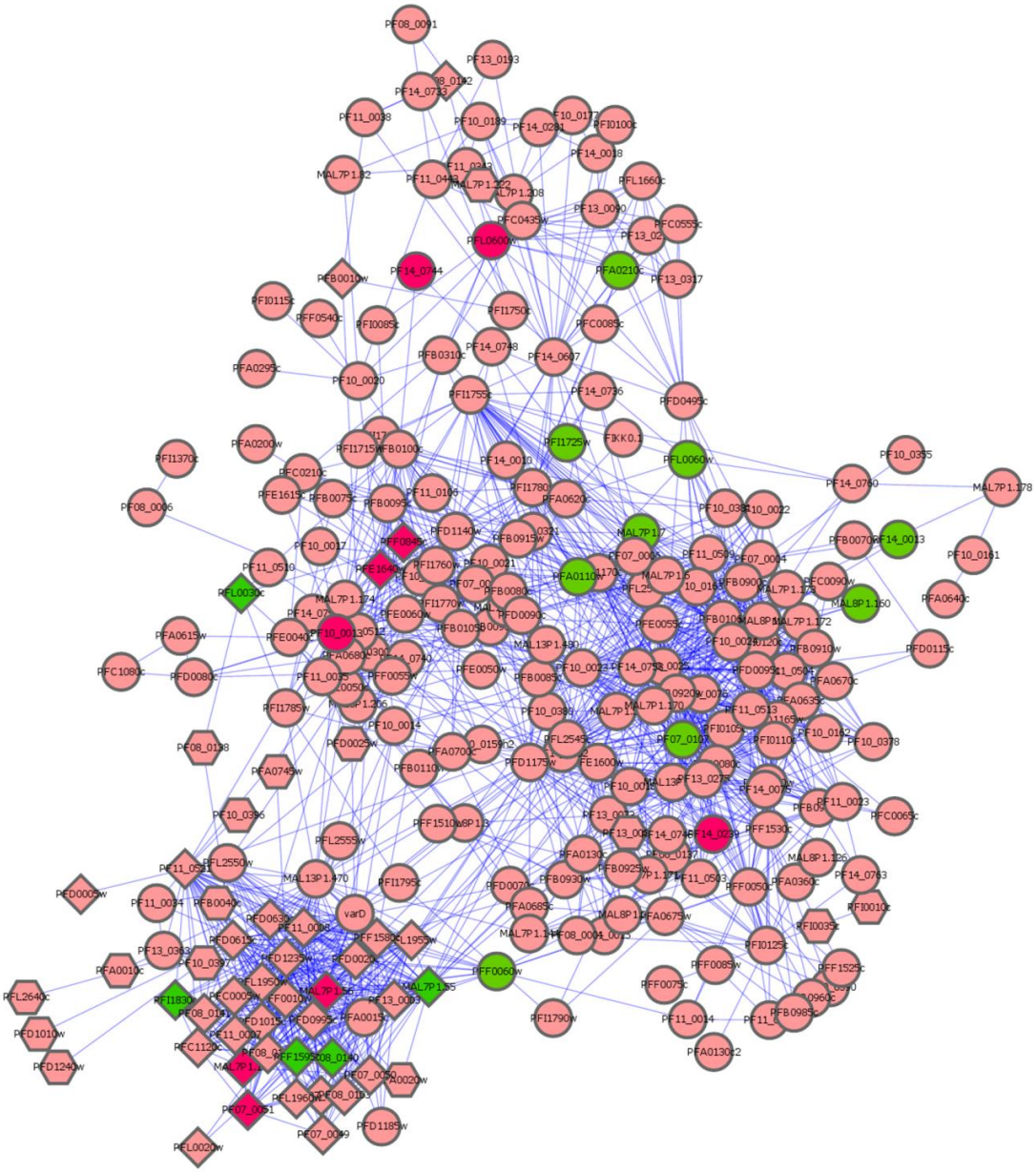


Figure 5.19. Protein-protein interaction network of exportome proteins including variable parasite surface antigens. Network was constructed by taking reported exportome proteins and variable surface antigens as an input list in StringDB V9.1(Franceschini et al., 2013). Figure shows only connected nodes in the network. Diamond shaped nodes represent PfEMP1, Hexagonal shaped nodes represent RIFINs, round shaped nodes represent exportome proteins, red color nodes represent up-regulated genes, green color nodes represent down-regulated genes in complicated isolates and pink color node represent stably expressed genes in both isolates (complicated and uncomplicated isolates.

The second gene (PF14_0239) interacts with 14 other exportome proteins which includes proteins like serine/threonine kinases (FIKK), Dnaj protein (PFB0920w), heat shock protein 40 (PFA0660w), type II, hyp proteins and PHIST a and b domain containing proteins. Interacting partners of this node have been predicted as performing important functions in the parasite. FIKK proteins are serine threonine kinases which have been proposed as phosphorylating the cytoskeletal proteins of infected host erythrocytes, thus altering the mechanical properties of infected RBCs (Nunes, et al., 2010). Altering the mechanical properties of infected RBCs may be critical for parasite survival during circulation, especially during conditions of disease perturbations. Another interacting partner is a heat shock protein 40 believed to perform cochaperone-like function in infected erythrocytes (Boddey and Cowman, 2013) which might be involved in chaperone-mediated protein translocation, folding and assembly (Botha, et al., 2007), more required in the complicated disease environment.

The third gene PFL0600w is a conserved *Plasmodium* protein with unknown function found to interact with 10 proteins, which includes five conserved *Plasmodium* protein with no known function, an ABC transporter encoding protein (MDR family protein), an erythrocyte membrane protein 3 encoding protein (PfEMP3)(PFB0095c), an ADP ribosylation factor (PF13_0090) and a *Plasmodium* exported hyp5 protein.

The protein-protein interaction network with the overlaid exportome and variant surface antigen transcriptome data was then subjected to MCL based clustering to find densely connected nodes. This identified 34 clusters of different sizes (**Figure 5.20**). The first eight clusters have been considered for analysis as genes in these clusters were seen to be highly connected with each other and were present in appreciable numbers. The variant surface antigen genes have mostly segregated into one cluster and exportome proteins into multiple separate clusters. However, a few clusters have both exportome proteins as well as variant surface antigens. Six of the eight clusters mentioned had differentially regulated genes (clusters 1-4, 6 and 8).

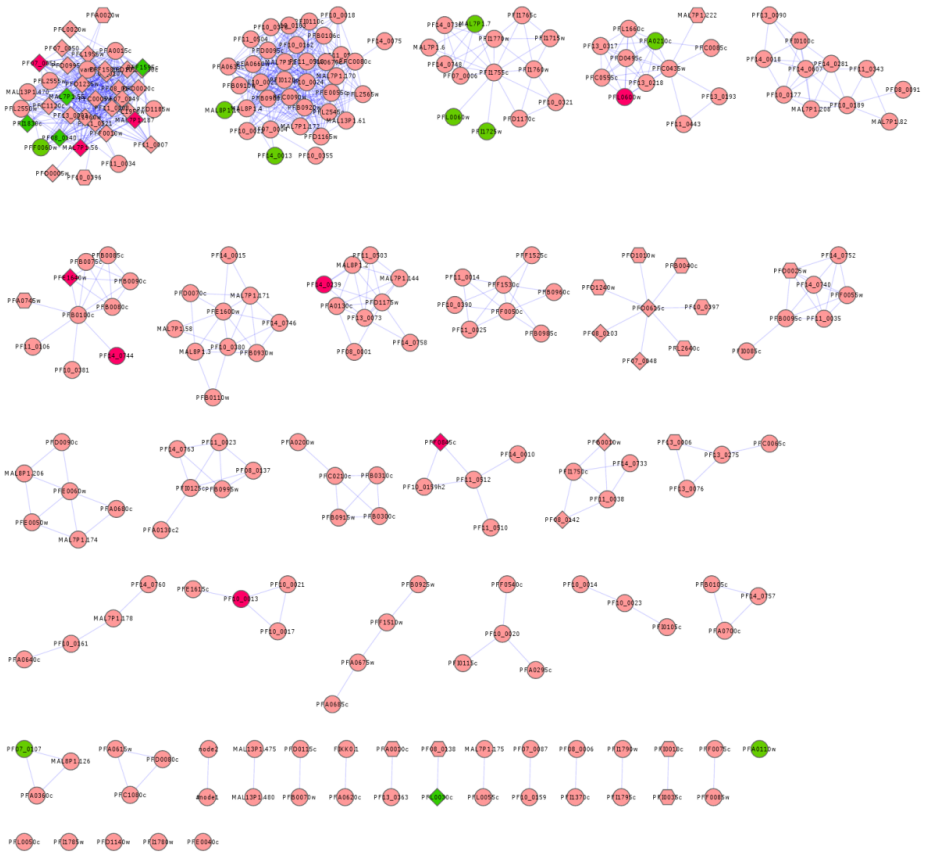


Figure 5.20. **Clusters of exportome network**. Clusters were generated by subjecting the whole network to MCL based clustering algorithm present in Cytoscape V 2.8.2 (Smoot et al., 2011). Figure shows 34 clusters of different sizes. Diamond shaped nodes represent PfEMP1, Hexagonal shaped nodes represent RIFINs, round shaped nodes represent exportome proteins, red color nodes represent up-regulated genes, green color nodes represent down-regulated genes in complicated isolates and pink color node represent stably expressed genes in both isolates (complicated and uncomplicated isolates).

Cluster 1 contains variant surface antigens with a few exportome proteins (**Figure 5.21A**). The later connects the VSA with other exportome proteins and includes three PHISTa domain containing proteins, two DnaJ domain containing proteins and one *Plasmodium falciparum* Maurer's cleft-2 transmembrane (PFMC-2TM) protein. Of these, the PFMC-2TM is down-regulated in the complicated isolates whereas the other five are not-differentially expressed at the level of the mRNA. There are four up-regulated PfEMP1 (*var* C and B groups), four down-regulated PfEMP1 (*var* B, B/C and B/A groups).

Cluster 2 contains 36 genes of which 2 were mapped as down-regulated (PF14_0013 and MAL8P1.160) (**Figure 5.21B**). Gene PF14_0013 encodes a DnaJ protein whereas MAL8P1.160 codes for a hyp protein. PF14_0013 interacts with two proteins, one of which is a protein containing PHISTc domain (MAL8P1.4) and the other an exported protein with unknown function. MAL8P1.160 interacts with four proteins of which two are PHISTc domain containing proteins the others a PF70 and a hyp2 protein. The second interacting partners of both the down-regulated genes cover the whole cluster. Members of this cluster include, FIKK kinases, a RESA

protein (PF11_0509), a RESA like protein with PHIST and DnaJ domains (PF10_0378), a merozoite surface protein, a plasmepsin protein (plasmepsin IV-PF14_0075), two type II HSP 40 and different class of hyp proteins.

Cluster 3 contains three down regulated genes linked through a common protein REX 3 (PFI1755c) (**Figure 5.21C**). The down-regulated genes are two exported proteins (PFL0060w and PFI1725w) with unknown function and a PHISTb domain containing protein. The cluster contains a hyp12 protein (MAL7P1.6), a REX-4 protein (PFI1760w), a sporozoite asparagine and threonine rich protein along with a few exported proteins with unknown function.

Each of the cluster numbers 4 and 8 contains one up-regulated gene (cluster 4- PFL0600w and cluster 8- PF14_0239) which along with its interacting partners have been described previously (**Figure 5.21D** and **5.21E**). Additionally, cluster number 4 contains a down-regulated gene which encodes for a conserved protein of unknown function. Whereas cluster 6 contains two up-regulated genes, of which one is a member of *var* A group (PFE1640w) and the other a *Plasmodium* exported protein with unknown function (PF14_0744). Both are connected through a knob associated histidine rich protein (KAHRP) (PFB0100c) (**Figure 5.21F**).

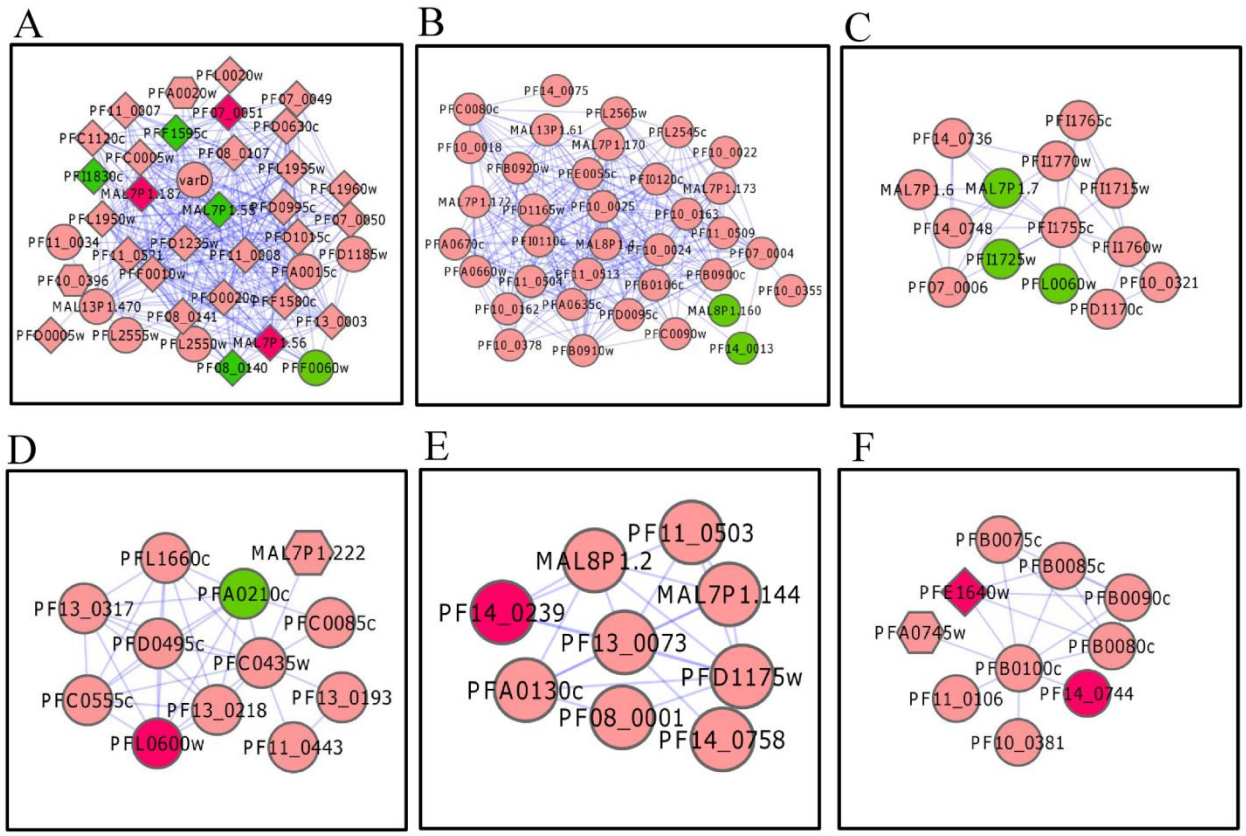


Figure 5.21. **Most significantly connected clusters of exportome proteins including variant surface antigens**. Top 6 clusters with mapped differentially regulated genes are shown. Diamond shaped nodes represent PfEMP1, hexagon shaped nodes represent RIFINs and round shaped nodes represent exportome proteins excluding variant surface antigens. Red nodes denote up-regulated genes, blue nodes denote down-regulated genes and pink nodes denote stable expressed genes. **A**: cluster 1; **B**: cluster 2, **C**: cluster 3, **D**: cluster 4, **E**: cluster 8 and **F**: cluster 6

5.2.6. **Real-time qPCR based validation of selected genes**

To confirm the microarray results, we examined five differentially regulated genes in addition to the *var* genes for validation using real-time qPCR. In the microarray hybridization, out of the five genes analyzed, three genes were up-regulated and two were down- regulated. Real-time qPCR data showed a similar expression pattern for all the five genes (**Figure 5.22A** and **Table 5.5**).

Group specific expression pattern of *var* genes was also confirmed through real-time qPCR experimentation. Group specific *var* primers (Rottmann et al., 2006) were utilized for analyzing the expression pattern of the var genes. Our microarray results show the overall up-regulation of *var* group B and C genes and down-regulation of *var* group A genes in PFC isolates (**Figure 5.22B**). Real-time qPCR experiments utilizing primer pair A2 which targets the conserved DBL1α region of all the 3D7 *var* genes and primer pair B1 which targets the conserved *ups* region of all the *var* B group genes confirmed the results of our microarray hybridization. The BC1 primer set designed to amplify a majority (17/22) of B group genes, 4 out of 13 C group genes, 1 out of 4 B/A genes and 2 out of 9 B/C group genes. The BC1 primer set detected up-regulation of RNA transcripts of *var* group B and C in PFC isolates.

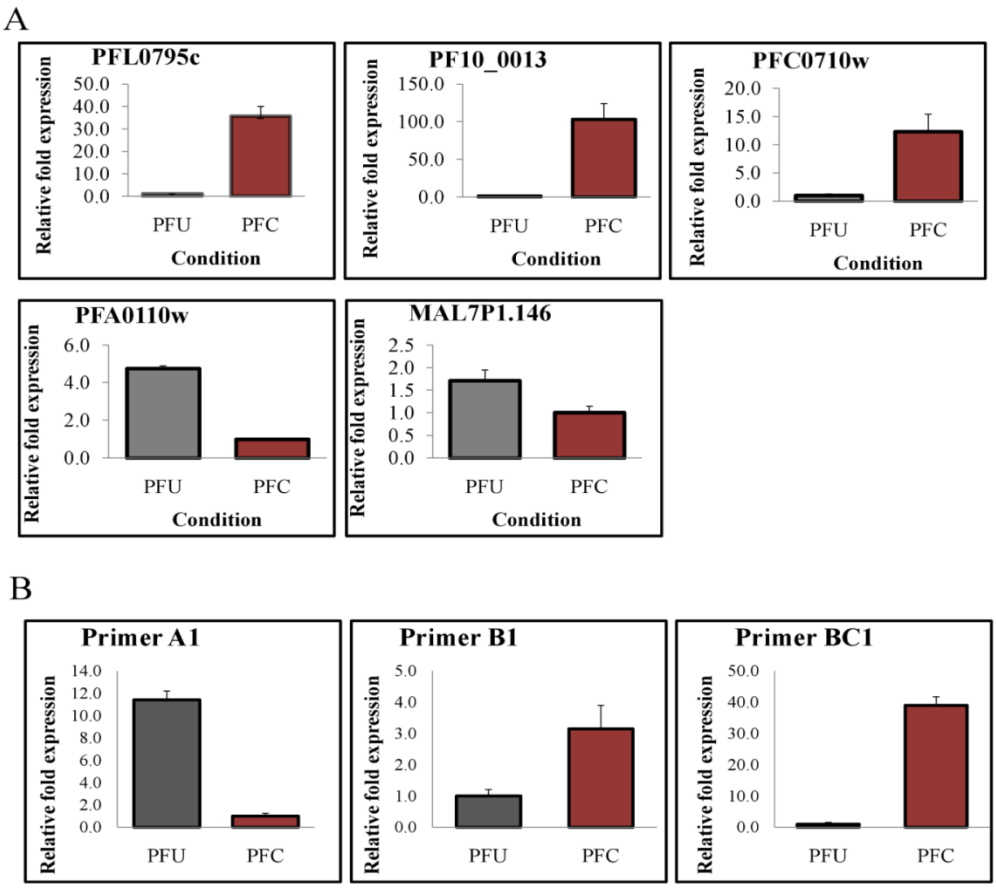


Figure 5.22. Quantitative real-time PCR based validation of few differentially regulated genes. Relative fold change expression was calculated using 2^{-ddCt} method. Seryl tRNA synthetase was used as a reference gene. Y-axis represent relative fold change expression (mean ± corrected S.D.; n=2) and X-axis represent conditions. Panel **A** shows the expression measurements of differentially regulated genes measured using gene specific primer; Panel **B** shows group specific expression pattern of *var* genes measured using group specific primer sets. PFU: Uncomplicated malaria isolates, PFC: Complicated malaria isolates.

Table 5.5. Comparison of microarray data with real-time qPCR data

Gene	Fold change expression in Microarray (Log_2 Fold Change)	Fold change expression in qPCR (Log_2)
PFL0795c	4.3	5.15
PF10_0013	4.6	6.69
PFC0710w	2.8	3.62
PFA0110w	-2.5	-3.75
MAL7P1.146	-3.8	-0.71

5.3. **Discussion**

A lot of effort has gone into understanding the molecular basis of complicated malaria pathogenesis in children. However, there is very scanty information from adult patients. These constitute a large proportion of the malaria infected population in areas of unstable malaria transmission. In addition, much of the *P. falciparum* biology has been understood through experiments utilizing parasites *in vitro* and information about the parasite biology *in vivo* is limited (Daily et al., 2005, Daily et al., 2007). Even these limited studies have identified the existence of differences in parasite biology *in vivo* in comparison to *in vitro* conditions (Daily et al., 2007). Lacunae exist in information about the parasite transcriptome in non-cerebral malaria related severity; specifically patients with symptoms like jaundice and/or renal failure. This study is based on the result of custom microarray hybridization experiments, which have highlighted certain differences in complicated disease states, part of which we are presenting in this paper.

To reduce partially the multiple variables naturally present in such a study, we utilized isolates which were collected from the same geographical region, from a single tertiary care hospital. All the isolates considered for this study had only ring stage parasites, as detected by microscopy. Transcriptome of all the samples showed a good positive Pearson correlation among themselves (0.79 - 0.95), which again confirms that the isolates contain parasites in similar developmental stages, and under analogous host pressure due to disease conditions. Presence of sexual stages in any of the isolates would have decreased the correlation with other isolates (Daily et al., 2005). Genes relating to most of the metabolic pathways did not differ in expression profile between the two clinical conditions. This suggests that pathways contributing to many basic processes of parasite biology behave quite similarly in both the clinical conditions considered (PFC and PFU). Gene Ontology analysis of up-regulated genes identified biological processes/molecular functions which are affected, these include regulation of intracellular pH, electron transport chain, cytochrome C oxidase activity, DNA packaging, chromosome organization, cell cycle, generation of precursor and metabolite energy. All the processes mentioned are directly dependent on the host physiological environment. Gene encoding Membrane transporters like PfAQP protein (PF11_0338), predicted to protect the parasite from osmotic shock (Hansen et al., 2002) and another putative drug transporter (PFE0785c), were also up-regulated in PFC isolates,

suggesting that the parasites in these complications could have altered requirements for transport across the parasite membranes. Up-regulation of genes like PfAQP (PF11_0338), oxidoreductase (PFC0475c) and decarboxylase (PFI0950w), whose products probably involved in sensing external environment and nutrient utilization which may provide an advantage to the parasite for its survival.

Systems Biology based network enrichment analysis has identified genes involved in processes like RNA metabolic process, gene expression, nitrogen compound metabolic process and RNA splicing. GO enrichment analysis has also identified chromosome organization process which has a direct influence on gene expression. Our observation is supported by others who have published that parasites *in vivo* and showed three distinct physiological states in one of these states, genes encoding histones and chromatin modifiers have shown differential expression pattern. This has been suggested as playing a critical role in deciding distinct transcriptional states (Daily, et al., 2007). Our data also shows similar expression profiles, suggesting enhanced transcriptional activity in the disease condition.

Parasites inside the RBCs access the host environmental signals by exporting proteins (exportome) which remodel the iRBCs (Cooke, et al., 2004, Lanzer, et al., 2006, Marti, et al., 2005 and Przyborski, et al., 2003). The impact of exported proteins on iRBCs is, that it changes many properties of iRBCs including permeability, rigidity and adherence. A sub-set of exportome protein encoding genes was identified to be differentially regulated in PFC isolates. Notably, up-regulation of a PHISTc domain encoding gene and down regulation of three PHISTb domain containing protein encoding genes suggests that although these proteins contains PHIST domain, both the PHIST sub-group perform different functions. It is in line with the prediction by a study where, PHIST domain containing proteins were reported for the first time (Sargeant, et al., 2006). A similar observation was made for hyp class of proteins in which genes encoding proteins with DnaJ domain were found to be down-regulated, one of these is a RESA protein (also a PHISTb subfamily member) whose disruption *in vitro* has been associated with decrease in iRBCs rigidity but interestingly, associated with slight increase in cytoadherence to CD36 under flow conditions (Cooke, et al., 2002, Maier, et al., 2008, Silva, et al., 2005). It would be too early to speculate about the effect of these down-regulated Dnaj domain containing proteins on parasite induced remodeling of iRBCs. Moreover, most of the differentially regulated

exported proteins are hypothetical in nature, with no known function. This generates difficulties in predicting anything about their specific role(s) in changing permeability and rigidity of iRBCs and its effect on parasite biology *in vivo*.

Complicated malaria disease outcome has been associated with up-regulation of specific groups of *var* gene super family members. This has clearly been demonstrated in case of PAM, where *var2csa* gene was found to be over expressed (Salanti, et al., 2004). In case of pediatric patients with complicated disease (specifically patients with cerebral malaria), majority of evidence suggests over expression of members of *var* A and *var* B/A groups. However, there are just a few reports which show over expression of *var* group B members (Kaestli, et al., 2006, Rottmann, et al., 2006). In this study, we have observed over expression of members of *var* B and *var* C groups in adult patients with complicated malaria. Each isolate (PFC) showed up-regulation of at least one member of the *var* B and/or *var* C groups. The array utilized in this study contains probes representing *var* genes of three laboratory strains (3D7, HB3 and IT4). Analysis identified a similar trend in expression of probes representing *var* genes of all the three strains. Each group of *var* genes has specific binding properties for host receptors (Janes, et al., 2011). PfEMP1 variants belonging to *var* group B, B/C, and C bind readily to CD 36. CD36 is a host receptor found to express in organs like liver and spleen but not in brain (Serghides, et al., 2003). This suggests that parasites expressing PfEMP1 variants of *var* B and *var* C group on the surface of iRBCs could bind to the hepatic microvasculature expressing the host receptor CD36. Interestingly, preliminary studies investigating perturbations in the host (PBMC) transcriptome profile under such disease conditions, (in a separate study by our group not part of this thesis) show good amount of CD36 transcripts, even though there is no up-regulation seen. So, the possibilities of interaction of *var* B and C with CD36 in complicated cases remain quite possible. Moreover, CD36 is also known to express on the surface of platelets which in turn binds to iRBCs to form clumps (Serghides, et al., 2003). These clumps might further increase malaria related severity by obstructing microvasculature of various organs. Although, primarily a mechanism for surviving circulatory disruption in spleen or other organs, an unsavory outcome is blocking of blood supply in the organs, leading to disease complications. As far as we are aware, this is the first study which shows the up-regulation of specific *var* group (B and C) members in adult patients with complicated (non-cerebral) disease symptoms. Importantly, significant amounts of transcripts have been detected for all the *var* groups in the isolates. The marked up-

regulation of the *var* groups B and C in PFC isolates suggests a link with the complicated disease conditions.

Data from studies suggest that type A RIFINs are exported through the PVM and display itself on the surface of iRBCs (Petter, et al., 2007), while type B RIFINs are not part of the exportome. *rifins* are shown to be clonally variant and antibodies from hyper-immune sera can recognize them suggesting that these proteins are immunogenic too (Abdel-Latif, et al., 2003). In this study, we have found that the type A *rifins* were up-regulated in all PFC isolates. Transcripts for only six members of type B *rifins* were detected and found to be up-regulated in two cases. Apart from this, in each PFC isolate, there was up-regulation of one to multiple members of *rifins* A. Over expression of type A *rifins* in PFC isolates suggests their possible involvement in enhancing the antigenicity which may provide mechanisms for host immune evasion. The *var* group genes encoding PfEMP1 are known to show mutually exclusive patterns of expression unlike *rifins* (Petter, et al., 2007). Presence of multiple RIFIN proteins on the iRBC surface increases the complexity of the molecules against which an immune response may be mounted. The molecules involved in the host response against such proteins which are up-regulated in the complicated disease state, may provide a clue to some of the factors initiating such a complication. Transcripts were detected from as many as 14 *stevor* genes. This observation supports previously published data that parasites *in vivo* express more number of *stevors* compared to *in vitro* conditions (Daily, et al., 2005).

The data from the MCL clustering of the PPi network and investigation of the clusters mentioned in the result section shows the connectivity between differentially regulated and non-differentially regulated *var* and exportome molecules. The clusters are text mining and co-expression based.

Cluster 1 shows a good connectivity between the variant surface antigens (VSA) and different members of the exportome. It is interesting to note that the interaction between PHIST a domain containing proteins (non-differentially regulated) and DnaJ domain containing proteins (non-differentially regulated) with up-regulated or down-regulated VSA genes. The interactions between up-regulated VSA and non-differentially regulated PHISTa proteins suggest the stable requirement of PHISTa like protein and probable involvement in surface expression of the up-regulated VSA. Both the *var* group proteins and the PHISTa proteins are specific for

Plasmodium falciparum. There are also clusters viz. (Cluster 2), which show interactions between stably expressed PHISTc domain proteins and exportome proteins like Plasmepsin IV protein (hemoglobin degradation/utilization) suggesting that proteins coding for basic fuctions of the parasite remain stably expressed even in the disease state.

Text mining naturally restricts part of the connectivity to data from published literature. In cluster number-6 a member of the *var* A group (PFE1640w) is connected to a *Plasmodium* exported protein with unknown function (PF14_0744) through KAHRP (PFB0100c). KAHRP is required for knob formation; this provides a platform for the presentation of the PfEMP1 encoded by the *var* genes (Oh, et al., 2000 and Waller, et al., 1999). Interestingly, in cluster 1which shows the connectivity between the variant surface antigens (VSA) and other exportome proteins, some of the VSA were up-regulated (*var* B&C groups) and connected to the exportome proteins. There was no connectivity of *var* genes with KAHRP. The data presented in this paper shows the up-regulation of the *var* B&C group genes in disease complications. Considering, that these also encode for PfEMP1 and should follow similar presentation on the iRBC surface as that from *var* A group, this could be a cause of concern. A possible explanation for the same could be the fact that only *var* A (*var*1CSA) and *var* 2CSA have been reported, as being expressed in association with KAHRP (Janes, et al., 2011 and Laishram, et al., 2012) and thus lack of association of *var* group B and *var* group C with KHARP is an outcome of data mining based approach.

In summary, this study shows the up-regulation of Variant Surface Antigens like the *var* B &C groups, members of the *rifins* and *stevor* and some members of the exportome in *P. falciparum* isolates from adult patients showing complicated disease manifested as hepatic and renal dysfunction. Although the function of most of the exportome proteins is unknown till date, our text mining based co-expression network clearly shows a connection between *var* B&C group and exportome family members in adult patients with complicated disease.

6.1. Introduction

Complete genome sequencing of hundreds of pathogenic and model organisms has provided enormous amounts of data that are now being analyzed to understand the nature of genes and proteins and their interactions in a cellular context (Jansen, et al., 2003). This part of understanding benefitted from the emergence and utilization of high throughput techniques which includes, but not limited to: gene expression, yeast two hybrid assay, RNA interference and mass spectrometry (Reed, et al., 2006). The data generated from the above mentioned techniques are huge and requires implementation of computational biology and the emerging disciplines of systems biology approaches to handle and interpret the data (Kitano, 2002 and Nurse, 2003). Gene expression microarray data is a form of high-throughput data which measures relative mRNA levels of thousands of genes in biological samples (Lee, et al., 2004). Most of the time microarray data are analyzed to find out differentially regulated genes in different disease states or developmental stages of an organism. However, there are other ways by which the same can be utilized to unearth many hidden information. One such example is generation of systems network through co-expression measurements of genes. Co-expression networks provide information about the genes that have similar expression pattern across a set of samples. Set of genes that show similar expression patterns are hypothesized to have a functional relationship (Freeman, et al., 2007). This information helps in predicting functions of many hypothetical proteins if functions of other proteins in the genes set are known. Many a time integration of differentially regulated genes information to such network helps in identifying active biological processes during a disease conditions in a more biologically meaningful way.

Gene co-expression based networks have successfully been utilized in identifying genes, biological processes and pathways that are involved in certain disease conditions in multiple organisms (Ghazalpour, et al., 2006, De Jong, et al., 2012., Miller, et al., 2010). In *P. falciparum*, many approaches have been applied to create and analyze networks with a major emphasis on predicting the functions of hypothetical proteins (Hu, et al., 2009, , Date, et al., 2006, Wuchty, et al., 2007, Zhou, et al., 2008). However, all of these have either employed genome wide datasets (transcriptomics, proteomics, yeast two hybrid) derived from laboratory strains in culture condition and/or have utilized computationally derived datasets.

In this chapter, we present the data from an *in vivo* parasite derived transcriptome dataset utilized to generate a weighted gene co-expression signed network. Highly connected genes have been identified as modules. Differentially regulated genes have been identified in *P. falciparum* clinical isolates showing complicated (PFC) compared to uncomplicated (PFU). These have been further mapped to genes in the modules in order to detect modules associated with disease status. Functional analysis of differentially regulated modules has been performed, enabling the identification of parasite biological processes connected with malaria related disease severity. Additionally, network concepts have been employed on each module to detect highly connected hub genes (differentially regulated or stably expressed) which might serve as potential candidates for intervention strategies.

6.2. Results and Discussion

6.2.1. Construction of a co-expression based systems network from transcriptome data of *Plasmodium falciparum* clinical isolates.

A weighted co-expression based *P. falciparum* systems network was constructed by utilizing the expression profiles of 21 clinical isolates from patients with clinical symptoms ranging from uncomplicated to different types of complicated malaria (**Table 6.1**). The framework for the construction of co-expression network was adopted from a previously published work **(**Zhang and Horvath, 2005). The detailed methodology for the same has already been summarized in Chapter 2 (Materials and Methods). An expression matrix was constructed by including expression measurements of only those genes whose corresponding probes were detected in at least 30% of the samples. Subsequently, hierarchical clustering of all the 21 samples was performed (**Figure 6.1**). This analysis resulted in identification of an outlier sample (PFC 14) which was not considered in further steps to construct the systems network. For the 20 samples which were considered for subsequent analysis, Pearson correlation coefficient was calculated for the expression values of genes pairs across microarray samples to construct a co-expression similarity matrix with pair-wise Pearson correlation measures. Using a soft thresholding procedure, the co-expression similarity matrix was then transformed into an adjacency matrix of connection strengths. This resulted in a weighted network (Zhang and Horvath, 2005). As mentioned previously, this adjacency matrix was generated by raising the co-expression similarity to a power $\beta = 16$ (**Figure 6.2**). This value was chosen as threshold based approximate scale free topology criteria for this dataset ($R^2 = 0.78$) (Zhang and Horvath, 2005).

Table 6.1. Clinical Characteristics of *Plasmodium falciparum* infected patient's samples.

Sl No	Patient ID	Clinical Presentation	Sl No	Patient ID	Clinical Presentation
1	PFU-03	Uncomplicated	12	PFC-14	CM, RF,T
2	PFU-04	Uncomplicated	13	PFC-15	J, T
3	PFU-05	Uncomplicated	14	PFC-17	J, RF
4	PFU-06	Uncomplicated	15	PFC-18	J, RF, T
5	PFU-07	Uncomplicated	16	PFC-19	J, RF
6	PFC-03	J, RF,T	17	PFC-21	RF, T
7	PFC-05	J, RF, T	18	PFC-22	J, RF, T & A
8	PFC-06	J	19	PFC-23	J
9	PFC-10	J, A	20	PFC-25	J, T
10	PFC-11	J, A	21	PFC-29	CM, A
11	PFC-12	J, A			

Abbreviations: CM, cerebral malaria; J, jaundice; RF, renal failure; A, anemia; SA, severe anemia; T, thrombocytopenia;

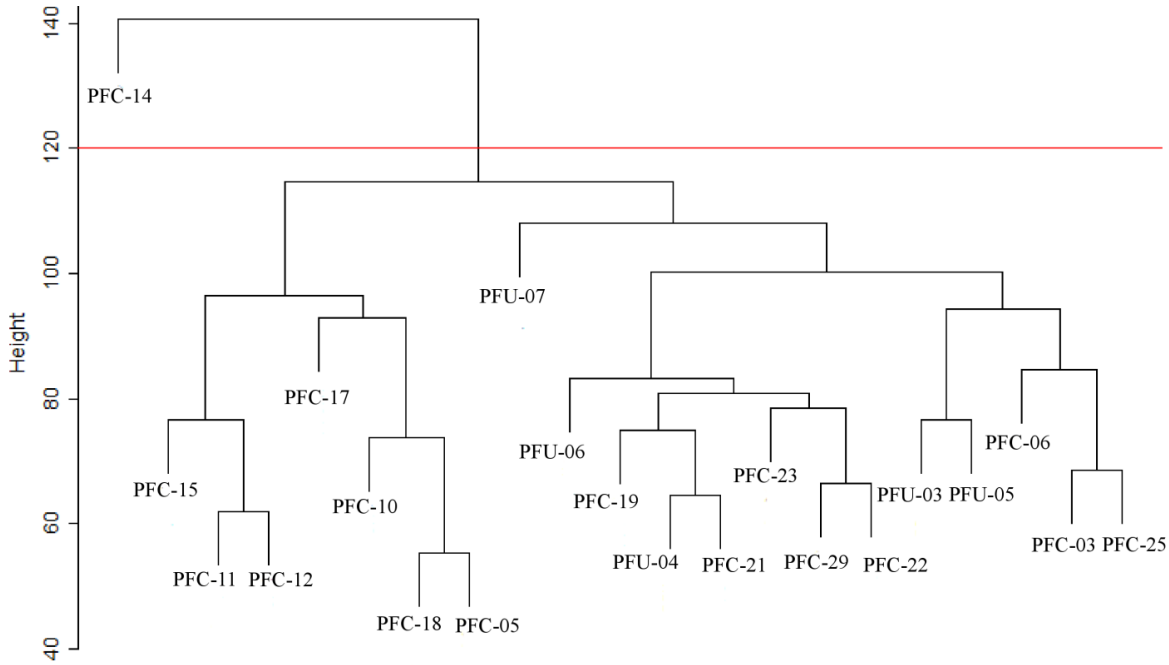


Figure 6.1. Clustering of samples based on their Elucidean distance to detect outlier(s). Sample PFC 14 was found to be an outlier and was removed from the analysis.

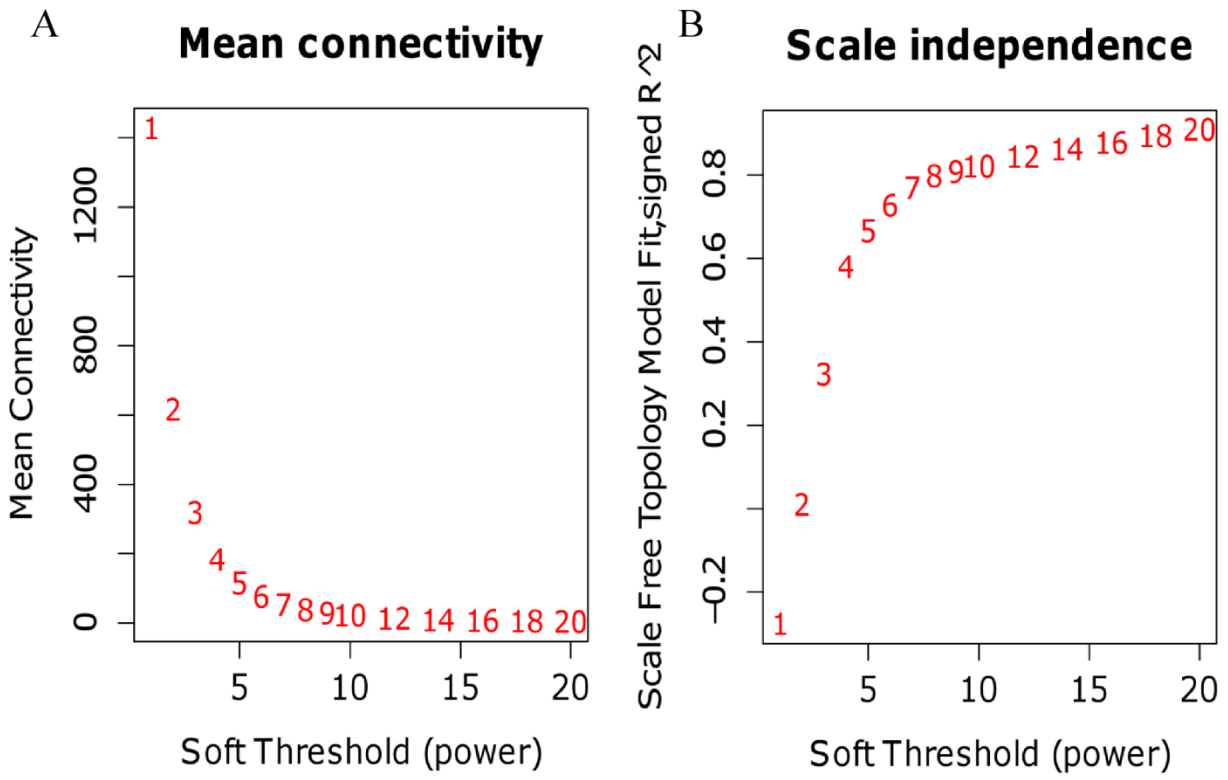


Figure 6.2. Various soft-thresholidng powers and corresponding network topology. This analysis provides the option to choose best soft-thresholding power to obtain a near scale free network. The best soft-thresholding power is the power where the curve starts to flatten. A, Graph represents the mean connectivity (y-axis) as a function of the soft-thresholding power; B, represents the scale free fit index (y-axis) as the function of soft-thresholding power.

This adjacency matrix with connection strengths was then transformed into a topological overlap matrix (TOM) by calculating the topological overlap (TO). The later measures the relative interrelatedness between two genes and their relationship with other genes in the network (Zhang and Horvath, 2005 and Miller, et al., 2010). A signed network was thus generated where all negative or positive values of correlation were preserved.

The expression values of 4857 genes across 20 samples were considered for the network construction. Groups of genes showing very similar pattern of connection strengths (modules) were then identified using average linkage hierarchical clustering based on TO measures. This resulted in identification of 27 modules and each module was designated by a color as per WGCNA convention (**Figure 6.3**). The size of modules in terms of number of genes ranged from 31 (White) to 586 (Turquoise) (**Table 6.2**, **Supplementary Table S6.1**). Subsequently, an eigenegne network was constructed to evaluate the relationship among modules. The eigengene is defined as a hypothetical representative gene of a module, whose expression profile is representative of majority of genes present in the module (Langfelder and Horvath, 2007). The relationship between 27 different modules is shown in the form of a cluster tree dendrogram in the **Figure 6.4**. By this process, 3 major clusters of modules with high inter modular connectivity were identified. The largest cluster contains 7 modules (Dark Turquoise, Dark Red, Green, Dark Green, Dark Orange, Tan and Royal Blue) and the other two clusters contain 6 modules each (Yellow, Blue, Light Cyan, Cyan, Pink, Light Green in one cluster and Salmon, Brown, Light Yellow, Midnight Blue, Magenta and Dark Grey in another cluster).

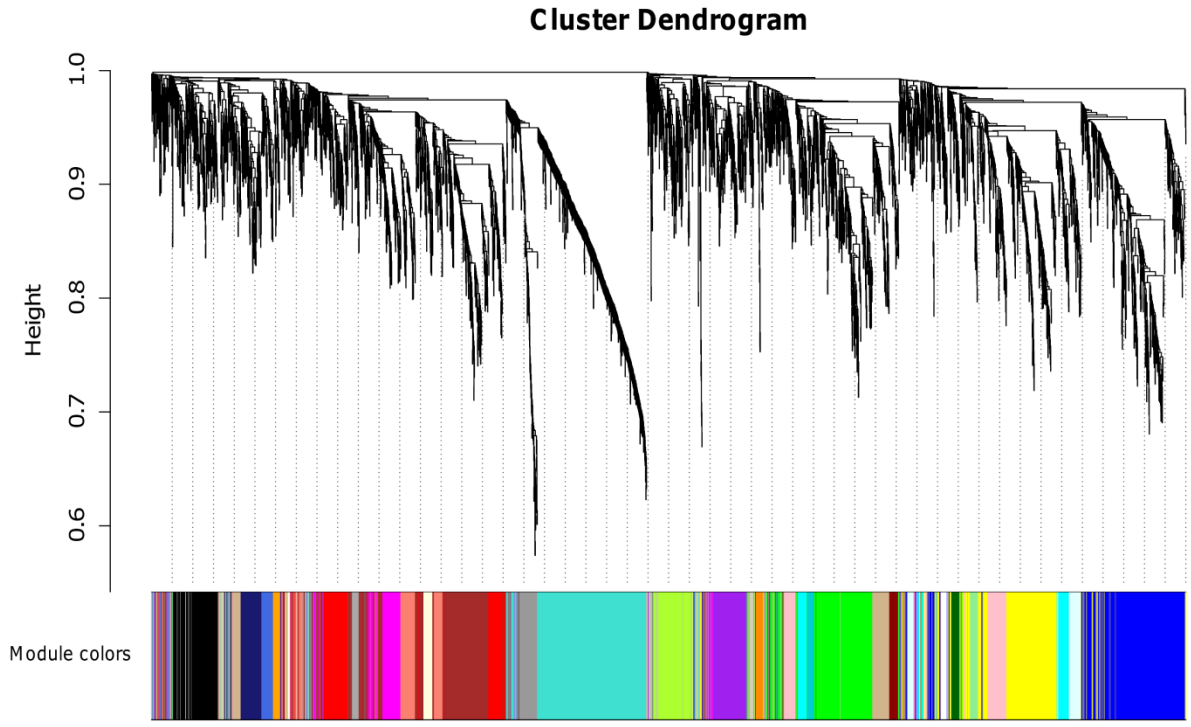


Figure 6.3. Hierarchical clustering dendrogram of genes in the network, partitioned into distinct modules. Each module was assigned with distinct color and represented in the horizontal bar below the dendrogram. Network was constructed using expression profiles of 4857 genes from 20 *P. falciparum* clinical isolates. The dendrogram was produced using average linkage hierarchical clustering taking topological overlap as a distance measure.

Table 6.2. Network modules identified after clustering analysis

Number	Name of the module	Total node (genes)
1	Grey 60	100
2	Cyan	136
3	Dark Orange	32
4	Dark Green	41
5	Orange	33
6	Light yellow	64
7	Brown	461
8	Turquoise	586
9	Light Cyan	103
10	Blue	523
11	Green	354
12	Dark Turquoise	40
13	Salmon	164
14	Purple	185
15	Pink	190
16	Yellow	426
17	Magenta	189
18	Midnight Blue	118
19	Black	199
20	White	31
21	Dark Grey	38
22	Dark Red	53
23	Royal Blue	59
24	Green Yellow	176
25	Red	210
26	Tan	164
27	Light Green	68

Table represents details of 27 modules detected during network analysis. Each module was assigned an arbitrary color based on WGCNA convention.

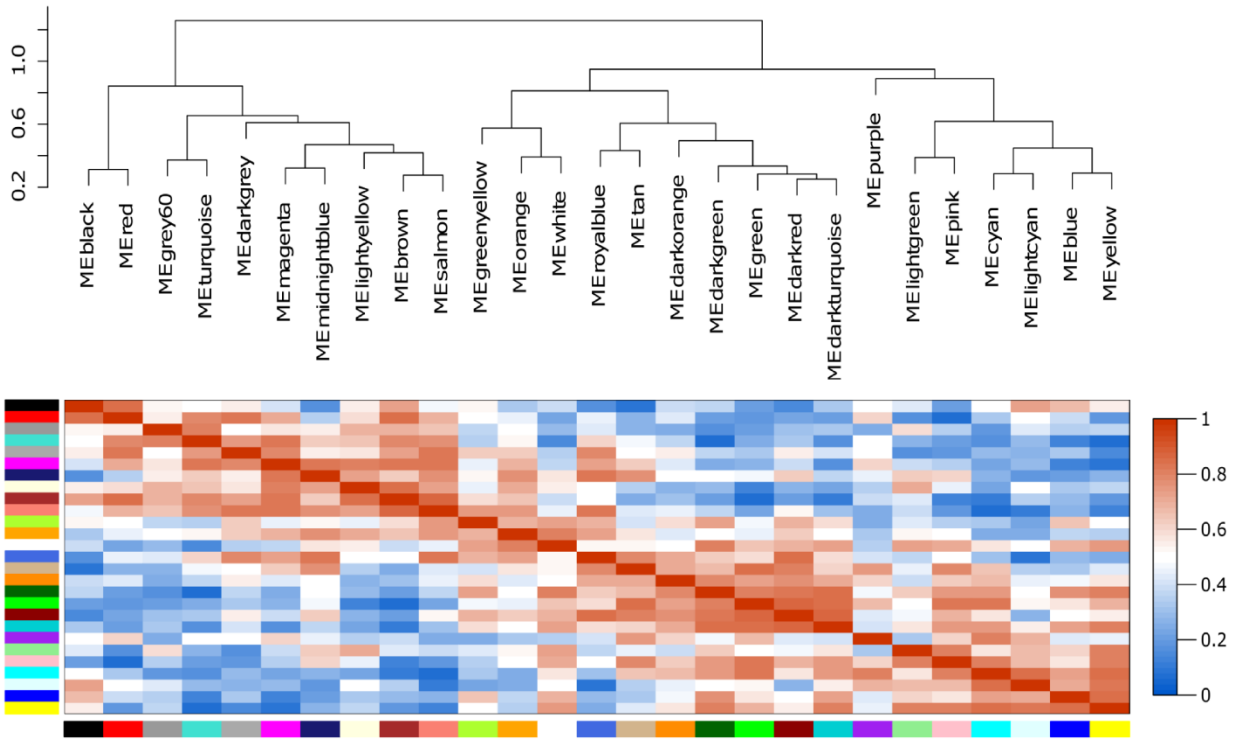


Figure 6.4. Eigengene network representing relationships between modules. Each module is represented by an eigengene (ME; Module Eigengene). A) Shows a hierarchical clustering tree of eigengene B) Shows a heatmap representing the eigengene adjacency.

6.2.2. Functional enrichment of modules and its biological significance

The generated modules were further analyzed by performing gene ontology (GO) enrichment analysis on genes in each module to ascertain the biological relevance of the same. This was done by using the online gene functional classification tool present in Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang Da, et al., 2009a & b). All the three GO categories; Biological Process: BP, Molecular Function: MF and Cellular Component: CC were seen as the result of this process. Most of the modules enriched to distinct GO terms, although a few modules shared GO terms (**Supplementary Table 6.2**).

The modules which enriched to Biological Process categories were Black (macro molecular Complex assembly, $P = 0.001$), Blue (oxidation reduction, $P = 0.0001$), Cyan (translation, $P = 0.02$), Dark Grey (exocytosis, $P = 0.034$), Green (mitochondrial transport, $P < 0.0001$), grey 60 (pathogenesis, $P = 0.019$), Light Cyan (ubiquitin dependent protein catabolic process, $P = 0.001$), Pink (chromosome organization, $P = 0.0002$), Purple (translation, $P < 0.0001$), Red (ion homeostasis, $P = 0.002$), Salmon (non-coding RNA metabolic process, $P = 0.0004$), Turquoise (pathogenesis, $P = 0.0018$) and Yellow (translation, $P = 0.0001$). A few modules shared the same GO terms. For example, Yellow, Purple and Cyan modules were enriched to same GO term i.e. translation. However, though the yellow and purple modules enriched to the same GO term, they are member of separate clusters and also show a lesser degree of correlation (< 0.05). This could indicate that the functional significance of the gene expression data in these modules could be different, though related to the process of translation.

This analysis demonstrates that the weighted gene co-expression network analysis (WGCNA) based network generation has the ability to segregate genes into distinct modules based on their expression patterns and shared biologically meaningful functions. Only the most significantly enriched Biological Process GO terms based on P values have been mentioned. There were modules which enriched to multiple GO terms of a particular GO category (**Supplementary Table 6.2**).

6.2.3. **Differentially regulated genes in two severe malaria conditions**

In order to find out differentially regulated genes in *P. falciparum* isolates from patients with non-cerebral malaria related complicated symptoms, the whole genome expression profile of the same was compared with expression data from parasites isolated from patients with uncomplicated malaria. This resulted in the identification of differentially regulated genes which could be further divided into two sets based on the patient's disease symptoms (**Table 6.3**). The first set contains 331 differentially regulated genes (JR-diffset) in *P. falciparum* isolates from patients with jaundice and renal failure as major symptoms (**Figure 6.5A**, **Supplementary Table 6.3**). The second set includes 178 differentially regulated genes (J-diffset) in *P. falciparum* isolates from patients with jaundice as main symptom (**Figure 6.5B**, **Supplementary Table 6.4**). All the differentially expressed genes were identified based on a volcano plot after applying a cut-off value of ≥ 2 fold over or under expression in complicated malaria samples compared to uncomplicated malaria at $P \leq 0.05$ (moderated t test). Further, both the gene sets were compared to identify common genes between them and unique genes present in each of the gene set (**Figure 6.6** and **Supplementary Table 6.5**). There were 153 genes which were common between the two gene sets. On the other hand 227 genes were unique to the JR-diffset and 178 genes were unique to J-diffset (**Supplementary Table 6.5**). Gene enrichment analysis of both differentially regulated genes sets were performed using functional annotation tool present in DAVID Bioinformatics Resources to obtain an idea about the biological functionality (Huang Da, et al., 2009a & b).

Table 6.3. Comparison of clinical isolate groups used for identifying differentially regulated genes.

Group 1 (n = 5)	Group 2 (n = 7)	Group 3 (n =6)
PFU-03	PFC-03	PFC-06
PFU-04	PFC-05	PFC-11
PFU-05	PFC-10	PFC-12
PFU-06	PFC-18	PFC-15
PFU-07	PFC-19	PFC-23
	PFC-21	PFC-25
	PFC-22	

Group 1 contains uncomplicated isolates, Group 2 contains complicated isolates isolated from patients with jaundice and renal failure as complication and Group 3 contains complicated isolates isolated from patients with jaundice alone as the main complication. Group 2 and Group 1 was compared to find out differentially regulated genes (JR-diffset) and Group 3 and Group 1 was compared to find out differentially regulated genes (J-diffset).

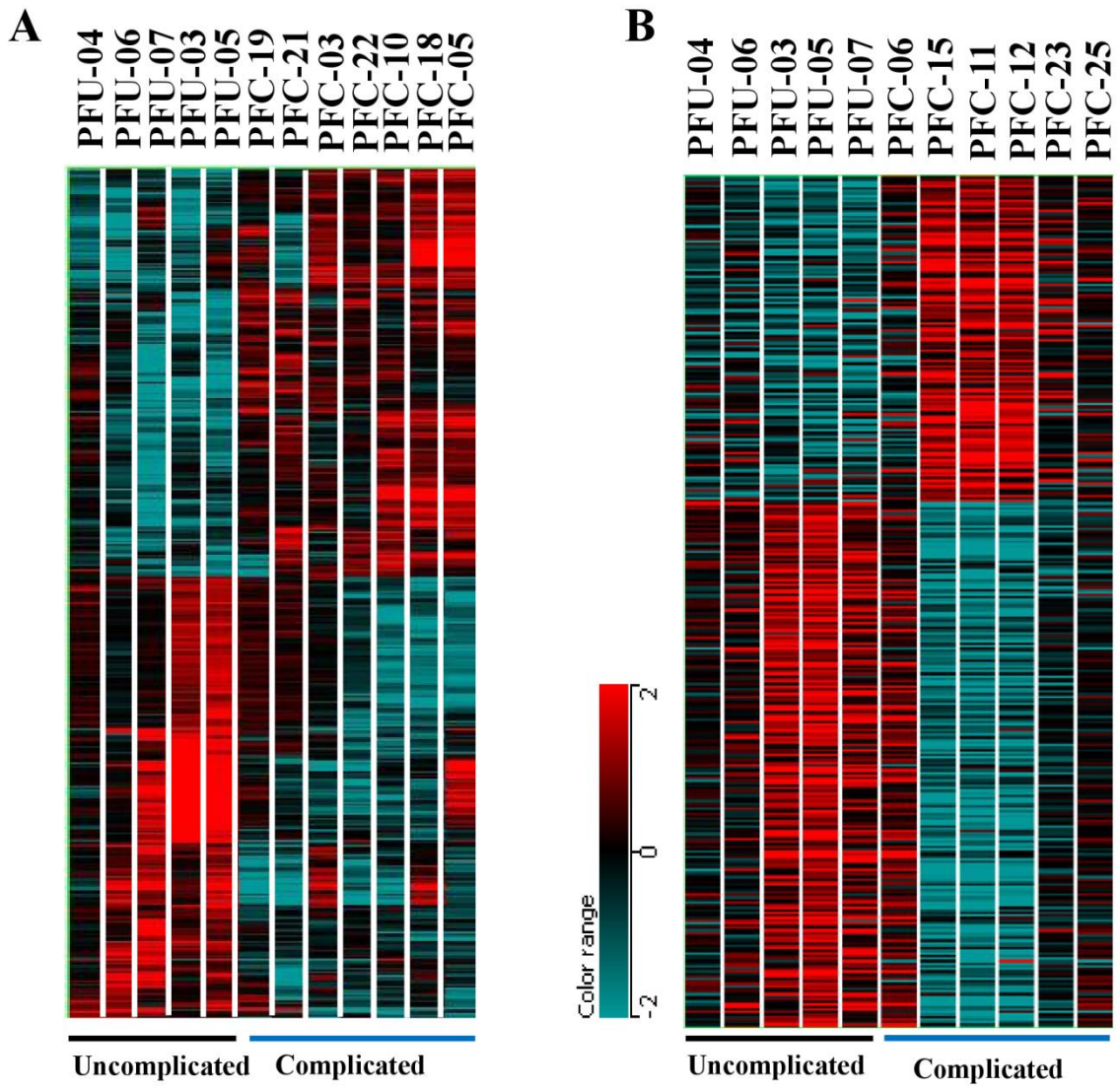


Figure 6.5. Differentially regulated genes in both gene sets. Panel **A)** Heatmap shows differentially regulated genes (JR-diffset) of *P. falciparum* isolated from patients with jaundice and renal failure compared to uncomplicated malaria. Panel **B)** Heatmap shows differentially regulated genes (J-diffset) of *P. falciparum* isolated from patients with jaundice compared to uncomplicated malaria. Genes were sorted based on Pearson centered distance measure of their expression measurement. All the genes were with $P \leq 0.05$.

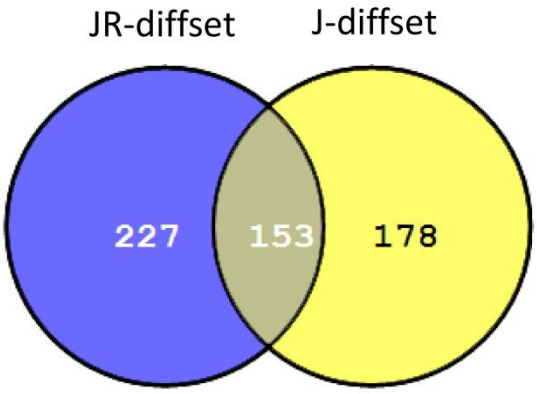


Figure 6.6. Comparison between two differentially regulated gene sets. Number in Venn diagram represents number of genes common and unique to both the gene sets. JR-diffset contains list of differentially regulated genes in parasite isolates from patients with jaundice and renal failure as symptoms and J-diffset contains list of differentially regulated genes in parasite isolates from patients with jaundice as a main symptom.

6.2.4. Complicated disease specific module identification

The network created was subsequently analyzed to identify functionally related differentially regulated genes and to associate them with the disease status based on the similarity in expression pattern. To investigate the enrichment of modules to differentially regulated genes, 558 differentially regulated genes were mapped to the same. This included 153 genes common to JR and J diffsets, 227 genes specific to the JR-diffset and 178 genes specific to the J-diffset. The percentage of nodes (genes) mapped as differentially regulated in modules range from 54% (Grey 60) to 0.47% (Red). However, the light green module did not have any such genes. Modules with at least 10% of their nodes mapped as differentially regulated were considered for further analysis (**Table 6.4 &Figure 6.7**). Differentially regulated nodes (genes) in each module were segregated based on their regulation status (i.e. up-regulated or down-regulated) and group to which they belong to i.e. common to JR and J diffsets, specific to JR-diffset or specific to J-diffset. A total of 12 modules were enriched to differentially regulated genes of which 7 modules (Cyan, Dark Orange, Dark Green, Light Cyan, Blue, Green, Dark Turquoise) were enriched with up-regulated genes and 5 modules were enriched with down-regulated genes (Grey 60, Orange, Light Yellow, Brown, Turquoise) (**Figure 6.7** and **Supplementary Table S6.1**).

Table 6.4. Modules with at least 10% of their genes mapped as differentially regulated

Number	Name of the module	Total nodes (genes)	Percentage of genes mapped as differentially regulated
1	Grey 60	100	54.00
2	Cyan	136	39.70
3	Dark Orange	32	37.50
4	Dark Green	41	31.70
5	Orange	33	21.21
6	Light yellow	64	18.75
7	Brown	461	17.13
8	Turquoise	586	16.89
9	Light Cyan	103	14.56
10	Blue	523	11.66
11	Green	354	10.16
12	Dark Turquoise	40	10.00

In order to gain further insight into the functional significance of each module enriched with up-regulated genes, functional enrichment analysis of modules was revisited (**Table 6.5** and **Supplementary Table S6.2**). Gene enrichment analysis of Green module shows that gene in this module were enriched to Biological process GO terms like mitochondrial transport, intracellular transport and oxidation reduction (P = 0.073). Modules Dark Orange and Dark Green did not enriched to any Biological Process GO term; however, our previous analysis of inter-modular correlation based on eigenegene relationship shows that all the three modules are present in the same cluster and showed a highly positive correlation (**Figure 6.4**). All the genes from these three modules were then made into a single gene list and functional enrichment analysis performed. Two additional Biological Process GO terms were seen to be enriched i.e. Peptide metabolic process and purine nucleotide metabolic process (ATP biosynthetic process) compared to the enriched GO terms only from the Green module.

Module Cyan was populated with up-regulated genes that are common between JR and J-diffsets and enriched for GO term translation (P = 0.027) (**Figure 6.7** and **Table 6.5**). Approximately 8% of the genes in the Blue module were mapped to those up-regulated in JR-diffset (Jaundice and Renal failure specific module. These enriched for GO term like oxidation-reduction, cytoskeletal reorganization and electron transport chain. Module like Light Cyan showed a similar gene expression profile like the blue module but enriched to GO term ubiquitin dependent catabolic process. The module dark Turquoise also specifically mapped to up-regulated genes from the JR-diffset, however the genes in this modules enriched to GO term RNA binding corresponding to the GO category molecular function.

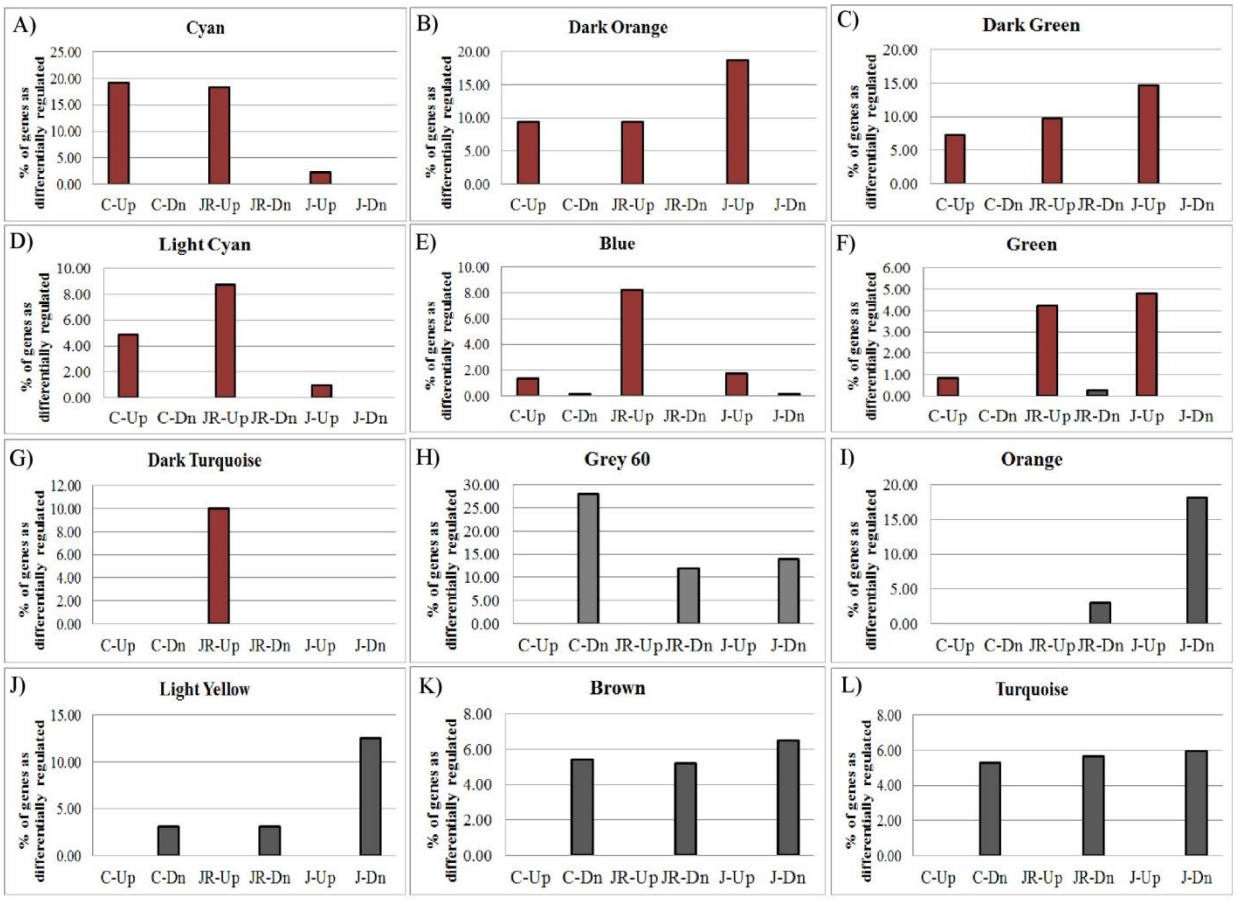


Figure 6.7. Modules with differentially regulated genes. Figure shows modules whose at least 10% of the genes were mapped as differentially regulated in complicated malaria isolates. X-axis represents % of genes represented as differentially regulated and Y-axis represents the group to which differentially regulated genes belongs to. **A-G**; up-regulated modules; **H-L**; down regulated modules. Abbreviations: C-Up, common up-regulated genes in JR and J diffsets; C-Dn, common down-regulated genes in JR and J diffsets; JR-Up, up-regulated genes unique to JR-diffset; JR-Dn, down-regulated genes unique to JR-diffset; J-Up, up-regulated genes unique to J-diffset; JR-Dn, down-regulated genes unique to J-diffset.

Table 6.5. Top gene ontology term of modules enriched with differentially regulated genes.

Module	Top GO term	Go term Category	*P* value
Up-regulated modules			
Cyan	Translation	BP	0.027
Dark Orange	N.S.E	N.S.E	N.S.E
Dark Green	N.S.E	N.S.E	N.S.E
Light Cyan	Ubiquitin-dependent protein catabolic process	BP	0.0016
Blue	Oxidation reduction	BP	1.47×10^{-4}
Green	Mitochondrial Transport	BP	2.49×10^{-5}
Dark Turquoise	Intracellular organelle lumen	CC	0.056
Down-regulated modules			
Grey 60	Pathogenesis	BP	O.019
Orange	Chromosome Organization	BP	2.11×10^{-4}
Light Yellow	N.S.E	N.S.E	N.S.E
Brown	Microtubule motor activity	MF	0.027
Turquoise	Glycerolipid biosynthesis	BP	0.017

Table lists only those modules whose at least 10% of the genes were mapped as differentially regulated. Abbreviations: BP, biological process,; N.S.E, not significantly enriched; MF, molecular function; CC, cellular component.

GO enrichment analysis of most of the up-regulated modules (Green, Dark Green, Dark Orange and Blue Module) suggests that the parasite is undergoing active respiration as genes involved in oxidation-reduction, electron transport chain, mitochondrial transport and ATP biosynthetic process are part of the up-regulated modules. It has been reported that asexual parasite *in vitro* rely exclusively on anaerobic glycolysis for energy (Lang-Unnasch and Murphy, 1998). This could be true of parasites supplied with ample amount of glucose in micro-aerophilic conditions. Studies suggest that patients with complicated disease from malaria could have altered substrate levels (LeRoux, et al., 2009). Glucose levels have been shown to drastically decrease in malaria infected patients and have been associated with severe malaria in children (Planche and Krishna, 2006). Again in the host the surrounding microenvironment can vary as far as oxygen and other necessary substrates are concerned. There is a strong possibility that the parasite for survival may modify its metabolic processes by utilizing alternative carbon sources and active respiration for its energy needs. One of the observations of Daily et al., 2007 from *in vivo* transcriptome studies of *P. falciparum* clinical isolates was that the parasites were in multiple physiological states. One of these states was associated with the starvation response where the parasites appeared to be involved in active respiration. This could be a logical interpretation of our observations, because, complicated disease could also imply lower level of nutrients and other requirements.

Alternatively, it could be argued that the active respiration process could be due to gametogenesis. It is accepted that gametocyte *in vitro* over express genes that are associated with reparation and mitochondrial function (Krungkrai, et al., 2000, Florens, et al., 2002, Tarun, et al., 2008, Young, et al., 2005). However, no gametocyte stages were detected in the samples used for this study although, there is always the possibility of the presence of very early gametocyte stages which cannot be morphologically distinguished from the ring stages. There is a suggestion that parasites *in vivo* tend to express certain sexual stage specific gene which are not seen in *in vitro* cultured parasites (Mackinnon, et al., 2009). There remains the possibility of such genes acting as over-all sensors of the environment with one of the possible outcomes as gemetogenesis.

The enrichment of Cyan module to translation, Light Cyan module to ubiquitin dependent catabolic process and Dark Turquoise to RNA binding GO terms indicate that there is

enhancement of both translation and degradation of proteins in parasites from patients showing complicated disease.

For the purposes of disease specific modules, the down-regulated gene cluster in modules were not considered. These modules enriched to chromosome organization, microtubule motor activity and glycerolipid biosynthesis (BP) and could be indicative of parasites under survival stress. Interestingly, the module grey 60 enriched to pathogenesis and was a down-regulated module (**Supplementary Table 6.2**). On further investigation, it was seen that this categorization (pathogenesis) is based on genes for PfEMP1 (PFI0005w, PF08_0141, MAL13P1.356) and Duffy binding receptor (PF10_0348). These genes have been found to be down-regulated in our dataset (PFC vs PFU) by the method of analysis adopted here by another analysis by which other *var* B and C group genes are seen as up-regulated (as described in Chapter 5). In our previous analysis *var* genes were analyzed sample wise rather than as average group.

6.2.5. Identification of hub genes from each up-regulated module

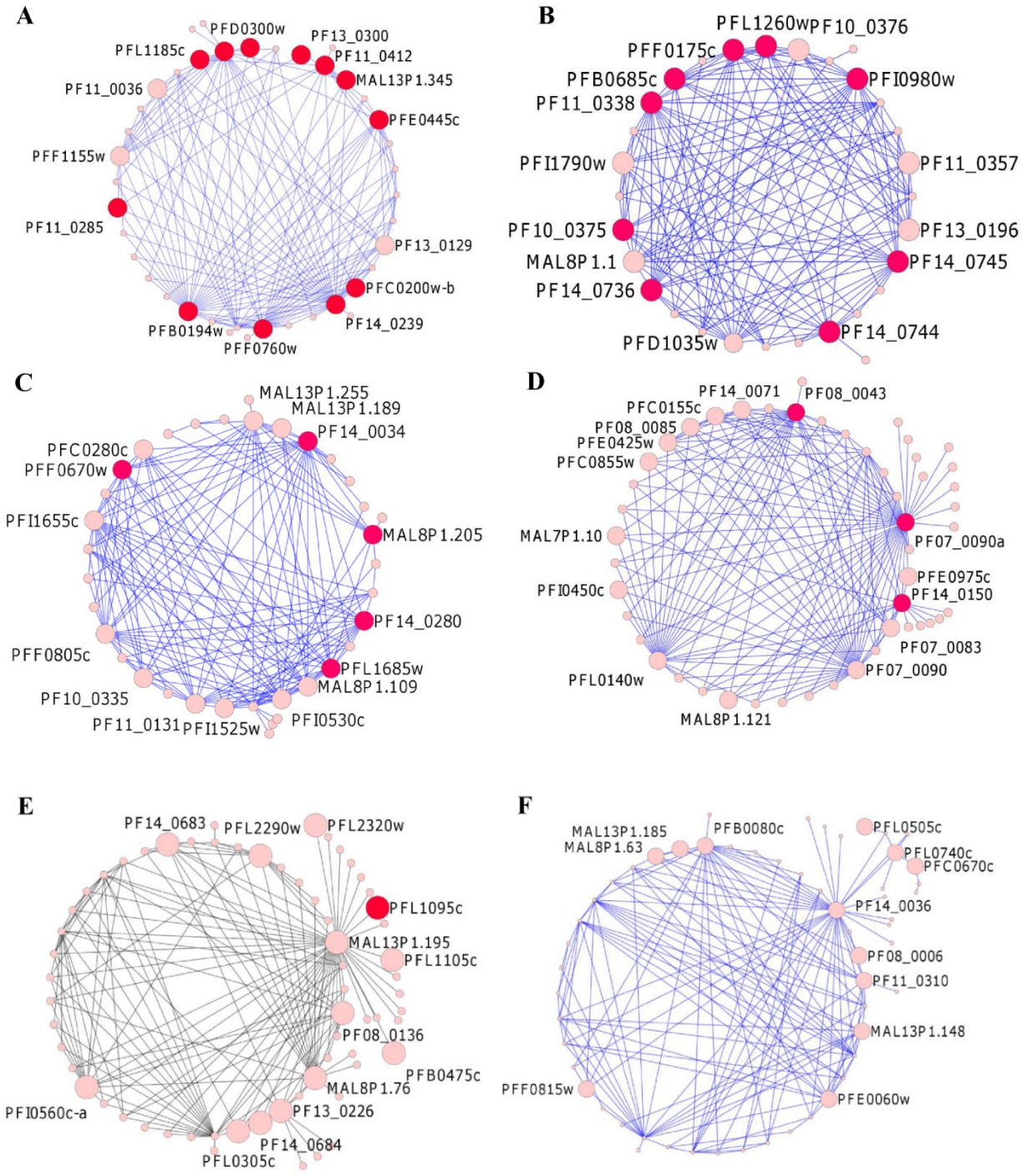
One of the concepts of systems network analysis is node connectivity measure (node degree) (Zhang and Horvath, 2005). The connectivity in a weighted network equals to the sum of the connection weights of a gene with other genes in the network (Horvath, 2008). A hub gene can therefore be defined as a gene with very high node connectivity (node degree). Since the WGCNA analysis identifies modules containing highly co-expressed genes a node can have two types of connectivity measures: one which measure the connectivity of a node with respects to whole network (whole network connectivity, kTotal) and the other which measures the connectivity with respect to the genes present in a particular module (intra-modular connectivity, kWithin). It has been felt that intra-modular connectivity measures gives more biologically meaningful information compare to its whole network counterpart during the selection of hub genes. The importance of such hub genes identified based on intra-modular connectivity measures has been reported (Ghazalpour, et al., 2006, Horvath and Dong, 2006, Oldham, et al., 2006, Carlson, et al., 2006, Gargalovic, et al., 2006).

In order to identify hub genes in each up-regulated module in our analysis, whole-network connectivity (kTotal) and intra-modular connectivity (kWithin) of each gene was calculated. Along with this, module membership of all genes in the network was also calculated. Module

membership is a "fuzzy" measure which defines the relationship between a gene with the module eigengene of a given module (Langfelder and Horvath, 2008). This measure is strongly related to intra-modular connectivity, as a gene with high module membership value in a particular module tends to have high intra-modular connectivity (hub gene). For each module, top 15 genes based on intra-modular connectivity were chosen (**Supplementary Table 6.6**). Although, we have considered only the up-regulated modules (i.e. at least 10% of the genes up-regulated) there is always the possibility of hub genes which show steady state expression but are hub genes. (**Supplementary Table 6.6**). **Figure 6.8** shows the highly connected genes in each up-regulated module.

To independently verify whether the top hub genes thus selected have any important biological role in parasite, the functions of all such genes with known annotations was correlated. The top three known annotated hub genes in the Cyan modules are, vacuolar ATP synthase subunit f (PF11_0412), cytochrome c heme lyase (PFL1185c) and Hexokinase (PFF1155w) of which PF11_0412 and PFL1185c are up-regulated in complicated cases (**Figure 6.8A**). Vacuolar ATP synthase otherwise known as V-type H^+ ATPase is a multi-subunit protein complex found in parasite digestive vacuoles (Saliba, et al., 2003). This protein complex activates the digestive vacuoles though active transport of protons which results in acidification of the vacuoles and generation of membrane potential across the vacuolar membrane. This is in turn necessary, for hemoglobin digestion, helps in accumulation of ions and drugs which could be involved in nutrient uptake and drug resistance phenomena (Yatsushiro, et al., 2005, Ginsburg, 2002, Moriyama, et al., 2003, Allen and Kirk, 2004). Hexokinase (PFF1155w) catalyzes the first step of glycolysis, the pathway on which *P. falciparum* relies for ATP generation. The first step results in the generation of Glucose 6 phosphate (G6P) which can have multiple fates. G6P can be utilized in the next step of glycolysis, can be channeled into pentose phostphate pathway (PPP), utilized for NADPH production (NADPH is a key molecule in the *P. falciparum* antioxidant defense) or can be utilized in nucleotide triphosphate biosynthesis pathways (Muller, 2004). This makes hexokinase a necessary molecule for parasite survival and it is considered as a potential drug target. To the best of our knowledge, the function of Cytochrome C heme-lyase in *P. falciparum* has not documented yet but studies in yeast, trypamosoma and leishmania suggest that it catalyzes the binding of heme to the apocytochrome C and higher expression of the same

has been associated with increased accumulation of apocytochrome C in mitochondria (Dumont, et al., 1991 and Kranz, et al., 2009).



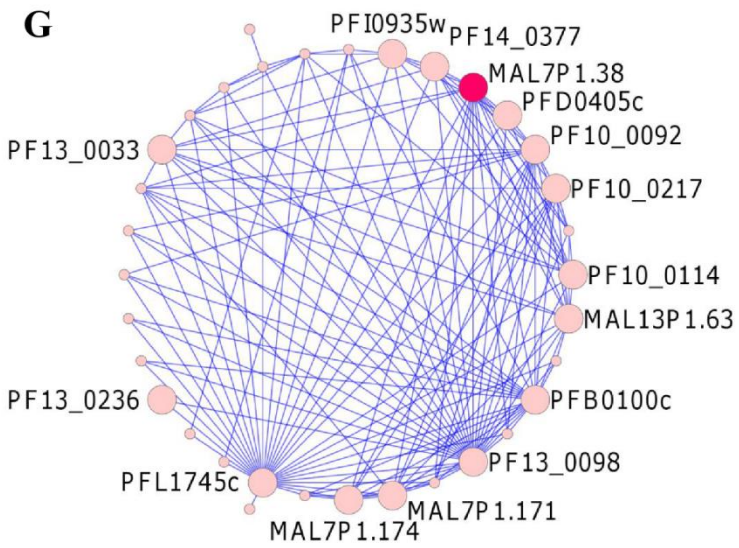


Figure 6.8. Network modules mapped with up-regulated genes and top connected nodes (hub genes). All the modules with at least 10% of their genes mapped with up-regulated genes. **A**, Cyan; **B**; Dark Orange; **C**, Dark Green; **D**, Light Cyan; **E**; Blue, **F**, Green; **G**, Dark Turquoise modules. Module network plots were generated by taking top 150 protein connections in each module in Cytoscape. Enlarged nodes in each module represents at most 15 top connected nodes based on node connectivity values (kWithin) in each modules. Enlarged red color nodes represent the up-regulated gene(s) in each module.

The three top hub genes in the Dark Orange module with known annotation are a long chain fatty acid elongation enzyme (PFI0980w), an aquaglyceroporin encoding protein (PF11_0338) and an acyl-CoA synthetase (PFB0685c) (**Figure 6.8B**). Interestingly all the three are up-regulated in PFC isolates. The product of the gene PFI0980w is an enzyme which converts Palmitic acid to Palmitoyl Coenzyme A, a key step in fatty acid biosynthesis (Mi-Ichi, et al., 2006). The Palmitic acid is a long chain essential fatty acid found in human serum that is essential for parasite intra-erythrocytic growth and proliferation (Mitamura, et al., 2000) as the *de novo* fatty acid biosynthetic pathway of the parasite do not have the capacity to generate long-chain fatty acids (Surolia and Surolia, 2001). The *P. falciparum* aquaglyceroporin (PF11_0338) is a multi-functional membrane channel protein which facilitates permeation of water, glycerol, erythritol,

urea, ammonia and sugar alcohols of up to five carbons long (Hansen, et al., 2002, Beitz, et al., 2004, Newby, et al., 2008, Song et al., 2012, Zeuthen, et al., 2006, Holm, et al., 2005). Reports suggests that this protein might play critical roles in osmotic protection of merozoite and ring stage parasites as well as acquisition of gelycerol from host serum glycerol pool for the use of ATP generation and phospholipid synthesis (Hansen, et al., 2002 and Promeneur, et al.,2006) and Considered a potential drug target (Newby, et al., 2008). The gene PFB0685c, encodes for an acyl-CoA synthetase (PfACS9), a member of multiple acyl-CoA synthetses present in *P. falciparum*. These enzymes transform long chain fatty acids to active CoA derivatives for metabolic utilization. These activated fatty acids serve multiple essential and regulatory functions which includes enzyme activation, protein acylation, protein transport, gene expression control (Groot, et al., 1976 and Glick, et al., 1987). This makes ACS activity essential for parasite growth and development as majority of fatty acids utilized is available from host milieu through ACS (Matesanz et al., 2003). The essentiality of this protein and the differences between the parasite and human PfACS sequences makes these enzymes potential molecular targets for therapeutic approaches (Matesanz et al., 2003). The nature of the protein products of these hub genes and their postulated biological roles makes the up-regulation of expression of such molecules understandable under complicated disease conditions.

Genes encoding DNA helicase (PF11_0131), protein phosphatase (MAL8P1.109) and AP2 transcription factors (PFF0670w) are three top hub genes in the dark green module (**Figure 6.8C**). Of these the first two are non-differentially regulated whereas the third is up-regulated. DNA helicases is well known for its essential role in catalyzing the unwinding of duplex nucleic acids required for processes involving DNA and RNA transactions (DNA for replication, transcription and recombination and many more events). Protein phosphatases are as essential as protein kinases to switch on/off the protein activities and play important roles in cell signaling which regulate different cellular functions. In *Plasmodium*, the protein phosphatases have been implicated in regulating parasite development and differentiation (Tewari, et al., 2010). Although the exact role of MAL8P1.109 is not known, its high degree of connectivity with the majority of the nodes in the modules indicates that it might be playing critical role(s) in this parasite. The PFF0670w is a transcription factor with an AP2 domain. This protein is a member of Apicomplexan AP2 (ApiAP2) family of DNA binding proteins (Balaji, et al., 2005 and Painter, et al., 2011). These are considered as the main family of transcription factors in *Plasmodium*

with suggested regulatory roles in gametocyte development (KafSack, et al., 2014), Ookinete formation (Yuda, et al., 2009), parasite development in liver (Iwanaga, et al., 2012), and *var* gene silencing (Flueck et al., 2010). The importance of up-regulated transcription factors such as this in parasite under disease stress with elevated transcriptional states is self-evident.

The top three known annotated hub genes of module Light Cyan are a DNA-directed RNA polymerase I/III subunit protein encoding gene (PF14_0150), a gene encoding ubiquitin conjugate enzyme (PF08_0085) and a small subunit rRNA processing protein (PF07_0083) (**Figure 6.8D**). Of these the first is up-regulated and the latter two are non-differentially regulated. The importance of RNA polymerase in biological systems is well known. Ubiquitin conjugate enzymes are key components of ubiquitin-proteasome system (UPS) where they catalyze the attachment of ubiquitin to the substrate protein ready for degradation. Ubiquitin-proteasome system is responsible for the degradation of majority of the intracellular proteins which is required for maintaining protein homeostasis and cell viability (Sjoerd and Wijik, 2010).

The identification of significant hub genes with known function/annotation in the selected modules let to the common observation that most of them have a vital role in the parasite survival and some are also considered to be potential drug targets. This gives credibility to the intra-modular connectivity measures which shows high biological relevance. Additionally, this emphasizes that other hub genes with unknown functions, annotated as hypothetical proteins could also be potential candidates for intervention studies.

In summary, a weighted gene co-expression network was constructed based on microarray hybridization data using clinical *P. falciparum* isolates from patients showing complicated or un-complicated disease symptoms. The network was further segregated into complicated disease specific modules and the inter-connectivity between such modules explored. Hub genes were identified based on densely inter-connected nodes and may provide insides into new molecules for use in disease intervention/prevention strategies.

7.1. Highlights

Understanding the transcriptome of the parasite is central to decode the parasite biology and hence malaria pathogenesis. Although, through high throughput transcriptomics experiments, we have gained enormous amount of information about the parasite transcriptome during its various developmental stages, information about its transcriptome in diseased human host still remains elusive. The main objective of the thesis was to investigate the *in vivo* transcriptome of *P. falciparum* clinical isolates from adult patients with differing disease conditions. During the thesis work, a custom *P. falciparum* cross strain 15K array (designed in our lab) was validated for its efficiency to detect transcripts from field isolates. Using a custom 244K strand specific *P. falciparum* microarray that was designed in order to select best probes for the custom 15K array, the diversity of natural antisense transcripts from field isolates was explored. This work for the first time unraveled the prevalence of natural antisense transcripts and its unique diversity in these isolates. Further, it demonstrated that some NATs change their expression status in different malaria related disease severity.

Subsequently, *in vivo* transcriptome of field isolates was investigated. Parasite transcriptomes from patients with uncomplicated malaria were compared with that of complicated malaria to decipher differences in parasite transcriptome in the two clinical conditions with special reference to genes encoding exportome proteins and variant surface antigens. This part of the study identified differences in gene expression between parasites from patients with uncomplicated (PFU) malaria as compared to those with complicated malaria (PFC).

In the last part of the thesis, a weighted gene co-expression based systems network was generated by incorporating gene expression profiles of field isolates. Further analyses of the network with integration of data from differential gene regulation analysis identified disease specific modules and thereby biological processes that seem to be perturbed in the diseased host. In addition, network parameters were employed to identify highly connected hub genes and subsequent analysis of the same suggest that many of them are potential candidate for intervention strategies and should be investigated further. To the best of our knowledge, this thesis reports for the first time about the transcriptome of *P. falciparum* clinical isolates from adult patients with diverse disease conditions. Ultimately, this thesis work has opened up the

hidden information of the parasite transcriptome inside the adult human host with complicated disease symptoms.

7.2. Specific Conclusions

1. *P. falciparum* custom cross strain array was validated for its efficiency to detect transcripts from *P. falciparum* field isolates.
2. A large percentage (91%) of the represented transcripts was detected from Indian *P. falciparum* patient isolates. Replicated probes and multiple probes representing the same gene showed perfect correlation between them suggesting good probe performance.
3. Additional transcripts could be detected due to inclusion of unique probes representing HB3 strain transcripts. Variant surface antigen transcripts were detected by optimized probes representing the VSA genes of three geographically distinct strains.
4. The 15K cross strain *P. falciparum* array has shown good efficiency in detecting transcripts from *P. falciparum* parasite samples isolated from patients.
5. Diversity of NATs in *P. falciparum* field isolates was explored using a custom 244K strand specific microarray.
6. A total of 797 NATs targeted against annotated loci have been detected. Out of these, 545 NATs are unique to this study compared to previously reported studies from *P. falciparum* laboratory strains.
7. The majority of NATs were positively correlated with the expression pattern of the sense transcript. However, 96 genes showed a change in sense/antisense ratio on comparison between uncomplicated and complicated disease conditions.
8. The antisense transcripts map to a broad range of biochemical/ metabolic pathways, especially pathways pertaining to the central carbon metabolism and stress related pathways.
9. Our data strongly suggests that a large group of NATs detected here are unannotated transcription units antisense to annotated gene models. The results reveal a previously unknown set of NATs that prevails in this parasite, their differential regulation in disease conditions and mapping to functionally well annotated genes.

10. The custom cross strain 15K array was employed to capture the expression profiles of parasite isolates from adult patient samples showing uncomplicated malaria or complicated malaria like hepatic dysfunction and/ renal failure. These are the most common manifestations seen in adults along with cerebral malaria.
11. Differential gene expression analysis found key genes involved in processes like regulation of intracellular pH, electron transport chain, cytochrome C oxidase activity, DNA packaging, chromosome organization, cell cycle, generation of precursor and metabolite energy. All the processes mentioned are directly dependent on the host physiological environment which suggests that parasites inside the diseased host adapt itself to the changing environment.
12. The impact of exported proteins on iRBCs **is** that it changes many properties of iRBCs including permeability, rigidity and adherence. Up-regulation of a PHISTc domain encoding gene and down regulation of three PHISTb domain containing protein encoding genes suggests that although these proteins contains PHIST domain, both the PHIST sub-group perform different functions.
13. Set of genes with unknown function belonging to exportome class of proteins were found to be differentially regulated in isolates from patients with complicated disease.
14. The data has been analyzed with reference to variant surface antigens, encoded by the *var*, *rifin* and *stevor* gene families.
15. Up-regulation of *var* group B and C genes whose proteins are predicted to interact with CD36 receptor in the host as also the up-regulation of group A *rifins* and many of the *stevors* were observed. This is contrary to most other reports from pediatric patients, with cerebral malaria where the up-regulation of *var* A group genes have been seen.
16. Presence of multiple RIFIN and STEVOR proteins on the iRBC surface increases the complexity of the molecules against which an immune response may be mounted. The molecules involved in the host response against such proteins which are up-regulated in the complicated disease state, may provide a clue to some of the factors initiating such a disease complication.
17. The differential expression of *var* group B and C in non–cerebral complicated malaria cases suggests that these molecules should also be considered before implementation of any anti-disease intervention strategy.

18. Systems network based approach was employed to identify disease specific modules populated with differentially regulated genes, subsequent analysis of which revealed biological processes that are likely to be perturbed during complicated malaria disease conditions.
19. In the concluding part of the work, sets of hub genes were identified based on their connectivity with other genes in the modules. Literature based analysis of the hub genes with annotated known function suggested that most of them are playing crucial role in parasite biology and some of them are either drug targets or projected as potential drug targets. The importance of majority of the known hub genes in parasite biology suggests that other hub genes annotated as hypothetical proteins are potential candidate for intervention strategies and should be investigated further.

7.3. Direction for future research (future perspectives)

- The results about NATs detailed here call for studies to deduce the possible mechanism of action of NATs, which would further help in understanding the *in vivo* pathological adaptations of these parasites.
- It would be interesting to see how the biological processes mapped with differentially regulated genes in complicated cases enhance the parasite virulence to cause disease severity.
- Dissecting the functional roles of differentially regulated exportome proteins will provide additional information about the association of the same with the iRBCs by permeating nutrient or displaying virulence factors.
- Systems network based analysis suggested that *in vivo* parasite transcriptome is dynamic and differs substantially in diverse disease conditions. However, it still needs to be investigated that how these differences contribute to the disease process.
- The importance of majority of the known hub genes in parasite biology suggests that other hub genes annotated as hypothetical proteins are potential candidate for intervention strategies and should be investigated further.

References

Abdel-Latif, M. S., Dietz, K., Issifou, S., Kremsner, P. G., and Klinkert, M. Q., 2003. Antibodies to Plasmodium falciparum rifin proteins are associated with rapid parasite clearance and asymptomatic infections. Infect Immun 71**,** 6229-6233.

Abdel-Latif, M. S., Khattab, A., Lindenthal, C., Kremsner, P. G., and Klinkert, M. Q., 2002. Recognition of variant Rifin antigens by human antibodies induced during natural Plasmodium falciparum infections. Infect Immun 70**,** 7013-7021.

Allen, R. J. and Kirk, K., 2004. The membrane potential of the intraerythrocytic malaria parasite Plasmodium falciparum. J Biol Chem 279**,** 11264-11272.

Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., and Menard, R., 2006. Quantitative imaging of Plasmodium transmission from mosquito to mammal. Nat Med 12**,** 220-224.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., and Young, I. G., 1981. Sequence and organization of the human mitochondrial genome. Nature 290**,** 457-465.

Atkinson, C. T. and Aikawa, M., 1990. Ultrastructure of malaria-infected erythrocytes. Blood Cells 16**,** 351-368.

Atroosh, W. M., Al-Mekhlafi, H. M., Mahdy, M. A., Saif-Ali, R., Al-Mekhlafi, A. M., and Surin, J., 2011. Genetic diversity of Plasmodium falciparum isolates from Pahang, Malaysia based on MSP-1 and MSP-2 genes. Parasit Vectors 4**,** 233.

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Jr., Treatman, C., and Wang, H., 2009. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res 37**,** D539-543.

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Treatman, C., and Wang, H., 2009. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Research 37**,** D539-D543.

Baer, K., Klotz, C., Kappe, S. H., Schnieder, T., and Frevert, U., 2007a. Release of hepatic Plasmodium yoelii merozoites into the pulmonary microvasculature. PLoS Pathog 3**,** e171.

Baer, K., Roosevelt, M., Clarkson, A. B., Jr., van Rooijen, N., Schnieder, T., and Frevert, U., 2007b. Kupffer cells are obligatory for Plasmodium yoelii sporozoite infection of the liver. Cell Microbiol 9**,** 397-412.

Balaji, S., Babu, M. M., Iyer, L. M., and Aravind, L., 2005. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Res 33**,** 3994-4006.

Balu, B., Chauhan, C., Maher, S. P., Shoue, D. A., Kissinger, J. C., Fraser, M. J., Jr., and Adams, J. H., 2009. piggyBac is an effective tool for functional analysis of the Plasmodium falciparum genome. BMC Microbiol 9**,** 83.

Balu, B., Shoue, D. A., Fraser, M. J., Jr., and Adams, J. H., 2005. High-efficiency transformation of Plasmodium falciparum by the lepidopteran transposable element piggyBac. Proc Natl Acad Sci U S A 102**,** 16391-16396.

Bartfai, R., Hoeijmakers, W. A., Salcedo-Amaya, A. M., Smits, A. H., Janssen-Megens, E., Kaan, A., Treeck, M., Gilberger, T. W., Francoijs, K. J., and Stunnenberg, H. G., 2010. H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3. PLoS Pathog 6**,** e1001223.

Baruch, D. I., Pasloske, B. L., Singh, H. B., Bi, X., Ma, X. C., Feldman, M., Taraschi, T. F., and Howard, R. J., 1995. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell 82**,** 77-87.

Behnke, M. S., Wootton, J. C., Lehmann, M. M., Radke, J. B., Lucas, O., Nawas, J., Sibley, L. D., and White, M. W., 2010. Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of Toxoplasma gondii. PLoS One 5**,** e12354.

Beitz, E., Pavlovic-Djuranovic, S., Yasui, M., Agre, P., and Schultz, J. E., 2004. Molecular dissection of water and glycerol permeability of the aquaglyceroporin from Plasmodium falciparum by mutational analysis. Proc Natl Acad Sci U S A 101**,** 1153-1158.

Ben Mamoun, C., Gluzman, I. Y., Hott, C., MacMillan, S. K., Amarakone, A. S., Anderson, D. L., Carlton, J. M., Dame, J. B., Chakrabarti, D., Martin, R. K., Brownstein, B. H., and Goldberg, D. E., 2001. Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite Plasmodium falciparum revealed by microarray analysis. Mol Microbiol 39**,** 26-36.

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W., and Clayton, D. A., 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell 26**,** 167-180.

Biggs, B. A., Kemp, D. J., and Brown, G. V., 1989. Subtelomeric chromosome deletions in field isolates of Plasmodium falciparum and their relationship to loss of cytoadherence in vitro. Proc Natl Acad Sci U S A 86**,** 2428-2432.

Billker, O., Lindo, V., Panico, M., Etienne, A. E., Paxton, T., Dell, A., Rogers, M., Sinden, R. E., and Morris, H. R., 1998. Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. Nature 392**,** 289-292.

Billker, O., Shaw, M. K., Margos, G., and Sinden, R. E., 1997. The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of Plasmodium berghei in vitro. Parasitology 115 (Pt 1)**,** 1-7.

Blythe, J. E., Surentheran, T., and Preiser, P. R., 2004. STEVOR--a multifunctional protein? Mol Biochem Parasitol 134**,** 11-15.

Blythe, J. E., Yam, X. Y., Kuss, C., Bozdech, Z., Holder, A. A., Marsh, K., Langhorne, J., and Preiser, P. R., 2008. Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. Infect Immun 76**,** 3329-3336.

Boddey, J. A. and Cowman, A. F., 2013. Plasmodium nesting: remaking the erythrocyte from the inside out. Annu Rev Microbiol 67**,** 243-269.

Botha, M., Pesce, E. R., and Blatch, G. L., 2007. The Hsp40 proteins of Plasmodium falciparum and other apicomplexa: regulating chaperone power in the parasite and the host. Int J Biochem Cell Biol 39**,** 1781-1803.

Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter,

S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C., and Barrell, B. G., 1999. The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum. Nature 400**,** 532-538.

Bozdech, Z., Llinas, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L., 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol 1**,** E5.

Bozdech, Z., Zhu, J., Joachimiak, M. P., Cohen, F. E., Pulliam, B., and DeRisi, J. L., 2003. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. Genome Biol 4**,** R9.

Breman, J. G., 2001. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. Am J Trop Med Hyg 64**,** 1-11.

Broadbent, K. M., Park, D., Wolf, A. R., Van Tyne, D., Sims, J. S., Ribacke, U., Volkman, S., Duraisingh, M., Wirth, D., Sabeti, P. C., and Rinn, J. L., 2011. A global transcriptional analysis of Plasmodium falciparum malaria reveals a novel family of telomere-associated lncRNAs. Genome Biol 12**,** R56.

Bull, P. C., Kortok, M., Kai, O., Ndungu, F., Ross, A., Lowe, B. S., Newbold, C. I., and Marsh, K., 2000. Plasmodium falciparum-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. J Infect Dis 182**,** 252-259.

Bull, P. C., Berriman, M., Kyes, S., Quail, M. A., Hall, N., Kortok, M. M., Marsh, K., and Newbold, C. I., 2005. Plasmodium falciparum variant surface antigen expression patterns during malaria. PLoS Pathog 1**,** e26.

Cappai, R., van Schravendijk, M. R., Anders, R. F., Peterson, M. G., Thomas, L. M., Cowman, A. F., and Kemp, D. J., 1989. Expression of the RESA gene in Plasmodium falciparum isolate FCR3 is prevented by a subtelomeric deletion. Mol Cell Biol 9**,** 3584-3587.

Carlson, M. R., Zhang, B., Fang, Z., Mischel, P. S., Horvath, S., and Nelson, S. F., 2006. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genomics 7**,** 40.

Caro, F., Miller, M., and DeRisi, J., 2012. Plate-based transfection and culturing technique for genetic manipulation of Plasmodium falciparum. Malaria Journal 11**,** 22.

Caro, F., Miller, M. G., and DeRisi, J. L., 2012. Plate-based transfection and culturing technique for genetic manipulation of Plasmodium falciparum. Malar J 11**,** 22.

Chaal, B. K., Gupta, A. P., Wastuwidyaningtyas, B. D., Luah, Y. H., and Bozdech, Z., 2010. Histone deacetylases play a major role in the transcriptional regulation of the Plasmodium falciparum life cycle. PLoS Pathog 6**,** e1000737.

Chakravorty, A., Klovstad, M., Peterson, G., Lindeman, R. E., and Gregg-Jolly, L. A., 2008. Sensitivity of an Acinetobacter baylyi mpl mutant to DNA damage. Appl Environ Microbiol 74**,** 1273-1275.

Chakravorty, S. J., Hughes, K. R., and Craig, A. G., 2008. Host response to cytoadherence in Plasmodium falciparum. Biochem Soc Trans 36**,** 221-228.

Chen, J., Sun, M., Hurst, L. D., Carmichael, G. G., and Rowley, J. D., 2005. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. Trends Genet 21**,** 326-329.

Chen, L. Y., 2013. Glycerol inhibits water permeation through Plasmodium falciparum aquaglyceroporin. J Struct Biol 181**,** 71-76.

Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M., and Saul, A., 1998. stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. Mol Biochem Parasitol 97**,** 161-176.

Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M., and Saul, A., 1998. stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. Molecular and Biochemical Parasitology 97**,** 161-176.

Chin, T. F. and Lach, J. L., 1965. Spectrophotometric determination of some quaternary compounds. J Pharm Sci 54**,** 1550-1551.

Chotivanich, K., Udomsangpetch, R., Dondorp, A., Williams, T., Angus, B., Simpson, J. A., Pukrittayakamee, S., Looareesuwan, S., Newbold, C. I., and White, N. J., 2000. The mechanisms of parasite clearance after antimalarial treatment of Plasmodium falciparum malaria. J Infect Dis 182**,** 629-633.

Conley, A. B. and Jordan, I. K., 2012. Epigenetic regulation of human cis-natural antisense transcripts. Nucleic Acids Res 40**,** 1438-1445.

Cooke, B. M., Glenister, F. K., Mohandas, N., and Coppel, R. L., 2002. Assignment of functional roles to parasite proteins in malaria-infected red blood cells by competitive flow-based adhesion assay. Br J Haematol 117**,** 203-211.

Cooke, B. M., Lingelbach, K., Bannister, L. H., and Tilley, L., 2004. Protein trafficking in Plasmodium falciparum-infected red blood cells. Trends Parasitol 20**,** 581-589.

Coppi, A., Tewari, R., Bishop, J. R., Bennett, B. L., Lawrence, R., Esko, J. D., Billker, O., and Sinnis, P., 2007. Heparan sulfate proteoglycans provide a signal to Plasmodium sporozoites to stop migrating and productively invade host cells. Cell Host Microbe 2**,** 316-327.

Cowman, A. F., Berry, D., and Baum, J., 2012. The cellular and molecular basis for malaria parasite invasion of the human red blood cell. J Cell Biol 198**,** 961-971.

Cui, I. and Cui, H., 2010. Antisense RNAs and epigenetic regulation. Epigenomics 2**,** 139-150.

Cui, L. and Miao, J., 2010. Chromatin-mediated epigenetic regulation in the malaria parasite Plasmodium falciparum. Eukaryot Cell 9**,** 1138-1149.

Daily, J. P., Le Roch, K. G., Sarr, O., Ndiaye, D., Lukens, A., Zhou, Y., Ndir, O., Mboup, S., Sultan, A., Winzeler, E. A., and Wirth, D. F., 2005. In vivo transcriptome of Plasmodium falciparum reveals overexpression of transcripts that encode surface proteins. J Infect Dis 191**,** 1196-1203.

Daily, J. P., Scanfeld, D., Pochet, N., Le Roch, K., Plouffe, D., Kamal, M., Sarr, O., Mboup, S., Ndir, O., Wypij, D., Levasseur, K., Thomas, E., Tamayo, P., Dong, C., Zhou, Y., Lander, E. S., Ndiaye, D., Wirth, D., Winzeler, E. A., Mesirov, J. P., and Regev, A., 2007. Distinct physiological states of Plasmodium falciparum in malaria-infected patients. Nature 450**,** 1091-1095.

Das, A., Holloway, B., Collins, W. E., Shama, V. P., Ghosh, S. K., Sinha, S., Hasnain, S. E., Talwar, G. P., and Lal, A. A., 1995. Species-specific 18S rRNA gene amplification for the detection of P. falciparum and P. vivax malaria parasites. Mol Cell Probes 9**,** 161-165.

Date, S. V. and Stoeckert, C. J., Jr., 2006. Computational modeling of the Plasmodium falciparum interactome reveals protein function on a genome-wide scale. Genome Res 16**,** 542-549.

de Jong, S., Boks, M. P., Fuller, T. F., Strengman, E., Janson, E., de Kovel, C. G., Ori, A. P., Vi, N., Mulder, F., Blom, J. D., Glenthoj, B., Schubart, C. D., Cahn, W., Kahn, R. S.,

Horvath, S., and Ophoff, R. A., 2012. A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. PLoS ONE 7**,** e39498.

Deitsch, K. W. and Wellems, T. E., 1996. Membrane modifications in erythrocytes parasitized by Plasmodium falciparum. Mol Biochem Parasitol 76**,** 1-10.

Dhingra, N., Jha, P., Sharma, V. P., Cohen, A. A., Jotkar, R. M., Rodriguez, P. S., Bassani, D. G., Suraweera, W., Laxminarayan, R., and Peto, R., 2010. Adult and child malaria mortality in India: a nationally representative mortality survey. Lancet 376**,** 1768-1774.

Divo, A. A., Geary, T. G., Davis, N. L., and Jensen, J. B., 1985. Nutritional requirements of Plasmodium falciparum in culture. I. Exogenously supplied dialyzable components necessary for continuous growth. J Protozool 32**,** 59-64.

Dolan, S. A., Miller, L. H., and Wellems, T. E., 1990. Evidence for a switching mechanism in the invasion of erythrocytes by Plasmodium falciparum. J Clin Invest 86**,** 618-624.

Donaldson, M. E. and Saville, B. J., 2012. Natural antisense transcripts in fungi. Mol Microbiol 85**,** 405-417.

Dumas, C., Chow, C., Muller, M., and Papadopoulou, B., 2006. A novel class of developmentally regulated noncoding RNAs in Leishmania. Eukaryot Cell 5**,** 2033-2046.

Duraisingh, M. T., Voss, T. S., Marty, A. J., Duffy, M. F., Good, R. T., Thompson, J. K., Freitas-Junior, L. H., Scherf, A., Crabb, B. S., and Cowman, A. F., 2005. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. Cell 121**,** 13-24.

Dzikowski, R. and Deitsch, K., 2006. Antigenic variation by protozoan parasites: insights from Babesia bovis. Mol Microbiol 59**,** 364-366.

Faghihi, M. A. and Wahlestedt, C., 2009. Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 10**,** 637-643.

Fang, J., Zhou, H., Rathore, D., Sullivan, M., Su, X. Z., and McCutchan, T. F., 2004. Ambient glucose concentration and gene expression in Plasmodium falciparum. Mol Biochem Parasitol 133**,** 125-129.

Fernandez, V., Hommel, M., Chen, Q., Hagblom, P., and Wahlgren, M., 1999. Small, clonally variant antigens expressed on the surface of the Plasmodium falciparum-infected erythrocyte are encoded by the rif gene family and are the target of human immune responses. J Exp Med 190**,** 1393-1404.

Flatt, C., Mitchell, S., Yipp, B., Looareesuwan, S., and Ho, M., 2005. Attenuation of cytoadherence of Plasmodium falciparum to microvascular endothelium under flow by hemodilution. Am J Trop Med Hyg 72**,** 660-665.

Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., Witney, A. A., Wolters, D., Wu, Y., Gardner, M. J., Holder, A. A., Sinden, R. E., Yates, J. R., and Carucci, D. J., 2002. A proteomic view of the Plasmodium falciparum life cycle. Nature 419**,** 520-526.

Flueck, C., Bartfai, R., Niederwieser, I., Witmer, K., Alako, B. T., Moes, S., Bozdech, Z., Jenoe, P., Stunnenberg, H. G., and Voss, T. S., 2010. A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. PLoS Pathog 6**,** e1000784.

Flueck, C., Bartfai, R., Volz, J., Niederwieser, I., Salcedo-Amaya, A. M., Alako, B. T., Ehlgen, F., Ralph, S. A., Cowman, A. F., Bozdech, Z., Stunnenberg, H. G., and Voss, T. S., 2009. Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog 5**,** e1000569.

Fong, J. S. and Good, R. A., 1971. Prevention of the localized and generalized Shwartzman reactions by an anticomplementary agent, cobra venom factor. J Exp Med 134**,** 642-655.

Franceschini, A., Hullugundi, S. K., van den Maagdenberg, A. M., Nistri, A., and Fabbretti, E., 2013. Effects of LPS on P2X3 receptors of trigeminal sensory neurons and macrophages from mice expressing the R192Q Cacna1a gene mutation of familial hemiplegic migraine-1. Purinergic Signal 9**,** 7-13.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., and Jensen, L. J., 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41**,** D808-815.

Frankland, S., Elliott, S. R., Yosaatmadja, F., Beeson, J. G., Rogerson, S. J., Adisa, A., and Tilley, L., 2007. Serum lipoproteins promote efficient presentation of the malaria virulence protein PfEMP1 at the erythrocyte surface. Eukaryot Cell 6**,** 1584-1594.

Frevert, U., Engelmann, S., Zougbede, S., Stange, J., Ng, B., Matuschewski, K., Liebes, L., and Yee, H., 2005. Intravital observation of Plasmodium berghei sporozoite infection of the liver. PLoS Biol 3**,** e192.

Frischknecht, F., Baldacci, P., Martin, B., Zimmer, C., Thiberge, S., Olivo-Marin, J. C., Shorte, S. L., and Menard, R., 2004. Imaging movement of malaria parasites during transmission by Anopheles mosquitoes. Cell Microbiol 6**,** 687-694.

Gallup, J. L. and Sachs, J. D., 2001. The economic burden of malaria. Am J Trop Med Hyg 64**,** 85-96.

Ganesan, K., Ponmee, N., Jiang, L., Fowble, J. W., White, J., Kamchonwongpaisan, S., Yuthavong, Y., Wilairat, P., and Rathod, P. K., 2008. A genetically hard-wired metabolic transcriptome in Plasmodium falciparum fails to mount protective responses to lethal antifolates. PLoS Pathog 4**,** e1000214.

Gardner, M., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R., Carlton, J., Pain, A., Nelson, K., Bowman, S., Paulsen, I., James, K., Eisen, J., Rutherford, K., Salzberg, S., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M., Vaidya, A., Martin, D., Fairlamb, A., Fraunholz, M., Roos, D., Ralph, S., McFadden, G., Cummings, L., Subramanian, G., Mungall, C., Venter, J., Carucci, D., Hoffman, S., Newbold, C., Davis, R., Fraser, C., and Barrell, B., 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419**,** 498 - 511.

Gardner, M. J., Shallom, S. J., Carlton, J. M., Salzberg, S. L., Nene, V., Shoaibi, A., Ciecko, A., Lynn, J., Rizzo, M., Weaver, B., Jarrahi, B., Brenner, M., Parvizi, B., Tallon, L., Moazzez, A., Granger, D., Fujii, C., Hansen, C., Pederson, J., Feldblyum, T., Peterson, J., Suh, B., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., White, O., Cummings, L. M., Smith, H. O., Adams, M. D., Venter, J. C., Carucci, D. J., Hoffman, S. L., and Fraser, C. M., 2002. Sequence of Plasmodium falciparum chromosomes 2, 10, 11 and 14. Nature 419**,** 531-534.

Gardner, M. J., Tettelin, H., Carucci, D. J., Cummings, L. M., Aravind, L., Koonin, E. V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D. C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G. G., Clayton, R., White, O., Smith, H. O., Fraser, C. M., Adams, M. D., Venter, J. C., and Hoffman, S. L., 1998. Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum. Science 282**,** 1126-1132.

Gargalovic, P. S., Imura, M., Zhang, B., Gharavi, N. M., Clark, M. J., Pagnon, J., Yang, W. P., He, A., Truong, A., Patel, S., Nelson, S. F., Horvath, S., Berliner, J. A., Kirchgessner, T. G., and Lusis, A. J., 2006. Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. Proc Natl Acad Sci U S A 103**,** 12741-12746.

Geary, T. G., Divo, A. A., Bonanni, L. C., and Jensen, J. B., 1985a. Nutritional requirements of Plasmodium falciparum in culture. III. Further observations on essential nutrients and antimetabolites. J Protozool 32**,** 608-613.

Geary, T. G., Divo, A. A., and Jensen, J. B., 1985b. Nutritional requirements of Plasmodium falciparum in culture. II. Effects of antimetabolites in a semi-defined medium. J Protozool 32**,** 65-69.

Genton, B., Al-Yaman, F., Betuela, I., Anders, R. F., Saul, A., Baea, K., Mellombo, M., Taraika, J., Brown, G. V., Pye, D., Irving, D. O., Felger, I., Beck, H. P., Smith, T. A., and Alpers, M. P., 2003. Safety and immunogenicity of a three-component blood-stage malaria vaccine (MSP1, MSP2, RESA) against Plasmodium falciparum in Papua New Guinean children. Vaccine 22**,** 30-41.

Ghazalpour, A., Doss, S., Zhang, B., Wang, S., Plaisier, C., Castellanos, R., Brozell, A., Schadt, E. E., Drake, T. A., Lusis, A. J., and Horvath, S., 2006. Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet 2**,** e130.

Gilson, P. R. and Crabb, B. S., 2009. Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites. Int J Parasitol 39**,** 91-96.

Ginsburg, H., 2002. Abundant proton pumping in Plasmodium falciparum, but why? Trends Parasitol 18**,** 483-486.

Golgi C., 1886. Sul' infezione malarica. Arch Sci Med Torino 10, 109-135.

Golgi C., 1889. Sul ciclo evolutivo dei parassiti malarici nella febbre terzana: diagnosi differenziale tra i parassiti endoglobulari malarici della terzanae quelli della quartana. Arch Sci Med Torino13, 173-196.

Goodman, C. D., Su, V., and McFadden, G. I., 2007. The effects of anti-bacterials on the malaria parasite Plasmodium falciparum. Mol Biochem Parasitol 152**,** 181-191.

Grassi B: Studi di uno Zoologo Sulla Malaria. Rome 1900.

Gravenor, M. B. and Kwiatkowski, D., 1998. An analysis of the temperature effects of fever on the intra-host population dynamics of Plasmodium falciparum. Parasitology 117 (Pt 2)**,** 97-105.

Gruring, C., Heiber, A., Kruse, F., Ungefehr, J., Gilberger, T. W., and Spielmann, T., 2011. Development and host cell modifications of Plasmodium falciparum blood stages in four dimensions. Nat Commun 2**,** 165.

Gunasekera, A. M., Patankar, S., Schug, J., Eisen, G., Kissinger, J., Roos, D., and Wirth, D. F., 2004. Widespread distribution of antisense transcripts in the Plasmodium falciparum genome. Mol Biochem Parasitol 136**,** 35-42.

Guttery, D. S., Holder, A. A., and Tewari, R., 2012. Sexual development in Plasmodium: lessons from functional analyses. PLoS Pathog 8**,** e1002404.

Haldar, K. and Mohandas, N., 2007. Erythrocyte remodeling by malaria parasites. Curr Opin Hematol 14**,** 203-209.

Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D., Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K., Buckee, C. O., Burrows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Corton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J., Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K. D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P., Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Rutter, S., Rutherford, K. M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J. E., Craig, A., Newbold, C., and Barrell, B. G., 2002. Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13. Nature 419**,** 527-531.

Hansen, M., Kun, J. F., Schultz, J. E., and Beitz, E., 2002. A single, bi-functional aquaglyceroporin in blood-stage Plasmodium falciparum malaria parasites. J Biol Chem 277**,** 4874-4882.

Harris, M. T., Walker, D. M., Drew, M. E., Mitchell, W. G., Dao, K., Schroeder, C. E., Flaherty, D. P., Weiner, W. S., Golden, J. E., and Morris, J. C., 2013. Interrogating a hexokinase-

selected small-molecule library for inhibitors of Plasmodium falciparum hexokinase. Antimicrob Agents Chemother 57**,** 3731-3737.

Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W., 1999. From molecular to modular cell biology. Nature 402**,** C47-52.

Hay, S. I., Okiro, E. A., Gething, P. W., Patil, A. P., Tatem, A. J., Guerra, C. A., and Snow, R. W., 2010. Estimating the global clinical burden of Plasmodium falciparum malaria in 2007. PLoS Med 7**,** e1000290.

Hayward, R. E., Derisi, J. L., Alfadhli, S., Kaslow, D. C., Brown, P. O., and Rathod, P. K., 2000. Shotgun DNA microarrays and stage-specific gene expression in Plasmodium falciparum malaria. Mol Microbiol 35**,** 6-14.

He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., and Kinzler, K. W., 2008. The antisense transcriptomes of human cells. Science 322**,** 1855-1857.

Heiber, A., Kruse, F., Pick, C., Gruring, C., Flemming, S., Oberli, A., Schoeler, H., Retzlaff, S., Mesen-Ramirez, P., Hiss, J. A., Kadekoppala, M., Hecht, L., Holder, A. A., Gilberger, T. W., and Spielmann, T., 2013. Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in Plasmodium falciparum protein export. PLoS Pathog 9**,** e1003546.

Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K., 2004. A Host-Targeting Signal in Virulence Proteins Reveals a Secretome in Malarial Infection. Science 306**,** 1934-1937.

Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K., 2004. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 306**,** 1934-1937.

Holm, L. M., Jahn, T. P., Moller, A. L., Schjoerring, J. K., Ferri, D., Klaerke, D. A., and Zeuthen, T., 2005. NH3 and NH4+ permeability in aquaporin-expressing Xenopus oocytes. Pflugers Arch 450**,** 415-428.

Horvath, S. and Dong, J., 2008. Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol 4**,** e1000117.

Hu, G., Cabrera, A., Kono, M., Mok, S., Chaal, B. K., Haase, S., Engelberg, K., Cheemadan, S., Spielmann, T., Preiser, P. R., Gilberger, T. W., and Bozdech, Z., 2010. Transcriptional

profiling of growth perturbations of the human malaria parasite Plasmodium falciparum. Nat Biotechnol 28**,** 91-98.

Huang da, W., Sherman, B. T., and Lempicki, R. A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37**,** 1-13.

Huang da, W., Sherman, B. T., and Lempicki, R. A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4**,** 44-57.

Imwong, M., Dondorp, A. M., Nosten, F., Yi, P., Mungthin, M., Hanchana, S., Das, D., Phyo, A. P., Lwin, K. M., Pukrittayakamee, S., Lee, S. J., Saisung, S., Koecharoen, K., Nguon, C., Day, N. P., Socheat, D., and White, N. J., 2010. Exploring the contribution of candidate genes to artemisinin resistance in Plasmodium falciparum. Antimicrob Agents Chemother 54**,** 2886-2892.

Iwanaga, S., Kaneko, I., Kato, T., and Yuda, M., 2012. Identification of an AP2-family protein that is critical for malaria liver stage development. PLoS ONE 7**,** e47557.

Janes, J. H., Wang, C. P., Levin-Edens, E., Vigan-Womas, I., Guillotte, M., Melcher, M., Mercereau-Puijalon, O., and Smith, J. D., 2011. Investigating the host binding signature on the Plasmodium falciparum PfEMP1 protein family. PLoS Pathog 7**,** e1002032.

Jensen, A. T., Magistrado, P., Sharp, S., Joergensen, L., Lavstsen, T., Chiucchiuini, A., Salanti, A., Vestergaard, L. S., Lusingu, J. P., Hermsen, R., Sauerwein, R., Christensen, J., Nielsen, M. A., Hviid, L., Sutherland, C., Staalsoe, T., and Theander, T. G., 2004. Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. J Exp Med 199**,** 1179-1190.

Jiang, H., Patel, J. J., Yi, M., Mu, J., Ding, J., Stephens, R., Cooper, R. A., Ferdig, M. T., and Su, X. Z., 2008. Genome-wide compensatory changes accompany drug- selected mutations in the Plasmodium falciparum crt gene. PLoS ONE 3**,** e2484.

Joannin, N., Abhiman, S., Sonnhammer, E. L., and Wahlgren, M., 2008. Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. BMC Genomics 9**,** 19.

Kaestli, M., Cockburn, I. A., Cortes, A., Baea, K., Rowe, J. A., and Beck, H. P., 2006. Virulence of malaria is associated with differential expression of Plasmodium falciparum var gene subgroups in a case-control study. J Infect Dis 193**,** 1567-1574.

Kaestli, M., Cockburn, I. A., Cortes, A., Baea, K., Rowe, J. A., and Beck, H. P., 2006. Virulence of malaria is associated with differential expression of Plasmodium falciparum var gene subgroups in a case-control study. J Infect Dis 193**,** 1567-1574.

Kafsack, B. F., Painter, H. J., and Llinas, M., 2012. New Agilent platform DNA microarrays for transcriptome analysis of Plasmodium falciparum and Plasmodium berghei for the malaria research community. Malar J 11**,** 187.

Kafsack, B. F., Rovira-Graells, N., Clark, T. G., Bancells, C., Crowley, V. M., Campino, S. G., Williams, A. E., Drought, L. G., Kwiatkowski, D. P., Baker, D. A., Cortes, A., and Llinas, M., 2014. A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature 507**,** 248-252.

Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28**,** 27-30.

Kapler, G. M. and Beverley, S. M., 1989. Transcriptional mapping of the amplified region encoding the dihydrofolate reductase-thymidylate synthase of Leishmania major reveals a high density of transcripts, including overlapping and antisense RNAs. Mol Cell Biol 9**,** 3959-3972.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J., Hume, D. A., Lipovich, L., Batalov, S., Engstrom, P. G., Mizuno, Y., Faghihi, M. A., Sandelin, A., Chalk, A. M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., and Wahlestedt, C., 2005. Antisense transcription in the mammalian transcriptome. Science 309**,** 1564-1566.

Kaviratne, M., Khan, S. M., Jarra, W., and Preiser, P. R., 2002. Small variant STEVOR antigen is uniquely located within Maurer's clefts in Plasmodium falciparum-infected red blood cells. Eukaryot Cell 1**,** 926-935.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler, and David, 2002. The Human Genome Browser at UCSC. Genome Research 12**,** 996-1006.

Kidgell, C., Volkman, S. K., Daily, J., Borevitz, J. O., Plouffe, D., Zhou, Y., Johnson, J. R., Le Roch, K., Sarr, O., Ndir, O., Mboup, S., Batalov, S., Wirth, D. F., and Winzeler, E. A., 2006. A systematic map of genetic variation in Plasmodium falciparum. PLoS Pathog 2**,** e57.

Kirchgatter, K. and 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**,** 16-23.

Kitano, H., 2002. Systems biology: a brief overview. Science 295**,** 1662-1664.

Kochar, D. K., Das, A., Kochar, S. K., Saxena, V., Sirohi, P., Garg, S., Kochar, A., Khatri, M. P., and Gupta, V., 2009. Severe Plasmodium vivax malaria: a report on serial cases from Bikaner in northwestern India. Am J Trop Med Hyg 80**,** 194-198.

Kochar, D. K., Kochar, S. K., Agrawal, R. P., Sabir, M., Nayak, K. C., Agrawal, T. D., Purohit, V. P., and Gupta, R. P., 2006. The changing spectrum of severe falciparum malaria: a clinical study from Bikaner (northwest India). J Vector Borne Dis 43**,** 104-108.

Kochar, D. K., Saxena, V., Singh, N., Kochar, S. K., Kumar, S. V., and Das, A., 2005. Plasmodium vivax malaria. Emerg Infect Dis 11**,** 132-134.

Kochar, D. K., Tanwar, G. S., Khatri, P. C., Kochar, S. K., Sengar, G. S., Gupta, A., Kochar, A., Middha, S., Acharya, J., Saxena, V., Pakalapati, D., Garg, S., and Das, A., 2010. Clinical features of children hospitalized with malaria--a study from Bikaner, northwest India. Am J Trop Med Hyg 83**,** 981-989.

Kraemer, S. M. and Smith, J. D., 2006. A family affair: var genes, PfEMP1 binding, and malaria disease. Curr Opin Microbiol 9**,** 374-380.

Krungkrai, J., Prapunwattana, P., and Krungkrai, S. R., 2000. Ultrastructure and function of mitochondria in gametocytic stage of Plasmodium falciparum. Parasite 7**,** 19-26.

Kumar, A., Dua, V. K., and Rathod, P. K., 2011. Malaria-attributed death rates in India. Lancet 377**,** 991-992; author reply 994-995.

Kyes, S., Christodoulou, Z., Pinches, R., and Newbold, C., 2002. Stage-specific merozoite surface protein 2 antisense transcripts in Plasmodium falciparum. Mol Biochem Parasitol 123**,** 79-83.

Kyes, S., Horrocks, P., and Newbold, C., 2001. Antigenic variation at the infected red cell surface in malaria. Annu Rev Microbiol 55**,** 673-707.

Kyes, S. A., Rowe, J. A., Kriek, N., and Newbold, C. I., 1999. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. Proc Natl Acad Sci U S A 96**,** 9333-9338.

Kyriacou, H. M., Stone, G. N., Challis, R. J., Raza, A., Lyke, K. E., Thera, M. A., Kone, A. K., Doumbo, O. K., Plowe, C. V., and Rowe, J. A., 2006. Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. Mol Biochem Parasitol 150**,** 211-218.

LaCount, D. J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J. R., Schoenfeld, L. W., Ota, I., Sahasrabudhe, S., Kurschner, C., Fields, S., and Hughes, R. E., 2005. A protein interaction network of the malaria parasite Plasmodium falciparum. Nature 438**,** 103-107.

Laishram, D. D., Sutton, P. L., Nanda, N., Sharma, V. L., Sobti, R. C., Carlton, J. M., and Joshi, H., 2012. The complexities of malaria disease manifestations with a focus on asymptomatic malaria. Malar J 11**,** 29.

Lang-Unnasch, N. and Murphy, A. D., 1998. Metabolic changes of the malaria parasite during the transition from the human to the mosquito host. Annu Rev Microbiol 52**,** 561-590.

Langfelder, P. and Horvath, S., 2007. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 1**,** 54.

Langfelder, P. and Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9**,** 559.

Lanzer, M., Wickert, H., Krohne, G., Vincensini, L., and Braun Breton, C., 2006. Maurer's clefts: a novel multi-functional organelle in the cytoplasm of Plasmodium falciparum-infected erythrocytes. Int J Parasitol 36**,** 23-36.

Lapidot, M. and Pilpel, Y., 2006. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep 7**,** 1216-1222.

Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G., and Mann, M., 2002. Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature 419**,** 537-542.

Lavazec, C., Sanyal, S., and Templeton, T. J., 2006. Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. Nucleic Acids Res 34**,** 6696-6707.

Lavazec, C., Sanyal, S., and Templeton, T. J., 2007. Expression switching in the stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. Mol Microbiol 64**,** 1621-1634.

Laveran, A., 1881. Un nouveau parasite trouvé dans le sang de malades attaints de fièvre palustre. Origine parasitaire des accidents de l'impaludisme. Bull Mém Soc Méd Hôpitaux Paris 17, 158-164.

Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C. M., and Casari, G., 2004. In search of antisense. Trends Biochem Sci 29**,** 88-94.

Lavstsen, T., Salanti, A., Jensen, A. T., Arnot, D. E., and Theander, T. G., 2003. Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. Malar J 2**,** 27.

Le Roch, K. G., Johnson, J. R., Ahiboh, H., Chung, D. W., Prudhomme, J., Plouffe, D., Henson, K., Zhou, Y., Witola, W., Yates, J. R., Mamoun, C. B., Winzeler, E. A., and Vial, H., 2008. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum. BMC Genomics 9**,** 513.

Le Roch, K. G., Johnson, J. R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S. F., Williamson, K. C., Holder, A. A., Carucci, D. J., Yates, J. R., 3rd, and Winzeler, E. A., 2004. Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. Genome Res 14**,** 2308-2318.

Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., and Winzeler, E. A., 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 301**,** 1503-1508.

LeRoux, M., Lakshmanan, V., and Daily, J. P., 2009. Plasmodium falciparum biology: analysis of in vitro versus in vivo growth conditions. Trends Parasitol 25**,** 474-481.

Liniger, M., Bodenmuller, K., Pays, E., Gallati, S., and Roditi, I., 2001. Overlapping sense and antisense transcription units in Trypanosoma brucei. Mol Microbiol 40**,** 869-878.

Llinas, M., Bozdech, Z., Wong, E. D., Adai, A. T., and DeRisi, J. L., 2006. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. Nucleic Acids Res 34**,** 1166-1173.

Llinas, M. and DeRisi, J. L., 2004. Pernicious plans revealed: Plasmodium falciparum genome wide expression analysis. Curr Opin Microbiol 7**,** 382-387.

Lopez-Barragan, M. J., Lemieux, J., Quinones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., and Su, X. Z., 2011. Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. BMC Genomics 12**,** 587.

Lopez-Rubio, J. J., Mancio-Silva, L., and Scherf, A., 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5**,** 179-190.

Lu, F., Jiang, H., Ding, J., Mu, J., Valenzuela, J. G., Ribeiro, J. M., and Su, X. Z., 2007. cDNA sequences reveal considerable gene prediction inaccuracy in the Plasmodium falciparum genome. BMC Genomics 8**,** 255.

Luo, C., Sidote, D. J., Zhang, Y., Kerstetter, R. A., Michael, T. P., and Lam, E., 2012. Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production. Plant J.

Mackinnon, M. J., Li, J., Mok, S., Kortok, M. M., Marsh, K., Preiser, P. R., and Bozdech, Z., 2009. Comparative transcriptional and genomic analysis of Plasmodium falciparum field isolates. PLoS Pathog 5**,** e1000644.

Maere, S., Heymans, K., and Kuiper, M., 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21**,** 3448-3449.

Mahan, M. J., Slauch, J. M., Hanna, P. C., Camilli, A., Tobias, J. W., Waldor, M. K., and Mekalanos, J. J., 1993. Selection for bacterial genes that are specifically induced in host tissues: the hunt for virulence factors. Infect Agents Dis 2**,** 263-268.

Maier, A., Rug, M., Oneill, M., Brown, M., Chakravorty, S., Szestak, T., Chesson, J., Wu, Y., Hughes, K., and Coppel, R., 2008. Exported Proteins Required for Virulence and Rigidity of Plasmodium falciparum-Infected Human Erythrocytes. Cell 134**,** 48 - 61.

Maier, A. G., Cooke, B. M., Cowman, A. F., and Tilley, L., 2009. Malaria parasite proteins that remodel the host erythrocyte. Nat Rev Micro 7**,** 341-354.

Maier, A. G., Rug, M., O'Neill, M. T., Brown, M., Chakravorty, S., Szestak, T., Chesson, J., Wu, Y., Hughes, K., Coppel, R. L., Newbold, C., Beeson, J. G., Craig, A., Crabb, B. S., and

Cowman, A. F., 2008. Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes. Cell 134**,** 48-61.

Mair, G. R., Braks, J. A., Garver, L. S., Wiegant, J. C., Hall, N., Dirks, R. W., Khan, S. M., Dimopoulos, G., Janse, C. J., and Waters, A. P., 2006. Regulation of sexual development of Plasmodium by translational repression. Science 313**,** 667-669.

Mair, G. R., Lasonder, E., Garver, L. S., Franke-Fayard, B. M., Carret, C. K., Wiegant, J. C., Dirks, R. W., Dimopoulos, G., Janse, C. J., and Waters, A. P., 2010. Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development. PLoS Pathog 6**,** e1000767.

Mamoun, C. B. and Goldberg, D. E., 2001. Plasmodium protein phosphatase 2C dephosphorylates translation elongation factor 1beta and inhibits its PKC-mediated nucleotide exchange activity in vitro. Mol Microbiol 39**,** 973-981.

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., O'Brien, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J. B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S. M., Marsh, K., Jiang, H., Su, X. Z., Amaratunga, C., Fairhurst, R., Socheat, D., Nosten, F., Imwong, M., White, N. J., Sanders, M., Anastasi, E., Alcock, D., Drury, E., Oyola, S., Quail, M. A., Turner, D. J., Ruano-Rubio, V., Jyothi, D., Amenga-Etego, L., Hubbart, C., Jeffreys, A., Rowlands, K., Sutherland, C., Roper, C., Mangano, V., Modiano, D., Tan, J. C., Ferdig, M. T., Amambua-Ngwa, A., Conway, D. J., Takala-Harrison, S., Plowe, C. V., Rayner, J. C., Rockett, K. A., Clark, T. G., Newbold, C. I., Berriman, M., MacInnis, B., and Kwiatkowski, D. P., 2012. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature 487**,** 375-379.

Marti, M., Baum, J., Rug, M., Tilley, L., and Cowman, A. F., 2005. Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. The Journal of Cell Biology 171**,** 587-592.

Marti, M., Baum, J., Rug, M., Tilley, L., and Cowman, A. F., 2005. Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. J Cell Biol 171**,** 587-592.

Marti, M., Good, R. T., Rug, M., Knuepfer, E., and Cowman, A. F., 2004. Targeting Malaria Virulence and Remodeling Proteins to the Host Erythrocyte. Science 306**,** 1930-1933.

Martin, R. E., Henry, R. I., Abbey, J. L., Clements, J. D., and Kirk, K., 2005. The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of Plasmodium falciparum. Genome Biol 6**,** R26.

Mendis, K., Rietveld, A., Warsame, M., Bosman, A., Greenwood, B., and Wernsdorfer, W. H., 2009. From malaria control to eradication: The WHO perspective. Trop Med Int Health 14**,** 802-809.

Mi-Ichi, F., Kita, K., and Mitamura, T., 2006. Intraerythrocytic Plasmodium falciparum utilize a broad range of serum-derived fatty acids with limited modification for their growth. Parasitology 133**,** 399-410.

Miao, J., Fan, Q., Parker, D., Li, X., Li, J., and Cui, L., 2013. Puf mediates translation repression of transmission-blocking vaccine candidates in malaria parasites. PLoS Pathog 9**,** e1003268.

Militello, K. T., Patel, V., Chessler, A. D., Fisher, J. K., Kasper, J. M., Gunasekera, A., and Wirth, D. F., 2005. RNA polymerase II synthesizes antisense RNA in Plasmodium falciparum. RNA 11**,** 365-370.

Militello, K. T., Refour, P., Comeaux, C. A., and Duraisingh, M. T., 2008. Antisense RNA and RNAi in protozoan parasites: working hard or hardly working? Mol Biochem Parasitol 157**,** 117-126.

Miller, J. A., Horvath, S., and Geschwind, D. H., 2010. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci U S A 107**,** 12698-12703.

Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K., 2002. The pathogenic basis of malaria. Nature 415**,** 673-679.

Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K., 2002. The pathogenic basis of malaria. Nature 415**,** 673-679.

Mitamura, T., Hanada, K., Ko-Mitamura, E. P., Nishijima, M., and Horii, T., 2000. Serum factors governing intraerythrocytic development and cell cycle progression of Plasmodium falciparum. Parasitol Int 49**,** 219-229.

Moazed, D., 2009. Small RNAs in transcriptional gene silencing and genome defence. Nature 457**,** 413-420.

Moriyama, Y., Hayashi, M., Yatsushiro, S., and Yamamoto, A., 2003. Vacuolar proton pumps in malaria parasite cells. J Bioenerg Biomembr 35**,** 367-375.

Mota, M. M., Hafalla, J. C., and Rodriguez, A., 2002. Migration through host cells activates Plasmodium sporozoites for infection. Nat Med 8**,** 1318-1322.

Muller, S., 2004. Redox and antioxidant systems of the malaria parasite Plasmodium falciparum. Mol Microbiol 53**,** 1291-1305.

Murphy, G. S. and Oldfield, E. C., 3rd, 1996. Falciparum malaria. Infect Dis Clin North Am 10**,** 747-775.

Murray, C. J., Rosenfeld, L. C., Lim, S. S., Andrews, K. G., Foreman, K. J., Haring, D., Fullman, N., Naghavi, M., Lozano, R., and Lopez, A. D., 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. Lancet 379**,** 413-431.

Nagaraj, V. A., Arumugam, R., Chandra, N. R., Prasad, D., Rangarajan, P. N., and Padmanaban, G., 2009. Localisation of Plasmodium falciparum uroporphyrinogen III decarboxylase of the heme-biosynthetic pathway in the apicoplast and characterisation of its catalytic properties. Int J Parasitol 39**,** 559-568.

Natalang, O., Bischoff, E., Deplaine, G., Proux, C., Dillies, M. A., Sismeiro, O., Guigon, G., Bonnefoy, S., Patarapotikul, J., Mercereau-Puijalon, O., Coppee, J. Y., and David, P. H., 2008. Dynamic RNA profiling in Plasmodium falciparum synchronized blood stages exposed to lethal doses of artesunate. BMC Genomics 9**,** 388.

Newby, Z. E., O'Connell, J., 3rd, Robles-Colmenares, Y., Khademi, S., Miercke, L. J., and Stroud, R. M., 2008. Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite Plasmodium falciparum. Nat Struct Mol Biol 15**,** 619-625.

Ni, T., Tu, K., Wang, Z., Song, S., Wu, H., Xie, B., Scott, K. C., Grewal, S. I., Gao, Y., and Zhu, J., 2010. The prevalence and regulation of antisense transcripts in Schizosaccharomyces pombe. PLoS ONE 5**,** e15271.

Niang, M., Yan Yam, X., and Preiser, P. R., 2009. The Plasmodium falciparum STEVOR multigene family mediates antigenic variation of the infected erythrocyte. PLoS Pathog 5**,** e1000307.

Niang, M., Bei, A. K., Madnani, K. G., Pelly, S., Dankwa, S., Kanjee, U., Gunalan, K., Amaladoss, A., Yeo, K. P., Bob, N. S., Malleret, B., Duraisingh, M. T., and Preiser, P. R.,

2014. STEVOR Is a Plasmodium falciparum Erythrocyte Binding Protein that Mediates Merozoite Invasion and Rosetting. Cell Host Microbe 16**,** 81-93.

Nielsen, M. A., Staalsoe, T., Kurtzhals, J. A., Goka, B. Q., Dodoo, D., Alifrangis, M., Theander, T. G., Akanmori, B. D., and Hviid, L., 2002. Plasmodium falciparum variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. J Immunol 168**,** 3444-3450.

Nunes, M. C., Okada, M., Scheidig-Benatar, C., Cooke, B. M., and Scherf, A., 2010. Plasmodium falciparum FIKK kinase members target distinct components of the erythrocyte membrane. PLoS ONE 5**,** e11747.

Nurse, P., 2003. Systems biology: understanding cells. Nature 424**,** 883.

Oakley, M. S., Kumar, S., Anantharaman, V., Zheng, H., Mahajan, B., Haynes, J. D., Moch, J. K., Fairhurst, R., McCutchan, T. F., and Aravind, L., 2007. Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic Plasmodium falciparum parasites. Infect Immun 75**,** 2012-2025.

Oh, S. S., Voigt, S., Fisher, D., Yi, S. J., LeRoy, P. J., Derick, L. H., Liu, S., and Chishti, A. H., 2000. Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. Mol Biochem Parasitol 108**,** 237-247.

Okoyeh, J. N., Pillai, C. R., and Chitnis, C. E., 1999. Plasmodium falciparum field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of glycophorin A. Infect Immun 67**,** 5784-5791.

Oldham, M. C., Horvath, S., and Geschwind, D. H., 2006. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc Natl Acad Sci U S A 103**,** 17973-17978.

Olszewski, K. L. and Llinas, M., 2011. Central carbon metabolism of Plasmodium parasites. Mol Biochem Parasitol 175**,** 95-103.

Olszewski, K. L., Morrisey, J. M., Wilinski, D., Burns, J. M., Vaidya, A. B., Rabinowitz, J. D., and Llinas, M., 2009. Host-parasite interactions revealed by Plasmodium falciparum metabolomics. Cell Host Microbe 5**,** 191-199.

Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Bohme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., Newbold, C., and Llinas, M., 2010. New insights into the

blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol Microbiol 76**,** 12-24.

Painter, H. J., Campbell, T. L., and Llinas, M., 2011. The Apicomplexan AP2 family: integral factors regulating Plasmodium development. Mol Biochem Parasitol 176**,** 1-7.

Pakalapati, D., Garg, S., Middha, S., Acharya, J., Subudhi, A. K., Boopathi, A. P., Saxena, V., Kochar, S. K., Kochar, D. K., and Das, A., 2013. Development and evaluation of a 28S rRNA gene-based nested PCR assay for P. falciparum and P. vivax. Pathog Glob Health 107**,** 180-188.

Pakalapati, D., Garg, S., Middha, S., Kochar, A., Subudhi, A. K., Arunachalam, B. P., Kochar, S. K., Saxena, V., Pareek, R., Acharya, J., Kochar, D. K., and Das, A., 2013. Comparative evaluation of microscopy, OptiMAL(R) and 18S rRNA gene based multiplex PCR for detection of Plasmodium falciparum &Plasmodium vivax from field isolates of Bikaner, India. Asian Pac J Trop Med 6**,** 346-351.

Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M., and Wirth, D. F., 2001. Serial analysis of gene expression in Plasmodium falciparum reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. Mol Biol Cell 12**,** 3114-3125.

Pelechano, V. and Steinmetz, L. M., 2013. Gene regulation by antisense transcription. Nat Rev Genet 14**,** 880-893.

Perez-Toledo, K., Rojas-Meza, A. P., Mancio-Silva, L., Hernandez-Cuevas, N. A., Delgadillo, D. M., Vargas, M., Martinez-Calvillo, S., Scherf, A., and Hernandez-Rivas, R., 2009. Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. Nucleic Acids Res 37**,** 2596-2606.

Petter, M., Bonow, I., and Klinkert, M. Q., 2008. Diverse expression patterns of subgroups of the rif multigene family during Plasmodium falciparum gametocytogenesis. PLoS ONE 3**,** e3779.

Petter, M., Haeggstrom, M., Khattab, A., Fernandez, V., Klinkert, M. Q., and Wahlgren, M., 2007. Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. Mol Biochem Parasitol 156**,** 51-61.

Planche, T. and Krishna, S., 2006. Severe malaria: metabolic complications. Curr Mol Med 6**,** 141-153.

Pologe, L. G. and Ravetch, J. V., 1988. Large deletions result from breakage and healing of P. falciparum chromosomes. Cell 55**,** 869-874.

Ponts, N., Fu, L., Harris, E. Y., Zhang, J., Chung, D. W., Cervantes, M. C., Prudhomme, J., Atanasova-Penichon, V., Zehraoui, E., Bunnik, E. M., Rodrigues, E. M., Lonardi, S., Hicks, G. R., Wang, Y., and Le Roch, K. G., 2013. Genome-wide mapping of DNA methylation in the human malaria parasite Plasmodium falciparum. Cell Host Microbe 14**,** 696-706.

Pradel, G. and Frevert, U., 2001. Malaria sporozoites actively enter and pass through rat Kupffer cells prior to hepatocyte invasion. Hepatology 33**,** 1154-1165.

Prasanth, K. V. and Spector, D. L., 2007. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev 21**,** 11-42.

Przyborski, J. M., Wickert, H., Krohne, G., and Lanzer, M., 2003. Maurer's clefts--a novel secretory organelle? Mol Biochem Parasitol 132**,** 17-26.

Raabe, C. A., Sanchez, C. P., Randau, G., Robeck, T., Skryabin, B. V., Chinni, S. V., Kube, M., Reinhardt, R., Ng, G. H., Manickam, R., Kuryshev, V. Y., Lanzer, M., Brosius, J., Tang, T. H., and Rozhdestvensky, T. S., 2010. A global view of the nonprotein-coding transcriptome in Plasmodium falciparum. Nucleic Acids Res 38**,** 608-617.

Radke, J. R., Behnke, M. S., Mackey, A. J., Radke, J. B., Roos, D. S., and White, M. W., 2005. The transcriptome of Toxoplasma gondii. BMC Biol 3**,** 26.

Ralph, S. A., Bischoff, E., Mattei, D., Sismeiro, O., Dillies, M. A., Guigon, G., Coppee, J. Y., David, P. H., and Scherf, A., 2005. Transcriptome analysis of antigenic variation in Plasmodium falciparum--var silencing is not dependent on antisense RNA. Genome Biol 6**,** R93.

Ramaprasad, A., Pain, A., and Ravasi, T., 2012. Defining the protein interaction network of human malaria parasite Plasmodium falciparum. Genomics 99**,** 69-75.

Reed, J. L., Famili, I., Thiele, I., and Palsson, B. O., 2006. Towards multidimensional genome annotation. Nat Rev Genet 7**,** 130-141.

Robinson, B. A., Welch, T. L., and Smith, J. D., 2003. Widespread functional specialization of Plasmodium falciparum erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol Microbiol 47**,** 1265-1278.

Rogerson, S. J., Hviid, L., Duffy, P. E., Leke, R. F., and Taylor, D. W., 2007. Malaria in pregnancy: pathogenesis and immunity. Lancet Infect Dis 7**,** 105-117.

Rogozin, I. B., Spiridonov, A. N., Sorokin, A. V., Wolf, Y. I., Jordan, I. K., Tatusov, R. L., and Koonin, E. V., 2002. Purifying and directional selection in overlapping prokaryotic genes. Trends Genet 18**,** 228-232.

Rottmann, M., Lavstsen, T., Mugasa, J. P., Kaestli, M., Jensen, A. T., Muller, D., Theander, T., and Beck, H. P., 2006. Differential expression of var gene groups is associated with morbidity caused by Plasmodium falciparum infection in Tanzanian children. Infect Immun 74**,** 3904-3911.

Rowe, J. A., Claessens, A., Corrigan, R. A., and Arman, M., 2009. Adhesion of Plasmodium falciparum-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. Expert Rev Mol Med 11**,** e16.

Rowe, J. A. and Kyes, S. A., 2004. The role of Plasmodium falciparum var genes in malaria in pregnancy. Mol Microbiol 53**,** 1011-1019.

Salanti, A., Dahlback, M., Turner, L., Nielsen, M. A., Barfod, L., Magistrado, P., Jensen, A. T., Lavstsen, T., Ofori, M. F., Marsh, K., Hviid, L., and Theander, T. G., 2004. Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. J Exp Med 200**,** 1197-1203.

Salanti, A., Staalsoe, T., Lavstsen, T., Jensen, A. T., Sowa, M. P., Arnot, D. E., Hviid, L., and Theander, T. G., 2003. Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria. Mol Microbiol 49**,** 179-191.

Salcedo-Amaya, A. M., van Driel, M. A., Alako, B. T., Trelle, M. B., van den Elzen, A. M., Cohen, A. M., Janssen-Megens, E. M., van de Vegte-Bolmer, M., Selzer, R. R., Iniguez, A. L., Green, R. D., Sauerwein, R. W., Jensen, O. N., and Stunnenberg, H. G., 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum. Proc Natl Acad Sci U S A 106**,** 9655-9660.

Saliba, K. J., Allen, R. J., Zissis, S., Bray, P. G., Ward, S. A., and Kirk, K., 2003. Acidification of the malaria parasite's digestive vacuole by a H+-ATPase and a H+-pyrophosphatase. J Biol Chem 278**,** 5605-5612.

Sargeant, T., Marti, M., Caler, E., Carlton, J., Simpson, K., Speed, T., and Cowman, A., 2006. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome Biology 7**,** R12.

Sargeant, T. J., Marti, M., Caler, E., Carlton, J. M., Simpson, K., Speed, T. P., and Cowman, A. F., 2006. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome Biol 7**,** R12.

Saxena, V., Garg, S., Tripathi, J., Sharma, S., Pakalapati, D., Subudhi, A. K., Boopathi, P. A., Saggu, G. S., Kochar, S. K., Kochar, D. K., and Das, A., 2012. Plasmodium vivax apicoplast genome: A comparative analysis of major genes from Indian field isolates. Acta Trop 122**,** 138-149.

Scherf, A., Lopez-Rubio, J. J., and Riviere, L., 2008. Antigenic variation in Plasmodium falciparum. Annu Rev Microbiol 62**,** 445-470.

Scherf, A. and Mattei, D., 1992. Cloning and characterization of chromosome breakpoints of Plasmodium falciparum: breakage and new telomere formation occurs frequently and randomly in subtelomeric genes. Nucleic Acids Res 20**,** 1491-1496.

Serghides, L., Smith, T. G., Patel, S. N., and Kain, K. C., 2003. CD36 and malaria: friends or foes? Trends Parasitol 19**,** 461-469.

Sharma, V.P., Jha, P., Dhingra, N., Jotkar, R.M., Peto, R., 2011. Malaria-attributed death rates in India. The Lancet 377, 994–995.

Shearwin, K. E., Callen, B. P., and Egan, J. B., 2005. Transcriptional interference--a crash course. Trends Genet 21**,** 339-345.

Shock, J. L., Fischer, K. F., and DeRisi, J. L., 2007. Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. Genome Biol 8**,** R134.

Shortt, H. E., Bray, R. S., and Cooper, W., 1954. Further notes on the tissue stages of Plasmodium cynomolgi. Trans R Soc Trop Med Hyg 48**,** 122-131.

Shortt, H. E., Fairley, N. H., and et al., 1949. The pre-erythrocytic stage of Plasmodium falciparum; a preliminary note. Br Med J 2**,** 1006-1008, illust.

Shortt, H. E. and Garnham, P. C., 1948a. The exoerythrocytic parasites of Plasmodium cynomolgi. Trans R Soc Trop Med Hyg 41**,** 705-716.

Shortt, H. E. and Garnham, P. C., 1948b. The pre-erythrocytic development of Plasmodium cynomolgi and Plasmodium vivax. Trans R Soc Trop Med Hyg 41**,** 785-795.

Shute, P. G. and Maryon, M., 1956. Is the malaria parasite within or upon the red blood corpuscle; with particular reference to the significance of stippling and other morphological changes observed in the host cell. Trans R Soc Trop Med Hyg 50**,** 139-149.

Siau, A., Silvie, O., Franetich, J. F., Yalaoui, S., Marinach, C., Hannoun, L., van Gemert, G. J., Luty, A. J., Bischoff, E., David, P. H., Snounou, G., Vaquero, C., Froissard, P., and Mazier, D., 2008. Temperature shift and host cell contact up-regulate sporozoite expression of Plasmodium falciparum genes involved in hepatocyte infection. PLoS Pathog 4**,** e1000121.

Siau, A., Toure, F. S., Ouwe-Missi-Oukem-Boyer, O., Ciceron, L., Mahmoudi, N., Vaquero, C., Froissard, P., Bisvigou, U., Bisser, S., Coppee, J. Y., Bischoff, E., David, P. H., and Mazier, D., 2007. Whole-transcriptome analysis of Plasmodium falciparum field isolates: identification of new pathogenicity factors. J Infect Dis 196**,** 1603-1612.

Silamut, K. and White, N. J., 1993. Relation of the stage of parasite development in the peripheral blood to prognosis in severe falciparum malaria. Trans R Soc Trop Med Hyg 87**,** 436-443.

Silva, M. D., Cooke, B. M., Guillotte, M., Buckingham, D. W., Sauzet, J. P., Le Scanf, C., Contamin, H., David, P., Mercereau-Puijalon, O., and Bonnefoy, S., 2005. A role for the Plasmodium falciparum RESA protein in resistance against heat shock demonstrated using gene disruption. Mol Microbiol 56**,** 990-1003.

Silvestrini, F., Alano, P., and Williams, J. L., 2000. Commitment to the production of male and female gametocytes in the human malaria parasite Plasmodium falciparum. Parasitology 121 Pt 5**,** 465-471.

Silvestrini, F., Bozdech, Z., Lanfrancotti, A., Di Giulio, E., Bultrini, E., Picci, L., Derisi, J. L., Pizzi, E., and Alano, P., 2005. Genome-wide identification of genes upregulated at the

onset of gametocytogenesis in Plasmodium falciparum. Mol Biochem Parasitol 143**,** 100-110.

Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S. S., Cox-Singh, J., Thomas, A., and Conway, D. J., 2004. A large focus of naturally acquired Plasmodium knowlesi infections in human beings. Lancet 363**,** 1017-1024.

Singh, V., Mishra, N., Awasthi, G., Dash, A. P., and Das, A., 2009. Why is it important to study malaria epidemiology in India? Trends Parasitol 25**,** 452-457.

Smith, J. D., Chitnis, C. E., Craig, A. G., Roberts, D. J., Hudson-Taylor, D. E., Peterson, D. S., Pinches, R., Newbold, C. I., and Miller, L. H., 1995. Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 82**,** 101-110.

Smith, J. D., Craig, A. G., Kriek, N., Hudson-Taylor, D., Kyes, S., Fagan, T., Pinches, R., Baruch, D. I., Newbold, C. I., and Miller, L. H., 2000a. Identification of a Plasmodium falciparum intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci U S A 97**,** 1766-1771.

Smith, J. D., Gamain, B., Baruch, D. I., and Kyes, S., 2001. Decoding the language of var genes and Plasmodium falciparum sequestration. Trends Parasitol 17**,** 538-545.

Smith, J. D., Subramanian, G., Gamain, B., Baruch, D. I., and Miller, L. H., 2000b. Classification of adhesive domains in the Plasmodium falciparum erythrocyte membrane protein 1 family. Mol Biochem Parasitol 110**,** 293-310.

Smith, T. G., Lourenco, P., Carter, R., Walliker, D., and Ranford-Cartwright, L. C., 2000. Commitment to sexual differentiation in the human malaria parasite, Plasmodium falciparum. Parasitology 121 (Pt 2)**,** 127-133.

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T., 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27**,** 431-432.

Snounou, G., Zhu, X., Siripoon, N., Jarra, W., Thaithong, S., Brown, K. N., and Viriyakosol, S., 1999. Biased distribution of msp1 and msp2 allelic variants in Plasmodium falciparum populations in Thailand. Trans R Soc Trop Med Hyg 93**,** 369-374.

Sorber, K., Dimon, M. T., and DeRisi, J. L., 2011. RNA-Seq analysis of splicing in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. Nucleic Acids Res 39**,** 3820-3835.

Spencer, C. A., Gietz, R. D., and Hodgetts, R. B., 1986. Overlapping transcription units in the dopa decarboxylase region of Drosophila. Nature 322**,** 279-281.

Spielmann, T. and Gilberger, T. W., 2010. Protein export in malaria parasites: do multiple export motifs add up to multiple export pathways? Trends Parasitol 26**,** 6-10.

Spielmann, T., Hawthorne, P. L., Dixon, M. W., Hannemann, M., Klotz, K., Kemp, D. J., Klonis, N., Tilley, L., Trenholme, K. R., and Gardiner, D. L., 2006. A cluster of ring stage-specific genes linked to a locus implicated in cytoadherence in Plasmodium falciparum codes for PEXEL-negative and PEXEL-positive proteins exported into the host cell. Mol Biol Cell 17**,** 3613-3624.

Springer, A. L., Smith, L. M., Mackay, D. Q., Nelson, S. O., and Smith, J. D., 2004. Functional interdependence of the DBLbeta domain and c2 region for binding of the Plasmodium falciparum variant antigen to ICAM-1. Mol Biochem Parasitol 137**,** 55-64.

Stephens, JWW., 1922. A new malaria parasite of man. Ann Trop Med Parasitol, 16, 383-388.

Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., Krueger, A., Pollok, J. M., Menard, R., and Heussler, V. T., 2006. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. Science 313**,** 1287-1290.

Su, X. Z., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. A., and Wellems, T. E., 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. Cell 82**,** 89-100.

Surolia, N. and Surolia, A., 2001. Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of Plasmodium falciparum. Nat Med 7**,** 167-173.

Tamez, P. A., Bhattacharjee, S., van Ooij, C., Hiller, N. L., Llinas, M., Balu, B., Adams, J. H., and Haldar, K., 2008. An erythrocyte vesicle protein exported by the malaria parasite promotes tubovesicular lipid import from the host cell surface. PLoS Pathog 4**,** e1000118.

Tarun, A. S., Peng, X., Dumpit, R. F., Ogata, Y., Silva-Rivera, H., Camargo, N., Daly, T. M., Bergman, L. W., and Kappe, S. H., 2008. A combined transcriptome and proteome survey of malaria parasite liver stages. Proc Natl Acad Sci U S A 105**,** 305-310.

Tewari, R., Straschil, U., Bateman, A., Bohme, U., Cherevach, I., Gong, P., Pain, A., and Billker, O., 2010. The systematic functional analysis of Plasmodium protein kinases identifies essential regulators of mosquito transmission. Cell Host Microbe 8**,** 377-387.

Tham, W. H., Payne, P. D., Brown, G. V., and Rogerson, S. J., 2007. Identification of basic transcriptional elements required for rif gene transcription. Int J Parasitol 37**,** 605-615.

Thomason, M. K. and Storz, G., 2010. Bacterial antisense RNAs: how many are there, and what are they doing? Annu Rev Genet 44**,** 167-188.

Trager, W. and Jensen, J. B., 1976. Human malaria parasites in continuous culture. Science 193**,** 673-675.

Trampuz, A., Jereb, M., Muzlovic, I., and Prabhu, R. M., 2003. Clinical review: Severe malaria. Crit Care 7**,** 315-323.

Trampuz, A., Osmon, D. R., Hanssen, A. D., Steckelberg, J. M., and Patel, R., 2003. Molecular and antibiofilm approaches to prosthetic joint infection. Clin Orthop Relat Res**,** 69-88.

Tuikue Ndam, N., Bischoff, E., Proux, C., Lavstsen, T., Salanti, A., Guitard, J., Nielsen, M. A., Coppee, J. Y., Gaye, A., Theander, T., David, P. H., and Deloron, P., 2008. Plasmodium falciparum transcriptome analysis reveals pregnancy malaria associated gene expression. PLoS ONE 3**,** e1855.

Ullu, E., Lujan, H. D., and Tschudi, C., 2005. Small sense and antisense RNAs derived from a telomeric retroposon family in Giardia intestinalis. Eukaryot Cell 4**,** 1155-1157.

van Ooij, C., Tamez, P., Bhattacharjee, S., Hiller, N. L., Harrison, T., Liolios, K., Kooij, T., Ramesar, J., Balu, B., Adams, J., Waters, A. P., Janse, C. J., and Haldar, K., 2008. The malaria secretome: from algorithms to essential function in blood stage infection. PLoS Pathog 4**,** e1000084.

Vanderberg, J. P. and Frevert, U., 2004. Intravital microscopy demonstrating antibody-mediated immobilisation of Plasmodium berghei sporozoites injected into skin by mosquitoes. Int J Parasitol 34**,** 991-996.

Vaughan, A. M., O'Neill, M. T., Tarun, A. S., Camargo, N., Phuong, T. M., Aly, A. S., Cowman, A. F., and Kappe, S. H., 2009. Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. Cell Microbiol 11**,** 506-520.

Vincensini, L., Richert, S., Blisnick, T., Van Dorsselaer, A., Leize-Wagner, E., Rabilloud, T., and Braun Breton, C., 2005. Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering Plasmodium falciparum proteins to the surface of its host cell. Mol Cell Proteomics 4**,** 582-593.

Vinetz, J. M., Dave, S. K., Specht, C. A., Brameld, K. A., Xu, B., Hayward, R., and Fidock, D. A., 1999. The chitinase PfCHT1 from the human malaria parasite Plasmodium falciparum lacks proenzyme and chitin-binding domains and displays unique substrate preferences. Proc Natl Acad Sci U S A 96**,** 14061-14066.

Vitoria, M., Granich, R., Gilks, C. F., Gunneberg, C., Hosseini, M., Were, W., Raviglione, M., and De Cock, K. M., 2009. The global fight against HIV/AIDS, tuberculosis, and malaria: current status and future perspectives. Am J Clin Pathol 131**,** 844-848.

Volkman, S. K., Sabeti, P. C., DeCaprio, D., Neafsey, D. E., Schaffner, S. F., Milner, D. A., Jr., Daily, J. P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M. T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D. B., Zainoun, J., Wiegand, R. C., Birren, B. W., Hartl, D. L., Galagan, J. E., Lander, E. S., and Wirth, D. F., 2007. A genome-wide map of diversity in Plasmodium falciparum. Nat Genet 39**,** 113-119.

Wagner, E. G. and Simons, R. W., 1994. Antisense RNA control in bacteria, phages, and plasmids. Annu Rev Microbiol 48**,** 713-742.

Waller, K. L., Cooke, B. M., Nunomura, W., Mohandas, N., and Coppel, R. L., 1999. Mapping the binding domains involved in the interaction between the Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) and the cytoadherence ligand P. falciparum erythrocyte membrane protein 1 (PfEMP1). J Biol Chem 274**,** 23808-23813.

Werner, A., 2005. Natural antisense transcripts. RNA Biol 2**,** 53-62.

Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W., and Williamson, D. H., 1996. Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. J Mol Biol 261**,** 155-172.

World Health, O., 2000. Severe falciparum malaria. Transactions of the Royal Society of Tropical Medicine and Hygiene 94**,** 1-90.

Wuchty, S. and Ipsaro, J. J., 2007. A draft of protein interactions in the malaria parasite P. falciparum. J Proteome Res 6**,** 1461-1470.

Yassour, M., Pfiffner, J., Levin, J. Z., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D. A., Friedman, N., and Regev, A., 2010. Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biol 11**,** R87.

Yatsushiro, S., Taniguchi, S., Mitamura, T., Omote, H., and Moriyama, Y., 2005. Proteolipid of vacuolar H(+)-ATPase of Plasmodium falciparum: cDNA cloning, gene organization and complementation of a yeast null mutant. Biochim Biophys Acta 1717**,** 89-96.

Yazdanbakhsh, M. and Sacks, D. L., 2010. Why does immunity to parasites take so long to develop? Nat Rev Immunol 10**,** 80-81.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13**,** 134.

Young, J. A., Fivelman, Q. L., Blair, P. L., de la Vega, P., Le Roch, K. G., Zhou, Y., Carucci, D. J., Baker, D. A., and Winzeler, E. A., 2005. The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. Mol Biochem Parasitol 143**,** 67-79.

Yu, M., Kumar, T. R., Nkrumah, L. J., Coppi, A., Retzlaff, S., Li, C. D., Kelly, B. J., Moura, P. A., Lakshmanan, V., Freundlich, J. S., Valderramos, J. C., Vilcheze, C., Siedner, M., Tsai, J. H., Falkard, B., Sidhu, A. B., Purcell, L. A., Gratraud, P., Kremer, L., Waters, A. P., Schiehser, G., Jacobus, D. P., Janse, C. J., Ager, A., Jacobs, W. R., Jr., Sacchettini, J. C., Heussler, V., Sinnis, P., and Fidock, D. A., 2008. The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. Cell Host Microbe 4**,** 567-578.

Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T., and Kaneko, I., 2010. Transcription factor AP2-Sp and its target genes in malarial sporozoites. Mol Microbiol 75**,** 854-863.

Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G. R., Janse, C. J., Waters, A. P., Kato, T., and Kaneko, I., 2009. Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. Mol Microbiol 71**,** 1402-1414.

Zeuthen, T., Wu, B., Pavlovic-Djuranovic, S., Holm, L. M., Uzcategui, N. L., Duszenko, M., Kun, J. F., Schultz, J. E., and Beitz, E., 2006. Ammonia permeability of the aquaglyceroporins from Plasmodium falciparum, Toxoplasma gondii and Trypansoma brucei. Mol Microbiol 61**,** 1598-1608.

Zhang, B. and Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4**,** Article17.

Zhou, Y., Ramachandran, V., Kumar, K. A., Westenberger, S., Refour, P., Zhou, B., Li, F., Young, J. A., Chen, K., Plouffe, D., Henson, K., Nussenzweig, V., Carlton, J., Vinetz, J. M., Duraisingh, M. T., and Winzeler, E. A., 2008. Evidence-based annotation of the malaria parasite's genome using comparative expression profiling. PLoS ONE 3**,** e1570.

Websites:

1. www.GlobalHealthFacts.org
2. National Vector Borne Disease Control Programme (India): www. nvbcdp.gov.in/maps
3. Medicines for Malaria Venture: www.mmv.org/malaria-medicines/prasite-lifecycle
4. Malaria Site: www.malariasite.com
5. World Malaria Report (2013): http://www.who.int/malaria/publications/ world _malaria _report _ 2013/en/
6. World Malaria Report (2012): http://www.who.int/malaria/publications/ world _malaria _report _ 2012/en/

Books:

Nagpal, B.N., Sharma, V.P., 1994. Indian Anophelines. Oxford & IBH. Co.
Grassi B: Studi di uno Zoologo Sulla Malaria. . Rome 1900.

List of Supplementary Tables

No.	Caption
Supplementary Table S3.1	Agilent 15K cross strain *P. falciparum* gene expression microarray probe annotation.
Supplementary Table S3.2	3D7 Transcripts not detected in this study although represented by specific probes.
Supplementary Table S3.3	3D7 transcripts detected in this study.
Supplementary Table S3.4	Transcripts detected in this study represented by probes specific only to HB3 strain.
Supplementary Table S3.5	*var* gene transcripts detected in this study. *var* genes transcripts detected for all the 3 representing strain is included in three separate excel sheets.
Supplementary Table S3.6	*rifins* gene transcripts detected in this study.
Supplementary Table S3.7	*stevor* gene transcripts detected in this study.
Supplementary Table S3.8	List of differentially regulated genes between two clusters
Supplementary Table S3.9	List of over expressed genes known to express specifically in gametocyte stages.
Supplementary Table S3.10	List of genes and their expression status whose protein products specifically found in gametocyte stage.
Supplementary Table S4.1	Sense and antisense transcripts detected in P. falciparum complicated and uncomplicated samples
Supplementary Table S4.2	List of genes in physical clusters
Supplementary Table S4.3	Expression pattern of sense and antisense transcript pairs in complicated and uncomplicated samples
Supplementary Table S4.4	Expression pattern of common genes in complicated and uncomplicated cases with sense, antisense or both sense and antisense transcripts.
Supplementary Table S4.5	Genes with only sense transcripts in uncomplicated samples, which expressed both sense and antisense transcripts in complicated samples.
Supplementary Table S4.6	Gene Ontology (GO) biological process term enrichment of genes with antisense transcripts in complicated cases.
Supplementary Table S4.7	Mapping of genes with antisense transcripts to biochemical/metabolic pathways.
Supplementary Table S4.8	Genes with antisense transcripts uniquely detected in this study.
Supplementary Table S4.9	Genes with antisense transcripts detected in our study and from other reported studies.
Supplementary Table S5.1	Differentially regulated genes in complicated malaria isolates.
Supplementary Table S6.1	List of genes in each module.
Supplementary Table S6.2	Enriched gene ontology terms in modules.
Supplementary Table S6.3	List of differentially regulated genes in JR-diffset.
Supplementary Table S6.4	List of differentially regulated genes in J-diffset.
Supplementary Table S6.5	Comparison between genes in JR-diffset and J-diffset.
Supplementary Table S6.6	List of hub genes for each up and down-regulated module.

List of Publications

1. Acharya, P., Pallavi, R., Chandran, S., Chakravarti, H., Middha, S., Acharya, J., Kochar, S., Kochar, D., **Subudhi, A**., Boopathi, A. P., Garg, S., Das, A., and Tatu, U., 2009. A Glimpse Into the Clinical Proteome of Human Malaria Parasites *Plasmodium falciparum* and *Plasmodium vivax*. Proteomics Clin Appl 3**,** 1314-1325.

2. Kochar, D. K., Das, A., Kochar, A., Middha, S., Acharya, J., Tanwar, G. S., Gupta, A., Pakalapati, D., Garg, S., Saxena, V., **Subudhi, A. K**., Boopathi, P. A., Sirohi, P., and Kochar, S. K., 2010. Thrombocytopenia in *Plasmodium falciparum*, *Plasmodium vivax* and Mixed Infection Malaria: a Study from Bikaner (Northwestern India). Platelets 21**,** 623-627.

3. Saxena, V., Garg, S., Tripathi, J., Sharma, S., Pakalapati, D., **Subudhi, A. K**., Boopathi, P. A., Saggu, G. S., Kochar, S. K., Kochar, D. K., and Das, A., 2012. *Plasmodium vivax* Apicoplast Genome: A Comparative Analysis of Major Genes from Indian Field Isolates. Acta Trop 122**,** 138-149.

4. Garg, S., Saxena, V., Lumb, V., Pakalapati, D., Boopathi, P. A., **Subudhi, A. K**., Chowdhury, S., Kochar, S. K., Kochar, D. K., Sharma, Y. D., and Das, A., 2012. Novel Mutations in The Antifolate Drug Resistance Marker Genes Among *Plasmodium vivax* Isolates Exhibiting Severe Manifestations. Exp Parasitol 132**,** 410-416.

5. Pakalapati, D., Garg, S., Middha, S., Kochar, A., **Subudhi, A. K**., Arunachalam, B. P., Kochar, S. K., Saxena, V., Pareek, R. P., Acharya, J., Kochar, D. K., and Das, A., 2013. Comparative Evaluation of Microscopy, OptiMAL((R)) and 18S rRNA Gene Based Multiplex PCR for Detection of *Plasmodium falciparum* & *Plasmodium vivax* from Field Isolates of Bikaner, India. Asian Pac J Trop Med 6**,** 346-351.

6. Pakalapati, D., Garg, S., Middha, S., Acharya, J., **Subudhi, A. K**., Boopathi, A. P., Saxena, V., Kochar, S. K., Kochar, D. K., and Das, A., 2013. Development and Evaluation of a 28S rRNA Gene-Based Nested PCR Assay for *P. falciparum* and *P. vivax*. Pathog Glob Health 107**,** 180-188.

7. Boopathi, P. A*., **Subudhi, A. K***., Garg, S., Middha, S., Acharya, J., Pakalapati, D., Saxena, V., Aiyaz, M., Chand, B., Mugasimangalam, R. C., Kochar, S. K., Sirohi, P., Kochar, D. K., and Das, A., 2013. Revealing Natural Antisense Transcripts from *Plasmodium vivax* Isolates: Evidence of Genome Regulation in Complicated Malaria. Infect Genet Evol 20, 428-443.
   * *These authors contributed equally to the study*

8. **Subudhi, A. K***., Boopathi, P. A*., Garg, S., Middha, S., Acharya, J., Pakalapati, D., Saxena, V., Aiyaz, M., Chand, B., Mugasimangalam, R. C., Kochar, S. K., Sirohi, P., Kochar, D. K., and Das, A., 2014. Natural Antisense Transcripts in *Plasmodium falciparum* isolates from patients with complicated malaria. Exp Parasitol 141, 39-54.
   * *These authors contributed equally to the study*

9. Boopathi, P. A*., **Subudhi, A. K***., Garg, S., Middha, S., Acharya, J., Pakalapati, D., Saxena, V., Aiyaz, M., Chand, B., Mugasimangalam, R. C., Kochar, S. K., Sirohi, P., Kochar, D. K., and Das, A., 2014. Dataset of Natural Antisense Transcripts in *P. vivax* Clinical Isolates Derived using Custom Designed Strand-Specific Microarray. Genomics Data 2, 199-201.
   * *These authors contributed equally to the study*

10. Kochar, D.K., Das, A., Kochar, A., Middha, S., Acharya, J., Tanwar, G.S., Pakalapati, D., **Subudhi., A.K**., Bopathi, P.A., Garg, S., Kochar, S.K. A prospective study on adult pateints of severe malaria caused by *Plasmodium falciparum*, *P. vivax* and mixed infection from Bikaner, northwest India. J Vector Borne Dis. (Manuscript Accepted)

11. **Subudhi, A. K***., Boopathi, P. A*., Pandey, I., Kohli, R., Middha, S., Acharya, J., Kochar, S. K., Kochar, D. K., and Das, A., *P. falciparum* Complicated Malaria: Modulation and Connectivity between Exportome and Variant Surface Antigen Gene Families (Manuscript Submitted)
   * *These authors contributed equally to the study*

12. **Subudhi, A. K***., Boopathi, P. A*., Garg, S., Middha, S., Acharya, J., Mugasimangalam, R. C., Kochar, S. K., Sirohi, P., Kochar, D. K., and Das, A., Devlopment of a Cross Strain Microarray Optimized for the Transcriptome Analysis of Patient Derived Isolates of *P. falciparum* (Manuscript Submitted).
   * *These authors contributed equally to the study*

13. **Subudhi, A. K***., Boopathi, P. A*., Pandey, I., Kohli, R., Middha, S., Acharya, J., Kochar, S. K., Kochar, D. K., and Das, A.. Disease specific modules and potential hub genes for intervention strategies: A co-expression network based approach for *Plasmodium falciparum* clinical isolates (Manuscript under Preparation).
   * *These authors contributed equally to the study*

14. **Subudhi, A. K***., Boopathi, P. A*., Garg, S., Middha, S., Acharya, J., Pakalapati, D., Saxena, V., Aiyaz, M., Chand, B., Mugasimangalam, R. C., Kochar, S. K., Sirohi, P., Kochar, D. K., and Das, A. An *in vivo* Transcriptome Dataset of Natural Antisense Transcripts from *P. falciparum* Clinical Isolates from Bikaner, North-West India (Manuscript under Preparation).

Abstracts and Conferences

1. "**Genome wide identification of natural antisense transcripts in *Plasmodium falciparum* clinical isolates**" FEBS EMBO-2014 Meeting, Palais Des Congress, Paris, France.

2. "**Genome wide identification of natural antisense transcripts in *Plasmodium falciparum* clinical isolates**" Recent Developments in Malaria Parasite Biology International Scientific Meeting 2013, International Center for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India.

3. "***Plasmodium falciparum* exportome - An *in vivo* transcriptome analysis of genes encoding exported protein from severe *P. falciparum* clinical isolates**" XI Symposium on Vector Borne Diseases 2011, Regional Medical Research Center for Tribals (RMRCT), Jabalpur, India.

4. "**Transcriptome data of *P. falciparum* and *P. vivax* parasites causing severe disease manifestations" – Systems Network and causal interpretation**. Contemporary Trends in Biological and Pharmaceutical Research 2011. BITS, Pilani, India.

5. "**Partial analysis of *P. vivax* transcriptome from adult patients suffering from severe malaria manifestations**" Recent Developments in Malaria Research 2011, International Centre for Genetic Engineering & Biotechnology (ICGEB), New Delhi, India.

6. "***In vivo* Transcriptome Study of Parasite Genes involved in Erythrocyte Sequestration and Immune Evasion from Severe *P. falciparum* field isolates**" EMBO Global Exchange Lecture Course on Molecular and Evolutionary Genetics of Malaria 2010. National Institute of Malaria Research, Dwarka, New Delhi.

7. "**Molecular Analyses of *Plasmodium vivax* Parasites Involved in Severe Manifestations**" 58th Annual Meeting of the American Society for Tropical Medicine and Hygiene (ASTMH) 2010, Marriott Wardman park hotel, Washington, DC**.**

8. "***In vivo* transcriptome analysis of molecules involved in severe manifestations in *Plasmodium falciparum***" Emerging Trends in Life Sciences Research 2009, BITS, Pilani, India.

9. "**Molecular Analyses of *Plasmodium vivax* Parasites Involved in Severe Manifestations**" Epigenetic and Genome Control 2009, CCMB, Hyderabad, India.

10. "**Global gene expression analysis of *P. vivax* parasites causing severe malaria and interaction with the host genome**" 13th Human Genome Organization Meeting 2008, Hyderabad, India.

Gene Expression Omnibus Accession Numbers

Gene Expression Omnibus Microarray Platform IDs

GPL16484: *Plasmodium falciparum* Custom 244K Array designed by Genotypic Technology Pvt. Ltd. and Prof. Ashis K Das (AMADID: 019056).

GPL18192: Agilent Custom *Plasmodium falciparum* 8 × 15K array designed by Ashis Das, Raja CM and Genotypic Technology Pvt. Ltd. (AMADID:24956).

Gene Expression Omnibus Microarray Expression Studies (Series IDs)

GSE44921: Natural Antisense Transcripts in *Plasmodium falciparum* Clinical Isolates.

GSE54253: A New Agilent 15K Cross Strain *Plasmodium falciparum* Microarray Optimized for The Transcriptome Analysis of Indian P. falciparum Isolates.

GSE59844: *Plasmodium falciparum* Complicated Malaria: Modulation and Connectivity between Exportome and Variant Surface Gene Families.

Biography of Prof. Ashis Kumar Das, PhD

Dr. Ashis Kumar Das Joined BITS, Pilani, as an Assistant Professor in the year 1998 and is currently a Professor in the Department of Biological Sciences, BITS, Pilani, Pilani Campus. He previously held the position of Dean, Research and Counseling Division (R & C) now renamed as Academic Research Division (ARD) and Group Leader, Biological Sciences Group (now Department). He completed his Masters from Ballygunj Science College, Kolkata, and obtained his PhD from National Institute of Immunology, New Delhi, India. He worked as a Post-Doctoral Fellow in the Department of Molecular Biology and Immunology, SHPH, Johns Hopkins University, Baltimore, USA and was a WHO fellow at the Malaria Branch, Center for Disease Control and Prevention, Atlanta, Georgia, USA. His current research interests include evolving diagnostic procedures for the human malaria parasites based on the parasite's 18s rRNA and 28 rRNA genes and understanding of molecules at host-parasite interface using Systems Biology approaches. He is also actively involved in isolating and characterizing plasmids from Lactic Acid Bacteria in order to identify and evaluate putative novel promoters with an intention of facilitating vector design. He has published research articles in the journals of international repute and has presented his research national and international conferences and as invited speaker. He has attracted funds from funding agencies like University Grant Commission (UGC), Council of Scientific and Industrial Research (CSIR) and Department of Biotechnology (DBT), and also from private organizations like Dabur. He has filed 3 Indian patents.

Biography of Amit Kumar Subudhi

Amit Kumar Subudhi has completed his B.Sc (Zoology Hons) from Sambalpur University, Samabalpur, Odisha and M.Sc. in Biotechnology from Alagappa University. On completion of his Masters, he joined Inter University Accelerator Center (formerly known as Nuclear Science Center) as a project assistant, where he studied the effect of heavy ion radiation on MCF 7 carcinoma cell lines. Subsequently, he joined the Department of Biological Sciences, BITS, Pilani, Pilani Campus as a project assistant in a DBT funded project to Prof. Ashis K Das and latter enrolled as a PhD student. During his Masters, he availed the Pansumpon Muthuramalingam and V.O. Chidambaram endowment scholarship. He has qualified Graduate aptitude test in Engineering (2006) and was awarded the Senior Research Fellowship (2010) from Council of Scientific and Industrial Research (CSIR), Govt. of India. During his PhD, he had a chance to visit the famous Wellcome Trust Sanger Institute to broaden his research skills. He has published research articles in journals of international repute and has presented his research work in national and international conferences and workshops. His current research interest includes understanding the transcriptome biology of malaria parasites, and epigenetic and non-coding RNA based genome regulation in eukaryotes especially in protozoan parasites.