The role of cis-regulatory elements in Plasmodium falciparum transcriptional regulation

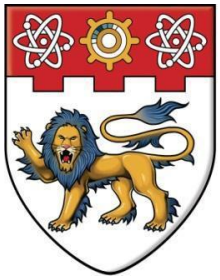
Lim, Eng How

2014

Lim, E. H. (2014). The role of cis-regulatory elements in Plasmodium falciparum transcriptional regulation. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/62533

https://doi.org/10.32657/10356/62533



NANYANG TECHNOLOGICAL UNIVERSITY

THE ROLE OF CIS-REGULATORY ELEMENTS IN *PLASMODIUM FALCIPARUM* TRANSCRIPTIONAL REGULATION

LIM ENG HOW

SCHOOL OF BIOLOGICAL SCIENCES

2014

THE ROLE OF CIS-REGULATORY ELEMENTS IN *PLASMODIUM FALCIPARUM* TRANSCRIPTIONAL REGULATION

LIM ENG HOW

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2014

Acknowledgement

I would first and foremost like to thank my supervisor and advisor for my thesis, Associate Professor Zbynek Bozdech, for his continued support for the work I have done in these 4 years. I thank him for supporting the many ideas that I constantly have and sharpening them consistently. I am grateful to him for supporting me to conferences where I have gained international experiences as a scientist. I would also like to thank Nanyang Technological University for the financial support provided in the form of scholarship. Next, I like to thank Prof Peter Preiser and Dr Su I-hsin for their advice during yearly thesis advisory committee meetings. People in the ZB lab, Dr Archna, Dev, Alok, Frances, Dr Annie Gao and Dr Karthik who has made the lab a more colorful place to work in. I'll like to thank the students whom I played a part to mentor, Tee Wanying, Chia Zhiqi, Kua Baohe, Chia Xinyi, Tan Xinyi and Jason Long, for their kind patience in bearing with me and helping me greatly in the work I do. I'll like to thank my family for their understanding of my plight as a PhD student. Finally and importantly, I'll like to thank Yanfen my beloved fiancée for her unending support when things are not going well, someone who is able to help me in experimental details and someone whom I can go to for advice.

Abstract

Malaria remains to be a global problem and has caused approximately 700,000 deaths per year. *Plasmodium falciparum* (*P. falciparum*) is a eukaryotic apicomplexan parasite and is one of the causative agents of malaria. The transcriptome of *P. falciparum* during the intraerythrocytic developmental cycle (IDC) is characterized by a highly regulated cascade of transcripts where 60-80% of the 5500 genes in the parasite genome are expressed at least once. However, mechanisms of transcriptional control are not well characterized in *P. falciparum*. In particular, there is a lack of functional *cis*-regulatory DNA motifs described in *P. falciparum*. We have utilized an algorithm to predict DNA motifs on co-regulated genes and tested their functionality by transfection assays. Interestingly, we have found an overrepresentation of motifs which repress transcription. We have also found that the repressive effects of several of motifs were stage specific and regulate transcription of groups of genes. Hence, this leads us to propose repression of transcription as a form of transcriptional regulation in *P. falciparum*.

Table of contents

List of publications

Mok, S., Liong, K.-Y., **Lim, E.-H.**, Huang, X., Zhu, L., Preiser, P. R. and Bozdech, Z. (2014), Structural polymorphism in the promoter of pfmrp2 confers Plasmodium falciparum tolerance to quinoline drugs. **Molecular Microbiology**, 91: 918–934. doi: 10.1111/mmi.12505

Gupta AP, Chin WH, Zhu L, Mok S, Luah Y-H, **Lim E-H** and Bozdech Z. (2013) Dynamic Epigenetic Regulation of Gene Expression during the Life Cycle of Malaria Parasite *Plasmodium falciparum*. **PLoS Pathog** 9(2): e1003170. doi:10.1371/journal.ppat.1003170

Conference presentation

25th Annual Molecular Parasitology Meeting, 2014, Marine Biology Laboratory, Woods hole, USA
Presentation title: The role of *Plasmodium falciparum* cis-regulatory elements in transcriptional regulation

9th Biology and Pathology of the Malaria Parasite (BioMalPar) 2013, EMBL, Heidelberg, Germany
Presentation title: The role of *Plasmodium falciparum* cis-regulatory elements in transcriptional regulation

Singapore Malaria Network Meeting (SingMalNet) 2012, Nanyang Technological University, Singapore
Presentation title: Structural polymorphism in the promoter of *Plasmodium falciparum* multidrug resistance-associated protein 2 (MRP2) confers resistance to mefloquine.

List of tables

List of figures

List of abbreviations

3’ untranslated region	3’ UTR
5’ untranslated region	5’ UTR
Adenosine and Thymine content	AT content
Acetylation lowers binding affinity protein family	Alba
Acetylation on histone 3 lysine residue 9	H3K9ac
Apicomplexan Apetala2-integrase transcription factors	ApiAP2
ATP-binding cassette	ABC
TFIIB Recognition Element	BRE
Blasticidin S deaminase	BSD
Carboxy-terminal heptapeptide repeat domain	CTD
Cyclic AMP-responsive element binding protein	CBP
CCCTC-binding factor	CTCF
Centromeric Histone 3	CenH3
Chromatin immunoprecipitation coupled to microarray analysis	ChIP-chip
Chromatin immunoprecipitation coupled to new generation sequencing	ChIP-seq
Comparative genomic hybridization	CGH
Cytoadherence linked asexual protein	CLAG
Dithiothreitol	DTT
Dimethyl sulfoxide	DMSO
Downstream Core Element	DCE
Downstream Promoter Element	DPE
Electrophoretic mobility shift assay	EMSA
Enhancer RNA	eRNA
Erythrocyte binding antigen	EBA
Ethylenediaminetetraacetic	EDTA
Finding Informative Regulatory Elements	FIRE
Firefly luciferase	Ffluc
General acetylation of histone 4	H4ac
General control nonderepressed 5	GCN5
General transcription factors	GTF
Heat shock protein	HSP
Heterochromatin Protein 1	HP1
High mobility group box B	HMGB
Histone acetyltransferase	HAT
Histone methyltransferase	HMT
Histone deacetylase	HDAC
Hours post invasion	hpi
human dihydrofolate reductase	hDHFR
Initiator	Inr
Insertions and deletions	Indels
Insulin growth factor 2	*Igf2*
Intraerythrocytic developmental cycle	IDC
Long non-coding RNA	lncRNA

Long intergenic non-coding RNA	lincRNA
Luria-Bertani	LB
Lysine	K
Merozoite surface protein	MSP
Motif Ten Element	MTE
Myb regulatory element	MRE
New generation sequencing	NGS
Nonidet P-40	NP-40
Nucleosome depleted regions	NDR
Oligonucleotides	Oligos
Phosphase buffered saline	PBS
Plasmodium berghei	*P. berghei*
Plasmodium falciparum	*P. falciparum*
P. falciparum erythrocyte membrane protein 1	PfEMP1
P. falciparum heat shock protein 86	PfHSP86
P. falciparum multidrug resistant protein 2	PfMRP2
Plasmodium vivax	*P. vivax*
Plasmodium yoelii yoelii	*P. y. yoelii*
Polyethylene glycol	PEG
Polycomb group proteins	PcG
Polycomb repressive complex	PRC
Positive transcription elongation factor b	P-TEF-b
Post translational modification	PTM
Preinitiation complex	PIC
Proliferating cell nuclear antigen	PCNA
Prx regulatory element (PRE) binding protein	PREBP
Red blood cells	RBC
Reticulocyte binding protein homologue 4	RH4
RNA polymerase II	RNAPII
X-core promoter element 1	XCPE1
Serial analysis of gene expression	SAGE
Single nucleotide polymorphism	SNP
SPE2 interacting protein	PfSIP2
Specific transcription factors	STF
TATA-box binding protein	TBP
TBP-associated factors	TAF
TBP-related factors	TRF
Tetramethylammonium chloride	TMAC
Transcriptional start site	TSS
Tri-methylation on histone 3 lysine 4	H3K4me3
Tri-methylation on histone 3 lysine 9	H3K9me3

1. Introduction

1.1 Malaria

The World Health Organization reported 219 million cases of malaria and causing 660,000 deaths per year [1]. It is a deadly disease which puts half of the world's population at risk. Malaria also causes great economic burden on the countries affected by the disease. The *Plasmodium* genus is the causative agent of malaria. It has over 200 species and its hosts include mammals, birds and reptiles. The 4 species which affects humans are *Plasmodium falciparum (P. falciparum), Plasmodium vivax (P. vivax), Plasmodium malariae* and *Plasmodium ovale*. Of which, *P. falciparum* is the most virulent. The most deadly manifestations include cerebral malaria, placental malaria and anemic malaria.

P. falciparum has a complex life cycle in 2 hosts: the female anopheles mosquito and the human (Figure 1.1). The parasite takes on numerous morphological forms and it utilizes multiple cell types, in its mosquito and human hosts. The stage in the lifecycle of the parasite that causes the most of the manifestations of the disease is the intra-erythrocytic developmental cycle (IDC). During the IDC, the parasite utilizes red blood cells (RBC) and undergoes multiplication. Briefly, a form of the parasite, the merozoite, invades into the RBC and develops into a ring. As the ring matures, it develops into a trophozoite. During the ring and trophozoite stage, the parasite obtains amino acids from the RBC through the digestion of hemoglobin from the RBC [2]. As the parasite develops further into a schizont, DNA is replicated and the parasite undergoes cell segregation into 8-32 daughter cells per schizont. The schizont then ruptures and releases merozoites to invade into other RBCs. The IDC is typically 48 hours.

Many drugs such as quinine, mefloquine, and chloroquine were effective against the parasite in the past but growing drug resistance has been a big problem against the effectiveness of these drug [3, 4]. Currently, artemisinin and its derivatives are the most effective drugs. However, there have been early signs of resistance as well [5]. There is no vaccine against malaria in clinical use at this point of time although many are in the phase of clinical trials [6]. The complex lifecycle of the parasite and ability of the parasite to undergo antigenic variation is an obstacle in vaccine development [7, 8].

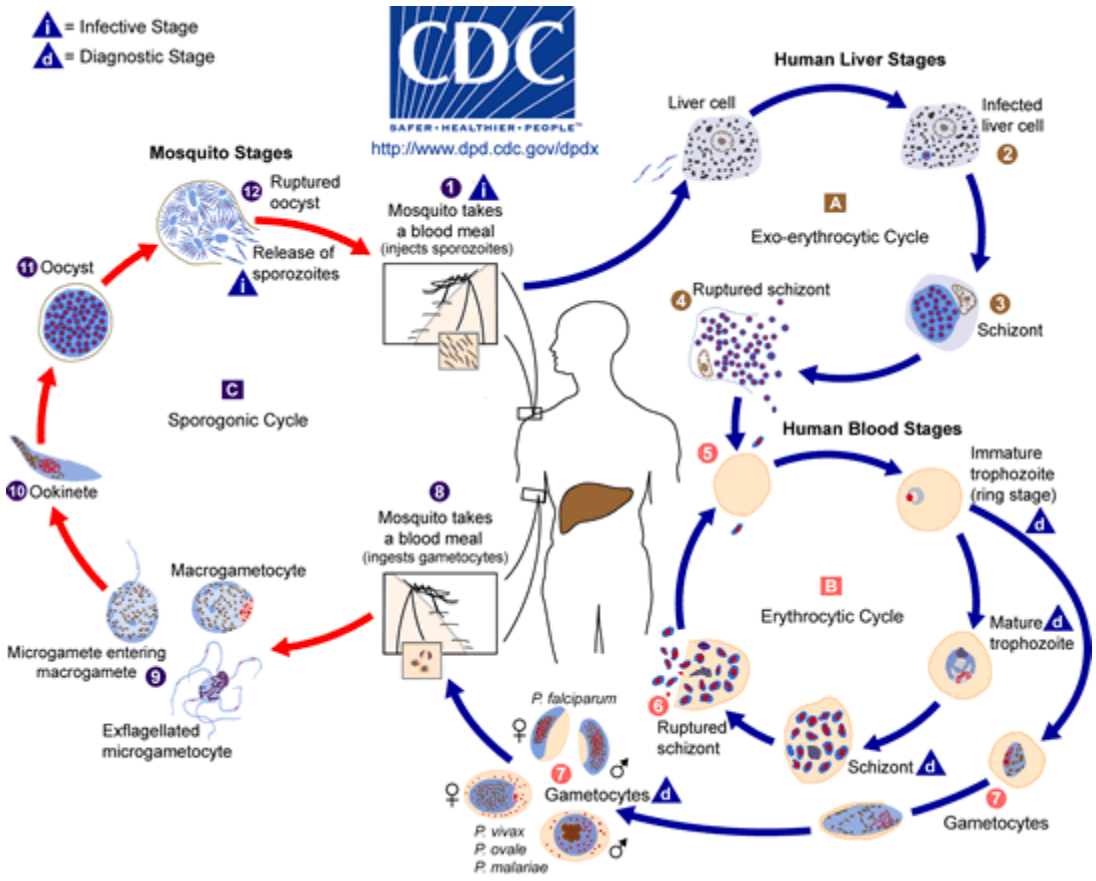


Figure 1.1 The *P. falciparum* life cycle.

The life cycle consists of 3 main stages: the human liver stages, the human blood stages (erythrocytic cycle) and the mosquito stages. In the human, the parasite exist mostly in an asexual form (except the gametocytes), while in the mosquito, it is in a sexual form. The human blood stages or the intraerythrocytic developmental cycle (IDC) takes approximately 48 hours. Source: (CDC website http://www.cdc.gov/malaria/about/biology/index.html).

1.2 *Plasmodium* genomics

1.2.1 *Plasmodium* genome

The genome sequence of *P. falciparum* [9] and *P. yoelii yoelii* (*P. y. yoelii*) [10] were first published using Sanger sequencing in 2002. Following which, the sequence of *P. vivax* was also published [11]. Experimental animal model parasites, *Plasmodium chabaudi* and *Plasmodium berghei*, were also sequenced subsequently [12]. *Plasmodium knowlesi*, which infects primates, was sequenced in 2008 [13]. The *P. falciparum* genome is a haploid genome of 23.3 megabases which consist of 14 chromosomes, a circular apicoplast genome and a linear mitochondrial genome. The average AT content of the genome is 80.6% and approximately 90% in the intergenic regions. The AT content of the *P. falciparum* genome is the highest among the *Plasmodium* species which have been sequenced. 52.6% of the genome is predicted to be protein coding. This is in contrast to the human genome where 98% of the genome is noncoding, indicating that the *P. falciparum* genome is compact in nature [14]. Approximately 5400 genes have been predicted in the genome. However, about 60% of the predicted genes have no known functions.

More than 3300 orthologues of the 5400 *P. falciparum* genes are found in the *P. y. yeolii* genome [10]. Further sequencing of the rodent malaria species has also revealed that there is a conservation of gene synteny in the central chromosome regions [12]. A difference in the sequence of *P. vivax* genome with its mammalian malaria *Plasmodium* species counterparts was its AT-content being the lowest of approximately 60% genome-wide average [11]. Again, a large proportion of *P. vivax* genes (77%) have orthologues in *P. falciparum, P. chabaudi* and *P. berghei*. It was found that gene families involved in antigenic variation are spread out randomly in the *P. knowlesi* genome [13]. This is in striking contrast with the

genomes of the *Plasmodium* species discussed so far where majority of gene families involved in antigenic variation lie in the subtelomeric regions of the genome.

Genetic diversity is important for parasite survival to be able to evade immune response and to gain resistance to drugs quickly. Earlier studies to characterize genomic diversity were largely done on a limited number of specific genes. A study attempted to characterize single nucleotide polymorphisms (SNP) in 10 different *P. falciparum* strains on 41 genetic loci [15]. The authors reported 39 SNPs and more than half of them are located in 3 out of the 41 loci. By means of microarray, another group had profiled variations in 14 field and laboratory strains of *P. falciparum* [16]. In this study, variations were found in 234 genes. Many surface antigens (e.g. circumsporozoite protein and merozoite surface proteins) and genes associated with drug resistance (e.g. chloroquine resistance transporter) are among the genes that are highly variable. Advances in sequencing technologies have accelerated research in characterizing genomic variations amongst various isolates of field parasites. In a study, 4 *P. vivax* strains isolated from different geographical locations were sequenced using new generation sequencing technologies (NGS) [17]. The authors reported that SNP diversity among the 4 *P. vivax* strains are twice as much as the 4 *P. falciparum* strains which are also geographically diverse. This may indicate that *P. vivax* is genetically more diverse than *P. falciparum* and may develop drug resistance more quickly. In more recent years, a study has identified the kelch propeller domain protein as a potential molecular marker of artemesinin resistance by means of NGS of clinically isolated field parasites which exhibit delayed clearance time by artemesinin following treatment with artemesinin [18].

1.2.2 Transcriptomics

Since the release of the genome sequence in 2002, genome-wide studies of *P. falciparum* has accelerated. Two studies on the transcriptome of culture adapted *in vitro P. falciparum* parasites using microarray technology profiled the steady-state accumulation of transcripts in the IDC [19, 20]. In one of the studies, transcripts were collected at every hour in the 48 hour IDC [19]. These studies have shown that more than 80% of the genes are transcriptionally regulated showing a periodic pattern with a peak of accumulated transcripts at least one point during the life cycle. This is in contrast to other eukaryotic cells, yeast [21] and human HeLa cells [22], where only 15% of the transcripts show difference in levels during the cell cycle. The time of peak accumulation of the transcript is correlated to the time in the IDC where the function of the gene is needed. For example, transcripts of genes involved in invasion display a peak in accumulation in the late schizont stage. At this stage, the schizonts are ready to rupture to release merozoites for invasion of new RBC.

In the other study of the *P. falciparum* transcriptome using microarray, 88% of the genes had transcripts detected in at least one point in the IDC [20]. This finding is similar to the previous study. They have also reported that the expression patterns of functional groups of genes correlates with the needs in the life cycle. For example, they found that there is a shift in accumulation of transcripts of genes related to protein synthesis in ring and trophozoite stage to cell surface structures in the schizont stage. This is also in agreement with the previous study.

With the recent development of NGS, the IDC transcriptome was characterized by the Illumina Genome Analyzer [23]. The temporal patterns of the transcripts were in high correlation with the earlier microarray data with Pearson correlation values of 0.7-0.85. The

study detected transcripts of 90% of genes throughout the IDC which is in agreement with the previous 2 microarray where a large proportion of the genes are transcriptionally regulated.

In another study, the transcriptomes of 3 culture adapted strains of *P. falciparum* (3D7, HB3 and Dd2) were profiled and compared [24]. It was found that the periodic pattern of transcript expression is remarkably similar amongst the 3 strains. This was in spite that 3D7, HB3 and Dd2 were isolated from geographical locations which are vastly apart and has different drug sensitivities to different anti-malarials. This is in contrast to the yeast where it was observed that there is 50% variability in expression of the genes even when growing under the same conditions [25]. This data has led authors to propose that *P. falciparum* has a "hard-wired" transcriptional program which is conserved from strain to strain. However, when in comparison to another *Plasmodium* species which also infects humans, *P. vivax*, approximately 33% of the genes show a change in expression pattern at varying degrees [26]. The authors suggested that an evolutionary change in the usage of the proteins between the 2 species could be a probable explanation for the change in timing of expression for some transcripts.

1.2.3 *Plasmodium* proteomics

The first genome-wide proteomic study of *P. falciparum* was published in the same year as the release of the genome sequence [27]. The authors utilized a multidimensional protein identification technology to identify proteins in 4 stages of the *P. falciparum* lifecycle, sporozoites, merozoites, trophozoites and gametocytes. Approximately 2400 proteins of the 5400 protein-coding genes predicted are identified and almost half of them are found in all 4 stages of the parasite life cycle. An interesting observation is the presence of proteins

involved in immune evasion (*var, rifin*) in the sporozoite stage. These genes were thought to be playing their roles only during the blood stages.

In order to assess of the roles played by different levels of gene regulation, quantitative proteomic approaches have since been developed to profile the levels of proteins in the IDC. The first of such studies was done by metabolic labeling of parasite proteins followed by a 2-dimensional gel approach [28]. When comparing the levels of 6 proteins against the levels of corresponding mRNA, a delay in peak abundance can be observed. Another study utilized a semi-quantitative mass spectrometry approach to study the relationship between protein levels and transcript levels [29]. Approximately 2500 proteins are detected in at least 1 out of the 7 stages studied. This number was similar in the first study described in the proteomics section. A striking observation that the authors made was that there was a high correlation between the proteome of a particular stage and the transcriptome of the stage preceding it.

To further characterize the relationship between transcript and protein, another study was done to interrogate the transcriptome and proteome of *P. falciparum* at high resolution in the IDC [30]. This is the largest fully quantitative proteomic data in *P. falciparum* so far, a total of 125 proteins were profiled. RNA and protein samples were harvested concurrently at 2 hour intervals in the IDC. The study revealed that, similar to mRNA profile, the protein levels show a peak in abundance at one point in the IDC. An important observation was that there is a delay in peak abundance of proteins compared to their corresponding transcripts by 6-16 hours. By mathematical modeling, the authors also hypothesized that transcript steady state levels can be directly related to protein translation.

1.2.4 *Plasmodium* transfection technologies

The ability to introduce exogenous DNA into a cell brings great advantages in the study of cell biology including gene function and control of gene expression [31, 32]. In *P. falciparum*, chemical mediated gene delivery remains to be relatively inefficient and the only effective method of gene transfer till date is through electroporation [33]. In the beginning, ring stage parasites were electroporated at high voltages and low capacitance [34]. It was found later that electroporation through low voltage and high capacitance achieved better efficiency, this method of electroporation has since been widely used [35]. The most recent method of *P. falciparum* gene transfer involves the preloading of DNA into uninfected RBCs by electroporation and subsequently, allowing parasites to infect into the DNA loaded RBCs, this is when parasites take up the DNA [36]. It was also found that preloading of RBCs is the most efficient method of gene transfer for *P. falciparum* and the efficiency is $1.08x10^{-5}$ [37]. In the more recent years, novel technologies have claimed to be able to conduct eletroporations with transfection buffers that result in high efficiencies for transfections in 96-well plate format [38]. A lipid nanoparticle has most recently been reported for efficient gene transfer to *P. falciparum*, serving as a potential alternative to electroporation methods [39]. However, these novel methods are yet to be widely used in the field and their effectiveness remains to be validated.

1.3 Gene expression regulation

Gene expression regulation can be achieved in several levels, namely, epigenetic regulation, transcriptional regulation (initiation and post-initiation) and post-transcriptional regulation. Figure 1.3 summarizes mechanisms in eukaryotic transcriptional regulation. Epigenetic regulation includes nucleosomal occupancy, the presence of histones (and its modification) in the genome, role of long non-coding RNAs (lncRNA) and the role of nuclear sublocalization.

Initiation regulation refers to the assembly of basal transcription factors and how transcription of genes is initiated at a specific time by specific transcription factors. Post-initiation regulation refers to transcript control at the level of transcript elongation and termination. Finally, post-transcriptional control refers to mRNA stability and translational control.

1.3.1 Epigenetic regulation

By definition, epigenetics refers to any means to hereditary memory other than changes to the underlying DNA sequences. Importantly, epigenetics regulates gene expression on top of the *cis*-regulatory elements embedded in DNA sequences. The study of DNA-packaging proteins, histones, and its modifications has dominated this field. The roles of lncRNA and nuclear sublocalizations have been also emerging themes in epigenetics.

1.3.1.1 Histones, histone modifications and modifiers

The eukaryotic DNA is packaged into nucleosomes where 147 bp of DNA is wrapped around an octamer of core histones [40]. Each octamer consists of 2 of each H2A, H2B, H3 and H4 core histones. Each histone core has an N-terminal tail which can be subjected to multiple post-translational modifications (PTMs) such as acetylation, methylation, ubiquitination, phosphorylation and ADP-ribosylation on several amino acid residues. In general, histones package DNA into transcriptionally silent heterochromatin and transcriptionally active euchromatin. Different combinations of core histone PTMs have been shown in eukaryotic systems to regulate transcription, this is also known as the ‘histone code’ [41]. In addition, the exchange of core histones for variant histones (H2A.Z, H2A.X, CenH3, H3.3 etc) also has consequences of transcriptional activity [42].

4 core histones and 4 variant histones have been found in the *P. falciparum* genome [43]. Canonical H3 has 2 variants, H3.3 and centromeric H3 (CenH3). Canonical H2A has 1 variant H2A.Z, canonical H2B has 1 variant H2Bv and finally H4 with no variants. Interestingly, *P. falciparum* has only 1 copy of each core histone in its genome. This is unlike other eukaryotes where multiple copies were found. Another difference with other eukaryotic systems is the absence of H1 linker histone in the *P. falciparum* genome. In the study, the authors also reported that mRNA and protein levels of core canonical histones peak in the late trophozoite and schizont stages. At least 50 PTMs (largely methylations and acetylations) were found on these histones [43, 44]. In addition, PTMs for active transcription were observed at higher abundances than repressive PTMs. PfH4 has also been found to be sumoylated in a later study [45]. Figure 1.2 summarizes the findings on epigenetics in *P. falciparum*.

1.3.1.2 Plasmodium euchromatin

Acetylation of histones has been generally associated with active gene transcription in eukaryotes [46]. Acetylation can affect gene transcription by neutralizing a positive charge on histones and thus increasing the accessibility of DNA to transcription associated proteins [47]. Another mechanism is through the binding of bromodomain-containing proteins to acetylated residues which affects transcriptional activity [48]. A histone acetyltransferase (HAT) has been fairly well characterized in *P. falciparum*. The *P. falciparum* homologue of the yeast general control nonderepressed 5 (GCN5) has been shown to preferentially acetylate lysine 9 (K9) and K14 of histone H3 [49]. Inhibition of PfGCN5 by curcumin reduces H3 lysine 9 acetylation (H3K9ac) on upstream regions of 4 genes and a reduction of steady state transcripts of the 4 genes was also observed [50]. The authors also showed that genome wide occupancy of PfGCN5 and H3K9ac correlates well. In another study, when parasites were

treated with anacardic acid, another HAT inhibitor, 5% of genes were differentially regulated [51]. On the other hand, when histone deacetylases (HDAC) were inhibited, approximately 60% of the genes were deregulated [52]. These studies underscore the importance of the balance of acetylation levels of histones in gene regulation.

To further access the role of histone modifications dynamics in regulating gene expression in the IDC, several genome-wide studies were done to profile genomic locations of various histone modifications and compare to gene expression. Together with H3K9ac, H3 lysine 4 tri-methylated (H3K4me3) is known to associate to active promoters in eukaryotes [53, 54]. In a study, authors reported that during ring stage, H3K9ac and H3K4me3 are present uniformly in ring stage active and ring stage inactive genes [55]. However, during the schizont stage, H3K9ac and H3K4me3 are enriched in upstream regions of schizont active genes but not for schizont inactive genes. In a later study, authors utilized chromatin immunoprecipitation coupled to NGS (ChIP-seq) to improve resolution in intergenic regions [56]. It was reported that enrichment of H3K9ac in upstream regions of genes correlates with the expression pattern but H3K4me3 is enriched in upstream regions of genes during trophozoite/schizont stage, regardless of transcriptional activity.

Even more recently, investigators used ChIP-chip to profile the dynamics of 12 histone modifications throughout the IDC [57]. Of these modifications, 8 modifications show an enrichment bias towards the 5’ upstream regions of genes, while the rest show no preference for both intergenic and intragenic regions. Enrichment of 8 histone modifications (H4K8ac, H4K16ac, H4ac4, H3K56ac, H3K9ac, H3K14ac, H3K4me3 and H4K20me1) have substantial correlation with transcriptional activity in varying degrees. Of which, only H4K8ac is enriched preferentially in the 5’ untranslated regions (5’UTRs) of genes, the rest

are enriched preferentially at the 5' ends of open reading frames or have no positional preference. This study has expanded our knowledge on histone modifications which associate with transcriptional activity in *P. falciparum*.

A variety of roles of histone modifications have not been explored in *Plasmodium*. Firstly, it has been reported in other eukaryotes that the positioning of the histone modification with respect to the gene elicits different effects. An example is the methylation of H3K9, an enrichment in the intragenic regions gives a positive effect on transcription but an opposite effect when in promoter regions [58]. Secondly, less common modifications such as ubiquitination and sumoylation has been reported for *P. falciparum* histones but their genome-wide occupancies and functions have not been reported [44, 45]. Lastly, binding partners of histones with specific modifications have been identified and thus eliciting functions have been identified in model eukaryotes [59, 60]. However, other than the binding partners of heterochromatin mark H3K9me3, partners of other modifications are still unknown in *Plasmodium* [61].

1.3.1.3 Plasmodium heterochromatin

Another area of active research in the epigenetics of *Plasmodium* is the silent heterochromatin structure located predominately in the telomeric and subtelomeric regions, marked by H3 lysine 9 tri-methylation (H3K9me3) [55]. This chromatin structure is almost exclusively associated with virulent gene families encoding for infected RBC surface proteins and merozoite surface proteins, involved in cytoadherence (*var, rif, stevor*) and invasion respectively [62]. With most of the genes in the virulent gene families kept silent in heterochromatin structure, the parasite then regulates expression of selected gene member(s) and occasionally switching gene members [63]. Variant expression of virulent genes allows

phenotypic diversity of its cell surface molecules which is key to parasite survival as it assists parasites in immune evasion.

The *var* multigene family encodes for 60 variants of *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) gene family which will be exported to the infected RBC surface after the parasite expresses it [63]. However, each parasite only expresses 1 or a few *pfemp1* genes at any one time and has the ability to switch members to evade the host immune system [64]. H3K9me3 is enriched and H3K9ac is depleted in the 5'UTRs and exon of the slient *var* genes [65, 66]. On the other hand, activating histone marks H3K9ac and H3K4me3 are enriched near the transcriptional start site (TSS) of the activated *var* gene [65]. The inactivation of NAD-dependent histone deacetylase Sir2A and Sir2B results in de-repression of different groups of *var* genes [67, 68]. This led to the possibility that removal of acetylation on histones by the Sir2 family sets the background for the repressive mark, H3K9me3, to be established. In addition, Heterochromatin Protein 1 (PfHP1) is recruited to H3K9me3 to establish heterochromatin at silent *var* genes [69]. As *var* genes are only expressed in the ring stage parasites, expression of active *var* gene is repressed in the late stage parasites. However, during the late stages, the active *var* gene is poised for activation in the next ring stage by H3K4me2 (histone 3 lysine 4 bi-methylation) at the TSS [65]. In addition, H3 lysine 36 trimethylation (H3K36me3) has also been implicated in silencing of *var* genes [70]. The knockout of a SET domain family histone lysine methyltransferase, PfSETvs (variant silencing), has led to the reduction of H3K36me3 on silenced *var* genes and consequently a de-repression of silent *var* genes.

A similar mechanism was observed to control the expression of erythrocyte invasion proteins. The ability to vary the expression of invasion proteins allows the parasite to switch invasion

pathways and thus adaptable to invade into various types of erythrocytes displaying different surface receptors [71]. Various studies have shown that H3K9me3 marks the silent invasion genes and, active marks such as H3K9ac and general acetylation of H4 (H4ac) marks the active invasion genes [72-74]. In these studies, when investigators 'forced' parasites to switch invasion pathways, the previously active invasion gene is repressed and another invasion gene is activated, this is accompanied with a switch in the enrichment of histone modifications in the upstream regions of the invasion genes. Genome-wide occupancy of H3K9me3 has also revealed that silenced members of invasion gene families such as *erythrocyte binding antigen* (*eba*) and *cytoadherence linked asexual protein* (*clag*) are in heterochromatin regions [61, 62].

1.3.1.4 Variant histones

H2A.Z plays the role of poising genes for activation in yeast and metazoans, and its acetylation status affects nucleosome stability [75, 76]. It was also reported that H2A.Z is present on both active and inactive genes [77]. In *P. falciparum*, it was found that H2A.Z is enriched in intergenic regions of euchromatin (both 5'UTR and 3'UTR) [56]. However, authors of the study reported that levels of enrichment do not correlate with gene expression patterns in the IDC as H2A.Z occupancy remains largely constant throughout the IDC. A similar observation was also reported in a different study investigating a small number of genes [78]. As H2A.Z colocalizes well with H3K9ac and H3K4me3, this led the authors to hypothesize that H2A.Z serves to mark intergenic regions and therefore a scaffold for binding of other proteins involved in epigenetics and transcription. Also, exclusion of H2A.Z from heterochromatin has also led the authors to speculate that H2A.Z prevents the spread of heterochromatin to euchromatic regions [56]. This is consistent with a separate study where authors did a cellular colocalization study and found that H2A.Z colocalizes well with

euchromatin marks (H3K4me3 and H3K9ac) but not heterochromatin marks [78]. In addition, co-immunoprecipitation experiments revealed that nucleosomes containing H2A.Z also contained H3K4me3 mark in that nucleosome [78].

In a more recent study, investigators found that H2B.Z (the same H2B variant previously named H2Bv by authors in [43]) also localizes to euchromatic intergenic regions and the occupancy correlates strongly with H2A.Z [79, 80]. This was further supplemented by a cellular colocalization study where H2B.Z localizes with H3K4me3 and H3K9ac, but not H3K9me3 [80]. By co-immunoprecipitation assays, the authors also found that H2A.Z and H2B.Z are found within the same nucleosome [79]. In addition, the same study also reported that H2A.Z and H2B.Z preferentially occupies upstream regions of genes with higher AT-content. It was also observed that the maximum expression of genes correlates well with the AT-content of the upstream regions of genes. This led the authors to speculate that the presence of H2A.Z and H2B.Z bivariant nucleosomes in promoters increases the strength of the promoter. The association of H2A.Z with H2B variant histone was also observed in other eukaryotic parasites, *Toxoplasma brucei* and *Toxoplasma gondii* [81, 82].

Although H2A.Z is depleted in the heterochromatin, studies have also shown a role of H2A.Z in the regulation of expression of the active *var* gene [78]. The authors have shown that accompanying the expression of the active *var* gene during ring stage, H2A.Z is enriched near the TSS of the *var* gene. However, this enrichment is not present in the schizont stage where *var* genes are not expressed. In addition, H2A.Z was found to be enriched in *var* introns regardless of expression. In a later study, it was revealed that H2B.Z functions in the same way together with H2A.Z in a bivariant nucleosome [80].

In addition to gene regulation, histones in *P. falciparum* have also been shown to have other functions. ChIP-seq revealed that PfCenH3 (centromere H3) localizes to a single region of 4-4.5kb on each of the 14 *P. falciparum* chromosomes and is present throughout the IDC [83]. An exception is chromosome 10 where the occupancy is 2.2kb. H2A.Z was observed to be present in the region as well, suggesting a role in aiding CenH3 in binding to the centromere [83]. However, heterochromatic mark H3K9me3 is not present. Immunofluorescence assays also revealed that the centromeres cluster together before replication, this occurred during the late ring and trophozoite stage [83]. After DNA replication and segregation during schizony, each nuclei has a distinct centromere cluster, the authors suggest that the centromeres are attached to the mitotic spindle. PfCenH3 was also shown to be able to complement the function of Cse4p (the yeast homologue of CenH3) and in this study, authors identified structural components in PfCenH3 which are important for centromere targeting and chromosome segregation [84].

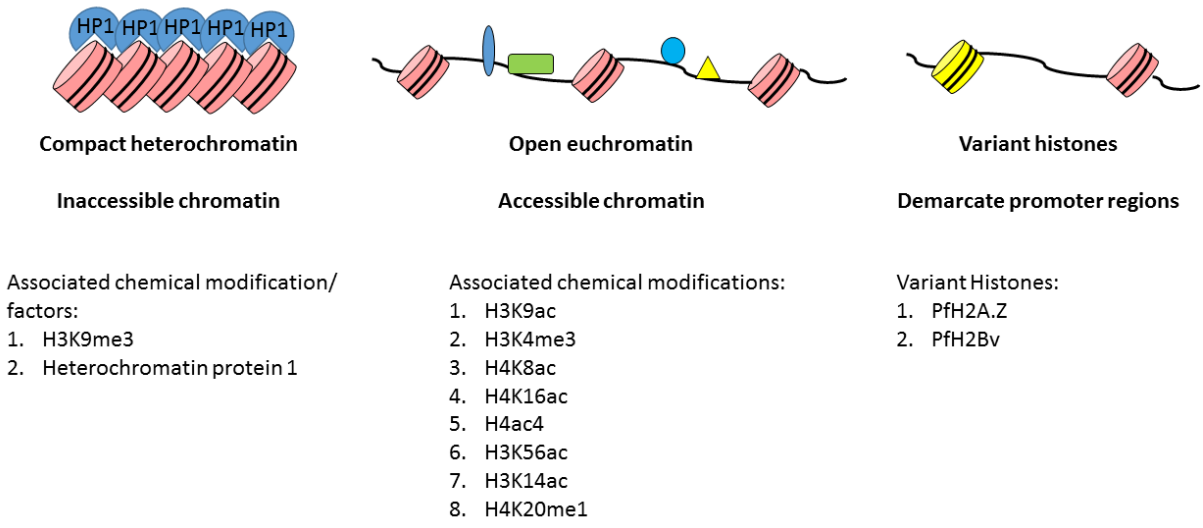


Figure 1.2 Epigenetic regulation in *P. falciparum*.
Heterochromatin and euchromatin has been extensively researched in *P. falciparum*. The associated chemical modification and important factors associated with the epigenetic regulation of *P. falciparum* has been summarized in this figure. In addition, 2 variant histones were characterized in *P. falciparum*.

1.3.1.5 High mobility group box proteins

4 members of the High mobility group box B (HMGB) proteins (PfHMGB1 to PfHMGB4) have been identified in the *P. falciparum* genome by sequence homology [85]. In eukaryotic organisms, the HMGB family proteins are known to regulate transcription by binding and distorting DNA structures [86]. Recombinant PfHMG1 and PfHMG2 have been shown to be able to bind to distort DNA structures and bend DNA sequences in *in vitro* experiments [85]. In addition, authors showed that both PfHMG1 and PfHMG2 localize mainly in the parasite nucleus during the blood stages and gametocyte stage [85]. While PfHMG1 is more highly expressed in the blood stages, expression of PfHMG2 is up regulated in the gametocyte stages [85]. The *P. y. yoelii* homologue, PyHMG2, has been implicated in oocyst formation [87]. Knock out of *pyhmg2* also causes the down regulation of 30 genes which, for most of the genes, transcripts and protein levels peak in gametocyte and oocyst stages [87].

1.3.1.6 Nucleosome occupancy

In addition to PTMs of histones, the positioning and dynamics of nucleosomes has an effect on gene expression. For example, it was shown that most promoters of budding yeast, *Drosophila* and humans have short stretches of DNA which are depleted of nucleosomes, known as nucleosome-depleted regions (NDRs) [88-90]. The NDR coincides with the TSS of the gene and, 2 well-positioned nucleosomes are directly upstream and downstream of the TSS which is also called the -1 and +1 nucleosome respectively [90]. It was also observed in yeast that regions which contain regulatory elements are relatively nucleosome depleted compared to both general coding and intergenic regions [91]. In the same study, the authors have also shown that promoters of actively transcribed genes are even more nucleosome depleted. It was also reported that in several promoters in yeast that upon activation by environmental cues, the nucleosomes near the TSS are evicted [92-94]. From these studies, it

was clear that the eviction of nucleosomes facilitates transcription factor binding and thus transcription. This is the case for genes which are usually not transcribed but are activated under certain environmental conditions. However, in the case of promoters of constitutively expressed genes, rearrangement of nucleosomes was not observed [95]. In addition, it was observed that 2 of such promoters of constitutively active genes contain sequences which disfavor the formation of nucleosomes [96, 97].

Genome-wide nucleosome occupancy maps have been generated for *P. falciparum* using both microarray and NGS. In one study, authors utilized ChIP-chip with H4 antibody [98]. Similar to studies in yeast, the authors found a general depletion of nucleosomes in 5'UTR and 3'UTR in contrast to coding regions. In addition, little changes in nucleosome occupancy are observed in the intergenic regions throughout the IDC. Greater fluctuations were reported for the coding regions but the fluctuations did not correlate well with gene expression.

In another study, the authors developed 2 complementary techniques to isolate genomic regions bound by nucleosomes and to isolate sequences in the genome unbound by nucleosomes and after which, coupling to NGS [99]. The authors profiled the nucleosome-free and nucleosome-bound regions in the IDC at single nucleotide-resolution. In general, a larger proportion of the genome is not bound by nucleosomes in the trophozoite stage but the genome is being occupied with nucleosomes increasingly as the parasite develops into a schizont [99]. This is consistent with a study which reported that the transcripts and protein levels of canonical histones peaked at late trophozoite and schizont stage [43]. In addition, a maximum of 50% of the genome is nucleosome bound, this is in contrast to 81% in yeast [88].

Similar to the previous study, the authors also found a general depletion of nucleosomes in the 5' and 3' UTRs of genes [99]. It has been shown in earlier studies that the presence of homopolymeric poly(dA:dT) tracts disfavors nucleosome formation *in vivo* and *in vitro* [100, 101]. The general depletion of nucleosomes may be explained by the extreme AT content in the intergenic regions of *P. falciparum* which reaches 90% and long stretches of poly(dA:dT) tracts occur frequently in the intergenic regions [9]. A reasonable deduction from the general depletion from 5' UTRs of genes (which likely contain promoter elements) is that the *P. falciparum* genome is in a transcriptionally permissive state. An extended study using the same techniques attempted to characterize the NDRs further [102]. The authors mapped nucleosome occupancy to the previously proposed TSS of about 2900 genes, where they found a 150-200 bp NDR around the TSS [103]. Further, they have also shown that genes which have a NDR at its TSS show a higher mRNA expression level compared to genes with a nucleosome at TSS. In addition, computationally predicted promoters were also found to be nucleosome depleted [104].

However, it must also be noted that the AT-richness of the *P. falciparum* genome may cause artifacts or biases in the amplification step prior to sequencing. Specifically, amplification of coding regions which has lower AT content when compared to regions with higher AT-content is more efficient [105]. In another study, authors used an amplification protocol optimized to reduce biases in AT-rich genomes and found that nucleosome occupancy is comparable between coding and intergenic regions [56, 105]. Hence, it is still an open debate on whether *P. falciparum* intergenic regions are nucleosome depleted compared to coding regions.

1.3.1.7 DNA methylation

The role of DNA cytosine methylation in gene regulation has been well documented in eukaryotes [106]. Its roles in transcription include genomic imprinting of silenced genes when CpG island near gene's TSS is methylated [107], regulation of enhancer function [108] and regulating enhancer-promoter function by affecting insulator function [109]. There is a lack of evidence of DNA methylation in the *P. falciparum* [110, 111]. However, it has been proposed that the lack of evidence was due to the detection limits by mass spectrometry based methods used as the methods used were optimized mainly for detection of methylations on CpG-dinucleotides on CpG islands in mammalian cells [112].

Nonetheless, the first evidence of DNA methylation in *P. falciparum* is in the *dihydrofolate reductase-thymidylate synthase* (*dhfr-ts*) gene [113]. The methylase-sensitive restriction digestion assay provided the sensitivity required to detect methylation in the coding sequence of the *dhfr-ts* gene in 4 *P. falciparum* strains. However, such studies are limited by the availability of restriction enzyme sites on the genomic sequence of interest and its incompatibility with genome-wide studies [114].

The first genome-wide evidence of DNA methylation in *P. falciparum* was recently published [115]. Authors proposed the presence of a putative *P. falciparum* DNA methyltransferase and verified its ability to methylate DNA *in vitro*. Upon optimizations of the mass spectrometry procedure to increase its sensitivity, the authors reported that 0.67% of cytosines in the genome are methylated in an asynchronous parasite culture. When comparing methylation status across the IDC, decrease in methylation was observed as parasite transits into the schizont stage [115]. The authors hypothesize that the loss of DNA methylation could be due to the replication of the genome during schizont stage. Authors also reported a

greater proportion of methylated cytosines in the gene bodies than in the 5' and 3' flanking regions of genes.

1.3.1.8 Non-coding RNA

1.3.1.8.1 Antisense transcripts

In other eukaryotes, antisense transcripts may affect transcription or translation at several levels. One of which is at transcription initiation level, where antisense transcripts inhibit transcription of genes via promoter competition, by effecting DNA methylation or affecting histone modifications [116-118]. On the other hand, antisense transcripts may form duplexes with sense mRNA, affecting mRNA stability or translation efficiency [119, 120]. It was reported that antisense transcripts are involved in the regulation of mRNA stability in another eukaryotic protozoan, *Dictyostelium discoideum* [121].

Earlier studies detected antisense transcripts in *P. falciparum* before the sequence of the genome was published. An example is the 4.4kb antisense transcript of *merozoite surface protein 2* (*msp2*) detected by northern blots during ring and early trophozoite stage, while the sense mRNA of *msp2* is detected in the mid trophozoite and schizont stages [122]. The authors found that the polyadenylated antisense transcript is a result of a read through from a neighboring gene. A genome-wide study at a later date made a similar observation that the transcript levels of a significant portion of antisense transcripts correlates well with its neighboring gene mRNA [123]. In another study, authors successfully knocked down the *clag9* at both transcript and protein level by artificially expressing the antisense transcript of *clag9* [124].

Genome-wide studies utilizing serial analysis of gene expression (SAGE) detected the widespread presence of antisense transcripts in the *P. falciparum* [125, 126]. SAGE is a quantitative method to determine the level of transcripts; in addition, it is able to provide information on the directionality of the transcripts surveyed. Approximately 12% of the transcripts surveyed are in the reverse direction and are distributed evenly in the genome [126]. However, it was found from another study that the antisense transcripts lie preferentially in the 3' ends of the coding sequences [123]. Strikingly, there was an inverse relationship between the levels of sense and antisense transcripts at 707 gene loci surveyed [123]. This may be indicative of a regulatory mechanism by antisense transcripts. In a separate study using nuclear run-on experiments, it was reported that antisense transcripts are synthesized by RNA polymerase II (RNAPII) [127].

Antisense transcripts have more recently been investigated by RNA-seq (new generation sequencing of cDNA derived from RNA) technology. Over 90% of the antisense transcripts are detected in the ookinete and gametocyte stages, suggesting that antisense transcripts play an important role in gene regulation in these stages [128]. The study also detected genes which have high levels of antisense transcript levels in gametocyte stage and high levels of sense mRNA in trophozoite and schizont stage. This is suggestive of antisense transcripts playing a role to suppress trophozoite and schizont sense mRNA during gametocyte stage. However, the mechanism of action of antisense transcripts in *P. falciparum* is still unreported.

1.3.1.8.2 RNA interference (RNAi) and small structured RNA

RNAi has been well characterized in many eukaryotic cells and even in a protozoan parasite, *Trypanosoma brucei*. RNAi mechanism plays a key role in post-transcriptional regulation and

is also exploited as a mean to gene silencing [129]. Briefly, double stranded RNA (dsRNA) is processed to form microRNA, and together with a protein complex, targets its corresponding mRNA for degradation. Incubation of *P. falciparum* parasite cultures with gene specific dsRNA have been reported to be able to knock down its corresponding sense mRNA [130, 131]. However, there are contradicting reports and furthermore, the enzymes required for processing of dsRNA such as Dicer and Arganaute have till date not been found in the *Plasmodium* genomes [132, 133]. In addition, *P. falciparum* specific microRNAs are not detected from the pool of small RNAs obtained from parasite infected RBC [134, 135]. Hence, the role of RNAi in regulating the fate of transcripts is still controversial. Nonetheless, human host RBC microRNAs have been reported to be imported into parasites during the IDC [136]. The levels of at least 2 microRNAs, miR-451 and miR-233, are elevated in sickle cell RBCs and the translocation of miR-451 has been shown to inhibit parasite growth [136].

By *in silico* means, the *P. falciparum* genome is predicted to contain structured RNA which are not tRNA and rRNA [137, 138]. The presence of these structured RNA in *P. falciparum* was also verified by northern blots and microarrays [137, 138]. These included RNA of known functions such as telomerase RNA, small nucleolar RNA (snoRNA) which are involved in processing of rRNA and splicesosomal small nuclear RNA (snRNA), and also structured RNA with unknown functions [138, 139]. Short non-coding transcripts of about 75-175 nucleotides were also reported to originate from centromeres of *P. falciparum* [140]. The transcription of centromeric ncRNAs is driven by a bidirectional promoter in the centromere and the centromeric ncRNAs are associated to the centromere itself [140]. However, the function of the ncRNAs in regulating centromere biology is still unknown.

1.3.1.8.3 Long non-coding RNA (lncRNA)

RNAs which are non-protein-coding and are more than 200 nucleotides long are typically classified as lncRNA [141]. Due to its ability to bind to specific gene loci by sequence complementarity coupled with the ability to bind to specific RNA-binding proteins, lncRNA plays a role in eukaryotic gene regulation in several ways. Firstly, lncRNA is able to induce epigenetic changes to chromatin at specific genomic locations by the recruitment of chromatin modifiers [142]. In addition, lncRNA is also able to affect transcription directly by recruiting transcription-associated factors to gene promoters and enhancers [143].

There is also mounting evidence of the role of lncRNA in *P. falciparum* gene regulation, in particular, virulent gene expression regulation. By surveying the trancriptome of the IDC using tiling microarray, investigators reported 60 putative lncRNA present and most of the putative lncRNA which are conserved in varying degrees in 6 different *Plasmodium* species are encoded in the telomeric and subtelomeric regions [144]. The authors also reported that the conserved lncRNAs were transcribed from at least 15 out of 28 chromosome ends and were all highly transcribed in the schizont stage. Thus, authors speculated that these telomeric lncRNA may recruit chromatin modifiers which regulate transcription or silencing of *var* genes. Further studies have shown that telomeric lncRNA localizes to a novel sub compartment in the nuclear periphery after transcription [145]. It is a separate compartment from the active *var* gene expression site and the nucleolus. The telomeric lncRNA is also able to form hairpin sturctures and is able to bind to factors in nuclear extract [145]. In an earlier study, it was reported that sense and antisense nonpolyadenylated lncRNA are transcribed from a bidirectional promoter from the *var* gene intron and is associated with chromatin [146].

In a more recent study, investigators utilized *P. falciparum* RNA-seq data and predicted 164 lncRNA [23, 147]. The predicted lncRNAs have a high dynamic fluctuation in expression levels in the IDC suggesting that they could play a regulatory role [147]. However, a mechanism of the role of lncRNA, for example its specific binding partners and effects, have not been reported for *P. falciparum*.

1.3.1.9 Nuclear sublocalization

By means of immunofluorescence and electron microscopy, it was shown that nascent mRNA and the active form of RNAPII colocalized at the same foci in an eukaryotic cell, presenting the presence of specific localities of a cell where active transcription takes place [148, 149]. These specific sites of active transcription have been termed 'transcription factories' and the number of foci per nucleus ranges from hundreds to thousands depending on cell types and differentiation states [150-152]. These foci also contain transcription factors, components of the splicing machinery and RNA processing factors [153-155]. In addition, it was also shown that distantly located gene loci which are templates of active transcription can localize in the same 'transcription factory' [156]. This has been further reinforced by the development of a technique which ligates spatially closely associated DNA, coupling to sequencing. It was shown that large distances of the chromosome have been looped out for the interaction of actively transcribed gene loci [157, 158].

The role of nuclear sublocalization in transcription has also been explored in *P. falciparum*. In a recent study, authors labeled nascent RNA with Bromouridine triphosphate (BrUTP) and revealed distinct foci with active transcription in the *P. falciparum* cell [159]. Authors reported that the number of foci is distinctly less than the mammalian cells. The location of foci is also developmentally regulated where a higher percentage of foci is located in the

nuclear periphery during ring stage but evenly distributed in the nucleus during trophozoites [159]. Similar to mammalian cells, the foci appears to be at areas of low chromatin density [160]. Colocalizations with other nuclear markers revealed that transcriptionally active foci lies within the H4ac stained area, which is a general euchromatin marker [159]. As expected, foci were in distinct compartments from silencing factor, PfSir2A. As the nucleolus is where transcription of rRNA genes takes place, it was surprising that the foci does not colocalize with PfNop1, a nucleolus marker [159]. In a separate study, investigators revealed that rDNA gene loci, which are present on subtelomeres of separate chromosomes, are localized in the nucleolus regardless of transcriptional status [161].

Studies were also carried out to characterize the locations of different histone modifications in a cell. Active marks such as H3K4me2 and H3K4me3 are generally diffuse throughout the whole nucleus but has different patterns, punctate and horseshoe respectively [162]. The repressive mark, H3K9me3, forms 2-3 foci at the nuclear periphery [162]. Localizations of histone modifications unstudied in *P. falciparum* were also explored [162]. An example is H3K79me3, an activating mark in other eukaryotes, has a punctate pattern and localizes to 5-7 foci. In addition, H4K20me3 which is a repressive mark localizes preferentially to the nuclear periphery and has a diffused pattern.

By means of immunostaining, there was a study which proposed a model to define the *P. falciparum* nucleus [163]. Nuclear pore stained by an antibody against PfNup100 (a nuclear pore protein), defines the outer limits of the nucleus. Antibody against CenH3 (the variant histone H3 which binds centromeres) defines the region outside DAPI stained region but within the boundaries of PfNup100. Consistent with previous data, H3K9ac stained region is

evenly within the DAPI stained region. In addition, the study also reported the localizations of 12 additional nuclear proteins [163].

Nuclear positioning has also been shown to be important in regulating virulent genes. Silent *var* genes which are enriched with H3K9me3 marks were shown to cluster at repressive centers at the nuclear periphery [164]. Even though the 5'UTR of the active *var* gene has been shown to contain H3K4me3 by ChIP analysis, the active *var* gene loci does not localize with H3K4me3 in colocalization studies [162]. Instead, the active var gene locus is located at H3K4me3 exterior edge [164]. In addition, it was also reported that during the activation of the *var* gene, the activated gene locus is relocated to another region of the nuclear periphery, a separate location from the silenced *var* genes [165]. However, the site of transcription of the active *var* gene is not at the nuclear pore and its associated nuclear structure is still unknown [166]. Such link between transcriptional activation and translocation of gene loci was also observed in an invasion gene, *Reticulocyte binding protein homologue 4 (PfRh4)* [167].

Chromosome conformation capture assays coupled to deep sequencing has been utilized in yeast to uncover interactions of chromosome regions which are large distances apart or even on separate chromosomes [168]. The first of such studies done in *P. falciparum* was published recently [169]. Consistent with cellular localization studies which showed that silent *var* genes clustered in the nuclear periphery, authors reported the interactions of the silent *var* genes through chromosome conformation capture [169]. They also detected the interactions of rDNA genes, which were previously shown to be localized in the nucleolus in colocalization studies [161]. In other eukaryotic systems, enhancers drive transcription from long distances away by looping out chromosomal regions and interacting with local promoter

regions [170]. However, other than the rDNA and *var* gene interactions described above, only few other long range interactions were detected in ring stage *P. falciparum* parasites [169].

It is becoming increasingly clear that the components on the nuclear envelope are involved in chromatin modeling processes and gene regulation [171]. Specifically, nuclear pore complexes (NPCs) were shown to associate with active genes and transcriptionally active foci [172]. By means of electron microscopy, dynamic changes in chromatin organization during the IDC were revealed [173]. In the ring stage, electron dense heterochromatin and electron sparse euchromatin is distributed evenly in the nucleus. Patches of heterochromatin was observed to spread evenly throughout the nucleus during the trophozoite stage. Clear segregation of euchromatin and heterochromatin was observed during the late schizont stage. In addition, nuclear pores cluster at a certain locality during the schizont stage and is only found beside euchromatin. The polarized localization of nuclear pores was also observed in a separate study done using immunofluorescence assay [166].

1.3.2 Transcriptional regulation

1.3.2.1 The core promoter and the general transcription machinery

Transcription of mRNA in eukaryotes begins with the assembly of the Pre Initiation Complex (PIC) over the core promoter region of a gene. The complex consists of RNA polymerase II (RNAPII) and the General Transcription Factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. Each of the GTFs and RNAPII in the PIC is a complex of individual proteins. Table 1.1 summarizes the protein components of each PIC element.

TFIIA	p35	MAL7P1.78
	p29	
	p12	PFI1630c, PFL2435w
TFIIB	p33	PFA0525w, PFE0415w
TFIID	TBP	PFE0305w, PF14_0267
	TAF1	PF1645w
	TAF2	MAL7P1.134
	TAF3	
	TAF4	
	TAF5	
	TAF6	
	TAF7	PFI1425w
	TAF8	
	TAF9	
	TAF10	PF1110w
	TAF11	
	TAF12	
	TAF13	
	TAF14	
TFIIE	p56	MAL7P1.86
	p34	MAL13P1.360
TFIIF	RAP30	
	RAP74	PF11_0458
TFIIH	p89/XPB	PF10_0369
	p80/XPD	PFI1650w
	p62	PFC1055w
	p52	PFL2125c
	p44	MAL13P1.76
	p40/CDK7	
	p38/ Cyclin H	PF13_0022
	p34	PF13_0279
	p32/ MAT1	PFE0610c
	p8/TFB5	PF14_0398
RNA pol II	RBP1	PFC0805w
	RBP2	PFB0715w
	RBP3	PFI1130c
	RBP4	PFB0245c
	RBP5	PF13_0341
	RBP6	PFC0155c
	RBP7	PF10_0269
	RBP8	PFL0665c
	RBP9	PFA0505c
	RBP10	PF07_0027
	RBP11	PF13_0023
	RBP12	MAL13P1.213

Table 1.1 The Pre-initiation complex (PIC) [174, 175].

In bold are the components of the PIC in the eukaryotic system. The protein members of each component are listed in the second column. The *P. falciparum* gene which is identified as the orthologue of each member of the components of the PIC is listed correspondingly in the third column. The members of the PIC components which do not have identified orthologues in the *P. falciparum* genome are colored in pink. Most of the PIC components were found in the genome, showing that the basal transcription machinery of *P. falciparum* is largely conserved. However, a large proportion of the TAFs which play a role in the recruitment of the PIC to the core promoter are not found in the *P. falciparum* genome.

The assembly of the PIC over core promoter region in eukaryotes is generally viewed as a sequential event where elements of the PIC bind in a specific order. The TATA box is first recognized by TFIID, after which, TFIIA and TFIIB binds to stabilize the binding of TFIID to the promoter region. RNAPII and TFIIF are then assembled to the PIC. This is followed by the binding of TFIIE and finally, TFIIH. The general functions of each element are as follows. TFIIA generally functions to stabilize the TFIID's binding to the core promoter region. In addition to the stabilizing function, TFIIB also recruits RNAPII and TFIIF. RNAPll is responsible for initiating, elongation and terminating transcripts. TFIIF is responsible for increasing the efficiency of transcription. TFIIE is involved in promoter clearance and recruiting TFIIH. TFIIH functions as a helicase for the unwinding of DNA during transcription.

1.3.2.1.1 TFIID complex

In model eukaryotes, the core promoter commonly refers to a region in the genome upstream of the coding sequence of a gene which contains specific DNA sequences competent of recruiting the PIC and initiating transcription [176, 177]. The TFIID is traditionally viewed to play a central role in recognizing the core promoter sequences. As seen in Table 1.1, the TFIID is a protein complex consisting of the TATA-binding protein (TBP) and approximately 13 different TBP-associated factors (TAFs) [178]. The TATA box is located 25-30 bp upstream of the TSS and is recognized by the TBP [179]. Although it has been previously held that a prerequisite for PIC recruitment to the core promoter is the recognition of TATA-box in the core promoter by TBP, it has become increasingly clear in the past decade that it is not the only mechanism for PIC recruitment. This is substantiated by evidence that only less than 10% of human promoters contain the TATA box [180]. Investigators have also shown that in yeast, the presence of the TATA-box in gene promoters

does not necessarily equate to active gene transcription [181, 182]. In human HeLa cells, TBP-independent transcription of interferon responsive genes has also been reported [183].

Studies have shown that several TAFs (TAF4, 5, 6, 8, 9, 10 and 12) form the TFIID core complex and, together in different combinations with other TAFs and TBP, can form a variety of TFIID complexes [184, 185]. It was also found that there are multiple genes encoding for TBP-related factors (TRFs) and variant forms of TAFs in higher eukaryotes [186]. In addition, some components of the TFIID complex are only expressed in specific cell types in a given organism [187, 188]. For example, in the differentiation of muscle cell precursors into myotubes, subunits of TFIID (TAF1, TAF4 and TBP) are expressed at low levels [189]. Instead, the complex is replaced with a TRF3/TAF3 complex. This allows a switch in the recognition of core promoter and activation of Myogenin, one of the key transcription factors which establishes myogenic differentiation [189]. TRF3 has also been implicated in haematopoiesis and investigators have shown that TRF3 binds to the promoter of an important transcription factor involved in development and depletion of TRF3 causes developmental defects [190]. As such, this allows a great flexibility in the number of permutations of TFIID or TFIID-like complexes that can be formed. This in turn gives specificity to the selection of core promoters that a particular complex can recognize and thus genes can be regulated spatially and temporally.

1.3.2.1.2 Core promoter elements

Advances in sequencing technologies have allowed a better view of core promoters in eukaryotes. By large scale sequencing of 5’ ends of cDNA, a clearer picture of genome wide TSSs has been defined [191, 192]. Based on the characteristic of the TSS, eukaryotic core promoters can be broadly defined into 2 groups: (1) Promoters which have clearly defined

TSS. In these promoters, the TSSs are only a few base pairs apart. The TATA box is commonly associated with these promoters. (2) Promoters which have a broad distribution of TSS over a 100 bp region. These promoters lack TATA boxes and typical core promoter elements but contain CpG islands [193]. In vertebrates, most promoters have a board distribution of TSS [193]. Promoters containing TATA box are commonly associated with the regulation of tissue-specific genes while TATA-less promoters are associated with ubiquitously expressed genes. However, they are exceptions in both cases [194, 195].

Flexibility in recognition complexes such as TFIID may suggest that there is a diversity of DNA motifs in the core promoter where the complexes can bind to. Core promoters were once viewed as providing a basal level of transcription and plays little role in spatiotemporal control of transcription. However in the past decade, in addition to the TATA-box, there are other motifs at the core promoter regions which play a role in the recruitment of the PIC to the core promoter have been found. Some examples of commonly known motifs are listed in Table 1.2, with its sequence, relative position from TSS and binding protein for each motif.

Motif in core promoter	Position (with respect to TSS)	Sequence	Binding protein
TATA box (TATA)	Metazoans: from -31 to -24 Yeast: from -40 to -100	TATAWAAR	TBP
Initiator (Inr)	from -2 to +5	YYANWYY in humans TCAKTY in Drosophila	TAF1/TAF2
Downstream promoter element (DPE)	from +28 to +34	RGWYVT in Drosophila	TAF6/TAF9
Motif Ten Element (MTE)	from +18 to +27	CSARCSSAAC in Drosophila	Undetermined
Upstream TFIIB-Recognition Element (BRE^u)	from -38 to -32	SSRCGCC	TFIIB
Downstream TFIIB-Recognition Element (BRE^d)	from -23 to -17	RTDKKKK	TFIIB
Downstream Core Element (DCE)	S_I: +6 to +11 S_{II}: +16 to +21 S_{III}: +30 to +34	S_I: CTTC S_{II}: CTGT S_{III}: AGC	TAF1
X core promoter element 1 (XCPE1)	from -8 to +2	DSGYGGRASM	Undetermined

Table 1.2 Motifs in the core promoter [174].
The motifs which are found in a typical eukaryotic core promoter are as shown. Degenerate nucleotides are presented according to the IUPAC nucleotide code. Most core promoters have a 'mix-and-match' variety of these motifs. The positions of the motifs stated are relative to the TSS. The following column shows the sequence of each motif and the final column shows the protein component of the PIC which binds to the motif.

The TSS is embedded in Initiator (Inr) motif and is bound by TAF1/TAF2 in the TFIID complex during transcription initiation [196, 197]. The "A" nucleotide in the motif is often the +1 start site with respect to the TSS. The Inr is conserved in the eukaryotes and has been described in humans, Drosophila and yeast [198]. Inr is present in both TATA-box containing

promoters and TATA-less promoters. The Inr has also been reported to be one of the most common motifs in promoters with clearly defined TSS [199].

The Downstream Promoter Element (DPE) lies about 30 bp downstream of the TSS. The DPE was first discovered in Drosophila but DPE is also found in humans, however, the consensus sequence in humans is still yet to be determined [200]. DPE functions in conjunction with Inr and the exact spacing between the 2 motifs is required for optimal function [200]. Approximately 2-20% of Drosophila genes contain DPE in their core promoter [201]. The DPE is recognized by TAF9/TAF6 in the TFIID [202]. In addition, DPE is found to be present on core promoters of *Hox* genes which are TATA-less [203].

Motif Ten Element (MTE) was first studied in *Drosophila* where the mutation of bases +18 to +22 with respect to the TSS abolishes MTE activity [204]. The authors also reported that MTE is also able to compensate loss of transcriptional activity due to mutations in the TATA-box and DPE. In addition, Inr is required in the function of MTE but not TATA-box and DPE. There has not been strong evidence of presence and function of MTE in humans as MTE has not been reported to be overrepresented in human promoters by using bioinformatics searches [205]. However, by using MTE on a *Drosophila* promoter, a study reported that MTE is recognized by human transcription factors [204]. In the same study, authors also identified a human promoter containing MTE but the binding partner of MTE has not been reported.

The TFIIB Recognition Element (BRE) was first reported as a stretch of sequences immediately upstream of the TATA-box which binds to TFIIB [206]. It was later found that TFIIB also binds to sequences immediately downstream of the TATA-box [207]. Hence they

are termed upstream BRE (BRE^u) and downstream BRE (BRE^d) respectively. The BRE functions together with the TATA box [208]. BRE is able to function both as a transcriptional activator and repressor depending on the context of the promoter BRE is in [206, 207]. It is one of the few motifs in the core promoter which is not recognized by the TFIID complex.

The Downstream Core Element (DCE) was first reported in the human β-globin promoter where mutations at 3 points on the promoter decreased transcription [209]. Authors also reported that DCE functions in conjunction with the Inr motif in the absence of the TATA-box. The DCE contains 3 sub-elements and functions in a TFIID-dependent manner [210]. In addition, authors also showed from UV photo crosslinking experiments that TAF1 is in close proximity to DCE. In addition, bioinformatics analysis revealed that DCE is present on many human promoters and occurs preferentially on promoters with the TATA-box [210].

The X Core Promoter Element 1 (XCPE1) is present on approximately 1% of the human promoters [211]. By itself, the XCPE1 exerts little transcriptional activation but activates transcription more efficiently with known activator binding sites such as NRF1, NF1 and SP1 [211]. Authors also reported that XCPE1 occurs frequently on TATA-less promoter.

A core promoter typically contains more than 1 type of motif and there is no one universal element which is present in all promoters [176, 177]. This allows for a large diversity of promoters which contains motifs in multiple permutations. A common theme that is evident in the examples listed above is that the motifs seldom function individually. The function of the motif often modulated or works in conjunction with other core promoter motifs. In addition, the function of a motif may be different or even opposite depending on the context of the promoter that the motif lies in.

Apart from the "mix-and-match" of motifs on the core promoter itself, studies have shown that enhancers add an additional layer of regulation by modulating the activity of the core promoter. In *Drosophila*, it has been shown that certain enhancers preferentially activate a promoter which contain TATA-box over the TATA-less promoters which contain DPE [212]. In addition, transcriptional activator VP16 preferentially activates TATA-box containing promoters while another transcriptional activator, Sp1, has a stronger preference for Inr containing promoters [213]. The function of enhancers in transcription regulation will be elaborated further in the later section.

1.3.2.2 Proximal promoters, enhancers, silencers and insulators

Apart from the core promoter, additional regulatory elements are proximal promoters, enhancers and silencers which are located further upstream of the core promoter [214]. The proximal promoter is located approximately -250 bp to -300 bp of the TSS. It serves as a binding site for sequence-specific transcription factors. However, it has been postulated that the proximal promoter alone does not always regulate transcriptional activity but rather serves as a docking platform for long-range enhancers to interact with the core promoter in proximity [215, 216].

Enhancers are regulatory elements which are positioned large distances away upstream of the TSS, within exons or introns, or downstream of the gene [217-219]. There is even mounting evidence of enhancers activating genes which are located on a different chromosome [220]. Enhancers are currently viewed as stretches of DNA sequences capable of binding to transcription factors, which together with the Mediator complex, enhance the efficiency of mRNA transcription [221].

A key question to answer is how does an element which is in such variable distances away from the gene loci regulate gene activity. By means of fluorescence *in situ* hybridization (FISH) assays and the recent development of chromosome capture technologies such as 3C, there have been evidence of physical interactions of distant elements in the genome [222]. The *β-globin* gene loci provided an excellent model in the study of long range interactions [223]. The *β-globin* gene locus is a gene cluster which contains the multiple variants of the hemoglobin β chain. These genes are regulated by enhancer elements which are located ranging from 25kb upstream to 20kb downstream of the locus. Transcription of the *β-globin* genes requires the presence of specific transcription factors, the erythroid transcription factor (GATA1) and the Kruppel-like factor 1 (KLF1), which enables the interactions between the distant enhancers and promoters of the *β-globin* genes [224, 225]. This is also known as the 'chromatin hub' where actively transcribed genes are clustered together. It was later found that such tissue specific interactions occur at other gene loci as well, such as the *kit* [226], *igf2* [227] and *T helper 2 cytokine loci* [228]. In addition, CCCTC-binding factor (CTCF) and Cohesin were also shown to mediate chromatin looping for long range physical interactions [229, 230].

The cyclic AMP-responsive element binding protein (CBP) and p300 possess acetyltransferase activity and have been implicated in transcriptional regulation [231]. An additional feature of enhancers is the presence of CBP and p300 [232, 233]. Histone modifications, H3K4me, H3K4me2 and H3K27ac, are also markers of enhancers [234]. Importantly, authors reported that histone modifications in enhancer regions are cell-type specific and are associated with enhancer activity, resulting in cell-type specific gene expression pattern [234]. This is in contrast to chromatin state of promoter regions where

histone modifications are invariant among different cell types. This observation was supported by other studies where H3K4me and H3K4me2 are present on cell-type specific enhancers in B cells [235] and breast cancer cells [236]. H3K4me and H3K4me2 have also been shown to facilitate transcription factor binding at enhancer sites as depletion of these modifications reduced transcription factor binding [237]. In addition, highly unstable nucleosomes containing histone variants H2A.Z and H3.3 are present on enhancer elements and is associated with increased transcription factor binding [238, 239]

The eukaryotic DNA is packaged into nucleosomes and may be inaccessible to transcription factors. Hence, the question of how transcription factors bind to enhancers or any regulatory element in the light of chromatin hindrance arises. Pioneer transcription factors are a special class of transcription factors which may engage target sites in chromatin first and subsequently allowing the assembly of other transcription factors. First, it was observed that pioneer transcription factors have enhanced ability to bind to compacted chromatin. PHA-4, a transcription factor in *Caenorhabditis elegans*, has been shown to bind to compacted chromatin at its enhancer sites and subsequently inducing its decompaction [240]. Next, pioneer transcription factors demonstrate the ability to induce changes to chromatin structure by inducing changes to histone modifications and DNA methylation [241-243].

An emerging theme is the discovery of 2 classes of ncRNA, the enhancer RNA (eRNA) and the long intergenic non-coding RNA (lincRNA). The eRNA was first reported in mouse cortical neurons where CBP binds to enhancers and RNAPII is recruited by CBP to a subset of these enhancers [244]. Authors showed that the eRNA is transcribed in the mediation of long range interactions where the enhancer is brought to close proximity of an active promoter and, possibly, playing a role to stabilize enhancer-promoter interactions [244].

However, the actual role of the eRNA is still unclear. An example of lincRNA is transcript from the 5' end of a *HOXA* gene locus which is required for the transcription of a cluster of *HOXA* genes in human cells [245]. Authors showed that the lincRNA brought the activated genes in close proximity through DNA looping and also serve as a scaffold for binding of an adapter protein and a methyltransferase. This led to the trimethylation of H3 histones on lysine 4 and hence activation of the target genes [245].

As its name suggests, silencers play the opposite role of enhancers to repress the transcription of genes. Similar to the properties of enhancers, silencers can influence transcription from long distances away from the target gene loci [246]. Silencers serve as binding sites for protein elements called repressors which in turn recruit co-repressors in order to elicit their activity [247].

Several models have been proposed for silencer function. The first is when the binding of a repressive transcription factor to silencer binding sites blocks the binding of an activator by binding to a nearby site [248] or competing for the same site [249]. Next, repressor binding may reduce chromatin accessibility by inducing repressive chromatin and thus preventing activating transcription factor binding [250]. In this example, the binding of a repressor, YY1, recruits members of the Polycomb group proteins (PcG) which induce repressive chromatin by deacetylation and methylation of histone H3 [250]. Repressor binding may also directly prevent the assembly of the PIC [251]. In this study done in yeast, repressor binding did not reduce activator binding but rather reduced the binding of TFIIB, TFIIE and RNAPII [251].

To prevent unspecific effects of enhancers or silencers on non-target genes, insulators are DNA elements on the genome which act as boundaries to block undesired effects. It has been reported that deletions and mutations on insulators disrupted gene expression patterns and thus caused defects in development [252]. By the nature of their effects, insulator elements are commonly divided into 2 groups: the enhancer-blocker which blocks the activating effects from enhancers on promoters and the barrier insulators which inhibit the spread of heterochromatin. Some insulators may exhibit characteristics of both.

Typically, insulators displaying enhancer-blocker effects elicit its function when it is placed in between the enhancer and promoter but not in flanking position [253]. In humans and mouse embryonic tissues, the imprinted expression of *Insulin growth factor 2* (*Igf2*) from the paternal locus is achieved by methylation of insulator elements in between the *Igf2* gene locus and its downstream enhancer [254]. It was also found that CTCF-binding sites are present on the insulator and, the methylation of the insulator element abrogates CTCF binding and insulator activity [109, 255]. One model proposed that the mode of action of CTCF in enhancer-blocking function is CTCF molecules binding at different genomic locations can interact with each other and tether to the nucleolus to form closed DNA loop domains which keeps promoters and enhancers in separate domains, preventing their interaction [256, 257]. This is also supported by evidence that insulator-insulator interactions may form distinct chromatin loops which disfavor enhancer-promoter interactions [257, 258]. However, the other model is based on evidence that insulators interact with promoters and enhancers physically and thereby, possibly disrupting the interaction between the promoter and enhancer [259, 260].

Barrier-insulators were first reported as DnaseI hypersensitive regions in between open and condensed chromatin [261]. Heterochromatin is characterized by H3K9me3 and H3K27me3, this is often accompanied with low levels of histone acetylation and the presence of HP1 [262]. In the *β-globin* gene example above, apart from having enhancer-blocking properties, the insulator described also displays properties of a barrier-insulator which is characterized by peak levels of histone acetylation and methylation at histone H3 lysine 4 [263, 264]. These histone modifications are the result of the recruitment of sequence-specific transcription factors and in turn recruiting HATs and HMTs [265]. This has led to the proposal that histone modifications at the insulator has prevented the spread of heterochromatin formation that is adjacent upstream of the insulator. Another possibility for the mechanism of the function of the insulator is from the observation that in some cases barrier-insulators are recruited to specific loci in the nucleus [256, 266]. Although the exact mechanism of how subnuclear localization prevents heterochromatin from spreading is still unclear, it is speculated that the microenvironment in which the insulator is recruited to may be disfavoring for heterochromatin to form.

Specific transcription factors (STFs) activate specific groups of genes upon developmental and environmental cues. In addition, these STFs recruit co-activators to relay regulatory signals to influence the activity of the PIC appropriately [267]. The most important group of co-activators is the Mediator complex, a complex containing approximately 30 protein subunits, which is found in to be conserved in many eukaryotic organisms including yeast, metazoans and plants [268-270]. The Mediator consists of 4 different modules, the ‘head’ which interacts with the RNAPII machinery, the ‘tail’ which interacts with activators, the ‘middle’ and the ‘kinase’ module which binds on the same part of the complex as where

RNAPII would bind [271, 272]. Hence, the kinase module is generally found on the inactive form of the Mediator complex as it may prevent Mediator from binding to RNAPII [273].

There are several mechanisms by which Mediator regulates transcription. The first involves the recruitment PIC. The first line of evidence is that electron microscopy reveals that Mediator makes physical contacts with the RNAPII [274]. It has also been shown that Mediator regulates the recruitment of TFIIH [275], and also likely TFIIB [276]. Mediator also regulates transcription by being involved in chromatin remodeling. Mediator has been reported to interact with p300, a histone acetyltransferase [277]. The complex then acetylates histones at the target locus, after which, p300 is replaced by TFIID, leading to the assembly of the PIC [277]. Importantly, tissue-specific transcriptional co-activators have been shown to mediate interactions between Mediator and chromatin modifying factors, leading to alterations to chromatin structure and transcriptional activation [278]. This allows tissue-specific activation of genes. Mediator has also been reported to induce heterochromatin through its interaction with G9a, a methyltransferase, causing the di-methylation of histone H8 on lysine 9 and has led to the repression of target genes [279].

In recent years, there is increasing studies which report the assembly of complete PIC but did not correlate with productive elongation of transcripts [280]. The association of Mediator with various elongation factors such as DSIF [281] and P-TEFb [282] have been observed to promote elongation. Specifically, Mediator subunit MED31 has been shown to have similar functions as elongation factors in yeast [283, 284]. From yeast studies, it was observed that after the PIC is released from the promoter for transcription, the Mediator remains attached to the activator for subsequent rounds of transcription [285]. The Mediator has also been implicated in promoter-enhancer interactions. By interactions with the cohesin complex

components, Mediator plays a role in mediating DNA loops which enables enhancers to interact with promoters [286]. Authors also reported that physical interactions between cohesin and Mediator forms a ring structure which facilitates the formation of DNA loops.

1.3.2.3 Promoter-proximal pausing of RNAPII

Promoter-proximal pausing of RNAPII is an emerging theme in the regulation of transcription. A large focus of research in transcriptional regulation has been traditionally focused on regulation on the level of transcription initiation. However, recent evidence has revealed that promoter pausing of the RNAPII plays a substantial role in transcriptional regulation as well. In early studies, it has been reported that the initiation of transcription does not necessitate the detection a full length transcript [287, 288].

The first observation of promoter pausing was in the *Drosophila heat shock protein* genes (*hsp* genes) where RNAPII accumulates just downstream of the *hsp* promoters and associates with a 20-60 nucleotide long transcript [289, 290]. It should be noted that RNAPII, in this case, pauses specifically in the promoter region and not in the gene body. The development of genome-wide techniques such as ChIP-chip and ChIP-seq has allowed the characterization of binding of RNAPII. In contrast to yeast [291] where RNAPII binding is uniform throughout the transcriptional unit (5' UTR, gene body and 3' UTR), RNAPII was observed to be accumulated near the TSS of a large proportion of genes in *Drosophila* [292, 293] and mammalian cells [294]. It was also found that, on the promoters of most protein-coding genes of human embryonic stem cells, paused RNAPII is accompanied with histone modifications which indicate of active transcription [294]. Global nuclear run-on sequencing revealed that these promoters which associate with paused RNAPII are capable of resuming transcription [295]. Several studies showed that approximately 30-90% of genes in mouse embryonic stem

cells, *Drosophila* cells and human lung fibroblast cells display promoter-proximal pausing [280, 295, 296]. It is also important to note that promoter pausing was observed in genes with a range of expression levels and less than 1% of genes which display pausing are inactive [295, 297]. This is indicative that promoter pausing is a mechanism for fine tuning the expression of active genes and unlikely to be a mechanism of the activation of inactive genes. In *Drosophila*, low nucleosomal occupancy was observed on the promoters of genes which display RNAPII promoter pausing [297]. In addition, depletion of pausing factor, NELF, resulted in loss of RNAPII at paused genes and increased levels of nucleosomes [298]. This suggests that paused RNAPII aids the maintenance of an open chromatin structure which is accessible to incoming transcription factors.

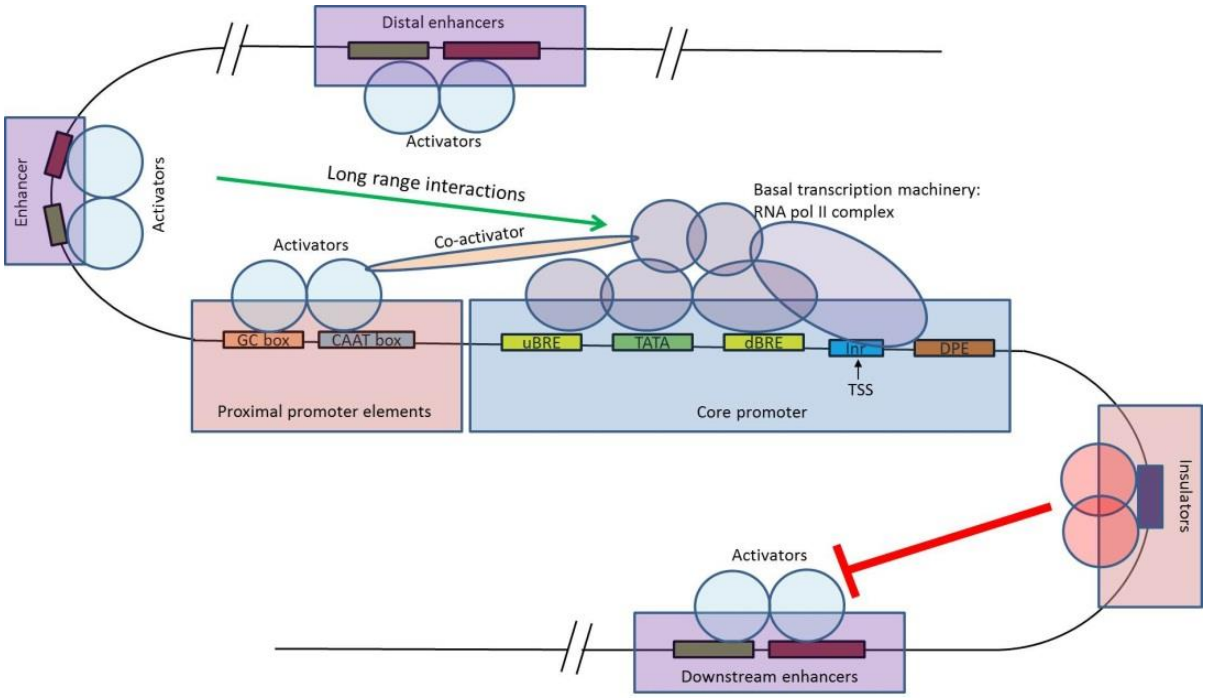


Figure 1.3 The cis-regulatory elements in a typical eukaryotic gene locus.
The core promoter serves as a platform for the assembly of the basal transcription machinery or the PIC. The assembly of the PIC is influenced by upstream regulatory elements like the proximal promoter, the enhancer and distal enhancers. The distal enhancer can lie in variable distances away from the core promoter and can even lie downstream of the gene. A final element is the insulator which blocks the long range interactions of an enhancer with the core

promoter to prevent inappropriate enhancer activation or prevents the spread of heterochromatin.

Earlier studies in *P. falciparum* transcriptional regulation revealed that the mechanisms are relatively similar to the eukaryotic model. It was demonstrated that transcription in *P. falciparum* is monocistronic as the transcript of the upstream gene terminates at a region before the initiation of the downstream gene [299]. Secondly, authors also demonstrated that transcription is dependent on RNAPII as α-amanitin inhibited transcript production [299]. In addition, it was also shown that transcription correlated with the binding of nuclear factors to specific sites near to the TSS [299, 300].

1.3.2.4 Basal transcription machinery in P. falciparum

P. falciparum homologues of all 12 subunits of the RNAPII and most of the components of the PIC have been identified (Table 1.1) [301]. Although there have been reports on notable exceptions as described in section 1.3.2.1.1, the recognition of the core promoter relies largely on the TFIID complex in eukaryotes. Therefore, an interesting exception in the GTFs in *P. falciparum* is TFIID where most of its components are not identified [175] (Table 1.1). In addition, the identification of common eukaryotic DNA motifs in the core promoter has not been reported in *P. falciparum* as well.

A notable exception is the *P. falciparum* homologue of the TBP (PfTBP) and its binding motif, the TATA-box. PfTBP has been shown to bind to TATA-box-like sequences near the TSS of 2 genes and localizes to the parasite nucleus [302]. Although there are no reports of genome-wide occurrences of the TATA-box in the *P. falciparum* genome, it will not be surprising to find an abundance of the TATA-box like sequences in a genome which is extremely AT-rich, especially in the intergenic regions [9].

It was also observed that PfTBP and PfTFIIE, which are components of the PIC, are assembled on promoter regions of IDC active genes regardless of stage specificity [303]. For example, a gene which transcript abundance peaks in the ring stage has the PIC assembled on its promoter regions both in ring and trophozoite stage. This was not observed for non-IDC specific genes, such as gametocyte and sporozoite specific genes [303]. Hence, authors proposed that the IDC genes are in a transcriptionally "poised" state where the PIC is assembled on promoters regardless of transcription status. However, it must be noted that this study was done on 6 specific genes and not a genome-wide study.

1.3.2.5 Gene-specific transcription regulation

In addition to the assembly of basal transcription machinery, gene transcription is spatiotemporally regulated by gene-specific transcription factors (STFs). Efforts have been made to search the genome for STFs since the genome was published in 2003. Although the basal transcription machinery or components in the PIC have been largely found in the *P. falciparum* genome, fewer than usual STFs were found and functionally annotated [304]. An earlier study searched 8 eukaryotic genomes including human, yeast and *P. falciparum* using 51 profile-hidden Markov models of specific transcription factors characterized in other organisms [301]. The search identified only 69 proteins in the *P. falciparum* and, 249 for *S. cerevisiae* and 205 for *S. pombe* as STFs. All 3 organisms have a genome size of about 5000 open reading frames (ORFs). It was then thought that post-transcriptional regulation and epigenetic regulation played a more significant role in transcript regulation as there was a lack of specific transcription factors.

This was held until the discovery of the presence of 27 proteins with the Apetala2-integrase (AP2) DNA binding domain in 6 *Plasmodium* species, now known as the Apicomplexan AP2 domain protein family (ApiAP2) [305]. The AP2 domain is commonly found in plant transcription factors [306]. A later study combined all earlier *in silico* predictions [307, 308], histones and histone modifying proteins [43], *in vitro* characterized PfMyb protein [309] and the ApiAP2 proteins [305]. The study reported a directory of transcription associated proteins reaching a number of 202 [175]. Out of this number, 73 are in the category of STFs, bringing the numbers closer to 169 STFs in yeast [310]. With about 90% of STFs belonging to the ApiAP2 and zinc finger families, there is a lack of variety of STFs when compared to the list in TRANSFAC (a database of eukaryotic transcription factors) [311]. This may suggest that *Plasmodium* may contain STFs which are unique from the other well studied eukaryotes.

1.3.2.6 Myb transcription factors

Despite the lack of knowledge of specific transcription factors, there were individual studies to characterize the proteins that were annotated as transcription factors. In an early study of *P. falciparum* transcription factors, alignment of the Myb domain of a eukaryotic transcription factor family led to the discovery of PfMyb1 and authors showed binding of PfMyb1 to a conserved eukaryotic Myb-regulatory element (MRE) [312]. Double stranded RNA knock down of PfMyb1 led to a 40% growth inhibition and a small group of genes were found to be differentially expressed [309]. Most of these genes contains the MRE and ChIP experiments also showed that the PfMyb1 binds to the promoter regions of these genes [309].

1.3.2.7 Prx regulatory element (PRE) binding protein (PREBP)

The PRE is a 102 bp DNA enhancer element on the upstream regions of the *pf1-cys-prx* gene [313]. The *pf1-cys-prx* gene encodes for an antioxidant protein which mRNA is highly

expressed during the trophozoite/schizont stages [314]. Using DNA-affinity purification, PREBP was identified to be binding to PRE and likely to regulate the specific transcription of the *pf1-cys-prx* gene [315]. PREBP is a novel transcription factor in *P. falciparum* which contains 4 K-homology domains which is known to bind to single stranded DNA or RNA [316]. Proteins containing the K-homology domain have not been described in *Plasmodium*. Transient overexpression of PREBP leads to increase in expression of the reporter gene when PRE is present in the promoter driving the expression of the reporter gene [315]. This provided evidence that PREBP induces the activity of the PRE-containing promoter in addition to its ability to bind to the PRE DNA element [315].

1.3.2.8 Acetylation lowers binding affinity (Alba) protein family

PfAlba3 is a DNA-binding protein which may play a role in transcription regulation [317]. Authors demonstrated that PfAlba3 inhibits transcriptional activity by binding to DNA *in vitro*, however, the binding is not sequence-specific. As its name suggests, acetylation of the protein lowers its binding affinity to DNA and 80% of the total amount of PfAlba3 in a cell is acetylated [317]. The endogenous PfAlba3 is localized in the nucleus throughout the IDC. In addition, co-immunofluorescence revealed the colocalization of PfAlba3 and PfSir2 during the ring and trophozoite stage [317]. Also, PfAlba3 was found to be localize at the telomeric clusters by both immunofluorescence and ChIP. The association of PfAlba3 at telomeric regions and interaction with PfSir2 suggests of its role in *var* gene regulation [317]. However, its exact role has not been elucidated.

1.3.2.9 ApiAP2 family

The largest and the most well characterized group of STFs in *P. falciparum* is the ApiAP2 protein family. DNA sequences which bind to individual ApiAP2 protein have been

discovered by protein-binding microarray [318, 319]. Further, authors also reported that the transcript profile of 2 ApiAP2 proteins are highly correlated with its putative target genes which are genes containing the predicted DNA motif in their 5' upstream regions [318, 319].

AP2-O is one of the ApiAP2 proteins which is found to be functional during the ookinite stage of *P. berghei* [320]. It was found that AP2-O activates various ookinite stage genes by binding to a 6 base motif (TAGCTA) in their promoter regions [320]. Another ApiAP2 protein, AP2-Sp has been shown to be essential for sporozoite formation in *P. berghei* [321]. In addition, AP2-Sp was able to bind to an 8-bp motif, TGCATGCA, which was found on most of the sporozoite specific genes [321]. Another ApiAP2 protein, AP2-L, has been characterized again in *P. berghei* [322]. Authors found that parasites depleted of AP2-L have a decreased ability to develop in hepatocytes [322]. In addition, the expression of several liver stage genes and membrane proteins were found to decrease after AP2-L depletion. However, authors did not report DNA sequences which bind to AP2-L.

The only ApiAP2 protein characterized in *P. falciparum* IDC is PfSIP2 (SPE2 interacting protein) which binds to the previously characterized SPE2 motif and is involved in subtelomeric heterochromatin formation [323]. Multiple copies of the SPE2 motif [(T/G)GTGC(A/G)(N)4(T/G)GTGC(A/G)] is found in approximately 2kb upstream of upsB *var* genes and is associated with gene silencing [324]. PfSIP2 is found to be enriched in arrays of SPE2 motifs upstream of upsB *var* genes and is within heterochromatin regions marked by H3K9me3 and HP1 by ChIP-chip assays [323].

Even more recently, PfAP2-G is characterized in *P. falciparum* as a key regulator of gametocytogenesis [325]. Authors reported higher transcript levels of *pfap2-g* in parasite

clones which produce more gametocytes and knock out of the gene ablates the ability to produce gametocytes [325]. It was also found that PfAP2-G binds to a specific motif on promoters of early gametocyte genes and thereby affecting their transcriptional activity [325]. Importantly, the authors proposed that *pfap2-g* is kept silent under heterochromatin in most parasites whereas in a small number of sexual committed parasites, *pfap2-g* is in permissive chromatin where it transcribed. The *P. berghei* homologue of PfAP2-G, PbAP2-G, was also characterized as a transcription factor playing a role in the commitment to gametocytogenesis which binds to the same DNA motif sequence as PfAP2-G [326]. In addition, authors also reported a secondary transcription factor, Pb-AP2-G2, which is required after sexual differentiation.

Notably, there are no reports of ApiAP2 proteins in the regulation of gene transcription in the IDC which may be due to the difficulty in observing a phenotype after gene disruption in the blood stages as the transcription factors may have overlapping functions. Although extensive work has been done on the ApiAP2 proteins, there has been no concrete evidence or identification of factors which control the transcriptional cascade during the IDC.

1.3.2.10 Cis-regulatory elements

Cis-regulatory elements are usually contained within the 5' UTR of gene transcripts. Due to the extreme AT-richness of intergenic regions in *P. falciparum*, it has been difficult to determine the lengths of UTRs by sequencing of transcript ends. In a more recent study, a group collated northern blot data from 105 transcripts and reported that *P. falciparum* 5' UTRs are about 600 – 1350 bp [327]. Regulatory elements in *P. falciparum* have been largely characterized by transient transfections of stepwise deletions of the 5' UTR of genes. These studies have largely shown a common theme in the *P. falciparum* promoters that a region in

the promoter region is important for basal transcription and another region which have an enhancer effect [313, 328-335]. In some cases, deletion of sequences resulted in increase in gene reporter activity, indicating of repressor elements [336-338]. In addition, sequences in promoter regions of yeast are not able to drive reporter gene transcription in *P. falciparum* [329]. Interestingly, it was also reported that promoter sequences from *P. vivax*, another species which infects humans, drives little or no transcription in *P. falciparum* [339]. This seems to suggest that the transcription-competent sequences upstream of *P. falciparum* genes are specific for *P. falciparum*, and possibly bind factors unique in *P. falciparum*. A list of *cis*-regulatory elements discovered in *P. falciparum* which have been functionally verified by various means have been summarized in Table 1.3.

There have been only few demonstrations of *Plasmodium* promoters controlling the temporal expression of gene transcripts. A notable example is a gametocyte-specific promoter, authors showed by means of stable transfections showed that 5' and 3' UTRs of a gametocyte specific gene (*pfs27*) are sufficient to drive the expression a reporter gene specifically in the gametocyte stage of *P. falciparum* [340]. In a separate study, it was also shown that the promoter of a liver stage specific gene is able to drive a reporter gene expression specifically in *P. berghei* liver stage and not other stages in the life cycle [341]. By a series of deletions of the promoter of *merozoite surface protein 2 (msp2)*, authors from another study showed that certain regions in the promoter are responsible for the correct temporal expression of the *msp2* transcript in the late schizont stage [338]. The promoter of a housekeeping gene expressed during the IDC of *P. falciparum*, *proliferating cell nuclear antigen (Pfpcna)*, was also demonstrated to be able to direct expression of reporter gene temporally in the late trophozoite stage [342].

In *Plasmodium*, the extreme AT-rich genome poses as a main hindrance to *in silico* motif searches. Similar to STFs identified by bioinformatics means, *in silico* identified motifs are largely left functionally uncharacterized with a few exceptions. A study reported a functional G-box motif in approximately 200 bp upstream of the TSS of *Pfhsp86* and is also found to be conserved in promoter regions of *hsp86* genes of other *Plasmodium* species [343]. In the study, the AlignAce algorithm was used to search upstream regions of *hsp* genes of several *Plasmodium* species. The authors could not find any heat shock elements which are commonly found in promoters of eukaryotic chaperone genes such as the CCAAT box and SP1 binding site in the promoter regions of all *Plasmodium hsp* genes.

Another study reported a few putative motifs enriched in gene clusters differentially regulated when treated with chloroquine [344]. By gel shift and reporter assays, another study identified several motifs that are involved in regulation of transcription of *var* genes [345]. A summary of the promoters characterized before year 2000 was reviewed and almost all promoters analyzed do not contain common eukaryotic motifs [346]. After year 2000, individual studies examining sequences in promoters reported of no common eukaryotic motifs as well [328, 341, 347-352]. Taken together, there are few motifs functionally verified and of the few that are verified, they are only applicable to a limited gene set. Hence, the picture of how *cis* regulatory motifs function to regulate gene transcription in *P. falciparum* is still unclear.

In the more recent years, algorithms have been developed to overcome difficulties posed by the AT-rich genome. Most of these algorithms work on the assumption that genes which are co-expressed are co-regulated, and therefore, have the same motif in its promoter regions. This approach has been used in the prediction of *cis*-regulatory motifs in yeast [353]. Each

algorithm differ from each other in some aspects to increase robustness and accuracy of the motifs identified. FIRE (Finding Informative Regulatory Elements) is an algorithm which utilizes unsupervised transcriptomic data to search for motifs in regions upstream of co-expressed genes [354]. On the other hand, GEMS (gene enrichment motif searching), searches for motifs not just according to transcriptomic data but also taking gene ontology into account to generate clusters of genes which have similar timing of transcript expression and similar gene function [335]. Another study utilized 3 different algorithms to search for motifs (MEME, Weeder and AlignAce) and reported motifs which are predicted by all 3 algorithms as significant [355]. In the most recent years, putative DNA motifs of each factor in the ApiAP2 transcription factor family was predicted by protein binding microarrays [356].

Using a separate approach, based on the assumption that important regulatory elements will be conserved in evolution, algorithms have also been developed to search for motifs in orthologous genes in the *Plasmodium* species. A study included the *P. y. yoelii* orthologs of the *P. falciparum* genes in the search of motifs in the upstream regions of genes [357]. Another algorithm, MDOS (motif discovery using orthologous sequences) uses the 5' and 3' flanking regions of orthologous genes in *P. falciparum*, *P. yoelii* and *P. knowlesi* for alignments [358]. This study yielded 38 and 11 motifs in the 5' and 3' flanking regions respectively. Authors also reported that multiple different motifs were found on *Plasmodium* promoters. A similar observation was made in a study where authors found the number of different motifs present on a single promoter is significantly higher than the number in yeast promoters [357]. Authors also proposed that this could be in line with the paucity of gene specific transcription factors found in *P. falciparum*, because the presence of multiple different motifs on a single promoter could mean that *P. falciparum* utilizes fewer

transcription factors by increasing the permutations of transcription factors on each promoter to achieve fine tuning of transcriptional regulation.

Following the discovery of the ApiAP2 transcription factor family, protein-binding microarray have been utilized to identify binding sequences for each of the 27 members of the family [318]. Authors have identified DNA sequences which bind to each AP2 domain in most of the members of the ApiAP2 family and have also verified several of the sequences by electrophoretic mobility shift assays (EMSA). In addition, they have also demonstrated that the average timing of transcript expression of the genes harboring the motif is similar to the timing of expression of the transcription factor which is predicted to bind to the motif. However, the real picture is more complex as the authors also demonstrated that several of the AP2 domains are able to bind to more than 1 motif and some of the ApiAP2 transcription factors contain more than 1 AP2 domain.

Motif sequence	Function of motif	Discovery method	Reference
Sequence motif found upstream of *kahrp* gene ACTGCATGTAGTGTAG T	Forms 3 complexes with rings nuclear extract. Forms 1 complex with schizont nuclear extract.	EMSA	[300] (1992)
5-bp sequence upstream of *PfGBP130* gene GTATT	Binds nuclear factors. Inclusion of 5 bp sequence increases reporter gene activity	Transient transfections, EMSA	[336] (1999)
24-bp sequence upstream of CDP-diacylglycerol synthase gene in *P. falciparum*	Forms sequence specific complex with crude nuclear extract. Unknown factor.	Transient transfections, EMSA	[333] (2002)
SPE2 [(T/G)GTGC(A/G)(N)4(T/G)GTGC(A/G)]	Binds PfSIP2 (ApiAP2 protein) in *P. falciparum* Silencing of *var* genes	EMSA and reporter assay	[324] (2003) [323] (2010)
G-box motif Upstream of *pfhsp86* promoter GCCCCGCGGAAAGGGGC	Conserved in hsp86 genes in *P. y. yoelii*, *P. vivax* and *P. berghei*	Transient transfections	[329] (2004)
MRE (T/G)AACNGN	Binds PfMyb1	EMSA	[312] (2004)
TATA box element (TATAA and TGTAA)	Binds to PfTBP and found on promoter regions of *kahrp* and *gbp-130*	EMSA and foot printing assays	[302] (2005)
102 bp enhancer region (PRE)	Binds transcription factor PREBP in *P. falciparum* Induces activity of PRE-containing promoter	Transient transfection reporter assay; EMSA; DNA footprinting	[313] (2008) [315] (2013)
TAGCTA	Activates ookinite genes in *P. berghei*; binds AP-O (ApiAP2 protein)	ChIP; validated by EMSA and reporter assay	[320] (2009)
TGCATGCA	Binds AP-Sp (ApiAP2 protein) in *P. berghei*; Present on upstream regions of sporozoite-specific genes	EMSA and reporter gene assay	[321] (2010)
(G/A)TG(T/G)(A/C)CA (C/T)	Binds PfAP2-G (ApiAP2 protein) in *P. falciparum* and *P. berghei* Induces transcription of early gametocyte genes	EMSA and reporter assay	[325] (2014)

Table 1.3 List of *cis*-regulatory elements in *P. falciparum*.

This is a list of DNA motifs which are discovered in *P. falciparum* and have been functionally verified. In this table, their functions and the methods which have been used to validate their functions are summarized. In addition, they have been ordered chronologically according to their year of discovery. However, this list does not include motifs which are predicted from *in silico* studies and motifs which have been discovered from protein-binding microarray studies of the ApiAP2 family of transcription factors. Strikingly, there are few reports of common eukaryotic motifs (other than the TATA box element) in *P. falciparum*.

1.4 Aim of thesis

The analysis of the *P. falciparum* transcriptome in 2003 has revealed that transcription in *P. falciparum* is highly regulated during the IDC [19]. Despite the recent uncovering of a large group of transcription factors in the *P. falciparum* genome, the ApiAP2 transcription factor family, there is still lack of direct evidence of how these transcription factors regulate transcription in the *P. falciparum* IDC [319]. Prior to this discovery, there have been only a few attempts to characterize transcription factors in the *P. falciparum* IDC [312, 317]. Recent advances in *in silico* methods have allowed the prediction of *cis*-regulatory elements in the *P. falciparum* AT-rich genome [335, 359]. However, these studies provided little or no functional studies on the motifs predicted. Hence, our knowledge of how *cis*-regulatory elements regulate transcription in the *P. falciparum* IDC is still very limited. It is of importance to understand how *P. falciparum* regulates its transcription in the IDC as the accurate developmental regulation of *P. falciparum* genes is required for parasite development, and hence the pathogenesis of the malaria disease.

Hence, this thesis aims to investigate the role of *cis*-regulatory elements in *P. falciparum* transcriptional regulation through the following 3 main objectives:

1.4.1 Characterization of the *P. falciparum multidrug resistant protein 2 (pfmrp2)* promoter

Previous study in our laboratory has generated isogenic clones of the *P. falciparum* 3D7 culture adapted strain [360]. In the study, a set of clones was reported to be drug resistant and another set was reported to be drug sensitive. In addition, a deletion in the 5' UTR of the *pfmrp2* gene in drug resistant clones was observed. Hence, we have conducted promoter deletion and transient transfection assays to characterize the promoter regions of the *pfmrp2* in both drug resistant and drug sensitive clones. In addition, we have established and optimized a method to characterize promoter activities which will be used in our motif screening studies in the following objective.

1.4.2 Prediction and validation of DNA motifs regulating IDC gene transcription

We have utilized a previously developed and published algorithm to predict DNA motifs on co-regulated genes in the *P. falciparum* IDC [359]. We have used the algorithm on an improved transcriptome of the IDC which was done using a microarray chip with higher representation of the *P. falciparum* genes [30]. A method for the cloning of highly AT-rich *P. falciparum* promoters and screening of motifs was developed in this study. By means of transient transfection assays, we have identified several motifs of high confidence for further studies in the next objective.

1.4.3 Further characterization on novel motifs 3, 6, 16, 17 and 20

We have identified motifs 3, 6, 16, 17 and 20 as highly confident motifs and characterized them in more detailed studies in this section. Mutagenesis assays have been done to mutate the motifs in promoters to understand the importance of each base in the motif. Additional promoters which contain each motif have been selected to investigate whether these motifs

regulate transcription on multiple promoters. Stable transfections were then done to characterize the stage specific activity of each motif and in other words, the role of the motif in regulating stage specific transcription. Finally, EMSAs were conducted to investigate the protein-binding capabilities of the motifs of interest.

2. Materials and methods

2.1 Parasite culture

Parasite culture was done according to an established protocol for continuous culturing of *P. falciparum* [361]. *In vitro P. falciparum* parasites were grown with human red blood cells in RPMI 1640 media (Gibco, USA) supplemented with 10 mg/l gentamycin (Gibco, USA), 2 g/l sodium bicarbonate (Sigma, Germany), 0.25% Albumax II (Gibco, New Zealand) and 0.1 mM hypoxanthine (Sigma, Germany). Parasite culture was kept in sterile 175 cm^2 flask (Nunc, Denmark), at 2% hematocrit and a maximum of 25% parasitemia. Cultures were incubated at 37^oC and gassed with mixture gas of 5% carbon dioxide, 5% oxygen and 90% nitrogen gas. Parasites were synchronized for ring stage parasites with 5% D-Sorbitol (Sigma, USA) at approximately 0 hpi and 12 hpi. Parasitemias were obtained by cell counting of blood smears stained with Giemsa (Sigma, Germany).

2.2 Transient transfections

200 µl of packed 20-25% parasitemia ring stage (8-12 hpi) infected RBC were electroporated with 50 µg of firefly luciferase construct and 50 µg of renilla luciferase construct, together with a transfection buffer, cytomix (120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4/KH_2PO_4, 25 mM Hepes, pH 7.6). Electroporations were done using Gene Pulser Xcell (Bio-rad) with settings, 310 V, 950 µF and 0 Ω in 2 mm cuvettes (Bio-rad, China). 50 ml of media with 200 µl of uninfected RBC were pre-prepared before the transfection and incubated at 37^oC. Electroporated parasites were placed in the media-uninfected RBC mixture immediately after transfection. Media was changed 1 hour and 8 hours post transfection and subsequently every 24 hours. Bioluminescence was measured at the next ring stage (8-12 hpi), 48 hours post transfection.

2.3 Stable transfections

Electroporations were done in the same way as transient transfections. 50 µg of pPLN-FFLuc with promoter of interest was transfected with 50 µg of pINT. Selection media containing 125 µg/ml of G-418 (Sigma, Germany), 2.5 µg/ml of blasticidin (Invitrogen, USA) and 2.5 nM of WR99210 was added 8 hours after invasion. Transfected parasites were kept under selection media for 4-5 weeks before parasites become resistant against all drugs. Parasites were then placed on media without selection drug for 3 weeks and placed on media with 2.5 µg/ml of blasticidin and 2.5 nM of WR99210 again for 1 week. This cycle was repeated twice. gDNA was extracted from schizont stage transfectants and tested by PCR on the 5' and 3' end of the integration site.

2.4 Parasite cloning by limiting dilution

Highly synchronized cultures of trophozoite stage integrants were placed on a gently shaking shaker and incubated at 37^{o}C for 12-16 hours. Shaking was to ensure that single infections occur in the next invasion. At 6-8 hpi in the next invasion, giemsa-stained smears were made and parasitemia was calculated by observing under the light microscope. It was ensured that more than 95% of the infected RBCs were singly infected.

Based on the assumption that there are 10^7 RBCs in 1 µl of RBC, the parasitized culture was diluted such that there will be 1 parasite in every 200 µl of 3% hematocrit. 100 µl of each diluted parasite culture was placed onto wells in a 96-well plate (Iwaki). By calculations, there will be 1 parasite in every 2 wells, thus 15 wells were filled for each integrant to obtain at least 5 clones. Media was changed every 2 days and placed in a gassed bag. Giemsa-

stained smears were made for each well from day 8 onwards. Parasites were in observable quantity by light microscopy on day 12.

Half or less than half of the wells contained parasites, which was indicative of successful cloning based on the calculations made when doing the dilutions. 5-10 out of 15 wells were positive of parasites after day 12. gDNA were extracted from positive clones and were tested for integration by PCR.

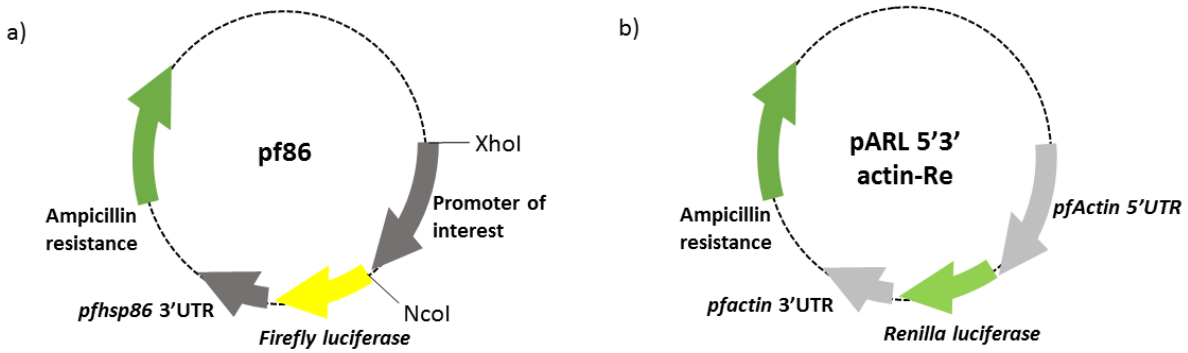
2.5 Constructing plasmids

2.5.1 PCR amplifications

PCR was carried out using PfuUltra II Fusion DNA Polymerase (Stratagene) and thermal cycling was done using Mastercycler® Pro (Eppendorf). PCR reactions were done in 50 µl volumes containing 1x PCR reaction buffer (supplied by manufacturer), 30 ng of 3D7 genomic DNA or plasmid DNA, 0.2 µM of forward and reverse primer each, 250 µM of each dNTPs, 60 mM of tetramethylammonium chloride (TMAC) and 1 µl of DNA polymerase enzyme. Cycling conditions were as follows, initial denaturation at 94^{o}C for 5 min, followed by 30 cycles of denaturation at 94^{o}C for 30 sec, primer annealing at 50^{o}C for 1 min, strand extension at 60^{o}C for 1 min/kb; after which, final extension at 60^{o}C for 5 min. Sizes of PCR products were verified by agarose gel electrophoresis. MinElute PCR purification kit (QIAGEN) was used to purify PCR products according to manufacturer's protocol. Purified PCR products were eluted in a volume of 20 µl of distilled water (dH_20). Purified PCR products were quantified by UV spectrometry.

2.5.2 Restriction digestion

The digestion reaction for PCR products contained 1 µg of PCR product, 1x digestion buffer (provided by New England Biolabs (NEB)) and 1 unit of respective restriction enzyme (NEB) in a 30 µl reaction. To obtain the plasmid backbone, pf86 and pPLN-ffluc plasmids were digested in the following reaction: 10 µg of plasmid, 1x digestion buffer (NEB) and 1 unit of appropriate restriction enzyme (NEB) (Figure 2.1). All digestion reactions were done by incubation at 37^{o}C for 16 hours. Plasmid backbones were resolved in a 1% agarose gel and then gel extracted using QIAquick gel extraction kit (QIAGEN) according to the manufacturer's protocol. Purified plasmid backbones were then eluted with 50 µl of dH_20. The digested PCR products were purified by MinElute PCR purification kit (QIAGEN) according to manufacturer's protocol. Purified PCR products were eluted in a volume of 20 µl of dH_20. Purified PCR products and plasmid backbones were quantified by UV spectrometry.



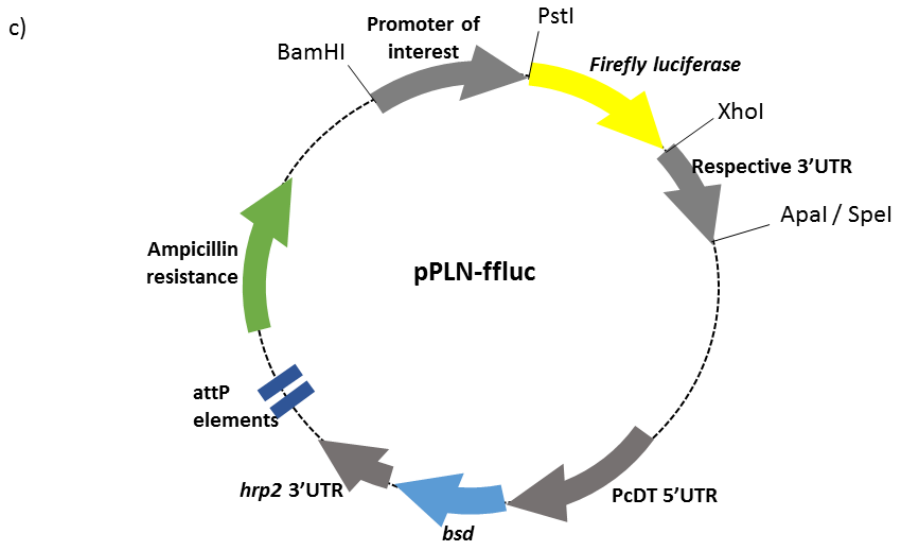


Figure 2.1 Plasmids that have been used in the studies in this thesis.
a) pf86. This plasmid has been used in transient transfection experiments and does not contain a *P. falciparum* drug selectable marker. This plasmid contains a firefly luciferase gene which is terminated by the 3'UTR sequences of the *pfhsp86* gene. The promoter of interest is cloned in between restriction enzyme sites, XhoI and NcoI. b) pARL-5'3'-Re. This plasmid contains a *renilla luciferase* gene which is driven and terminated by 5' and 3' UTR sequences of *pfactin* respectively. This plasmid is co-transfected with pf86 in transient transfection experiments as a normalizing control. c) pPLN-ffluc. This plasmid is used for stable transfections/integration experiments. It contains a *P. falciparum* drug selectable marker *blasticidin-S-deaminase (bsd*). It also contains the *attP* sites which are used for integration into the *cg6* locus in the Dd2 *attB* strain *P. falciparum* parasites. This plasmid also carries the *firefly luciferase* gene and the promoter of interest is cloned in between BamHI and PstI restriction sites. The *firefly luciferase* gene will be flanked at the 3' end by the 3' UTR of the respective promoter of interest using restriction sites XhoI and ApaI or SpeI.

2.5.3 Phosphorylation

For the construction of mutant promoters with the motif deleted in Figure 4.8, phosphorylation of 5' hydroxyl groups of the PCR amplified fragments was required for the ligation of fragments. Phosphorylation was carried out using T4 Polynucleotide Kinase (Fermentas). 20 pmol of PCR fragment was phosphorylated in a 20 µl reaction mixture of 50 mM Tris-HCl (pH 7.6 at 25°C), 10 mM $MgCl_2$, 5 mM DTT, 100 µM spermidine, 1 mM ATP

and 10 units T4 Polynucleotide Kinase. Reaction mixture was incubated at 37^{o}C for 20 min and inactivated by heating at 75 ^{o}C for 10 min.

2.5.4 Ligation

In all ligations, 30 ng of plasmid backbone was used. A 1:3 molar ratio of plasmid backbone to insert was used. Ligations were carried out using LigaFast™ Rapid DNA Ligation system (Promega) in the following reaction mixture: 30 ng of plasmid backbone, 1:3 molar ratio of insert DNA, 30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 5% PEG and 3 units of ligase enzyme in a total reaction volume of 10 µl. The reaction mixture was then incubated at 23 ^{o}C for 5 min (1 fragment ligation) or 15 min (2 fragment ligation).

2.5.5 Bacteria transformation

10 µl of ligation reaction mixture was added to 100 µl of DHα strain *E. coli* chemically competent cells (Agilent Technologies). The competent cells together with the ligated products were then left on ice for 20 min, after which heat shock at 42^{o}C was carried out for 1 min and left on ice again for 5 min. 1 mL of Luria-Bertani (LB) broth with no ampicillin was added to the competent cells and shaken at 220 rpm, 37^{o}C, for 45 min. Bacteria cells were spun down at 6000xg for 5 min and were spread on LB agar plate with 100 µg/mL of ampicillin. LB agar plates were then incubated for 12-16 hours at 37^{o}C.

2.5.6 Verification of plasmids

Colonies from the LB agar plates which were incubated for 12-16 hours, were picked and inoculated into 3 ml of LB broth supplemented with 100 µg/ml of ampicillin. The inoculated bacteria were then incubated for 12-16 hours at 37^{o}C, shaking at 220 rpm. Plasmids were then extracted from bacteria using QIAprep Miniprep Plasmid Purification Kit (QIAGEN)

according to manufacturer's protocol. Plasmids were finally eluted in 40 µl of dH_20. Plasmids were then restriction digested with appropriate restriction enzymes. Fragments were resolved on 1% agarose gels to verify the sizes of the fragments. Clones which were tested positive sent for capillary sequencing (Axil Scientific).

2.5.7 Sequence verification

Primers were designed such that the entire sequence that was cloned into the vector backbone is covered in the capillary sequencing. Sequenced reads were aligned with published 3D7 reference genome sequence (PlasmoDB) using Bioedit programme. 12 clones of the same construct were sequenced and deviations in sequence were recorded for each clone. In transient and stable transfections for comparisons between wild type promoter and mutant promoter where motif was deleted, only clones with exact same sequence as the reference genome were used in the transfection assays.

2.6 Dual luciferase assays

Assays were done with Dual-Luciferase® Reporter Assay System (Promega, USA) 48 hours post transfection on ring stage parasites (8-12 hpi). Infected RBC were lysed using 1:10 volume ratio of 0.1% saponin (Sigma) dissolved in phosphate buffered saline (PBS) incubated for 5 min at room temperature. The lysate was washed with PBS twice. After which, Passive Lysis Buffer (Promega, USA) was used to lyse the parasites. Subsequently, the cell debris was spun down and the 20 µl of the supernatant containing luciferase protein were added to firefly luciferase substrate and reading was taken. A mixture which contains renilla luciferase substrate and an agent to stop the firefly luciferase activity was added and a second measurement was taken for the renilla luciferase activity.

2.7 Nucleic acids isolation

To isolate RNA, parasites were harvested individually by centrifugation at 2500 rpm for 5 min and washed with PBS twice. Pelleted infected RBC were quickly frozen and stored at -80^{o}C until further treatment. To 200 µl of infected RBC, 2 ml of TRIzol (Invitrogen, USA) was added to lyse RBC and parasites, subsequently 400 µl of chloroform (Fisher Scientific, USA) was also added. Upon centrifugation, the aqueous phase was separated from the organic phase. Ice cold isopronanol (Fisher Scientific, USA) was added to the aqueous supernatant and kept at -20^{o}C overnight to precipitate the RNA. Upon centrifugation, a pellet containing the RNA emerged and was washed with 70% ethanol (Merck, USA). After air-dying at room temperature, the pellet was resuspended in RNase-free water and stored at -80^{o}C.

To isolate DNA, 500 µl of infected RBCs were pelleted and 5 ml of 0.1% saponin (dissolved in 1x PBS) was added to lyse the RBCs. It was then incubated for 5 min at room temperature. Parasites were harvested by centrifugation at 4500 rpm for 5 min. Parasite pellet was washed in 1x PBS twice. Parasites were lysed using lysis buffer containing 40 mM Tris-HCl, 80 mM EDTA and 2% Sodium dodecyl sulfate (SDS). After through mixing, parasites and lysis buffer was incubated at 37°C for 2 hours. Equal volume of phenol was then added and aqueous phase was separated from the organic phase by centrifugation at 4500 rpm for 10 min. Chloroform extraction was preformed twice by adding equal volume of chloroform, centrifuging at 4500 rpm and separation of aqueous phase. gDNA was finally precipitated with 2.5x volume of absolute ethanol. Precipitated gDNA was obtained by centrifugation at 12,000 rpm for 20 min. gDNA was washed twice with 70% ethanol.

2.8 cDNA synthesis

DnaseI treatment of RNA samples were first carried out in a 10 µl reaction mixture containing 5 µg of total RNA, 1 unit of Rnase-free DnaseI (Fermentas), 10 mM Tris-HCl, 2.5 mM $MgCl_2$ and 0.1 mM $CaCl_2$. DnaseI treatment was carried out by incubating at 37°C for 30 min and DnaseI heat inactivation for 65°C for 10 min with the addition of 5 mM of Ethylenediaminetetraacetic acid (EDTA). 2 µl of the DnaseI treatment reaction mixture was used directly for reverse transcription. The 20 µl reaction mixture for reverse transcription consists of 2 µl of DnaseI-treated RNA, 1 pmol/µl of gene specific primers, dNTP mix (dATP, dTTP, dGTP and dCTP each at 1mM concentration) (Fermentas), 200 units of RevertAid H Minus MMuLV Reverse transcriptase (Fermentas), 50 mM Tris-HCl, 50 mM KCl, 4 mM $MgCl_2$ and 10 mM DTT. Sequence of primers for reverse transcription of reference gene transcript (*pfl0900c – P. falciparum arginyl-tRNA synthase*) and *firefly luciferase* transcript are provided in Table 2.1. The reaction mixture was then incubated at 42°C for 1 hour to carry out reverse transcription. Heat inactivation of reverse transcriptase was then carried out by incubation at 70°C for 10 min. After heat inactivation, 1 µl of the reverse transcription reaction mixture was used directly for real time quantitative PCR (qPCR).

2.9 Real Time Quantitative PCR

Real Time-qPCR was done using the Lightcycler® SYBR Green I PCR Master Mix (Roche) with the LightCycler® 480 System. Forward and reverse primers were designed to amplify 117 bp and 251 bp of the reference gene (*pfl0900c*) and *firefly luciferase* respectively. The qPCR reaction mixture consists of 1 µl of the reverse transcription reaction mixture, 1 µM of each forward and reverse primer and 2x master mix (which contains Taq DNA polymerase, reaction buffer and dNTP mix). The cycling conditions are as follows: 1 cycle of pre-

incubation at 95°C for 5 min, 45 cycles of amplification at 95°C for 10 sec, 55°C for 10 sec and 70°C for 10 sec. This was followed by 1 cycle of 95°C for 5 sec and 65°C for 1 min. Acquisition of the melting curve occurred at 97°C. After which, the threshold cycle (Ct) was obtained for each primer pair and each cDNA sample from different time points.

The relative quantification of *firefly luciferase* (*ffluc*) transcripts driven by the wild type construct and mutant construct in the transient transfection assays in Chapter 4 and 5 are as follows. Ct values for the *ffluc* transcript and the reference gene transcripts from cells transfected with constructs with the *ffluc* gene driven by the wild type promoter and constructs with the *ffluc* gene driven by the mutant promoter. In both samples, the *ffluc* transcripts were normalized by subtracting the Ct value of the reference gene transcripts from the Ct value of *ffluc* transcripts. Next, the quantity of *ffluc* transcripts driven by the mutant promoter was expressed as a fold change relative to the quantity of *ffluc* transcripts driven by the wild type promoter by the ΔΔCt method. ΔΔCt is calculated by subtracting normalized *ffluc* Ct of the wild type promoter sample from normalized *ffluc* Ct of the mutant promoter sample. Fold change was then calculated by $2^{-\Delta\Delta Ct}$. Plasmid copy numbers were also calculated in a similar way where 10 ng of gDNA was used as a template for qPCR instead of cDNA. The fold change calculated by for quantity of *ffluc* transcripts were then normalized by copy number differences. For clarity, fold change was then presented on a log scale in Figure 4.10. Hence, if there was no difference in quantity of *ffluc* transcripts between samples obtained from wild type and mutant promoters, log (fold change) is 0. If there is a greater amount of *ffluc* transcripts in samples from mutant promoters compared to samples from wild type promoters, log (fold change) is greater than 0. If there is a smaller amount of *ffluc* transcripts in samples from mutant promoters compared to samples from wild type promoters, log (fold change) is less than 0.

The relative quantification of *ffluc* transcripts in the different stages (rings, trophozoites and schizonts) for the stable transfectants was calculated in Chapter 5 as follows. There will be 6 samples for each motif to be tested. Namely, ring stage (wild type), trophozoite stage (wild type), schizont stage (wild type), ring stage (Del), trophozoite stage (Del) and schizont stage (Del). In each sample, the *ffluc* transcripts were normalized by subtracting the Ct value of the reference gene transcripts from the Ct value of *ffluc* transcripts. The *ffluc* transcripts in the Del integrant were then expressed as a fold change relative to the *ffluc* transcripts in the WT integrant of its respective stage (ring, trophozoite or schizont) by the ΔΔCt method. For example, ΔΔCt during ring stage is calculated by subtracting normalized *ffluc* Ct of the WT integrant during ring stage from normalized *ffluc* Ct of the Del integrant during ring stage. The same was done for the other 2 stages for each set of integrants. Following which, for each stage, fold change of Del integrant sample relative to WT integrant sample was calculated by $2^{-\Delta\Delta Ct}$. *ffluc* copy numbers were also calculated in a similar way where 10 ng of gDNA was used as a template for qPCR instead of cDNA. The fold change calculated for the quantity of *ffluc* transcripts were then normalized by copy number differences.

Primer name	Oligo sequence	Purpose
PFL0900c_RT	CCAGAACAGTCGATAACATAGG	Reverse transcription of *pfl0900c* transcript
Ffluc_RT	TTTACATAACCGGACATAATCATAG	Reverse transcription of *firefly luciferase* transcript
PFL0900c_qPCR_F	AAGAGATGCATGTTGGTCATTT	Forward primer for qPCR of *pfl0900c* cDNA
PFL0900c_qPCR_R	GAGTACCCCAATCACCTACA	Reverse primer for qPCR of *pfl0900c* cDNA
Ffluc_qPCR_F	CTGTTTCTGAGGAGCCTTCAGG	Forward primer for qPCR of *firefly luciferase* cDNA
Ffluc_qPCR_R	GGGTGTAATCAGAATAGCTGATGTAG	Reverse primer for qPCR of *firefly luciferase* cDNA

Table 2.1 Primers used for reverse transcription and real time qPCR experiments. "PFL0900c_RT" and "ffluc_RT" were used for gene-specific reverse transcription from RNA to cDNA. The rest of the primers were used in real time qPCR experiments to quantify *ffluc* transcripts in individual samples.

2.10 Electrophoretic mobility shift assays

2.10.1 Nuclear protein extraction

0.5 ml of parasitized RBCs containing 30% schizonts were lysed by incubating in 5 ml 0.1% saponin for 5 min to release parasites from RBC. Parasites were washed twice in 1x PBS and the parasite pellet was then incubated with 300 µl of cytoplasmic lysis buffer which contains 20 mM Hepes pH 7.9, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.65% Nonidet P-40 (NP-40), 1 mM dithiothreitol (DTT) and 1x protease inhibitor (Roche), for 10 min on ice. Nuclear fraction was then obtained by centrifugation at 5000 rpm for 5 min. The supernatant was discarded and the pellet was washed 3 times in cell lysis buffer.

A study reported that treatment with DnaseI and Micrococcal nuclease was necessary for release of certain nuclear proteins tightly bound to DNA and chromatin associated nuclear factors [61]. Hence, the nuclear pellet was treated with DnaseI and Micrococcal nuclease to extract as much nuclear proteins as possible. The resultant nuclear pellet obtained after lysis with cell lysis buffer was resuspended in 200 µl of low salt buffer (20 mM Tris HCl pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 15 mM NaCl, 60 mM KCl, 5 mM $MnCl_2$, 300 mM sucrose, 0.4% NP-40, 1 mM DTT, 1x protease inhibitor, 300 units of Micrococcal nuclease (Fermentas) and 20 units of DnaseI (Fermentas)). The resuspended nuclear pellet was incubated at 37°C for 20 min and was centrifuged at 13,000 rpm for 10 min. The supernatant was collected as the low salt fraction and the pellet was washed twice in low salt buffer.

The resultant pellet was resuspended in 200 µl of high salt buffer (20 mM Hepes pH 7.9, 1 M KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT and 1x protease inhibitor) and was vortexed for 30 min at 4°C. This was followed by centrifugation at 13,000 rpm for 10 min. The supernatant was collected as the high salt fraction and the pellet was washed twice in high salt buffer. The low salt and high salt nuclear fractions were then combined. The combined nuclear extract were concentrated using 3K molecular weight cutoff column protein concentrators (Thermo Scientific) to a 100 µl volume.

2.10.2 Binding reactions

Binding and gel shift assays were carried out using the LightShift Chemiluminescent EMSA kit (Thermo Scientific). 50 bp oligonucleotides (oligos) were designed with the motif of interest in the middle of the oligo with approximately 20 bp sequence of endogenous promoter flanking both sides of the motif. 50 bp oligos with the motif omitted from the sequence were also designed. Oligos were in 2 forms, one with the biotin moiety attached at

the 5' end and the other without biotin. The oligos designed are summarized in Table 2.2. Binding reactions were carried out in 10 mM Tris, 50 mM KCl, 1 mM DTT, 2.5% glycerol, 10 mM $MgCl_2$, 50 ng/µl poly(dI.dC) DNA, 0.05% NP-40 and 10 µg of concentrated nuclear extracted, in a 20 µl reaction. Oligos with biotin attached are provided in 20 fmol concentration and competitor DNA (if any) was in 4 pmol concentration (200 times in excess). The single stranded oligos were first annealed by heating equimolar concentrations of forward and reverse oligos at 95^{o}C for 5 min, after which the heat block is switched off and allowed to cool slowly overnight to room temperature. Appropriate annealed oligos were added to the binding reaction and incubated at room temperature for 20 min.

A native 6% polyacrylamide gel was prepared in 0.5x TBE buffer. The gel was pre-run at 100 V for 1 hour in a vertical electrophoresis unit (Bio-rad) using 0.5x TBE as the running buffer. After binding reactions were completed, 5 µl of loading buffer was added and sample was loaded into the pre-ran polyacrylamide gel. Electrophoresis was carried out until the bromophenol blue dye front reached ¾ of the gel. A nylon positively charged membrane (GE Healthcare Life Sciences) was soaked in 0.5x TBE. DNA from the TBE polyacrylamide gel was transferred onto the nylon membrane using an electrophoretic transfer unit (Bio-rad) at 100 V for 30 min. DNA that was transferred onto membrane was cross-linked by using UV light. DNA was then visualized on the membrane by using chemiluminescence detection and exposure on X-ray film.

Oligo name	Oligo sequence
Motif3_F	TAGATAATAGTAAATATATTTCTATATTATATATATCTTTACATTTT
Motif3_R	AAAATGTAAAGATATATATAATATAGAAATATATTTACTATTATCTA
Motif3del_F	TAGATAATAGTAAATATATTTATATATATCTTTACATTTT
Motif3del_R	AAAATGTAAAGATATATATAAATATATTTACTATTATCTA
Motif6_F	TGCAATATAATTTTATACTTATATAGGAAGAAATTAAAATGTAACGGT
Motif6_R	ACCGTTACATTTTAATTTCTTCCTATATAAGTATAAAATTATATTGCA
Motif6del_F	TGCAATATAATTTTATACTAGAAATTAAAATGTAACGGT
Motif6del_R	ACCGTTACATTTTAATTTCTAGTATAAAATTATATTGCA
Motif16_F	ATAGAATAATTTATTGTACGTGCATGCTATTTTTTATTTTTATTTTT
Motif16_R	AAAAATAAAAATAAAAAATAGCATGCACGTACAATAAATTATTCTAT
Motif16del_F	ATAGAATAATTTATTGTACGTATTTTTTATTTTTATTTTT
Motif16del_R	AAAAATAAAAATAAAAAATACGTACAATAAATTATTCTAT
Motif20_F	TTTGTAAAACTTTAATAAATAAAACATGTATAAAGTGTACAATCTGTCA
Motif20_R	TGACAGATTGTACACTTTATACATGTTTTATTTATTAAAGTTTTACAAA
Motif20del_F	TTTGTAAAACTTTAATAAATATAAAGTGTACAATCTGTCA
Motif20del_R	TGACAGATTGTACACTTTATATTTATTAAAGTTTTACAAA

Table 2.2 Oligos used for EMSAs.

The above oligos were designed and synthesize to determine protein binding activity of the motifs 3, 6, 16, 17 and 20. Approximately 25 bp sequences flanking the motif were obtained from the endogenous promoter sequence. Oligos with and without the motif were designed. Unlabeled and 5' biotin labeled versions of each oligo were synthesized.

3. Characterization of the *P. falciparum multidrug resistance-associated protein 2 (pfmrp2)* promoter

3.1 Background and rationale

pfmrp2 is one of the 16 ATP-binding cassette (ABC) transporters encoded in the *P. falciparum* genome. ABC transporters are transmembrane proteins which actively pump out structurally diverse groups of drugs which have been shown to lead to drug resistance in multiple organisms [362, 363]. In *P. falciparum*, several of the ABC transporters are associated with drug resistance and they will be described in this paragraph. SNPs and copy number increases in *P. falciparum* multidrug resistant protein 1 (PfMDR1) has been associated with increased quinolone-based antimalarial drug resistance *in vivo* [364]. The MRP family is a subclass of the ABC transporter family and the *P. falciparum* genome encodes for 2 *mrp* genes, *pfmrp1* and *pfmrp2*. It has been suggested that that MRPs play a role in redox metabolism specifically in transport of glutathione and glutathione conjugates across membranes [365]. *pfmrp1* has been implicated in drug resistance as the knock out of the gene has resulted in increased sensitivity to multiple anti-malarial compounds [366]. In addition, 2 different mutations of *pfmrp1* have been associated with parasites causing recurrent infections and increase in resistance to anti-malarials in Africa and the Thai-Myanmar border [367-369]. *pfmrp2* is the less studied homologue of the 2 MRP transporters in the genome. Although SNPs in the *pfmrp2* were reported in culture adapted laboratory *P. falciparum* strains, they were not associated with drug resistance [370].

By means of limiting dilution, several isogenic clones from the 3D7 parasite culture were isolated [360]. Work in our laboratory was focused on 2 clones which have different drug sensitivities to chloroquine and mefloquine. It has been shown that the mode of action of both

chloroquine and mefloquine are associated with the inhibition of clearance of toxic free heme which is a by-product of the haemoglobin digestion [371]. Clone 11C is more drug sensitive and clone 6A is more drug resistant, relative between the 2. Clone 11C and 6A has IC50 for mefloquine at 15.4 nM and 37.5 nM respectively [360]. In addition, 11C and 6A has IC50 for chloroquine at 31.4 nM and 58.6 nM respectively. Comparative genomic hybridization (CGH) has also been done to characterize the genome wide differences between the 2 clones. In addition, gene expression analysis by microarray has revealed that 23 genes are differentially expressed between 2 clones. *pfmrp2* stands out among the 23 differentially regulated genes as many of which were previously found to be highly transcriptionally variable [372] and *pfmrp2* is the only gene that has a link with drug resistance. As MRPs contribute to drug resistance by facilitating the efflux of drugs, differential expression of PfMRP2 may allow the transporter to be expressed at a parasite stage where efflux of chloroquine and mefloquine can be more efficient and thus more effective clearance of the drug from the parasite [366, 373]. CGH analysis has also revealed a deletion in 5' upstream region of *pfmrp2*. Upon further investigation by PCR and sequencing analysis, a 4.1 kb deletion of the upstream region was observed in the drug resistant clone, 6A (Figure 3.1b). The experimental work described so far has been performed by Sachel Mok.

The differential expression of *pfmrp2* due to the deletion in the promoter region may be the cause of the increased resistance of clone 6A to anti-malarial drugs. In addition, this *in vivo* phenomenon of deletion of a promoter region causing a change in gene expression profile is a good platform to characterize *cis*-regulatory motifs which define the stage specificity of *P. falciparum* genes in the IDC. Hence, it is of interest to further characterize the *pfmrp2* promoter to identify regions in the promoter which are important in transcription regulation and how it led to gene expression changes and ultimately, drug resistance.

3.2 4.1kb deletion in the *pfmrp2* promoter led to changes in gene expression profile

Gene expression analysis by microarray was done on clones 11C and 6A in order to characterize differentially expressed genes between the 2 clones which may have led to drug resistance to anti-malarial drugs. In our study, RNA was obtained from highly synchronized cultures of clones 6A and 11C at 8 hour intervals in the 48 hour IDC. Gene expression profiles of 5500 genes in both clones were analyzed by microarray. The *pfmrp2* expression profile of drug sensitive clone 11C peaks during the ring stage and declines as the parasite develops into trophozoite and schizont stage (Figure 3.1a, red line). This expression profile concurs with the *pfmrp2* expression profile previously characterized in an earlier microarray study [19]. However, the *pfmrp2* expression profile changes dramatically in the drug resistant clone 6A. In clone 6A, *pfmrp2* transcript is repressed during the ring stage and peaks towards the trophozoite/schizont stage (Figure 3.1a, blue line). Interestingly, 6A was more resistant to chloroquine and mefloquine than 11C in the trophozoite/schizont stage [360], this may be due to the increased expression of *pfmrp2* during the trophozoite/schizont stage in 6A.

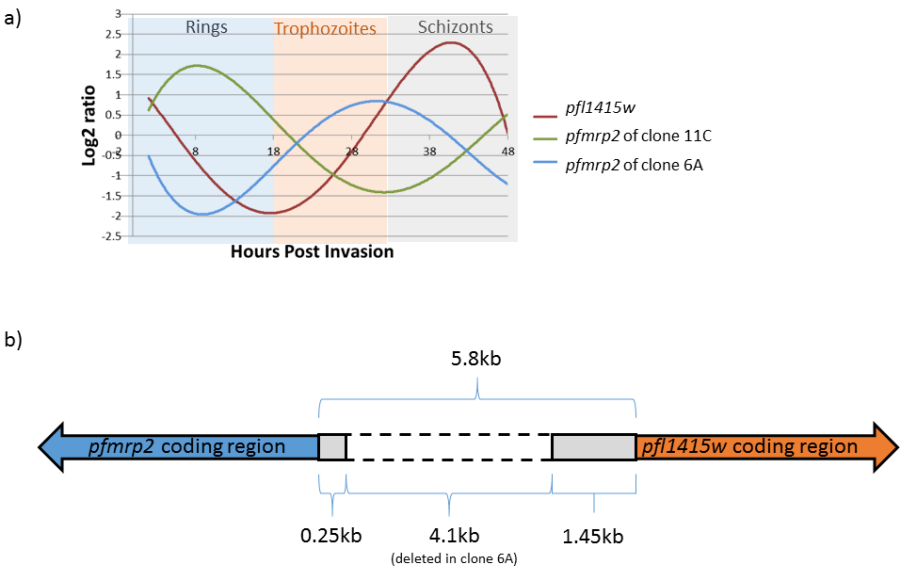


Figure 3.1 Changes in gene expression pattern associated with structural polymorphisms in 5' upstream regions.

a) Endogenous gene expression patterns of 11C *pfmrp2* and 6A *pfmrp2*. Total mRNA was isolated from clones 6A and 11C in 8 hour intervals throughout the 48 hour IDC. cDNA was synthesized and gene expression patterns were analyzed using microarray. In 11C, the *pfmrp2* gene is most highly abundant in the ring stage and in 6A, during the trophozoite/schizont stage. b) Expression profile of *pfl1415w*. This figure was adapted from [374]. Abundance of *pfl1415w* peaks during the schizont stage for both clones 11C and 6A. c) Structural polymorphisms in the *pfmrp2* 5' upstream regions. The intergenic region between the 2 genes in clone 11C is the same as the published sequences. However, a 4.1 kb region was deleted in clone 6A, the resultant intergenic space between *pfmrp2* and *pfl1415w* is now 1.7 kb. Blue: *pfmrp2* gene body; Orange: *pfl1415w* gene body; Grey and dotted region: intergenic space between the 2 genes in clone 11C; Dotted region: intergenic sequences deleted in clone 6A; Grey only: intergenic space between *pfmrp2* and *pfl1415w* in 6A after 4.1 kb deletion.

It was also subsequently found that a 4.1 kb region was deleted in the upstream regions of the 6A *pfmrp2* gene and the exact location of the deletion had been mapped (Figure 3.1c). Interestingly, the junctions at both ends where the deletion had occurred contained homopolymeric poly(dA:dT) tracts which may have facilitated the deletion of sequences in between. The intergenic space between *pfmrp2* and its upstream gene, *pfl1415w*, is 5.8 kb in

the wild type clone 11C. However, after deletion of the 4.1 kb intergenic space, approximately 1.7 kb is left in between *pfmrp2* and *pfl1415w* in clone 6A. In addition, *pfmrp2* and *pfl1415w* are arranged in a head-to-head gene arrangement. The gene expression profile of *pfl1415w* was that of a schizont specific gene, with peak expression during the late schizont stages and minimal expression during the ring stage (Figure 3.1b, pink line). The expression profile of *pfl1415w* is similar in both clones 6A and 11C (Figure 3.1b, pink and blue line). Remarkably, the *pfmrp2* expression profile of clone 6A had switched to schizont stage specific after a 4.1 kb intergenic space deletion (Figure 3.1a, blue line). With only 1.7 kb intergenic space between the 2 genes, it is possible that schizont stage-specific *cis*-regulatory elements in close proximity to *pfl1415w* is brought to closer proximity to *pfmrp2*, leading to the change in stage specificity of *pfmrp2* (Figure 3.1c).

Further, we attempted to test this putative schizont specific *cis*-regulatory element by a reporter gene assay. We hypothesized that the 6A *pfmrp2* putative promoter would drive reporter activity maximally in the schizont stage and ring stage for 11C *pfmrp2* putative promoter. First, the entire 1.7 kb upstream region of *pfmrp2* in clone 6A that is remaining after the 4.1 kb deletion was cloned into a construct driving *firefly luciferase (ffluc)* reporter gene (Figure 3.2a,b). For fair comparison, 1.7 kb upstream region of *pfmrp2* in clone 11C was also cloned to drive the *ffluc* gene in a separate construct. As a normalizing control, a construct containing *renilla luciferase* driven by a putative promoter that contained sequences 2 kb upstream of *Pfactin* gene and also 1 kb of the *Pfactin* gene's 3' terminating sequence was made (Figure 3.2c). Constructs containing either 6A or 11C *pfmrp2* putative promoters driving *ffluc* were each co-transfected with the pARL-5'3'Actin-Re construct in ring stage parasites. Transfected parasites were placed back in culture and were assayed for ffluc and

renilla luciferase activity in the subsequent cycle at ring (8-12 hours post invasion (hpi)), trophozoite (20-24hpi) and schizont (32-36 hpi).

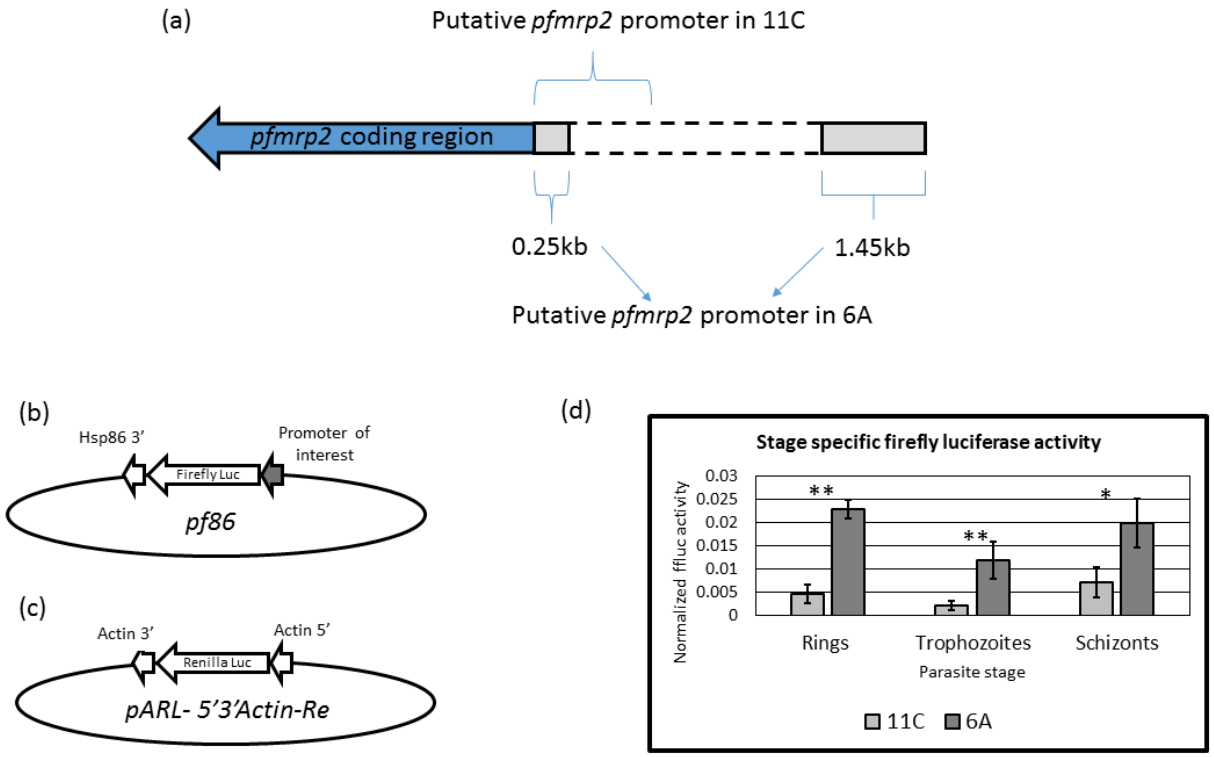


Figure 3.2 Stage specific promoter activities.

a) Schematic representation of 6A and 11C *pfmrp2* putative promoter regions. Due to a 4.1 kb deletion in the intergenic region between *pfmrp2* and upstream gene, *pfl1415w*, an intergenic space of 1.7 kb was left in between the 2 genes in clone 6A. Hence, the entire 1.7 kb was considered as a putative promoter for *pfmrp2* in clone 6A. For a fair comparison of promoters with the same lengths, 1.7 kb upstream of *pfmrp2* gene in clone 11C (which does not exhibit the deletion) was considered to be the putative promoter. Both putative promoters were separately cloned on to a construct (pf86) driving *firefly luciferase (ffluc)* gene. b) pf86 and pARL-5'3'Actin-Re. pf86 is a construct containing the *firefly luciferase* gene with a cloning site upstream of the gene for the cloning of promoter of interest. It contains the *pfhsp86* 3' UTR as terminating sequences immediately downstream of the *ffluc* gene. pARL-5'3'Actin-Re is used as a normalizing control in all transient transfection expreiments. The construct contains a *Renilla luciferase* gene driven by 2 kb 5' UTR upstream of *PfActin* gene and 1 kb of 3' UTR PfActin as a terminating sequence. c) Constructs containing *ffluc* and 11C and 6A putative *pfmrp2* promoters were separately co-transfected with pARL-5'3'Actin-Re in ring stage parasites (8-12 hpi). Parasites were allowed to develop as per usual cell culture conditions and were harvested

at 3 time points in the next parasite cycle, (8-12 hpi), trophozoite (20-24 hpi) and schizont (32-36 hpi). Dual luciferase assays were carried out and activities for ffluc and renilla luciferase were obtained. Ffluc activity was normalized against Renilla luciferase activity for each sample and plotted. Ffluc activities driven by the 6A *pfmrp2* promoter were higher than the 11C *pfmrp2* promoter in all stages. However, the assay was not able to recapitulate the stage specific gene expression as seen from the microarray. Statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days.

It is evident that the 6A *pfmrp2* promoter drove higher levels of *ffluc* activity when compared to the 11C *pfmrp2* promoter in all stages (Figure 3.2d). However, the reporter assay did not recapitulate the stage specific activity of the putative promoters in general. The 11C *pfmrp2* promoter drove minimal activity during trophozoite stage, intermediate activity during ring stage and increasing towards the later stages with maximal activity during the schizont stage (Figure 3.2d). On the other hand, minimum activity driven by the 6A *pfmrp2* promoter was in the trophozoite stage and comparable maximal activity during the ring and schizont stages (Figure 3.2d). The activity profile across the 3 stages did not correlate well with the endogenous gene expression of *pfmrp2* for both clones 11C and 6A. As the constructs were transiently expressed in the parasites and were not kept under drug selection pressure, there could be a complex interplay between loss of plasmids and, accumulation and degradation of Ffluc in the parasites. In addition, there may be other factors such as post-transcriptional regulation. These could be reasons why the reporter assay utilized here was not able to fully recapitulate the stage specificity of the putative promoters.

3.3 Defining *cis*-regulatory elements in the *pfmrp2* promoter

In order to define *cis*-regulatory elements present in the putative *pfmrp2* promoters of both clones 6A and 11C, step-wise deletion of the putative *pfmrp2* promoters were carried out and

subsequently tested for their ability to drive transcription of the reporter gene (*ffluc*) in transient transfections and dual-luciferase assays. We started from 2 kb upstream of *pfmrp2* of each clone and successively deleting segments off the 5' end of putative promoter regions (Figure 3.3a). We expected that deleting certain regions would result in changes (increase or decrease) in reporter gene activity more than others and that would indicate that important *cis*-regulatory elements are contained within those regions. As experiments in section 3.2 indicated, these assays were not suitable for evaluation of stage specificity. Hence, constructs were transfected during ring stage and assayed next cycle during the ring stage only.

In general, we observed that there would be a length of promoter which drives maximum activity and successively shorter promoters drive less activity. For the constructs that have been tested, 2 kb of the 11C putative *pfmrp2* promoter and 1.5 kb of the 6A putative *pfmrp2* promoter drove maximal activity (Figure 3.3a). Interestingly, the 2 kb and 1.8 kb constructs of the 6A putative *pfmrp2* promoter did not drive maximum activity even though they were the constructs containing the longest promoters of 6A (Figure 3.3a). A reason could be that these 2 promoters contained segments of coding region of the upstream gene (*pfl1415w*) and exclusion of those sequences in the 1.5 kb 6A putative *pfmrp2* promoter drove maximal activity (Figure 3.3a). This could indicate that repressive elements may be contained within the coding regions. The maximal activity of the 6A putative *pfmrp2* promoter was more than 3 times higher than that of clone 11C (Figure 3.3a). In addition, the maximal activity of the 11C putative *pfmrp2* promoter was equal to the activity driven by the 1 kb 6A putative *pfmrp2* promoter (Figure 3.3a).

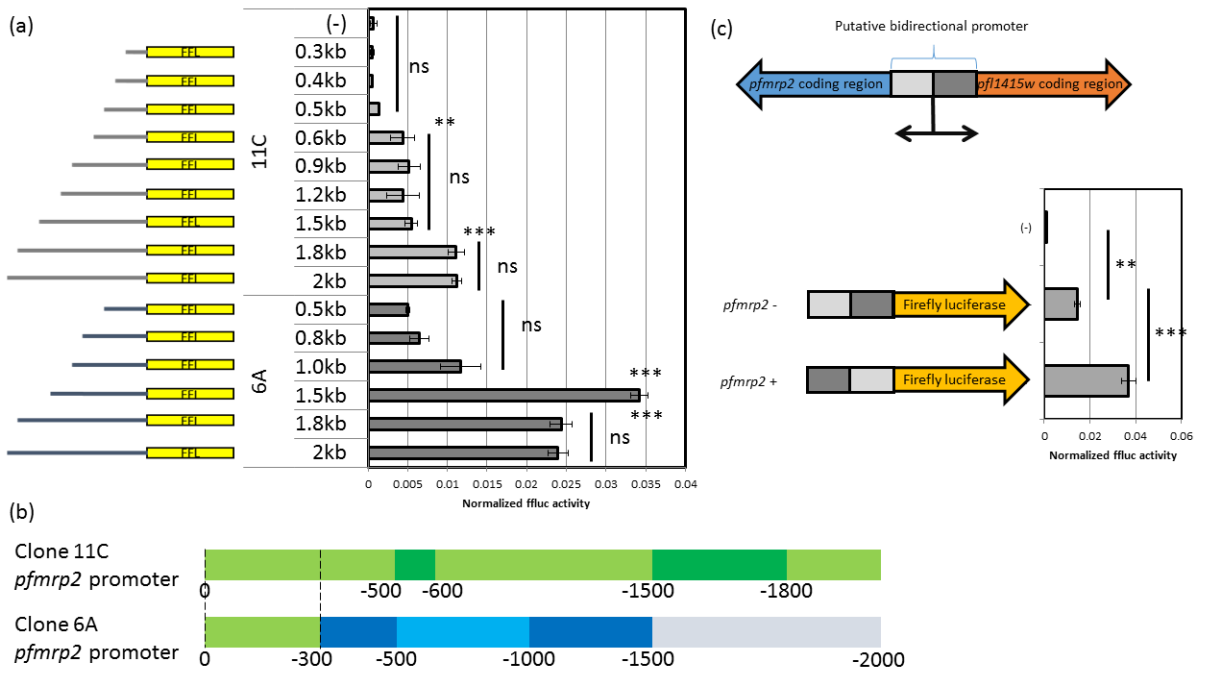


Figure 3.3 Promoter activity by transient transfections.

a) Serial deletions and transfections of *pfmrp2* 5' upstream regions of clones 6A and 11C. Yellow boxes represent *ffluc* gene on the construct. Beginning with 2 kb upstream from the *pfmrp2* translation start site, 5' upstream regions of both 6A (in dark grey) and 11C (in grey) were successively truncated from the 5' end, fragments were cloned into vector pf86, driving transcription of *ffluc*. Constructs containing *ffluc* were co-transfected with pARL-5'3'Actin-Re into ring stage parasites and parasites were harvest during the next ring stage. Parasites were lysed and assayed for ffluc and renilla activity. The normalized ffluc activity for each construct was recorded and plotted as shown. Deletion of certain segments of DNA resulted in large changes in activity. b) Schematic of regions on putative promoter which are important for transcription. As depicted in the transient transfections, deletion of some segments of DNA resulted in large changes in ffluc activity. These regions may be important for transcription (darker shades of color). In clone 11C, important regions are between -500 to -600 bp and between -1500 to -1800 bp (highlighted as dark green). In clone 6A, as the first 300 bp upstream of *pfmrp2* is identical to clone 11C, it is depicted in light green. Important regions in 6A putative promoter include are between -300 to -500 bp and between -1000 to -1500 bp (highlighted as dark blue). As the intergenic space between *pfmrp2* and *pfl1415w* in clone 6A is only approximately 1.7 kb, constructs containing 1.8 kb and 2 kb upstream sequences contain parts of *pfl1415w* coding sequences. Inclusion of those sequences had a repressive effect (grey region). c) Bidirectional promoter. The 1.7 kb putative promoter of clone 6A was cloned in both directions to drive *ffluc* transcription in pf86. *pfmrp2+* is the direction driving transcription of *pfmrp2* in clone 6A and *pfmrp2-* is the direction driving transcription of *pfl1415w* in clone 6A.

Transfections were done in ring stage and assayed during the next ring stage. (-) represent background levels of activity where no ffluc construct was transfected into parasites. Both directions of the promoter drive ffluc activity that is above background levels, demonstrating the bidirectional promoter property of the 1.7 kb sequence. *pfmrp2*+ drove ffluc activity 2 times higher than *pfmrp2*-. Statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days.

Minimum activities were driven by the 0.5 kb or shorter putative *pfmrp2* promoter of both clones, although this may be limited by the detection limit of the assay utilized (Figure 3.3a). For the 11C putative *pfmrp2* promoter, the 0.3 kb and 0.4 kb constructs drove background levels of activity when compared to the negative control where no ffluc construct was transfected (Figure 3.3a). Comparatively, the minimum activity which is driven by the 6A 0.5 kb putative *pfmrp2* promoter is almost 5 times higher than the minimum activity driven by the 11C 0.5 kb putative *pfmrp2* promoter (Figure 3.3a). It must be noted that approximately first 0.3 kb is identical between the 2 clones and when 0.2 kb of sequences further upstream of *pfmrp2* was included into the 0.5 kb 6A construct due to the structural polymorphism deletion of 4.1 kb, increased activity by 5 times. Hence, an important transcriptional element may be present in this 0.2 kb of the 6A putative *pfmrp2* promoter (between -300 to -500 bp).

In the 11C putative *pfmrp2* promoter, inclusion of 100 bp (between -500 to -600 bp) increased activity by almost 4 times and inclusion of 300 bp sequences (between -1.5 kb and -1.8 kb) increased activity by 2 times, indicating the presence of *cis*-regulatory elements in those regions (Figure 3.3a). In the 6A putative *pfmrp2* promoter, inclusion of 500 bp sequences (between -1 kb and -1.5 kb) increased activity by almost 3.5 times (Figure 3.3a). All in all, the regions harboring putative *cis*-regulatory elements which play a role in driving transcription are summarized in a schematic (in darker shades of colors) (Figure 3.3b).

After the 4.1 kb deletion in intergenic space in clone 6A, only 1.7 kb of intergenic space is in between *pfmrp2* and *pfl1415w*. As the 2 genes are arranged in a head-to-head gene arrangement, it is possible that the 1.7 kb acts as a bidirectional promoter which drives transcription of both genes. To test this hypothesis, the putative bidirectional promoter was separately cloned into 2 constructs driving *ffluc* in both orientations. Namely, "*pfmrp2* +" is the construct promoter cloned in the orientation driving transcription of endogenous *pfmrp2* and "*pfmrp2* -" has the promoter in opposite orientation, which means it is driving transcription of *pfl1415w* in the endogenous allele. Firstly, the putative promoter was capable of driving transcription of reporter gene in both orientations, confirming the hypothesis of being a bidirectional promoter (Figure 3.3c). Even though the *pfmrp2* promoter for clone 6A drives transcription in both orientations, it was observed that reporter gene activity driven by the 2 orientations were not equal (Figure 3.3c). The promoter in the *pfmrp2* orientation drove reporter gene transcription more than 2 times more strongly than the promoter in the opposite orientation (Figure 3.3c). This may be due to the proximity of *cis*-regulatory elements with respect to the gene location. A potential caveat in this analysis is that the promoter was not driving the transcription of 2 genes simultaneously but in opposite orientations on separate constructs.

3.4 Conclusions and discussion

By cloning of *in vitro* culture adapted *P. falciparum* parasites, we have found a parasite clone which has a 4.1 kb deletion in the 5' upstream region of the *pfmrp2* gene. This deletion has caused a shift in timing of transcription to the later stages of the parasite IDC and thus leading to increased resistance to anti-malarial drugs. In addition, the deletion in the 5' upstream region has led to the artificial construction of a promoter which drives more

transcriptional activity for *pfmrp2* than the promoter before the deletion. Regions which are important for transcription in promoters of both the clones were also identified.

There have been many reports of variability in transcription in parasites which has led to phenotypic plasticity. Transcriptional variability has been observed among different strains of culture adapted parasites [24] and subclones from the same culture adapted *P. falciparum* cultures [372]. In addition, *in vivo* parasites in the field with distinct transcriptional profiles have been associated with different pathological states including artemisinin resistance [375] and causing severe manifestations of malaria [376]. It is clear that transcriptional variability exists in parasites *in vivo* and *in vitro*, and such variability leads to phenotypic variations; this allows the parasite population to be adaptable in different host conditions thus leading to different disease pathogenesis. In cancer, genetic diversity has led to tumor heterogeneity which results in different responses to chemotherapy [377].

A possible reason for this transcriptional variability could be its unique genome. The *P. falciparum* genome, which is 80-90% AT-rich, is one of the most distinct features. In addition, it is not just the base composition of the genome *per se*, but homopolymeric poly(dA:dT) tracts are interspersed throughout the genome and especially frequent in the intergenic regions. It has been reported that homopolymeric tracts which are more than 10 bases long cover 5% of the genome [378]. Repetitive sequences in the genome have been known to be regions of genomic instability and hypothesized to be a major driving force for DNA sequence evolution [379]. Mechanisms utilized by repetitive sequences to drive sequence instability include slippage misalignments during DNA replication and this results in deletions and duplication of bases which in turn cause frameshift mutations [380]. This is observed in DNA polymerase performing PCR reactions *in vitro* as well [381]. A second

mechanism is the long range slippages, due to sequence homology of stretches of repetitive sequences, transient misalignment between 2 stretches of homologous repetitive could lead to duplication or deletion of sequences in between the 2 stretches (Figure 3.4b). This was observed in bacterial cells where one of the mutational hotspots are deletions formed between repeats of more than 8 bases [382, 383]. Pathogenic bacteria have utilized such mechanisms to activate or inactivate genes or even entire gene loci for selective advantage [384].

Repetitive sequences in *P. falciparum* have been analyzed genome-wide. The largest stretches of repeats lie in regions near to the telomeres and approximately 9% of coding regions contain stretches of repeats, a number that is higher than any organism reported [385]. Analysis of genomic sequences of *P. falciparum* clones and *P. reichenowi* revealed that stretches of repetitive sequences enabled the multiplication of gene loci [385]. In addition, it was also reported that the low complexity sequences show diversity among *P. falciparum* isolates and clones [386] and also across different *Plasmodium* species [387]. Investigators have also reported of increased occurrences of SNPs in regions which are close to low complexity sequences [387]. Even more recently, a group has proposed a 2-step approach of how culture adapted parasites have multiplied a gene loci involving flanking poly(dA:dT) tracts and leading to increase in resistance to an inhibitor against dihydroorotate dehydrogenase [388]. Taken together, there is mounting evidence of low complexity repetitive sequences driving genome instability in the *P. falciparum* genome as a possible mechanism of parasite adaptability and survivability.

The studies on low complexity repeats so far have been mostly done on coding regions, possibly due to its importance as a protein sequence or may be due to the difficulty in obtaining reliable sequence data from intergenic regions where homopolymeric tracts are

abundant causing artifacts in PCR and sequencing. In this study, we have shown that structural polymorphisms have led to the construction of a mutant promoter which has altered promoter activity. It was interesting to note that two polyA tracts flank the 4.1 kb intergenic space that is deleted in clone 6A, named tract 1 and tract 2 (Figure 3.4a). A possible mechanism involving DNA polymerase could have occurred during replication leading to the deletion of sequences in between the 2 tracts (Figure 3.4). An accidental slippage misalignment of the 2 tracts could have occurred (Figure 3.4b). Tract 2 on the template strand could have misaligned onto tract 1 on the newly synthesized strand. As the intervening sequences between the 2 tracts in the template strand could have been looped out and the sequences were not included into the newly synthesized DNA. Hence, 4.1 kb was deleted from the intergenic space between *pfmrp2* and *pfl1415w* (Figure 3.4c).

This study has also characterized the *pfmrp2* promoters of clones 6A and 11C in defining important regions important for transcription (Figure 3.3b). These regions may contain DNA motif sequences which can bind to transcription factors to define the stage specific expression of genes. There have not been reports of DNA motifs defining stage specificity of genes in the IDC so far. Further work can be done to narrow down the important regions to specific DNA sequences and binding partners of these motifs can also be identified.

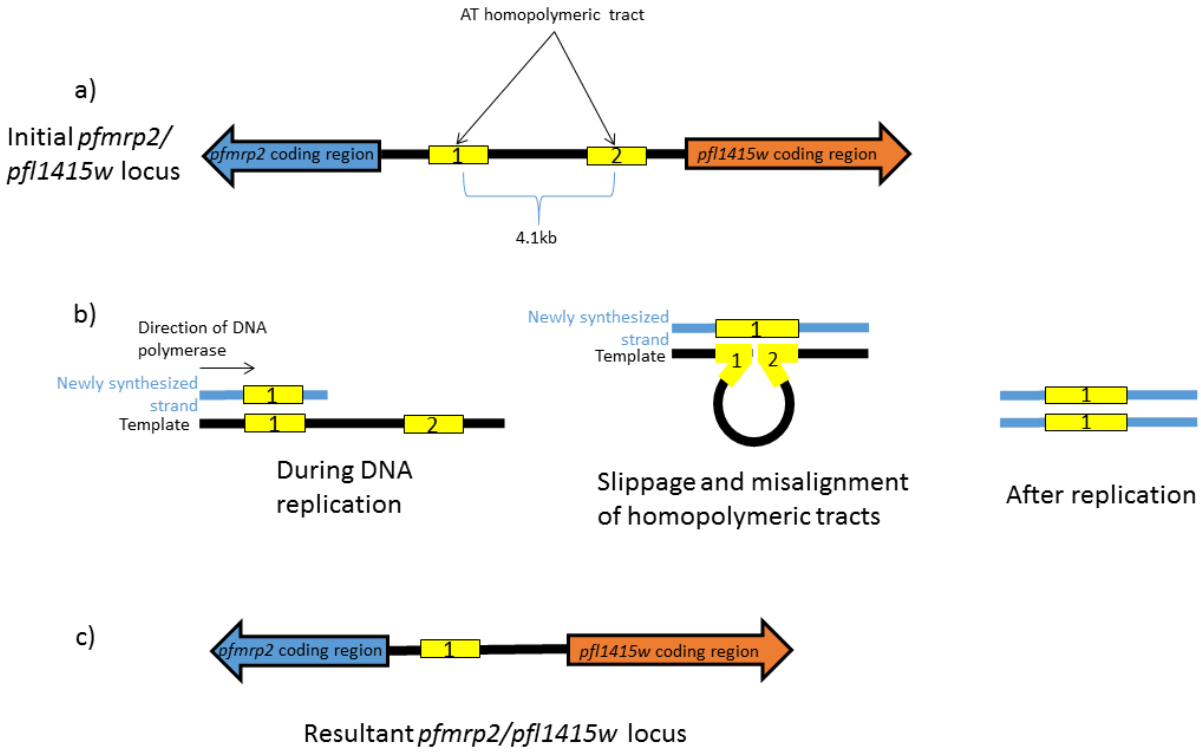


Figure 3.4 Proposed mechanism of the deletion of 4.1 kb intergenic space between *pfmrp2* and *pfl1415w* in clone 6A.

a) Orginal gene locus of *pfmrp2/pfl1415w*. A total of 5.8 kb intergenic space existed between the 2 genes in the original locus. The 2 homopolymeric poly(dA:dT) tracts of interest (labeled as 1 and 2) flank a region of 4.1 kb. b) During DNA replication, the newly synthesized strand (blue) is synthesized using the original DNA as template (black). Accidental slippage occurs during DNA replication and tract 2 of the template strand misaligns with tract 1 on the nascent strand due to sequence homology. This causes the looping out of sequences in between the 2 tracts on the template. As a result, the sequences that were looped out were not read by the DNA polymerase and therefore nor included the DNA of daughter cells. c) The resultant locus of *pfmrp2/pfl1415w* after replication with slippage misalignment. 4.1 kb sequence between tract 1 and tract 2 were deleted from the genome for clone 6A.

4. Prediction and validation of DNA motifs regulating IDC gene transcription

4.1 Background and rationale

The *P. falciparum* transcriptome in the IDC is characterized by a highly organized cascade of transcripts, this is found by using various experimental procedures [19, 20, 23, 30]. Each gene has a cyclic pattern of transcript expression where the abundance of transcripts peak at a certain point of the IDC and, as for most genes, the peak coincides with the need of the function of the gene. As such, it is apparent that *P. falciparum* requires an intricate mechanism of transcriptional regulation. Due to the difficulty in annotating the genome, the search for specific transcription factors (STFs) in the *P. falciparum* genome for eukaryotic transcription factor homology has yielded disappointing results [301, 304]. On the other hand, the *P. falciparum* genome contains a full complement of proteins involved in epigenetic regulation [43, 175]. Hence, much attention and progress has been made in the area of transcription regulation at the level of epigenetic regulation in the past decade.

Discovery of the ApiAP2 transcription factor family, which contains the plant transcription factor AP2 domain in the *Plasmodium* species genomes has sparked new interest in transcription regulation. However, much work has been done in characterizing STFs playing roles in other stages of the *Plasmodium* life cycle such as the liver [322], mosquito [320] and gametocyte [325] stages. The STFs characterized in the IDC were for a specific gene [315], the *var* gene family [323] or for a limited gene group [309]. Although much progress has been made through those studies, direct evidence of how IDC genes are transcriptionally regulated is still very limited.

In eukaryotes, STFs function by binding to specific DNA motifs on gene promoters or enhancers to elucidate its influence on transcription, whether to activate or repress transcription. There have been reports of computer algorithms searching the 5' upstream regions of genes for specific DNA sequences enriched in different gene groups. Genes have been grouped according to their transcript abundance patterns based on the assumption that co-expressed genes are co-regulated and thus, would be regulated by the same set of transcription factors and their promoters would, therefore contain the same DNA motifs [335, 359, 389]. Although a large list of motifs has been generated from these computational analyses, little or no validations have been done on the functionality of these motifs.

Hence, in this study I utilized one of the computer algorithms developed to analyze the 5' upstream regions of *P. falciparum* IDC genes using an updated transcriptome. Although the algorithm, Finding Informative Regulatory Elements (FIRE), has been used to analyze DNA motifs in *P. falciparum* [359], the study had used the transcriptome of approximately 2700 genes [19]. The transcriptome has been updated by using a microarray chip with oligonucleotides representing all 5554 *P. falciparum* ORFs [390]. The updated transcriptome that I have used for the motif prediction analysis in this study contains transcript abundance profiles of 4670 genes [30]. After which, I devised a quick method to screen the motifs for their function in *P. falciparum* IDC *in vitro* by transient transfections. The results from the initial screens were validated by real time quantitative PCR (qPCR).

4.2 FIRE motif prediction

Each gene was assigned a phase value (from π to $-\pi$), according to the transcript abundance profile of each gene from the updated transcriptome [30]. With the gene IDs on one column and corresponding phase values in the next column, the data was submitted into the FIRE

algorithm. The FIRE algorithm then clustered the genes in a continuous fashion with respect to its respective phase value. In other words, FIRE placed genes with similar transcript abundance profiles (depicted by similar phase values) together in the same cluster. FIRE reported a total of 26 clusters from the data provided; they are depicted as columns (Figure 4.1a). From left to right, the clusters contain genes which transcripts are more abundant in ring on the left to the clusters containing genes which transcripts are more abundant in schizont stage on the right. After which, the algorithm searched the 1 kb sequences upstream of the translational start site of genes in each cluster for motifs (7-10 bp DNA sequences) which are over representation or under representation in each cluster. The predicted motifs are reported next to the columns of clusters and whether or not the motif is enriched in the particular cluster, it is reflected in different colors. Yellow depicts that particular motif is overrepresented in a particular cluster of genes and blue depicts underrepresented.

FIRE predicted a total of 24 motifs (Figure 4.1a). Motifs 1, 2 and 3, are overrepresented in clusters containing genes which transcripts are abundant in ring stage but are underrepresented in clusters containing schizont stage genes (Figure 4.1a). Motif 8 is overrepresented in many clusters containing genes which transcripts are more abundant in the later half of the IDC while being underrepresented in clusters containing genes belonging to the first half of the IDC (Figure 4.1a). From motif 10 to 16, they appear to be progressively overrepresented in the clusters containing genes in the later stages (Figure 4.1a). Motif 15 and 16 are overrepresented in clusters containing late schizont and early ring stage genes; and underrepresented in clusters containing late ring and trophozoite genes (Figure 4.1a). Finally, similar to motif 8, motif 22 is overrepresented in many clusters containing genes which transcripts are more abundant in the later half of the IDC while being underrepresented in clusters containing genes belonging to the first half of the IDC (Figure 4.1a). There were also

motifs which are overrepresented in 2 distinctly different clusters in the life cycle. An example is motif 6 where it is overrepresented in an early ring stage cluster and also in a trophozoite cluster (Figure 4.1a).

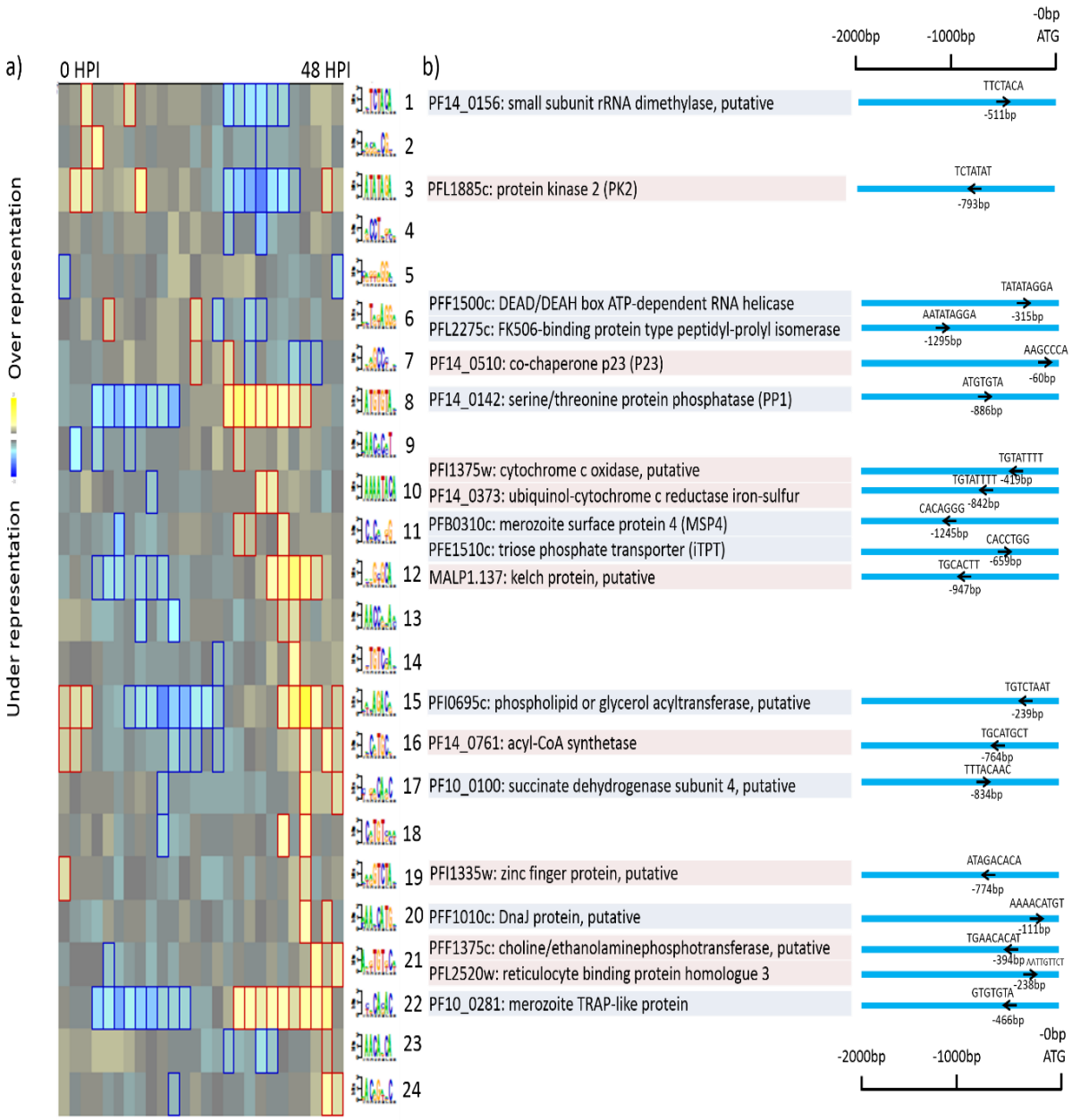


Figure 4.1 Motifs predicted by FIRE algorithm.

a) 24 motifs predicted by FIRE and its associated clusters. The data of 4670 *P. falciparum* genes and the associated transcription profile represented by phase (-π to π) was submitted to the FIRE algorithm. FIRE then clusters the genes continuously into 26 different clusters represented in columns. The early ring stage abundant genes on the left and progressively, late

schizont stage abundant genes on the right. FIRE searches the 1 kb sequences upstream of translational start site of the genes in each of the 26 clusters for DNA sequences motifs which are overrepresented or underrepresented in the cluster. The 24 motifs reported are represented in rows. Therefore, for every motif in each cluster will be represented by a box. For example, if the motif is overrepresented in the 1 kb upstream region in genes of a particular cluster, the box will appear yellow and if underrepresented, the box will appear blue. b) The list of gene promoters chosen for motif screening. The schematic of 2 kb upstream of translational start site of the respective genes is represented as the blue horizontal lines on the right most of the figure. The exact motif sequence is as shown and the orientation of the motif is depicted by the orientation of the black arrows.

In order to validate the functionality of these motifs, we had to first devise a strategy to select the motifs and promoters of genes to be tested. First, overly degenerate motifs such as motifs 2, 4 and 5 where more than half of the nucleotide positions were degenerate were not tested. In addition, motifs which show only slight overrepresentation such as motifs 9, 13, 14, 18, 23 and 24 were also left out of the analysis. For the remaining 14 motifs, either 1 or 2 promoters of genes in the overrepresented clusters were picked for further analysis. The selected genes are listed beside their respective motifs (Figure 4.1b). The positions, orientations and the exact sequences of the motifs are shown in the schematic diagram beside the gene names (Figure 4.1b).

4.3 Optimization of motif screening strategies

We had known from experiments in section 3.2 and previous studies that transient transfections were not a suitable tool in defining stage specific effects due to compounding factors such as transfection efficiencies leading to copy number differences and rate of loss of plasmid. However, a merit of transient transfections was that the motifs could be screened relatively quickly and motifs of high confidence in the initial screen could be focused on in the further experiments. The general idea of transient transfection screening would be to

create 2 constructs (one construct with the motif of interest and the other, without the motif of interest) driving a reporter gene activity and comparing the activities driven by the 2 constructs.

The first strategy devised to screen the motifs was to first amplify a segment of the promoter of interest from the translation start site to the position of the motif on the promoter (Figure 4.2a). Thus, by designing appropriate primers, motifs can be either included (blue arrow) or excluded (red arrow) into the amplified region (Figure 4.2a). Subsequently, the amplified region either with or without the motif is cloned into plasmid pf86 (introduced in section 3.2) driving the transcription of a *firefly luciferase* reporter gene (*ffluc*). Hence, the activity driven by construct containing the motif at the very end of the promoter region could be compared with the construct containing the promoter without the motif. As detailed in section 3.3, transfections were done in ring stage parasites and assayed in the ring stage of the subsequent cycle.

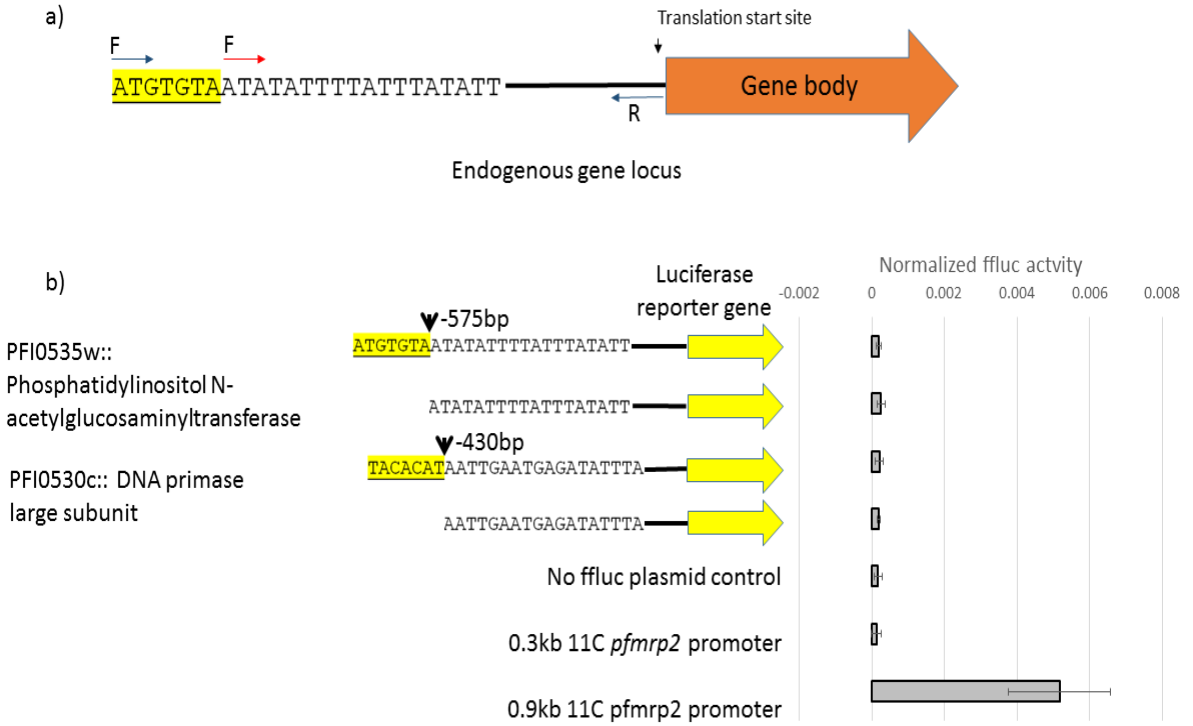


Figure 4.2 Cloning strategy.

a) PCR strategy. The PCR template is gDNA of the 3D7 *P. falciparum* clone. The orange arrow is the gene body and the sequences upstream of the gene from translational start site to the position of the predicted motif (or just before) will be cloned into the vector containing the *ffluc* gene, pf86. The primers designed are represented in blue and red arrows. The blue primers were used to amplify the promoter with the motif present at the very end of the promoter. This will be done by including the motif sequences at the very end of the blue forward (F) primer. For PCR amplification of the promoter without the motif, the same blue reverse (R) primer is used but the motif sequences were not included in the red forward primer. Hence, the motif sequence will not be included in the promoter without the motif sequence. b) Transfections and dual luciferase experiments. 2 genes (*pfl0535w* and *pfl0530c*) were selected to test out this protocol. The constructs with and without motif were prepared as described in figure 4.2a. The end sequences and motifs are as shown, the motif sequences are highlighted in yellow and underlined. The promoters are cloned into pf86 to drive the transcription of *ffluc*. As negative controls, the 0.3 kb 5' upstream of clone 11C *pfmrp2* promoter (from section 3.3) and equal volume of water was used in place of plasmids were used. The 0.9 kb 5' upstream of clone 11C *pfmrp2* promoter was used as positive control for the experiments. Transfection and assays are done as described in Chapter 2. For the promoters that were tested, they drove only background levels of activity which was indicated by the levels of ffluc activity driven by 0.3 kb

5' upstream of clone 11C *pfmrp2* promoter and no plasmid negative control. The error bars represent data from 3 independent transfections done on 3 separate days.

The 0.9 kb 5' upstream of clone 11C *pfmrp2* promoter driving ffluc activity construct used in section 3.3 was used as a positive control (Figure 4.2b). In addition, the 0.3 kb 5' upstream of clone 11C *pfmrp2* promoter driving ffluc activity construct which has been shown to drive background levels of activity and a transfection done without the any plasmid containing ffluc were used as negative controls. Motif 8 was first tested using this approach and 2 promoters (*pfl0535w* and *pfl0530c*) which contain the motif were selected. However, the results showed that all 4 constructs whether with or without motif drove only background levels of activity which was similar to the 0.3 kb 5' upstream of clone 11C *pfmrp2* promoter driving ffluc activity construct and the no plasmid control (Figure 4.2b). Hence, it was not conclusive as to whether or not the inclusion of the motif made a difference in transcriptional activity.

As FIRE searches 1 kb sequences 5' upstream of translational start sites of genes, the motifs reported will be in only within the 1kb region. Using the approach that had just been described will limit the length of promoter to be tested to less than 1 kb, which may not be appropriate as short promoters may not drive enough ffluc activity such that it can be compared between constructs. Furthermore, the context of the surrounding sequences may be needed by motifs to function correctly. Hence, a new strategy needs to be devised.

The next strategy was to first clone 2 kb 5' upstream of the gene of interest into the pf86 construct (Figure 4.3a). After which, appropriate primers were designed to amplify the entire plasmid with the exception of the motif sequence of interest to create the mutant promoter where the motif is deleted from the promoter sequence (Figure 4.3b). The primers used were

designed with appropriate restriction enzyme sites such that the plasmids could be ligated and transformed into bacteria cells with ease. Using this cloning strategy, we were able to obtain 2 constructs; one with the wild type promoter sequence and the other with the exact same 2 kb promoter sequence with the exception of the motif sequence deleted. However, restriction enzyme sites were used for ligation in this strategy and this would mean that an additional 6 bases were added into the promoter sequence, in place of the motif of interest. Hence, the sites need to be tested if they would cause any effect in transcriptional activity.

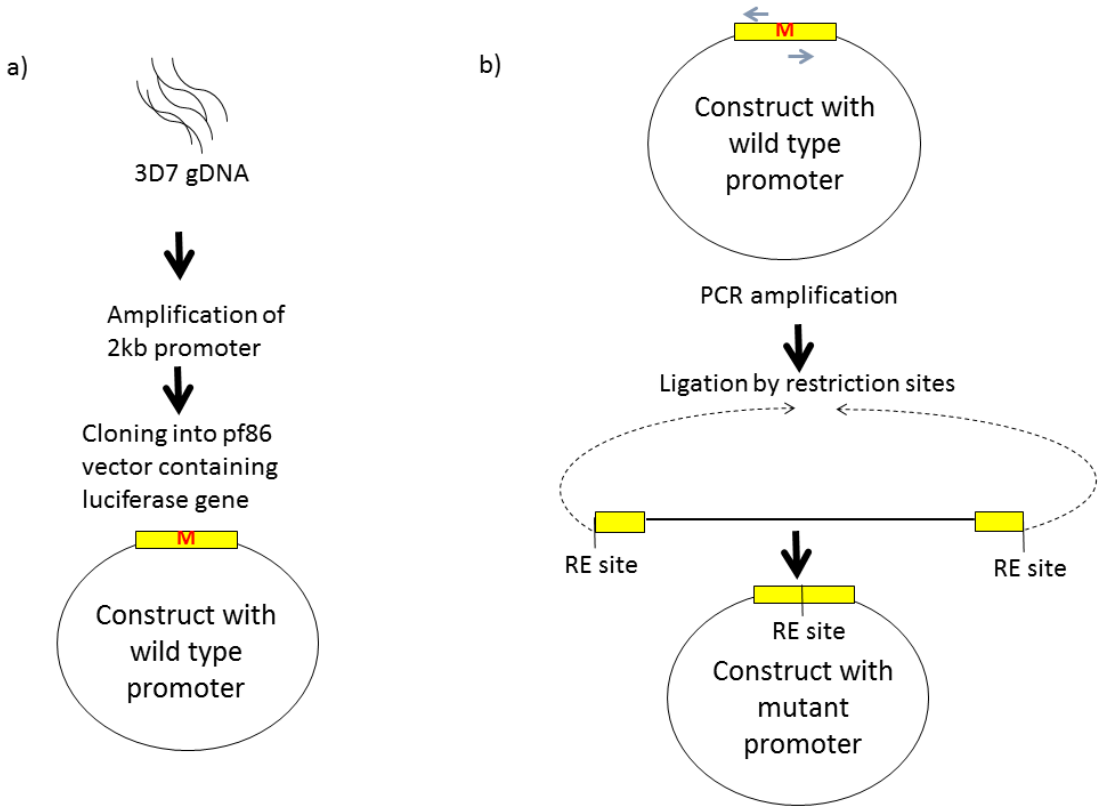


Figure 4.3 Cloning strategies.

a) Cloning of the wild type promoter construct. Approximately 2 kb upstream of the gene of interest was amplified by PCR using primers with restriction enzyme sites XhoI and NcoI. The PCR products were then restriction digested and subsequently ligated into the pf86 vector backbone. b) Cloning of the mutant promoter. The mutant promoter with the motif deleted is created from the construct containing the wild type promoter. Primers containing separate restriction sites PstI, HindIII and SpeI were used to amplify the entire plasmid. Sequences of the motif were not included in the primers. The positions of the primers are as indicated in the

diagram where it is adjacent to the position of the motif (blue arrows). The PCR product was then digested with its respective restriction enzymes. The digested products were ligated and transformed into Dh5α competent cells.

An excellent platform to test this strategy out would be the *pfhsp86* promoter. Although the function of PfHsp86 has not been studied in *P. falciparum*, it is one of the members of a family of chaperone and co-chaperone proteins which have been well studied in *P. falciparum*. In other organisms, chaperones play an important role in facilitating the correct folding, assembly and localization of proteins [391]. PfHsp90 has been found to associate with a large chaperone complex which includes PfHsp70 and inhibition of PfHsp90 has caused a developmental block in parasites [392]. It has also been demonstrated that PfHsp90 plays a role in the protection from the developmental block caused by febrile temperature treatment at 41^{o}C [393]. PfHsp40s are co-chaperones and have been shown to be up regulated upon heat shock [394, 395]. Upon heat stress, PfHsp40 colocalizes with chaperone protein PfHsp70 and increases the activity of PfHsp70 [396]. Although Hsp86 has not been studied in *P. falciparum*, it was shown to be important during mouse embryonic development [397]. The G-box motif is a palindromic GC-rich sequence found on the *pfhsp86* promoter and has been found to induce transcriptional activity by transient transfection assays [329]. The original pf86 construct contains the *pfhsp86* promoter and is thus, a convenient platform to use. 3 different sets of primers each set containing 3 different restriction enzyme sites (PstI, HindIII and SpeI) for ligation were designed for the deletion of the G-box as described in Figure 4.3b. As expected, all G-box deletions (G-box del) resulted in a decrease in ffluc activity when compared to the wild type *pfhsp86* promoter (pf86) as reported in previously published data [329] (Figure 4.4, G-box del). In addition, the 3 constructs with the G-box deleted but containing 3 different enzyme sites drove ffluc activity to a similar level (Figure 4.4, G-box del (PstI, HindIII, SpeI)). To further substantiate whether the inclusion of

restriction enzyme sites had an effect on the construct's transcriptional activity, 3 additional constructs were made. Each containing the 3 restriction enzyme sites, however in this case, none of the nucleotides on the *pfhsp86* promoter were altered. Hence, the resulting construct would only have additional nucleotides added (the restriction enzyme sites), and apart from that, is identical to the wild type promoter. However, it was observed that the inclusion of additional nucleotides from the restriction enzyme sites caused a significant increase in activity when compared to the wild type *pfhsp86* promoter (Figure 4.4, no del (PstI, HindIII, SpeI)). This was undesirable as the restriction enzyme sites included had elicited effects which may complicate the results when used in the motif screening experiments.

Hence, a strategy needs to be devised such that no restriction enzyme site will be added during the ligation process in the making of the construct with the motif deleted (Figure 4.3b). Ligation without enzyme sites is an inefficient process which would entail tedious bacteria clone screening. However, it was possible with the phosphorylation of 5' hydroxyl groups of the PCR products with T4 Polynucleotide Kinase and subsequently ligating the 2 ends. This was done after amplification of the wild type construct (Figure 4.3b). The result was that with no nucleotides included, the activity driven by the construct was similar to the wild type *pfhsp86* promoter (Figure 4.4, no del (no site)). In addition, the G-box deletion with no restriction enzyme sites added caused a similar decrease in ffluc activity when compared to the wild type *pfhsp86* promoter as seen previously when restriction sites were added (Figure 4.4, G-box del (no site)).

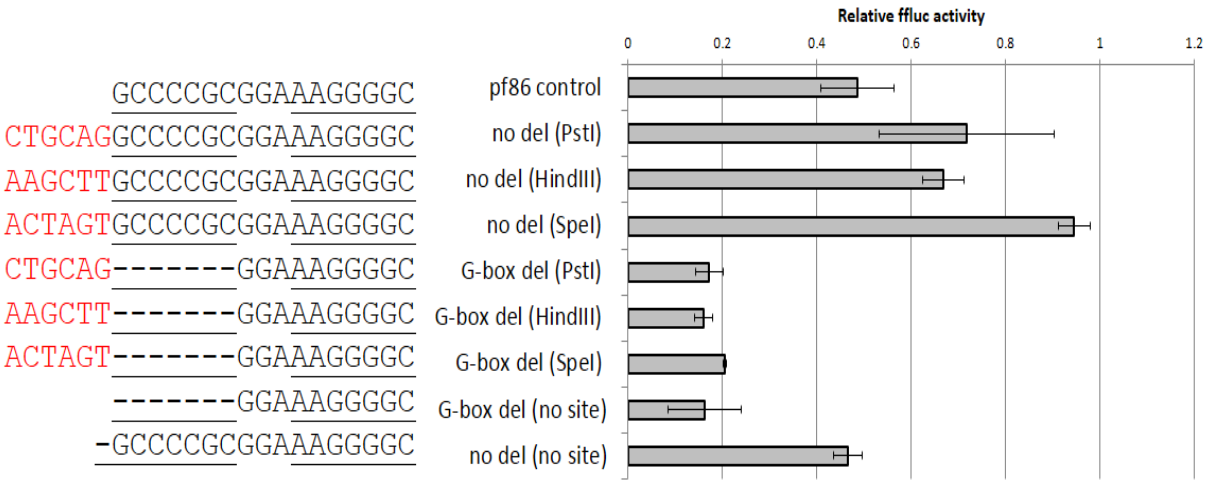


Figure 4.4 Testing of cloning strategy in Figure 4.3.

All the constructs listed contained the 1.8 kb 5' upstream region of *pfhsp86* driving the transcription of ffluc, varied only at the vicinity of the G-box motif (underlined). The variations in each construct are as listed. pf86 control contains the wild type 1.8 kb upstream region of *pfhsp86* with the G-box motif fully intact. In order to test the effects of different restriction enzyme sites, constructs with additional sequences of the restriction enzyme sites (PstI, HindIII and SpeI) adjacent to the G-box motif were added but the G-box motif remains intact (no del (PstI, HindIII and SpeI)). Half of the palindromic sequence of the G-box motif was deleted in the G-box del constructs. Constructs were ligated with the respective restriction enzyme sites (G-box del (PstI, HindIII and SpeI)). The G-box del construct was also created with no restriction enzyme site (G-box del no site). This done by phosphorylating the 5'OH groups of the PCR products using T4 polynucleotide kinase in Figure 4.3b and subsequently ligating the ends. No restriction enzyme sites were added when designing the primers for this amplification. By the same method, a construct was created as a negative control. This construct had a one base deletion adjacent to the G-box motif (in order to differentiate construct altered by amplification and wild type construct) but the G-box motif is intact and no restriction sites were added. All transfections and assays were done according to Chapter 2. The inclusion of enzyme sites in constructs increased in activity when compared to the pf86 control. Deletion of G-box motif in constructs decreased activity when compared to the pf86 control when sites were added or not. When there was no inclusion of enzyme sites, the activity was similar to the pf86 control. The error bars represent data from 3 independent transfections done on 3 separate days.

As this strategy appears to be feasible, 3 motifs (motif 3, 7 and 20) were selected for initial testing (Figure 4.2). Constructs were created as described in Figure 4.3 and mutant constructs were created by ligation without restriction enzyme sites. Constructs were sequenced by

capillary sequencing over the region containing the motif to ensure the presence of motif in the wild type construct and absence of motif in the mutant construct. In addition, *ffluc* genes on all clones were sequenced to ensure that differences in activity were not due to mutations on the *ffluc* gene. 3 construct clones of wild type and mutant construct were tested for each promoter (Figure 4.5). Motif 3, 7 and 20 were tested in the context of promoters for *pfl1885c*, *pf14_0510* and *pff1010c* respectively. For motif 20, all clones of the mutant constructs displayed a higher activity compared to all the clones of wild type constructs (Figure 4.5). However, such clear evidence could not be observed in motif 3 and motif 7. There was up to 35% difference in ffluc activity driven by different clones of the same construct (wild type or mutant). This would mean that if the true difference in activity between the wild type construct and the mutant construct was subtle (<35%), it would not be detectable by this assay. For example, when comparing clones of motif 3, mutant clone 4 displayed a higher ffluc activity when compared to all wild type clones. However, there is no significant difference when comparing mutant clone 2 to the wild type clones. Hence, the data could be inconclusive when using this strategy.

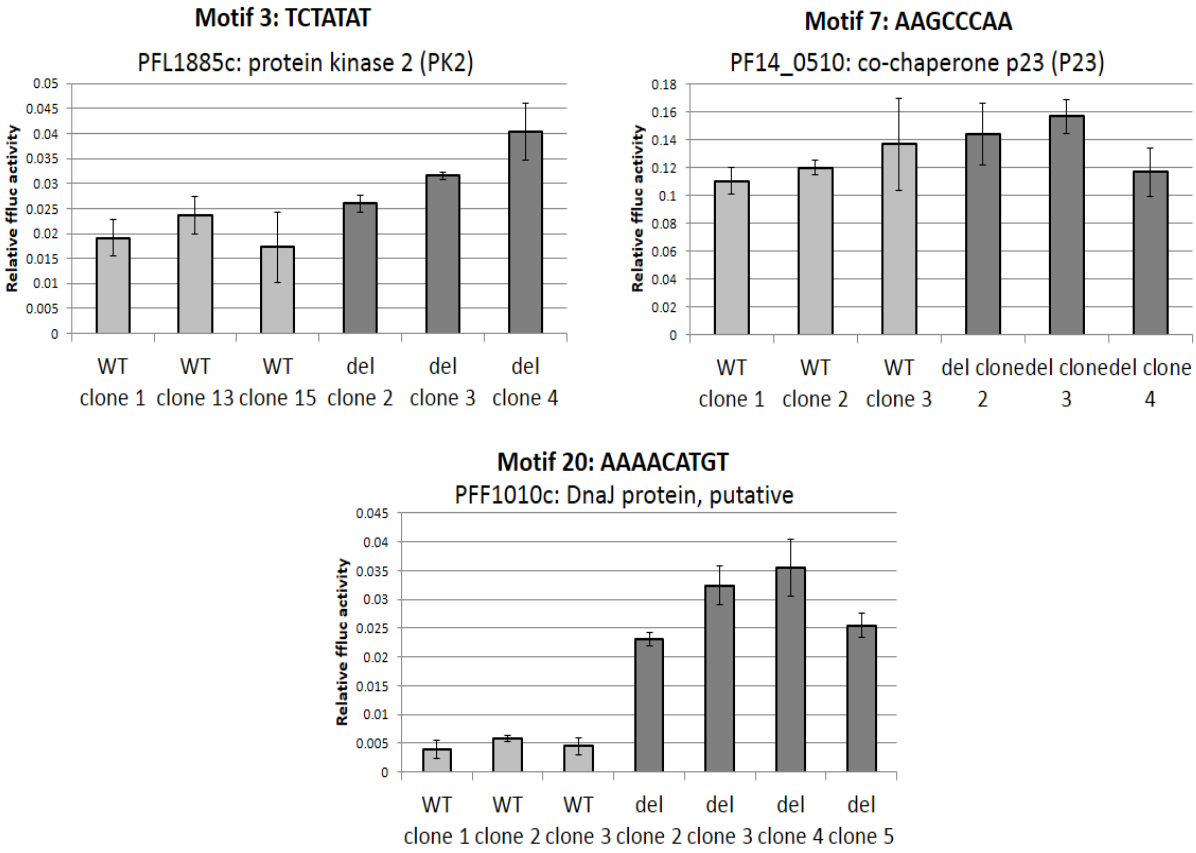


Figure 4.5 Screening of motifs 3, 7 and 20 and variations in activity among clones.
Constructs containing wild type and mutant 2 kb upstream regions of *pfl1885c*, *pf14_0510* and *pff1010c* were created using cloning strategy described in Figure 4.3. After sequencing over the motif position to ensure presence of motif in wild type construct and absence of motif in mutant construct, 3-4 clones of wild type and mutant constructs of each promoter were selected for transient transfections. Transient transfections and dual luciferase assays were carried as described in Chapter 2. Among the clones of the same promoter, a difference in ffluc activity of up to 35% could be observed. The error bars represent data from 3 independent transfections done on 3 separate days.

4.4 Transcriptional variation among clones

In order to investigate the reason behind the variation in ffluc activity driven by different clones of the same promoter (Figure 4.5), 12 clones of the 2 kb wild type promoter of *pfl1885c* used for screening motif 3 were thoroughly sequenced in its entire 2 kb promoter region. After which, transient transfection assays were done using each of the 12 clones.

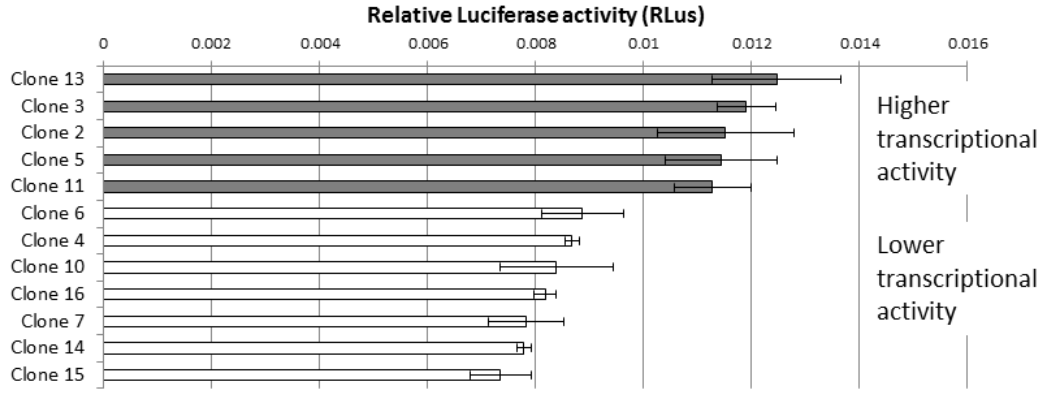
It was observed that there are 10 homopolymeric poly(dA:dT) tracts of more than 15 bases spread along the entire length of the 2 kb promoter region. The layout of the tracts in the entire 2 kb region of the promoter is shown in Figure 4.7. They were either dinucleotide repeats of AT or repeats of A or T. These sequences are in plenty in the *P. falciparum* genome especially in the intergenic regions. Sequencing of the clones has revealed that the only differences among the clones were the insertions and deletions (indels) in these tracts. The sequences of the clones were compared to the 3D7 genome sequence and the indels were recorded for each clone at each individual tract (Figure 4.6b). It is interesting to note that deletions were more common. Another observation was that the number of indels on AT tracts is only in even multiples while indels in A or T tracts can be in both even and odd multiples. These observations were characteristic of local slippages during the amplification causing template and nascent DNA strands to be misaligned, which results in indels.

There is a maximum variation of approximately 40% when comparing the ffluc activity driven by the clone with the highest activity to the clone with the lowest activity (Figure 4.6a). When the activities of the clones were ordered in descending order of ffluc activity, a clear segregation of clones which drove higher activity and clones which drove lower activity could be observed (Figure 4.6a). The only difference among the clones was the length of the tracts due to indels along the 2 kb region and no other mutations were observed.

To further investigate whether the variations in the tracts could be directly linked with the variations in transcriptional activity displayed by the different clones, the total number of base deletions for every tract were summed up for each clone and summarized in a box plot (Figure 4.6c). It is observed that a higher number of bases deleted from the promoter was associated with the group of clones which displayed lower ffluc activity when compared the

group of clones which displayed higher ffluc activity (Figure 4.6c). In other words, the group of clones which displayed lower ffluc activity had, in general, shorter poly(dA:dT) homopolymeric tracts. This difference was statistically significant.

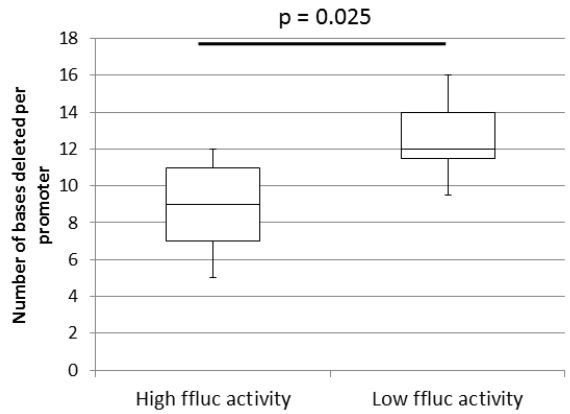
a)



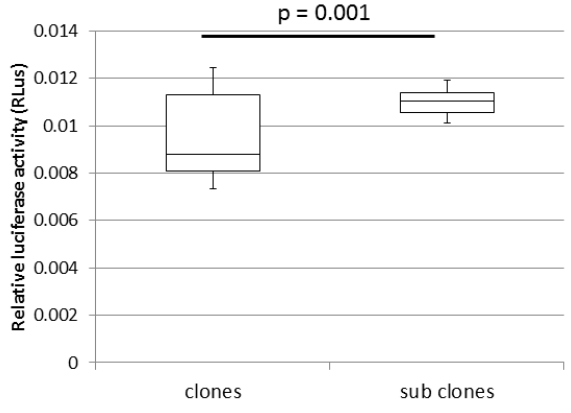
b)

Tract	1	2	3	4	5	6	7	8	9	10
Length	15	20	35	16	27	17	27	31	22	32
Type	T	AT	AT	A	AT	T	AT	AT	T	AT
Position	-1874	-1720	-1575	-1507	-986	-951	-828	-662	-555	-255
	Number of bases inserted or deleted									
Clone 13	0	-2	-2	-3	0	0	-2	0	0	-4
Clone 3	0	0	0	0	0	0	0	-2	-5	0
Clone 2	-1	0	0	0	0	-1	-2	-4	-1	0
Clone 5	-1	0	-4	0	0	-1	2	-2	-1	0
Clone 11	-1	0	-4	-1	0	0	-2	0	-4	2
Clone 6	0	0	-2	0	0	-1	-2	0	-8	0
Clone 4	0	0	0	-2	0	-1	0	0	-2	-2
Clone 10	0	-2	-2	-1	0	-1	0	0	-2	-4
Clone 16	0	0	-4	-2	0	2	-2	0	-4	-6
Clone 7	-1	0	2	-2	-2	0	-4	-2	-6	0
Clone 14	0	0	-4	0	0	0	-2	0	-5	-2
Clone 15	0	-2	0	0	0	0	-4	0	-5	0

c)



d)



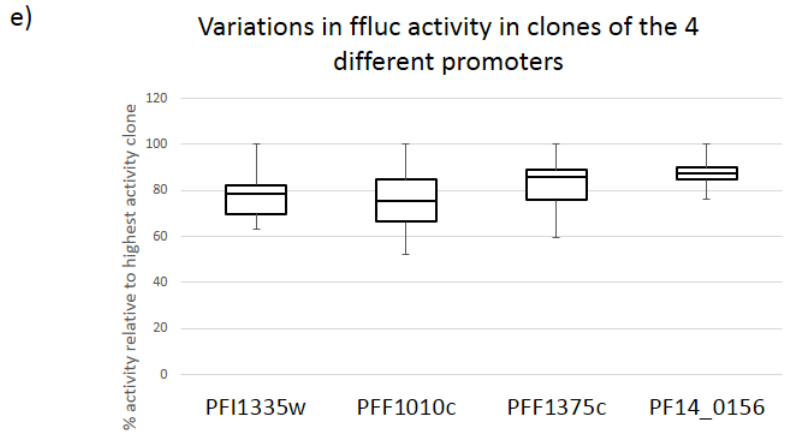


Figure 4.6 Tract length polymorphisms and transcriptional variations.

a) Variations in ffluc activity driven by the clones containing 2 kb 5' upstream region of *pfl1885c*. 12 clones of the construct containing 2 kb 5' upstream region of *pfl1885c* driving *ffluc* gene were transfected and assayed as described in Chapter 2. The clones are arranged in order of descending ffluc activity. A variation in ffluc activity of approximately 40% was observed between the clone which drives the highest ffluc activity (clone 13) and the clone which drives the lowest ffluc activity (clone 15) with clones driving intermediate levels of ffluc activity in between. The clones can be divided into 2 groups: one driving higher ffluc activity and the other driving lower activity. The error bars represent data from 3 independent transfections done on 3 separate days. b) The 2 kb 5' upstream region of *pfl1885c* in all the 12 clones were thoroughly sequenced and indels were observed in the lengths of 10 poly(dA:dT) tracts among the 12 clones. The length, position on promoter and composition of the 10 tracts were as stated. The number of bases deleted or inserted (with respect to the 3D7 reference genome sequence) at each tract of every clone was recorded. c) Box plot comparing the number of bases deleted in the high activity group to the low activity group. Number of bases deleted for each clone were summed up as recorded in (b) and plotted in a box plot for each high activity group and low activity group. The clones in the lower activity group had a statistically significant higher number of bases deleted from the tracts on the promoter summed up together by 2-tailed student's t-test (p=0.025). d) Box plot comparing the variance in activity of clones of the construct containing 2 kb 5' upstream region of *pfl1885c* and sub clones of clone 11. The sequences of 2 kb 5' upstream region of *pfl1885c* in the sub clones of clone 11 are identical to each other. A significantly greater variation can be observed in the activity driven by the clones when compared to the sub clones by f-test (p=0.001). e) Transcriptional variance in clones of 4 other promoters. For each promoter, the activity driven by the clone which drives the highest activity is set at 100% and the activity of rest of the clones are expressed as a percentage it. Of the 4 promoters tested, the clones of each promoter displayed a variation in activity of 20-50%.

To further confirm that the difference in transcriptional activity in the different clones is due to the indels that are associated with promoter on individual clones, one of the bacteria clones was chosen to be sub-cloned. Plasmid was extracted from 12 sub-clones and the 2 kb region was again sequenced. It was observed that the entire 2 kb sequence was identical for 12 sub-clones. The plasmids from the sub-clones were then transfected into parasites and assayed. The normalized ffluc activities were summarized in a box plot (Figure 4.6d). For comparison, the ffluc activities driven by the individual clones shown in the bar graphs in Figure 4.6a are also plotted as a box plot (Figure 4.6d). There was a significantly smaller spread of ffluc activities driven by the sub-clones as compared to the activities driven by the clones. This confirms that the variations in activity were due to the polymorphisms in tract lengths in different clones.

The reasons for tract length polymorphisms could be due to slippages during PCR amplification when cloning the promoter into the construct or it could be that the variations in tract lengths already existing in the parasite population. It could even be contributions from both. SNPs and structural polymorphisms involving large deletions or amplifications of genomic loci have been reported to be present within in clonal parasite populations [360, 388, 398]. However, polymorphisms in poly(dA:dT) tract lengths have not been reported for *P. falciparum*, this could due to sequencing difficulties over these regions. In addition, it was observed that tract length polymorphisms could ultimately lead to variations in transcriptional activity of up to 35% (Figure 4.5 and 4.6a) and thus may be a mechanism by which the parasite population varies its transcription. Furthermore, the variations in ffluc activity were observed in clones of 4 other promoters as well (Figure 4.6e).

Hence, it would be interesting to test if the polymorphisms in tract lengths were artifacts created during the PCR amplification process or it is due to a mixed population of tract lengths was present in the gDNA where PCR was performed. To test this, PCR amplification of the same 2 kb promoter region was done on 2 types of templates. The first template was the 3D7 parasite gDNA, the same as previously described in Figure 4.6a. The second template was a plasmid from one of the clones in Figure 4.6a. By using a plasmid templete for PCR amplification, it is ensured that amplification is from only a single template and if any variations resulted from the amplification, it would have been because of the PCR process. As the *P. falciparum* genome size is 23 Mbp and the plasmid is 7 kb, the template to be amplified is $3x10^3$ fold more in the plasmid per unit mass. Hence, 50 ng of gDNA and 10 pg of plasmid were used for PCR amplification to account for template copy number differences. After PCR amplification, the PCR products from both templates were gel purified, ligated into pf86 and subsequently transformed into DH5α competent cells. Plasmids were purified from 12 clones from each template and the 2 kb *pfl1885c* promoter region was sequenced. Variations from reference 3D7 sequence were recorded for each of the 10 tracts in all 12 clones from both templates was summarized in box plots (Figure 4.7). Greater variations could be seen in the clones that were generated from PCR products from gDNA and statistical significance was achieved for 6 out of the 10 poly(dA:dT) tracts. However, there were also variations observed in the tracts of the clones which were generated from PCR products of the single template. This was especially evident in tracts at position -1575 bp and -986 bp where the variations in the tract length in both gDNA and single template were almost equal. This could be because the adjacent tracts (position -1507 bp and -951 bp) were positioned very closely. In conclusion, PCR amplifications play a role in creating tract length polymorphisms. But, there could be polymorphisms existing in the

parasite population as PCR products from gDNA displayed greater variation than from the single template in some of the tracts.

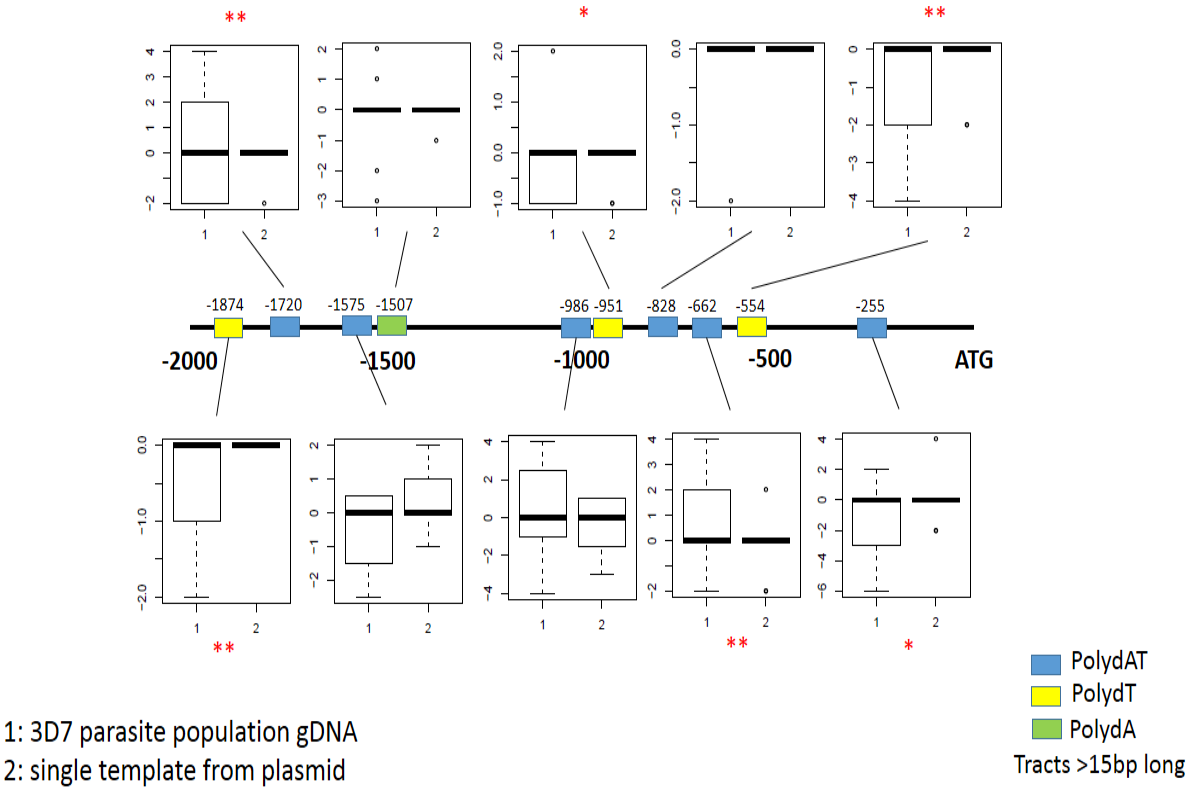


Figure 4.7 Comparison of tract length polymorphisms from PCR using 3D7 gDNA and single template plasmid.

Positions of poly(dA:dT) homopolymeric tracts length greater than 15 bp of various base compositions are presented on the 2 kb 5' upstream region of *pfl1885c*. PCR amplifications of 2 kb 5' upstream region of *pfl1885c* were done on using 3D7 gDNA and plasmid clone 11 (Figure 4.6a and b), PCR products were then restriction digested and cloned into the pf86 vector. Clones from the PCR products of the 2 templates were analyzed by sequencing of the 2 kb 5' upstream region of *pfl1885c* on the constructs. The tract length polymorphisms at each tract on every clone were recorded. Indels recorded for the clones from PCR products from the 3D7 gDNA template were in reference to the 3D7 reference sequence. Indels recorded for the clones from PCR products from plasmid template were in reference to the sequence of 2 kb 5' upstream region of *pfl1885c* on clone 11 (Figure 4.6b). The variations for each tract of the 2 kb 5' upstream region of *pfl1885c* on each clone are summarized in a box plot, 1: PCR products from 3D7 gDNA, 2: PCR products from plasmid. The extent of variations was compared between the clones generated from the PCR products of both templates by f-test. There is significantly greater variations observed on 6 tracts of 10 tracts (repeats >15bp) at positions -

255 bp, -554 bp, -662 bp, -951 bp, -1720 bp, -1874 bp (*p<0.05, **p<0.01) in clones generated from 3D7 gDNA compared clones generated from PCR from a single template plasmid.

4.5 Final screen

As seen from experiments in Figure 4.7, PCR amplifications create artifacts in tract length polymorphisms in the constructs created, and importantly, tract length polymorphisms has an effect on ffluc activity (Figure 4.6). Although PCR cannot be avoided completely, it would be beneficial to reduce the PCR steps when making constructs for the motif screening to avoid the complications in tract length polymorphisms. In the strategy devised in Figure 4.3, 2 rounds of PCR was used to create the mutant construct with the motif deleted (one round for making the WT construct, another round to amplify the mutant construct from the WT construct). To reduce the number of PCR steps, a new strategy was devised (Figure 4.8). The protocol for creating the wild type promoter construct is as per Figure 4.3. For the mutant promoter construct, 2 sets of primers were used to amplify 2 fragments of the 2 kb promoter. 5’ fragment amplified by primers 5’F and 5’R; 3’ fragment amplified by primers 3’F and 3’R (Figure 4.8). Motif of interest lies in between the 2 fragments and is not amplified. Primers 5’R and 3’F were designed such that the motif sequence is not included in both of the fragments. The 2 PCR fragments were phosphorylated, restriction digested and ligated into the pf86 vector. In this way, PCR amplification will be only be performed once for both the wild type and mutant promoter as both promoters will be amplified directly from gDNA (Figure 4.8). This was in contrast with the two-step approach in Figure 4.3 where the WT promoter was first amplified from gDNA and the mutant construct was obtained from amplification of the WT construct. Furthermore, the strategy reduced the length of PCR products which in turn reduced the number of mutations introduced. In addition, as polymorphisms on the tracts of the promoters cannot be avoided completely, the 2 kb promoter region cloned onto the pf86 construct was sequenced thoroughly such that only

wild type and mutant constructs with promoter sequences which match the 3D7 reference sequence completely were used for transfections and further analysis of the motifs.

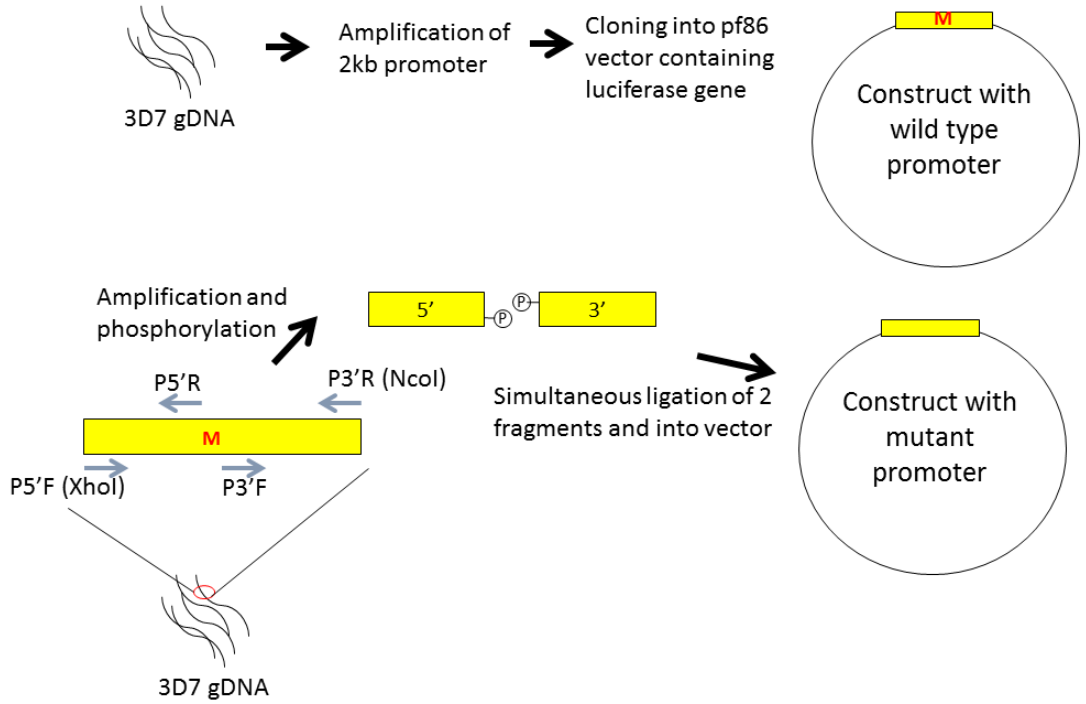


Figure 4.8 Final cloning strategy for motif screening.

In order to screen the predicted motifs in the context of its native promoter, a length of 2 kb upstream of the genes to be tested was cloned into the pf86 construct. The wild type promoter was cloned as it was described in Figure 4.3a. The motif of interest is represented by red "M" in the yellow box which represents the 2 kb promoter. For the construction of the mutant promoter with the motif deleted, primers were designed as represented by the blue arrows. The motif sequence was not included in the primers P5'R and P3'F. P5'F(XhoI) together with P5'R was used to amplify the 5' fragment of the 2 kb promoter. P3'F and P3'R(NcoI) was used to amplify the 3' fragment of the 2 kb promoter. The fragments were phosphorylated on their 5'OH groups, 5' fragment was digested using XhoI and 3' fragment was disgested using NcoI. The fragments were ligated simultaneously with the pf86 vector backbone. Constructs were sequenced thoroughly over the 2 kb upstream sequences and sequences were aligned to the 3D7 reference sequence. Only clones that had the exact sequence as the 3D7 reference genome were used in subsequent transfections.

This approach was again first tested on G-box motif in the *pfhsp86* promoter. 4 different constructs were created. The first construct was the wild type *pfhsp86* promoter (Figure 4.9, *hsp86* promoter WT). Second, the *pfhsp86* promoter with the first half of the G-box motif deleted (Figure 4.9, *hsp86* promoter G-box del). Third and fourth, deletions of 1 and 6 base deletion from 6 bases upstream of the G-box motif to serve as negative controls (Figure 4.9, *hsp86* promoter 1 base del and 6 base del respectively). As expected, deletion of the G-box motif resulted in a decrease in activity by more than 50% (Figure 4.9). In addition, the negative controls where random bases were deleted did not result in any changes in activity (Figure 4.9). This was a good indication of the strategy.

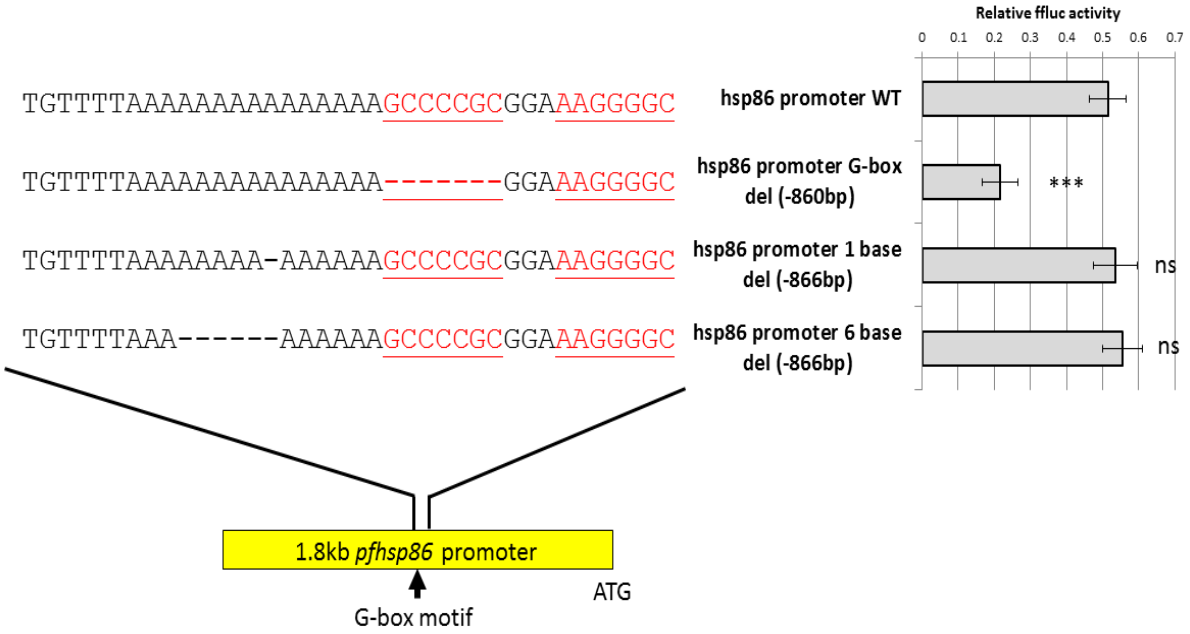


Figure 4.9 Screening strategy tested on G-box motif in pfhsp86 promoter.

For all 4 constructs, the entire 1.8 kb upstream of the *pfhsp86* gene was included in the constructs except for base deletions in the vicinty of the G-box motif (-860 bp with respect to the translational start site). *Hsp86* promoter WT contains the wild type 1.8 kb upstream of the *pfhsp86* gene with no alterations on the promoter. *hsp86* promoter G-box del contains the 1.8 kb upstream of the *pfhsp86* gene with the first half of the G-box motif deleted. hsp86 promoter 1 base del construct contains the 1.8 kb upstream of the *pfhsp86* gene with 1 base deleted at position -866 bp. hsp86 promoter 6 base del construct contains the 1.8 kb upstream of the

pfhsp86 **gene with 6 bases deleted at position -866 bp. Transfections and assays were done as described in Chapter 2. Deletion of G-box motif resulted in decrease in ffluc activity. 1 base and 6 base random deletions did not have any significant effect in ffluc activity. Statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days.**

This cloning strategy was then used to clone wild type and mutant promoters of genes listed in Figure 4.1 for the screening of the respective motifs. The normalized ffluc activity results are summarized in Figure 4.10a. The motif numbers are listed on the y-axis, the numbers correspond to the motifs numbered in Figure 4.1. 2 promoters containing motifs 6, 10, 11 and 21 had been screened so the results are represented as "a" and "b" for the 2 promoters respectively. Renilla activity normalized ffluc activity driven by the mutant promoter construct where the motif was expressed as a ratio over the renilla activity normalized ffluc activity driven by the wild type construct. After which, the ratio was expressed on a log scale on Figure 4.10a. If mutant construct drives higher activity than the wild type construct, the log ratio would be greater than 1. If mutant construct drives lower activity than the wild type construct, the log ratio would be less than 1. If both activities are equal, the log ratio would be approximately 1.

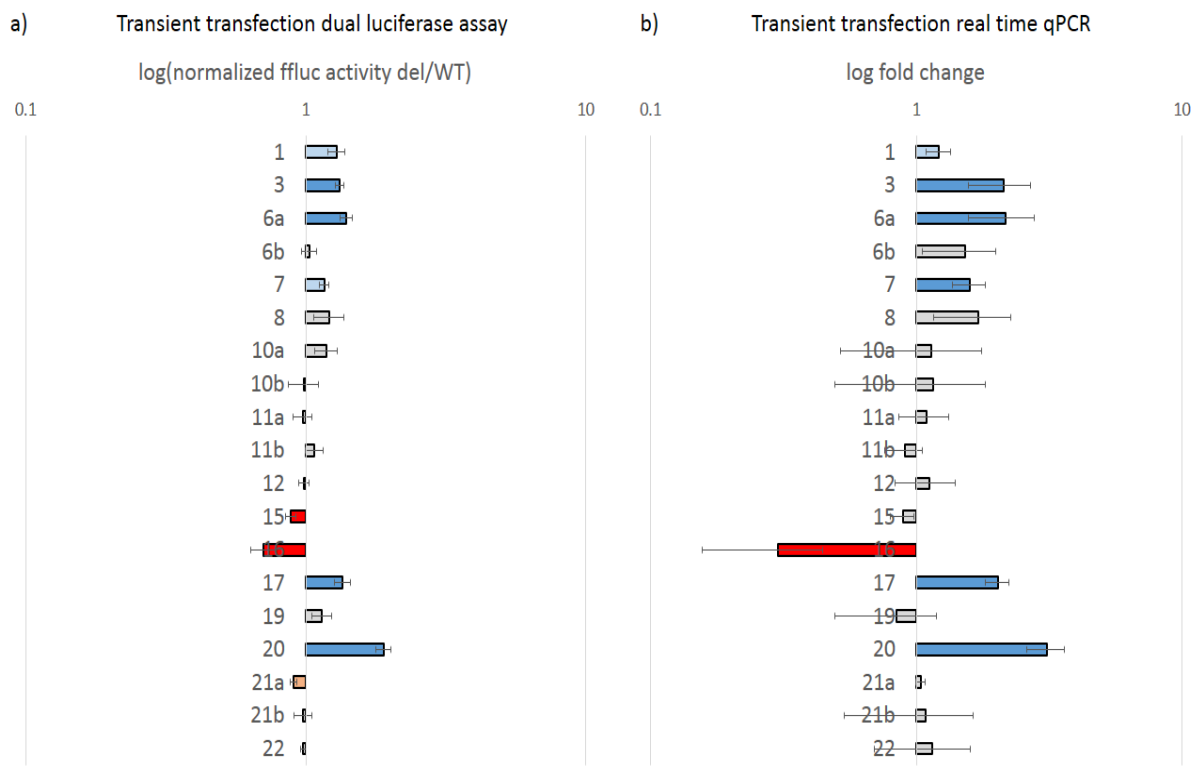


Figure 4.10 Screening of motifs predicted by FIRE.

The numbers represent the motif number listed in Figure 4.1. There were 2 promoters tested for motifs 6, 10, 11 and 21, they are represented as "a" and "b". Constructs were created as described in Figure 4.8. Transient transfections were carried out as described in Chapter 2, 50 µg of pf86 (containing promoter of interest and *ffluc* gene) was co-transfected with pARL-5'3'-Re (containing *renilla luciferase* gene). After transfection, half of the parasites were lysed for dual luciferase assays (renilla luciferase and ffluc activity measured). Total RNA and gDNA were harvested from the other half of the sample. The data is presented on a log scale graph. Student's t-test was used to define whether the activity or relative ffluc transcripts from the mutant constructs were statistically different from the wild type constructs. Grey bars, statistically insignificant. Dark red/blue, $p<0.01$. Light red/blue, $p<0.05$. Error bars represent 3 independent transfections done on 3 separate days. a) Dual luciferase assays. Ffluc activity was normalized by renilla luciferase activity for each transfection/assay. Normalized ffluc activity driven by the mutant promoter construct was expressed as a ratio over the normalized ffluc activity driven by its respective wild type promoter. The ratio for each motif is presented on a log scale graph. Deletion of motifs 1, 3, 6, 7, 17 and 20 increased ffluc activity driven significantly. Deletion of motif 15, 16 and 21 decreased ffluc activity driven significantly. b) Real time quantitative PCR. Reverse transcription PCR was done for the total RNA harvested using primers specific for *ffluc* gene and *pfl0900c* as a reference gene. Real time qPCR was then carried out to obtain a Ct value for the *ffluc* gene and *pfl0900c* which is representative of quantity of mRNA transcribed from wild type and mutant promoter constructs. The Ct for *ffluc*

gene in each sample was normalized with its corresponding Ct for *pfl0900c* ($Ct_{(norm)} = Ct_{(ffluc)} - Ct_{(PFL0900c)}$). The fold change of transcripts obtained from mutant constructs relative to the transcripts obtained from wild type constructs was calculated by the ΔΔCt method, Fold change = $2^{-(Ct(mutant\ norm) - Ct(wild\ type\ norm))}$. Using a similar method, relative copy numbers of the *ffluc* gene were calculated from real time qPCR of gDNA isolated from each sample. The transcript fold change is normalized using the relative copy numbers of the ffluc gene. The fold change is presented on a log scale graph. Deletion of motifs 1, 3, 6, 7, 17 and 20 increased ffluc transcripts expressed significantly. Only deletion of motif 16 decreased transcripts expressed significantly.

Overall, 9 motifs out of the 16 motifs screened displayed an effect which was statistically significant when deleted from the promoters of the genes used in the screening (Figure 4.10a). Six motifs (motifs 1, 3, 6, 7, 17 and 20) showed an increased level of ffluc activity when deleted from their respective promoters (Figure 4.10a). This indicated that these motifs played a role in repressing transcriptional activity. On the other hand 3 motifs (motifs 15, 16 and 21) showed a decrease in activity when deleted from their respective promoters (Figure 4.10a). This is indicative that these motifs played a role in inducing transcriptional activity. In addition, 3 other motifs (motif 8, 10 and 19) showed a slight increase in activity when the motif is deleted but did not achieve statistical significance (Figure 4.10a). Deleting motif 10, 11, 12 and 22 did not have any effect in ffluc activity from our assays (Figure 4.10a). This could be mean that these motifs have no effect on transcription or it was out of the detection limit of the transient transfection assays.

There were 2 promoters that were utilized to represent motif 6: *pff1500c* and *pfl2275c.* mRNA abundance for *pff1500c* peaks during the ring stage and *pfl2275c* peaks during the trophozoite stage (Supplementary figure 2). Interestingly, deleting motif 6 from *pff1500c* promoter region resulted in increased activity while deleting the same motif from *pfl2275c* did not have any effect (Figure 4.10a). This difference may be explained by the difference in the stage by which the 2 genes are abundant in. Also, 2 promoters were tested for motif 21.

Motif 21 is found to be inducing for *pff1375c* but no effect on *pfl2520w* (Figure 4.10a). Although both genes are highly abundant in late schizont/ early ring stages and the same motif is predicted for both genes, the exact motif sequence is highly degenerate among the 2 (Figure 4.1b and Supplementary Figure 2). Hence, a difference in the motif sequence may have caused a difference in the effects elicited. Motif 10 and 11 did not have any effect when deleted from the 2 promoters containing each motif, although deletion of motif 10 increased activity of the promoter of *pfi1375w* slightly but was not statistically significant (Figure 4.10a).

In order to validate the results from the luciferase activity assays, mRNA and gDNA of parasites that were transfected for the dual luciferase assays were harvested alongside. cDNA for *ffluc* and reference gene (*pfl0900c*) were synthesized and the quantity of transcripts were analyzed by qPCR. The normalized quantity of *ffluc* transcripts from parasites transfected with the mutant promoter construct were expressed as a fold change relative to the normalized quantity of *ffluc* transcripts from parasites transfected with the wild type promoter construct. Fold change was again normalized to the relative plasmid copy numbers. Relative plasmid copy numbers was found to be approximately equal between wild type and mutant promoter constructs (Supplementary Figure 1). Results from the qPCR were similar to the dual luciferase assays with the exception of motif 15 and 21 having no effect when tested using the qPCR method but were shown to be inducing using the dual luciferase assays (Figure 4.10b).

4.6 Conclusion and discussion

In this chapter, a strategy for quick screening of *cis*-regulatory motifs in *P. falciparum* had been optimized (Figure 4.8). In addition, we have preliminarily identified 6 motifs which play

a repressive role in transcription and 3 motifs which play an inducing role (Figure 4.10). Of which, motifs 3, 6, 16, 17 and 20 were selected for further characterizations as they were shown to have a more significant effect when deleted from their respective promoters in both dual luciferase and qPCR assays (Figure 4.10, $p<0.01$). The methods used will be discussed here but the full discussion on the implications of results in the context of transcription in *P. falciparum* will be discussed together in Chapter 6.

It is inconclusive for the rest of the motifs as they could be playing a role but not detectable in our assays (Figure 4.10). A limitation could be the detection limit of our assays, some of the predicted motifs could be playing subtle but combinatorial roles. In our assays, motifs are individually deleted and it is possible that some subtle changes in activity/transcript level were not detected. The next limitation is the choice of gene promoters. It could be that the motif is not playing a role on the promoter that was chosen but plays a role in other promoters. A third possibility is that the motif plays a stage specific role. As seen in Chapter 3, transient transfection assays are not suitable for elucidating stage specific effects. The stage specific effects could have been confounded by other factors due the nature of the assay such as rapid loss of plasmids and thus left undetected. The final caveat is the bias of the FIRE algorithm, the algorithm may have biases for certain motifs and other motifs which play a role in transcription may not be detected by FIRE. However, this strategy is still a quick way to identify high value targets for further and more thorough investigations.

The investigations on the association of tract length polymorphisms and transcription presented in this chapter are valuable in both methodological standpoint and also its consequence in parasite transcription *in vivo*. Here, we have shown that poly(dA:dT) tract length polymorphisms cause a variation of 35-40% in transcriptional activity (Figure 4.6a).

The tract length polymorphisms may be caused by both polymerase slippages during PCR and variations in tract length existing in 3D7 parasite mixed population (Figure 4.7).

Slippage during PCR has been reported to introduce errors for repetitive sequences and this occurs for many polymerases including proofreading polymerases [381, 399-401]. Slipped-strand mispairing has been proposed to explain on how indels occur in the PCR amplification (Figure 4.11) [379]. In the PCR process, extension temperature goes up to 72^{o}C for optimal activity of most polymerases. At high temperatures, the hydrogen bonds between the nascent strand and template strands are unstable causing both strands to disassociate occasionally. This is especially true for AT-rich repeats as there are only 2 hydrogen bonds between Adenine and Thymine nucleotide pairs in contrast to 3 hydrogen bonds in Guanine and Cytosine pairs. Depending on how the stands re-anneal back onto each other, nucleotide(s) can be inserted or deleted to the repeats (Figure 4.11).

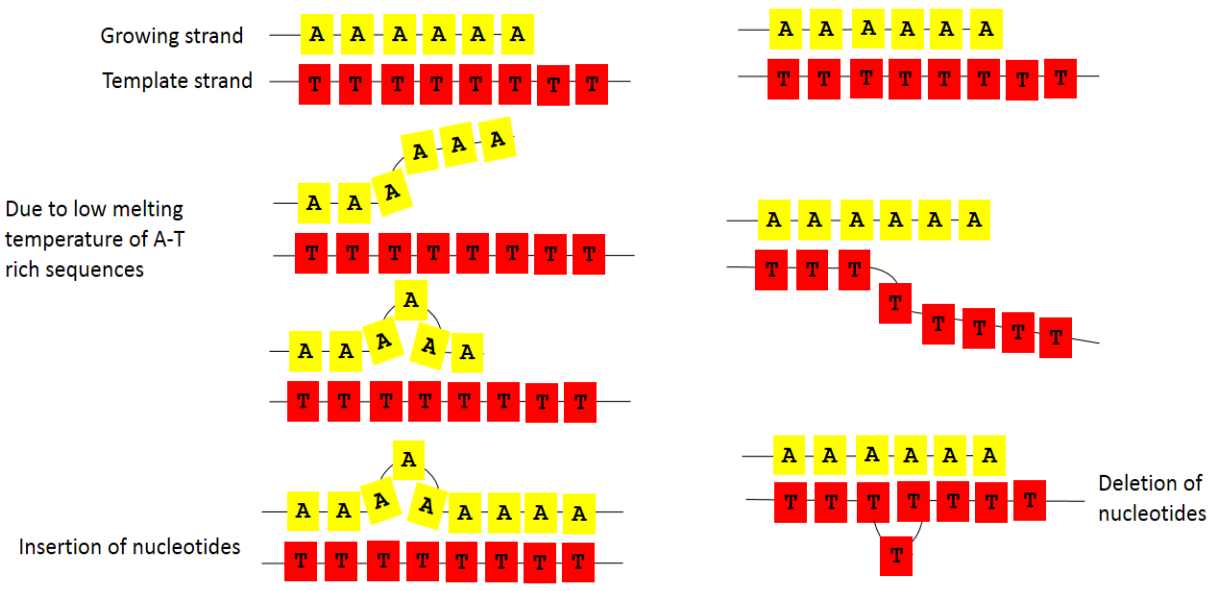


Figure 4.11 Slipped strand mispairing causing indels in repetitive seqeunces.
During the extension phase of the PCR process, samples are heated to approximately 72°C for optimal activity of the polymerase. This poses instability to the hydrogen bonds between the nascent and template strands. This is even more so in A-T pairs as there are only 2 hydrogen

bonds in contrast to 3 hydrogen bonds in G-C pairs. As such, poly(dA:dT) tracts with long stretches of A and T nucleotides are particularly prone to slipped strand mispairing. The instability causes the nascent strand to dissociate from the template strand. When the temperature lowers during the PCR cycle, the 2 strands may anneal back together but may not anneal accurately. If the nascent strand folds back on itself during the process of reannealing (right panel) it will cause an insertion of nucleotides. On the other hand, if the template strand folds back, deletion of nucleotides will occur. Number of nucleotides inserted or deleted will vary according to the number of nucleotides folded back upon.

As these poly(dA:dT) tracts or repeats are aplenty in the *P. falciparum* (especially in intergenic regions), variations to the PCR protocol have been suggested. Thermal instability of Adenine-Thymine pairs has not just caused inaccuracies in PCR amplification by slippages but has been a hindrance to PCR amplification in *P. falciparum* in general. It has been shown that reducing extension temperatures from 72^{o}C to 60^{o}C has allowed amplification of large fragments of AT-rich sequences [402] and also increased coverage in *P. falciparum* whole genome amplifications for NGS [105, 403]. Reduction of extension temperatures reduces the chances of Adenine-Thymine pairs from dissociating from each other, thus increasing the efficiency of PCR. In the same way, reducing temperatures in PCR may reduce occurrences of dissociation of nascent and template strands, resulting in reduced slippages. However, this could mean that the temperature for polymerase extension may be suboptimal, leading to reduction in PCR yield. Hence, polymerase extension efficiency needs to be taken into account when optimizing temperatures as well.

Addition of small amounts of tetramethylammonium chloride (TMAC) has been shown increase thermal stability of AT-rich sequences and has increased yield of PCR amplifications from AT-rich sequences [404]. Increased thermal stability leads to decreased dissociation of nascent and template strands and thus, less occurrence of slippages. This has been shown to be effective in *P. falciparum* genome as well [105]. However, TMAC inhibits

the efficiency of some polymerases [105]. Hence, the amount of TMAC needs to be balanced as well. In addition, usage of highly processive polymerases has also improved amplifications of AT-rich sequences [105, 405]. New generation polymerases have DNA-binding domains attached which increases the processivity of the polymerase may reduce the occurrences of strand displacement.

In summary, these optimizations reduced the chances of displacement of the nascent and template strand during PCR and thus improved the overall amplification of the *P. falciparum.* However, there have not been reports of whether these optimizations improved the extent of slippages during amplification. In the amplifications done in the experiments in this thesis, extension temperatures were at 60°C, primer annealing temperatures of 50°C for most primer sets, 60 mM of TMAC was added and a highly processive polymerase (Pfu Ultra II fusion, Stratagene) was used. We have shown that although we were able to amplify most of the promoters that were used in our motif screening study, slipped-strand mispairing still occurred during PCR (Figure 4.7). In addition, we also provided evidence that tract length polymorphisms cause variations in transcriptional activity of a promoter region (Figure 4.6a). Hence, constructs used in promoter analysis assays needs to be sequenced thoroughly for meaningful interpretation of data.

Since the publication of the genome, efforts have been made to characterize genetic diversity in *P. falciparum* in search of drug targets and vaccine candidates. One study had conducted genome-wide sequencing of 16 parasites which are geographically diverse including isolates from South America, Africa and Asia [406]. The authors have identified 46,937 SNPs and have found that more SNPs associate with cell surface molecules which mediate antigenic variation while genes encoding for metabolic enzymes have almost no SNPs. By comparing

sequences of drug resistant and drug sensitive strains, authors also reported of genomic loci which may associate with chloroquine and pyrimethamine resistance [406]. In another study, authors surveyed 3539 genes (approximately 65% of predicted genes) from 4 isolates and have identified 3918 SNPs [407]. As antigenic genes are under selection by the host immune system, they tend to accumulate more SNPs. By focusing on highly polymorphic genomic loci, authors have identified previously functionally unknown genes as potential antigenic genes which may be novel vaccine candidates [407]. In a more recent study, 4 geographically diverse isolates of *P. vivax*, another parasite of the *Plasmodium* genus, were sequenced and authors reported that SNPs occur more frequently in *P. vivax* in comparison to *P. falciparum* [17]. *P. falciparum* clone 3D7 was obtained from NF54 parasite strain by limiting dilution methods and NF54 was in turn isolated from a patient in the Netherlands [408, 409]. Even though 3D7 has been originally isolated from one patient from a single geographical location, the culture adapted 3D7 parasite has been cultured in *in vitro* for multiple cycles since 1987 [408]. It would not be surprising to find variations within *in vitro* cultures of 3D7 and two recent publications have shown that prolong culture of *in vitro* parasites can lead to accumulation of mutations and step-wise gain of resistance to antimalarial drugs [388, 398]. It is likely that variations are even more frequent in low complexity sequences in the genome as they are prone to errors made by DNA polymerase during replication [380]. Studies in low complexity sequences in *P. falciparum* have largely focused on limited loci in the coding regions, authors have found that they are regions of recombination hotspots and have indicated of slippages during replication [385, 386]. However, variations in low complexity sequences such as long tracts of AT-rich sequences which are aplenty in *P. falciparum* intergenic regions are not well studied, possibly due to technical difficulties. It would be interesting to study these tracts and how they affect gene expression the parasite.

5. Further characterizations on novel motifs 3, 6, 16, 17 and 20

5.1 Background and introduction

Motifs predicted by FIRE have been screened as described in Chapter 4. Motifs 3, 6, 16, 17 and 20 out of the 15 motifs that were screened have been shown to have the most significant effects on promoter activity in our assays (Figure 4.10). Although it is still inconclusive as to whether the rest of the motifs play a role in transcription as they may have been undetected in our assays, these 5 motifs are valuable targets as they have achieved high levels of statistical significance in both dual luciferase assays and real time qPCR assays (Figure 4.10). Hence, further validations and investigations will be carried out on these motifs.

First, site directed mutagenesis on the motifs of interest will be carried out. These experiments would be able to validate the deletion assays to rule out the possibility that the deletion of the motif caused an unspecific effect by creating another motif in the process of deletion or any other unspecific effect due to our methodology. On the other hand, these experiments will also identify important nucleotides in the motif sequence.

Next, a motif would likely play a role in multiple promoters as it would regulate the transcription of multiple genes. Although FIRE has predicted the presence of these motifs on multiple promoters, but it has not been shown whether the motifs play a role in the promoters other than the one that was selected for the screening (Figure 4.10). Hence, it is important to elucidate whether these motifs play a role in other promoters to substantiate its role in transcription.

A key limitation of transient transfection assays is its inability to characterize stage specific effects of the promoter/motifs. Stable transfections would allow a reporter gene driven by a promoter of interest (with or without the motif) to be stably expressed in parasites. An excellent platform would be the Bxb1 mycobacteriophage integrase mediated integration of reporter gene into the parasite genome [410].

DNA motifs regulate gene transcription through the binding of specific DNA-binding protein elements which are likely to be specific transcription factors (STFs). A final characterization of the motifs would be to detect protein binding activity mediated by the motifs.

5.2 Site directed mutagenesis

Constructs used for the following site directed mutagenesis experiments were made as described in Figure 4.8. Primers were designed accordingly to make alterations to the motif of interest in the 2 kb promoter region. Constructs were tested using the same transient transfection and dual luciferase assay protocol as in the screening of the motifs (Figure 4.10).

Motif 3 is overrepresented in clusters containing genes of transcripts which peak in early and late ring stage, as well as late schizont stage genes (Figure 5.1a). On the other hand, it is underrepresented in genes which are abundant in the trophozoite and early schizont stages (Figure 5.1a). The motif predicted “ATATAGA” is one of the least degenerate motifs in the list of motifs predicted. The promoter that was used in the initial screening is the promoter of the protein kinase 2 gene (*pfl1885c*). The motif lies at -793 bp relative to the translational start site of the *pfl1885c* coding sequence (Figure 5.1b). As a negative control, when a random 7-nucleotide sequence was added adjacent 5’ to the predicted motif, there was no change in activity driven (Figure 5.1c, Mut1). When the motif is deleted, there was an almost

50% increase in activity, similar to the results seen in experiments done in figure 4.10 (Figure 5.1). In addition, when the 7-nucleotide motif was replaced with the random 7-nucleotide sequence, an increase in activity to a similar extent as when the motif was deleted was observed (Figure 5.1, Mut2). Finally, to test if the nucleotide "C" is important in the motif, when it was mutated into "A", an intermediate increase in activity of about 25% was observed (Figure 5.1, Mut3).

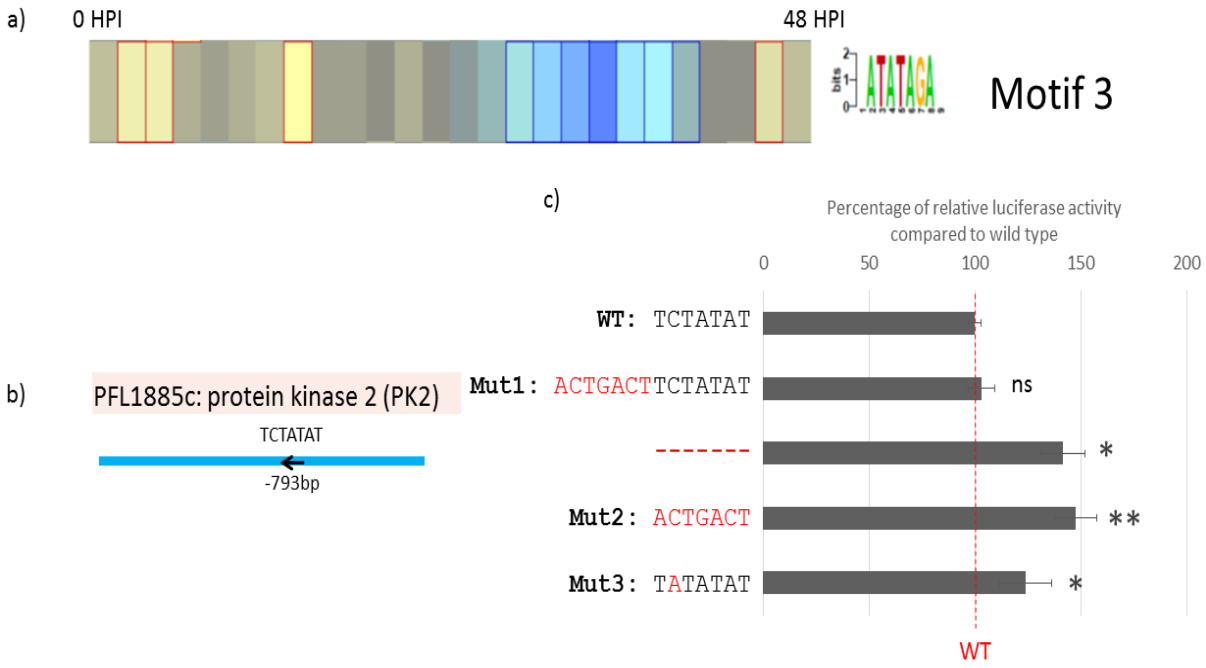


Figure 5.1 Motif 3 mutagenesis.

a) FIRE output for motif 3. Motif 3 is overrepresented in the 1 kb upstream regions of genes which mRNA abundance peaks during the late schizont and ring stages. It is also underrepresented in clusters of genes with maximal mRNA abundance during the schizont stages. b) Schematic diagram of 2 kb upstream of *pfl1885c*. *pfl1885c* is one of the genes which are in the overrepresented clusters. The 2 kb upstream regions of *pfl1885c* was used in the in initial screening of the motif 3. Motif 3 is located at -793 bp with respect to the translational start site. c) Site directed mutagenesis of motif 3 in the 2 kb upstream regions of *pfl1885c*. Constructs were made as described in Figure 4.8. Transient transfections and dual luciferase assays were done as described in Chapter 2. Normalized ffluc activity of each mutant promoter construct is presented in a percentage relative to the activty driven by the wild type promoter of *pfl1885c*. The wild type sequence of the motif is as stated (WT). As a negative control, a random

sequence which is used to mutant the motif was placed adjacent to the motif (mut1). The actvity driven was the same as the wild type promoter construct. As reported in Figure 4.10, deletion of motif 3 resulted in increased ffluc activity driven by about 50%. Scrambled sequence mutation of motif 3 resulted in increased ffluc activity driven to a similar level as when the motif was deleted (mut2). Mutation of "C" to "A" resulted in approximately 25% increase in activity (Mut3). Statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days.

Motif 6 overrepresented in clusters containing genes abundant during the ring and trophozoite stages (Figure 5.2a). However, it is underrepresented in clusters containing genes which transcripts are abundant during the schizont stages (Figure 5.2a). The promoter that was used in the initial screening is the 2 kb putative promoter region of the DEAD/DEAH box ATP-dependent RNA helicase gene (*pff1500c*). The motif lies at -315 bp relative to the translational start site of the *pff1500c* coding sequence (Figure 5.2b). Similarly, the negative control containing random nucleotides adjacent to the motif sequence did not have any effect on ffluc activity (Figure 5.2c, mut1). Deletion of the motif and scrambling the entire motif sequence increased activity of the promoter by about 35%, similar to what was observed in Figure 4.10a (Figure 5.2c, mut2). Mutation of the 2 guanine bases in the motif increased the activity driven by the promoter by 35%, a level similar to activity driven when the motif was deleted or sequence scrambled (Figure 5.2c, mut3). In addition, addition of 1 and 2 more copies of the motif decreased the promoter activity with an additive effect (Figure 5.2c, mut2x and mut3x). This further substantiates that this motif plays a repressive role in transcription.

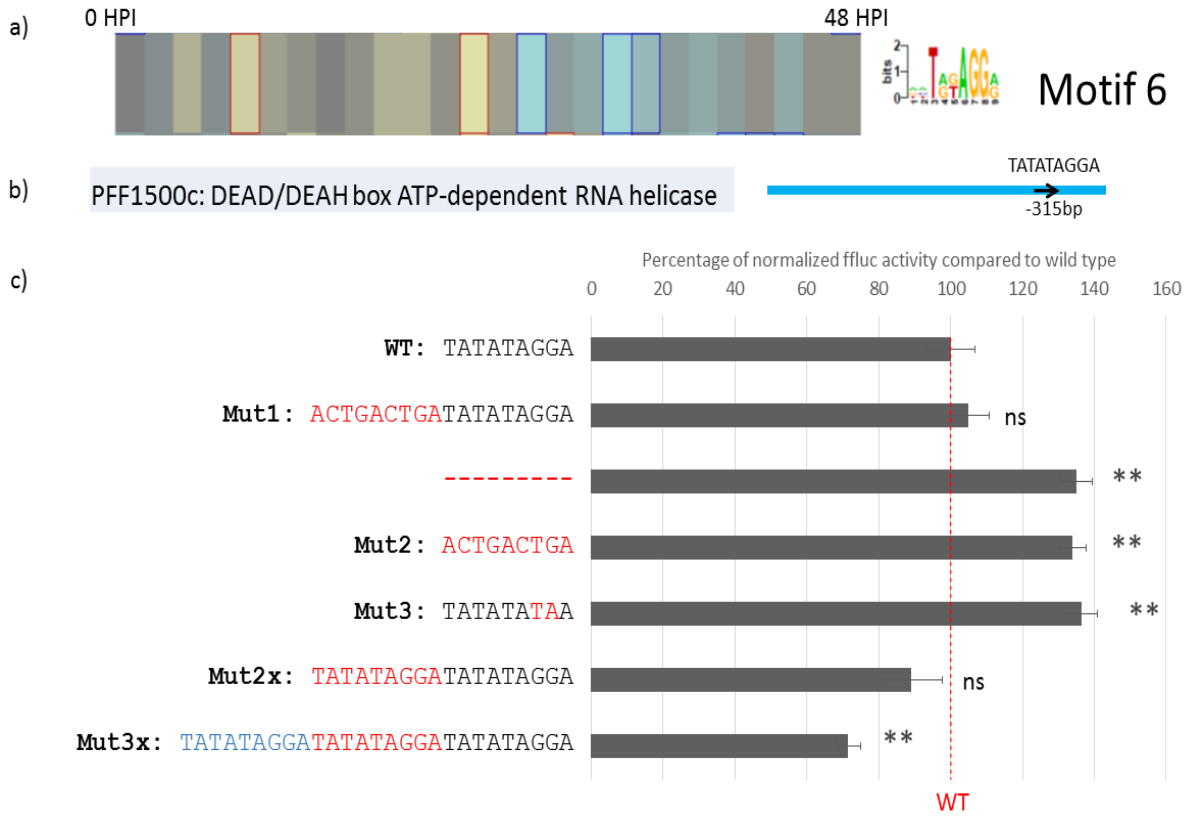


Figure 5.2 Motif 6 mutagenesis.

a) FIRE output for motif 6. Motif 6 is overrepresented in the 1 kb upstream regions of genes which mRNA abundance peaks during the mid to late ring stages. It is also underrepresented in clusters of genes with maximal mRNA abundance during the schizont stages. b) Schematic diagram of 2 kb upstream of *pff1500c*. *pff1500c* is one of the genes which are in the overrepresented clusters. The 2 kb upstream regions of *pff1500c* was used in the initial screening of the motifs. Motif 6 is located at -315 bp with respect to the translational start site. c) Site directed mutagenesis of motif 6 in the 2 kb upstream regions of *pff1500c*. Constructs were made as described in Figure 4.8. Transient transfections and dual luciferase assays were done as described in Chapter 2. Normalized ffluc activity of each mutant promoter construct is presented in a percentage relative to the activty driven by the wild type promoter of *pff1500c*. The wild type sequence of the motif is as stated (WT). As a negative control, a random sequence which is used to mutant the motif was placed adjacent to the motif (mut1). The actvity driven was the same as the wild type promoter construct. As reported in Figure 4.10, deletion of motif 6 resulted in increased ffluc activity driven by about 40%. Scrambled sequence mutation of motif 6 resulted in increased ffluc activity driven to a similar level as when the motif was deleted (mut2). Mutation of "GG" to "TA" resulted in approximately 40% increase in activity as well (mut3). When 1 additional copy of the motif was included in the promoter sequence, ffluc activity was reduced but was statistically insignificant (mut2x). When 2 copies of the motif was included, the activity driven was significantly reduced by about 30% (mut3x). Statistical test

used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days.

Motif 16 is a motif which induces transcriptional activity (Figure 4.10). It has been shown that deletion of the motif causes a reduction in the transcriptional activity of the promoter (Figure 4.10). Motif 16 is overrepresented in the clusters containing genes which mRNA are more abundant during the schizont and early ring stages (Figure 5.3a). In contrast, it is underrepresented in the genes which are abundant during the late ring and trophozoite stages (Figure 5.3a). The promoter that was utilized in the initial screening is the 2 kb putative promoter region of the acyl-CoA synthetase gene (*pf14_0761*). The motif lies at -764 bp relative to the translational start site of the *pf14_0761* coding sequence (Figure 5.3b). Negative control, where a scrambled sequence was included adjacent of motif 16, did not show induce any significant difference in activity when compared to the wild type promoter (Figure 5.3c, mut1). Deletion of motif and scrambling of sequence in the motif reduced activity by almost 50% (Figure 5.3c, mut2). Nucleotides "TGC" are repeated in the motif "TGCATGCT", hence to test the importance of these 3 nucleotides in this motif, they were mutated in 2 different constructs (Figure 5.3c, mut3 and mut4). Both constructs resulted in a decrease in approximately 50% of activity compared to the wild type construct. Mutation of the first "T" nucleotide in the motif sequence did not affect the activity of the promoter significantly (Figure 5.3c, mut5). The inducing role played by motif 16 is further seen when stepwise increases in activity was observed when 1 and 2 additional copies of the motif were added into the promoter sequence (Figure 5.3c, mut2x and mut3x).

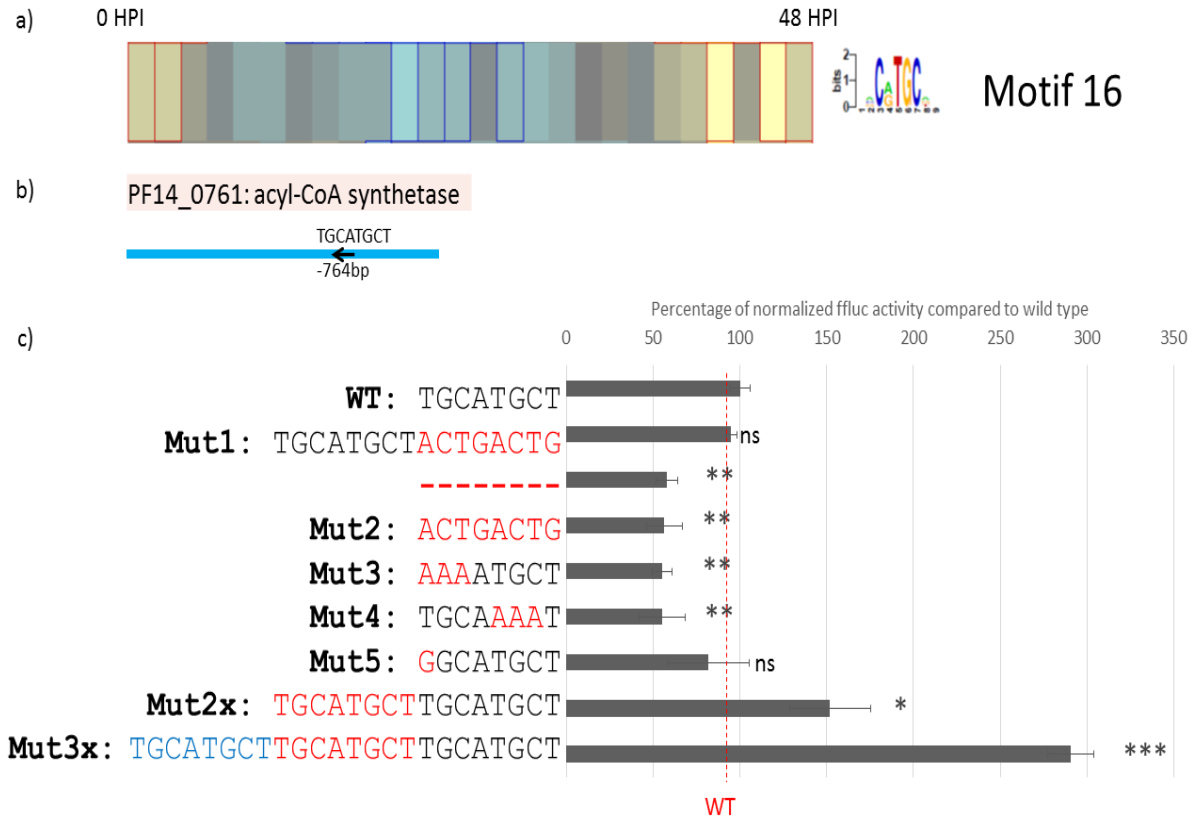


Figure 5.3 Motif 16 mutagenesis.

a) FIRE output for motif 16. Motif 16 is overrepresented in the 1 kb upstream regions of genes which mRNA abundance peaks during the schizont and early ring stages. It is also underrepresented in clusters of genes with maximal mRNA abundance during the late ring and trophozoite stages. b) Schematic diagram of 2 kb upstream of *pf14_0761*. *pf14_0761* is one of the genes which are in the overrepresented clusters. The 2 kb upstream regions of *pf14_0761* was used in the in initial screening of the motifs. Motif 16 is located at -764bp with respect to the translational start site. c) Site directed mutagenesis of motif 16 in the 2 kb upstream regions of *pf14_0761*. Constructs were made as described in Figure 4.8. Transient transfections and dual luciferase assays were done as described in Chapter 2. Normalized ffluc activity of each mutant promoter construct is presented in a percentage relative to the activty driven by the wild type promoter of *pf14_0761*. The wild type sequence of the motif is as stated (WT). As a negative control, the random sequence which is used to mutant the motif was placed adjacent to the motif. The actvity driven was the same as the wild type promoter construct (mut1). As reported in Figure 4.8, deletion of motif 16 resulted in decreased ffluc activity driven by about 40%. Scrambled sequence mutation of motif 16 resulted in decreased ffluc activity driven to a similar level as when the motif was deleted (mut2). Mutation of "TGCATGCT" to "AAAATGCT" resulted in approximately 40% decrease in activity as well (mut3). Mutation of "TGCATGCT" to "TGCAAAAT" resulted in decrease in activity by 40% (mut4). Mutation of "TGCATGCT" to "GGCATGCT" resulted in a slight decrease in activity but was statistically insignificant

(mut5). When 1 additional copy of the motif was included in the promoter sequence, ffluc activity was increased by approximately 50% (mut 2x). When 2 copies of the motif was included, the activity driven was increased by almost 200% (mut3x). Statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days.

Motif 17 was shown to have a repressive effect (Figure 4.10). Motif 17 is overrepresented in the clusters containing genes which mRNA are more abundant during the schizont stages (Figure 5.4a). In contrast, it is underrepresented in the genes which are abundant during the late ring and trophozoite stages (Figure 5.4a). The promoter that was used in the initial screening is the putative promoter region of the succinate dehydrogenase subunit 4, putative gene (*pf10_0100*). The motif lies at -843 bp relative to the translational start site of the coding sequence of *pf10_0100* (Figure 5.4b). As reported in the previous chapter, deletion of the motif resulted in an increase in ffluc activity driven by approximately 50% (Figure 4.10). When motif 17 was replaced by a scrambled sequence, a similar increase in ffluc activity driven as when the motif was deleted was observed (Figure 5.4c, mut1). However, in contrast to the previously described motifs, addition of multiple copies of the motif 17 did not result in any effect on the activity driven (Figure 5.4c, mut2x and mut3x).

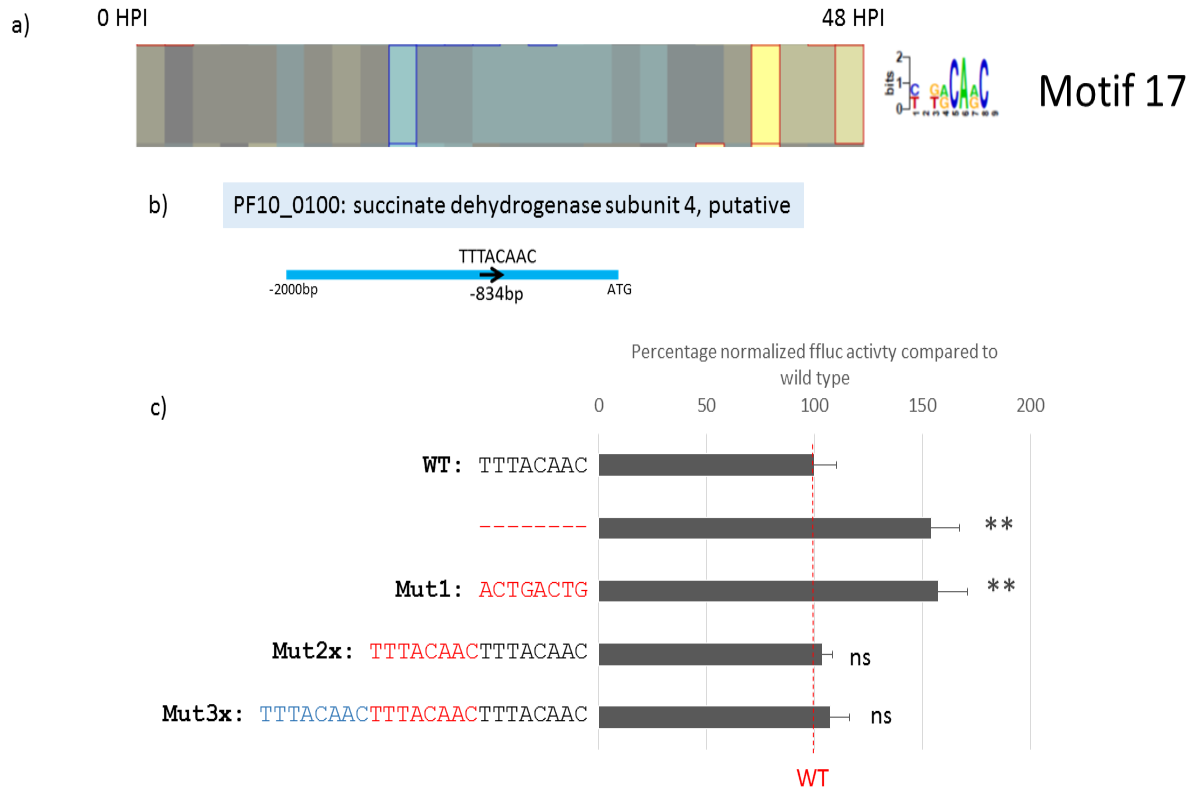


Figure 5.4 Motif 17 mutagenesis.

a) FIRE output for motif 17. Motif 17 is overrepresented in the 1 kb upstream regions of genes which mRNA abundance peaks during the schizont and early ring stages. It is also underrepresented in clusters of genes with maximal mRNA abundance during the late ring and trophozoite stages. b) Schematic diagram of 2 kb upstream of *pf10_0100*. *pf10_0100* is one of the genes which are in the overrepresented clusters. The 2 kb upstream regions of *pf10_0100* was used in the in initial screening of the motifs. Motif 17 is located at -834 bp with respect to the translational start site. c) Site directed mutagenesis of motif 17 in the 2 kb upstream regions of *pf10_0100*. Constructs were made as described in Figure 4.8. Transient transfections and dual luciferase assays were done as described in Chapter 2. Normalized ffluc activity of each mutant promoter construct is presented in a percentage relative to the activty driven by the wild type promoter of *pf10_0100*. The wild type sequence of the motif is as stated (WT). Deletion of motif 17 resulted in decreased ffluc activity driven by about 50%. Scrambled sequence mutation of motif 17 resulted in increased ffluc activity driven to a similar level as when the motif was deleted (mut1). When 1 and 2 additional copies of the motif was included in the promoter sequence, ffluc activity was unaffected (mut2x and mut3x). Statistical test used was the 2-tailed Student's T-test. ns, $p>0.05$, *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$. The error bars represent data from 3 independent transfections done on 3 separate days.

Motif 20 had the strongest repressive effect observed from the screening assays (Figure 4.10). Motif 20 is overrepresented in the clusters containing genes which mRNA is more abundant during the schizont stages (Figure 5.5a). In contrast, it is underrepresented in the genes which are abundant during the ring and trophozoite stages (Figure 5.5a). The promoter that was used in the initial screening is the putative promoter region of the DNAJ protein, putative gene (*pff1010c*). The motif lies at -111 bp relative to the translational start site in the *pff1010c* coding sequence (Figure 5.5b). Deletion and scrambled sequence of the motif resulted in more than 100% increase in activity (Figure 5.5c, mut1). Mutation of "AAAA" of the motif into "CCCC" resulted in a modest increase in activity driven of about 40% (Figure 5.5c, mut2). However, the mutation of "CATG" to "CCCC" in the motif resulted in increase in activity of about 90% (Figure 5.5c, mut3). Hence, mutational analysis of each of the nucleotides in "CATG" was carried out. Individual mutation of "A", "T" and "G" resulted in increases in levels of activity that was similar to when "CATG" was mutated together (Figure 5.5c, mut5, 6, 7). However when "C" was mutated individually, it did not result in any changes in activity driven when compared to the wild type promoter (Figure 5.5c, mut4). When additional copies of motif 20 were added to the promoter sequence, it resulted in additive decreases in activity when compared to the wild type promoter (Figure 5.5c, mut2x and mut3x). This provides additional evidence of the repressive nature of this motif.

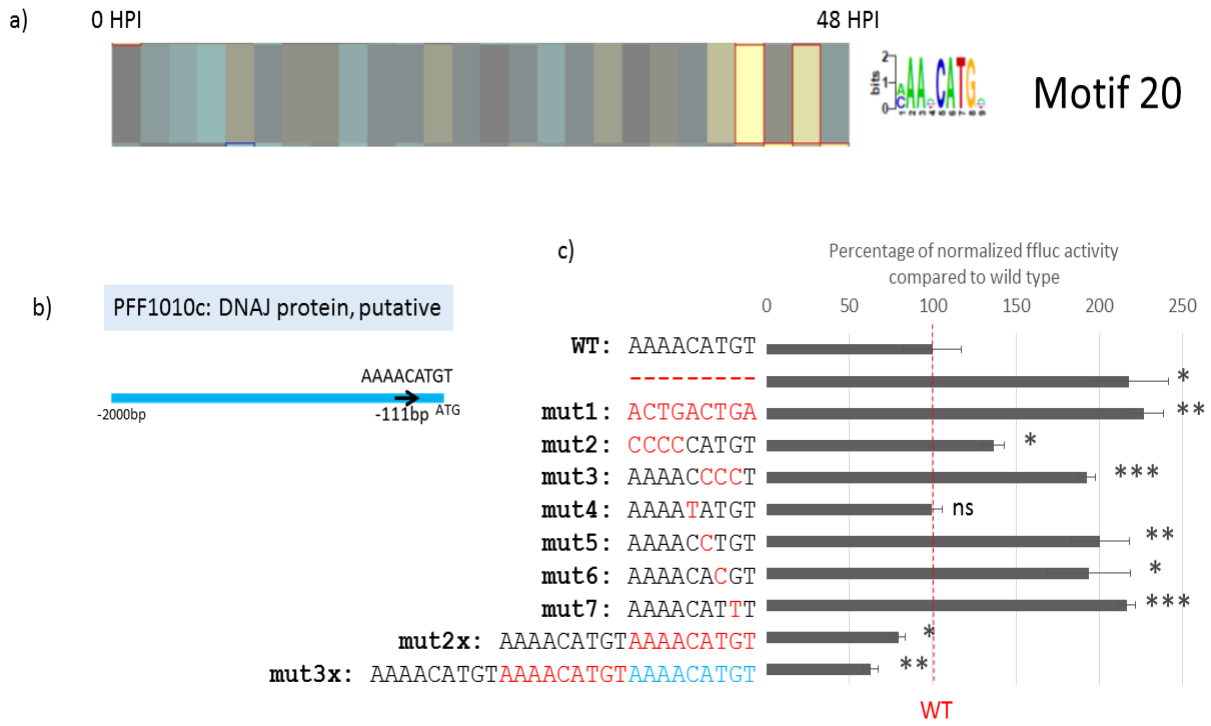


Figure 5.5 Motif 20 mutagenesis.

a) FIRE output for motif 20. Motif 20 is overrepresented in the 1 kb upstream regions of genes which mRNA abundance peaks during the mid-late schizont stages. b) Schematic diagram of 2 kb upstream of *pff1010c*. *pff1010c* is one of the genes which are in the overrepresented clusters. The 2 kb upstream regions of *pff1010c* was used in the in initial screening of the motifs. Motif 20 is located at -111 bp with respect to the translational start site. c) Site directed mutagenesis of motif 20 in the 2 kb upstream regions of *pff1010c*. Constructs were made as described in Figure 4.8. Transient transfections and dual luciferase assays were done as described in Chapter 2. Normalized ffluc activity of each mutant promoter construct is presented in a percentage relative to the activty driven by the wild type promoter of *pff1010c*. The wild type sequence of the motif is as stated (WT). As reported in Figure 4.8, deletion of motif 20 resulted in increased ffluc activity driven by approximately 100%. Scrambled sequence mutation of motif 20 resulted in increased ffluc activity driven to a similar level as when the motif was deleted (mut1). Mutation of "AAAACATGT" to "CCCCATGT" resulted in approximately 40% increase in activity (mut2). Mutation of "AAAACATGT" to "AAAACCCCT" resulted in increase in activity by approximately 90% (mut3). Mutation of "AAAACATGT" to "AAAATATGT" did not affect ffluc activity driven (mut4). Mutation of "AAAACATGT" to "AAAACCTGT", "AAAACACGT", "AAAACATTT", all 3 mutations increased ffluc activty driven to a level that was similar as when the motif was deleted (mut5, 6, 7). When 1 additional copy of the motif was included in the promoter sequence, ffluc activity was decreased by approximately 20% (mut2x). When 2 copies of the motif was included, the activity driven was increased by about

30% (mut3x). Statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days.

5.3 Motifs have similar effects on multiple promoters

The FIRE algorithm had predicted motifs based on the presence of specific DNA sequences present on upstream sequences of genes based on their gene expression patterns. However, only one gene out of all the genes reported to contain the particular motif in its upstream sequences was chosen for the screening done in Chapter 4 (Figure 4.1 and 4.10). Hence, to determine whether the motifs plays a role in transcriptional regulation in multiple genes with a similar expression pattern, other genes which harbor the motif in its 5' upstream regions were tested. 5 or 6 additional promoters of motifs 3, 16 and 20 were tested. In the same way as the screening strategy (Figure 4.8), motifs were deleted from 2 kb upstream of genes and promoter activity was compared to its respective wild type promoters by transient transfections and dual luciferase assays.

For motif 3, 5 additional promoters were chosen. The mRNA of these genes were all highly abundant during the late schizont/early ring or ring stages (Figure 5.6b). Both orientations of the motif were observed in the 5 promoters that were chosen (Figure 5.6a). Out of the 5 promoters, deletion of the motif from 4 of the promoters resulted in increase in activity of approximately 1.5 fold (Figure 5.6a). This was consistent with the repressive effect observed in the initial screening (Figure 4.10) and additional mutagenesis on motif 3 (Figure 5.1c). Interestingly, a significant decrease in activity was observed when motif 3 was deleted from the promoter of *mal7p1.144*. This may indicate that the same motif may have different effects in different sequence context in different promoters.

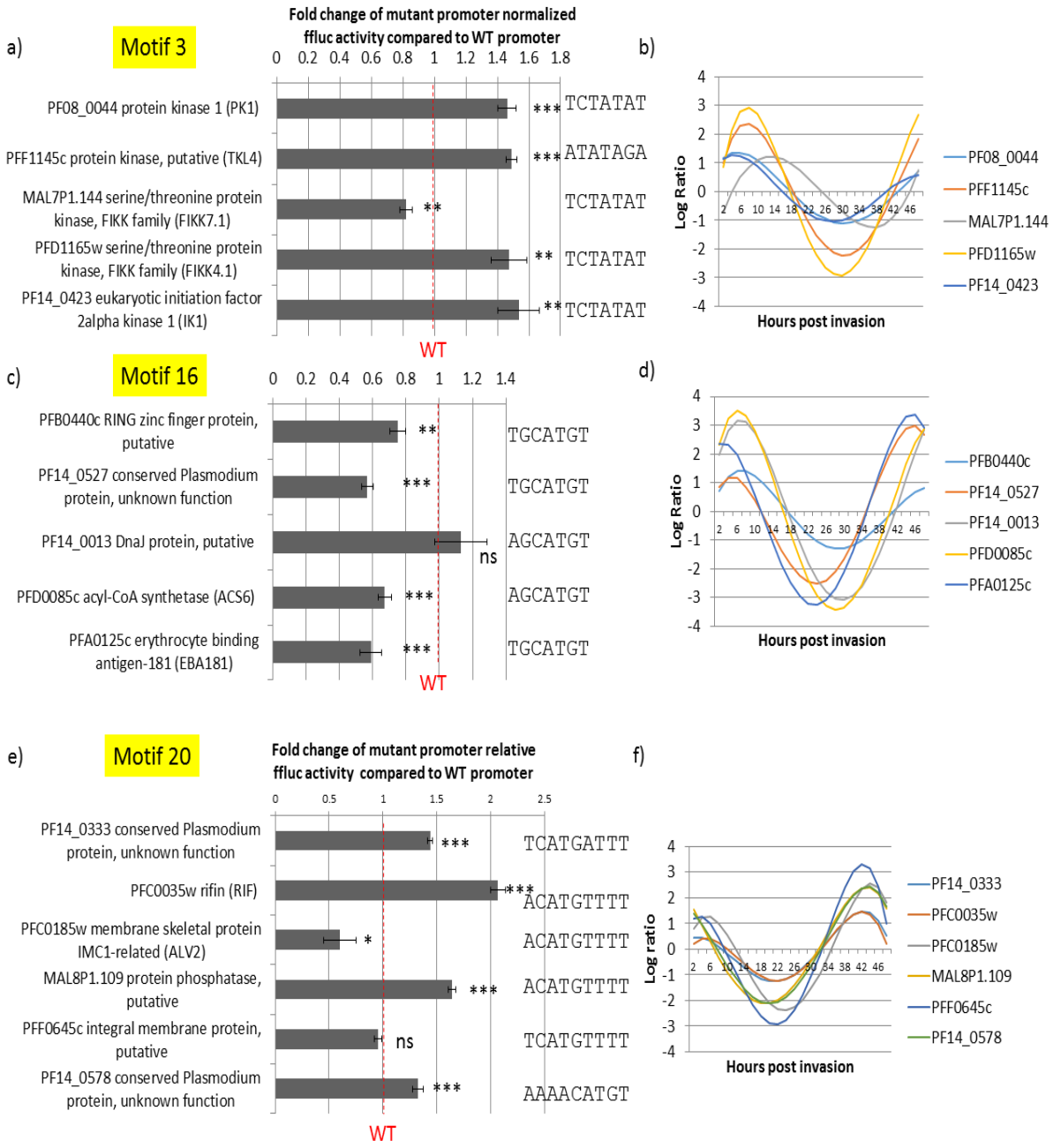


Figure 5.6 Additional promoters tested for motif 3, 16 and 20.

2 kb upstream of the genes named with respect to the translational start site was cloned into plasmid pf86, driving transcription of the *ffluc* gene. Mutant promoters of each promoter were cloned into pf86 as described in Figure 4.8. a, c, e) Transient transfection results of motif 3, 16 and 20. Transient transfections and dual luciferase assays were done as described in Chapter 2. Renilla normalized ffluc activity for each mutant promoter was expressed as a fold change over its respective wild type promoters. The exact motif sequence for each promoter is as listed. For motif 3, deletion of the motif resulted in increase in ffluc activity in 4/5 of the genes tested, decrease in ffluc activity for 1 of the genes. Deletion of motif 16 resulted in decrease in activity in 4/5 of the genes tested, no change in activity for 1 of the genes. Deletion of motif 20 in 4/6 gene

promoters resulted in increase in activity, no change in 1 of the genes and decrease in one of the genes tested. Statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. The error bars represent data from 3 independent transfections done on 3 separate days. b, d, f) mRNA expression profiles of the genes tested. The expression profiles of the gene promoters tested were obtained from microarray analysis [30]. Total RNA was harvested in 2 hour intervals from Dd2 *in vitro* cultures, cDNA synthesized and hybridized on a microarray chip containing 70-mer oligonucleotides which represent all the genes in the *P. falciparum* genome. Log ratios of selected genes at all time points were selected for additional screening were plotted on the graphs shown.

Motif 16 is overrepresented on genes which mRNA abundance peaks during the late schizont/early ring stages. Motif 16 has been previously shown in this thesis that it induces transcriptional activity as the deletion/mutation of it caused a decrease in ffluc activity driven by the promoter (Figure 4.10 and 5.3c). Hence, 5 additional promoters which contain motif 16 were chosen to test if motif 16 also plays an inductive role in the transcription in these promoters. The mRNA of the genes of the 5 additional promoters peaks during the late schizont/early ring stage (Figure 5.6d). Deletion of motif from the promoter regions of 4 out of 5 of the genes chosen resulted in a decrease in ffluc activity driven (Figure 5.6c). This provided further evidence that motif 16 plays an inductive role in transcription in multiple promoters. An exception is the promoter of the putative DNAJ gene (*pf14_0013*) where the deletion of motif 16 did not result a significant alteration in activity (Figure 5.6c).

Motif 20 has been shown to play a repressive role in transcription as deletion and mutation of the motif resulted in increase in levels of ffluc activity driven (Figure 4.10 and 5.5c). To further validate the role of this motif, 6 other gene promoters were chosen. FIRE algorithm had predicted motif 20 to be overrepresented on the upstream regions of genes which are abundant during the late schizont stages (Figure 5.5a). Accordingly, the mRNA abundance peaks of the 6 genes chosen were during the late schizont stages (Figure 5.6f). Deletion of

motif 20 in 4 out of the 6 gene promoters tested resulted in an increase in ffluc activity driven (Figure 5.6e). This is consistent with the repressive role in transcription seen in the promoter of *pff1010c* in the initial screening. However, there are 2 exceptions. Deletion of motif 20 from promoter of *pfc0185w* resulted in approximately 40% decrease in ffluc activity compared to its wild type promoter (Figure 5.6e). In addition, no change in activity was observed when motif 20 was deleted from the promoter of *pff0645c* (Figure 5.6e).

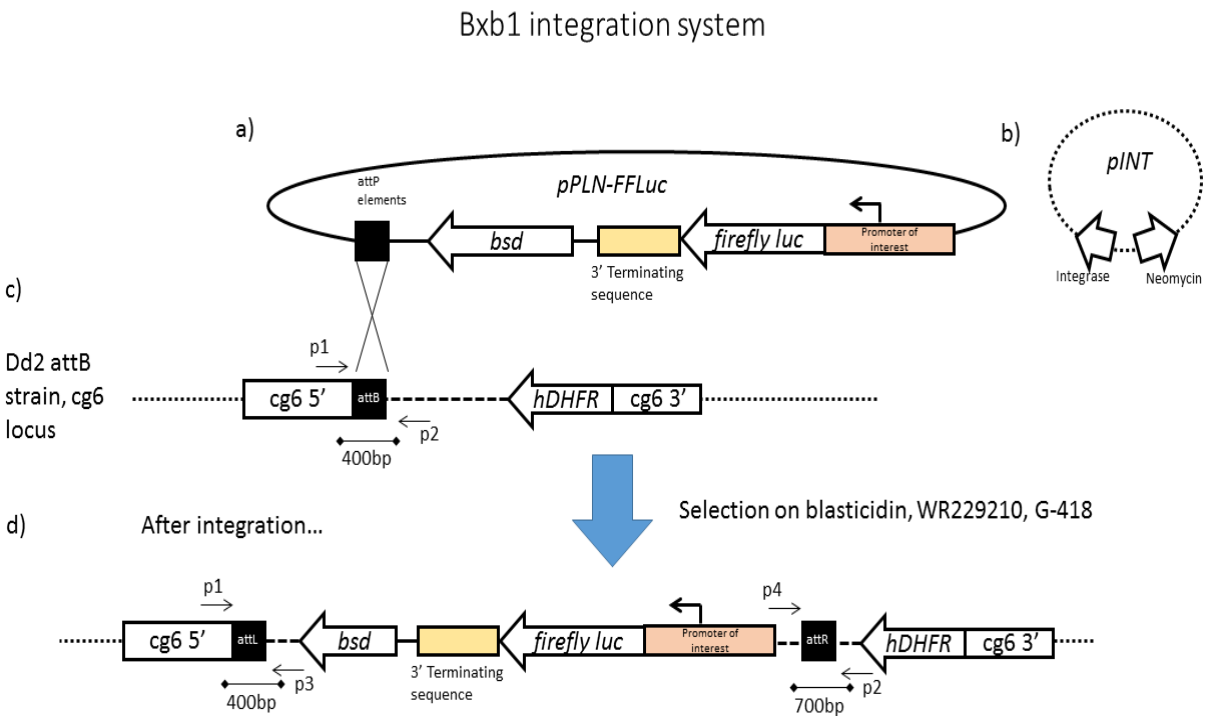
In summary, the repressive or inductive nature of the motifs previously characterized in the mutational assays could be seen in multiple gene promoters as well (Figure 5.6). Interestingly, exceptions could be seen in the additional promoters chosen (Figure 5.6). This underscores the complexity of the transcriptional network where the nature of a motif is not just dependent of its actual sequence by also in the context by which the motif lies in.

5.4 Motifs have stage specific effects

As described previously, transient transfections were done in ring stage parasites and assays were done during the next ring stage after reinvasion. In addition, previous experiments done in Chapter 3 where parasites were harvested and assayed for luciferase activity during trophozoite and schizont stage of the next cycle did not recapitulate the stage specific activity of the gene promoter (Figure 3.2d). In order to understand whether the motifs screened from transient transfection assays play a role in stage specific transcriptional regulation, the reporter gene together with the promoter of interest needs to be stably maintained in the parasite.

For this purpose, the *Bxb1* mycobacteriophage integrase system for *in vitro P. falciparum* cultures was utilized [410]. This system utilizes the Bxb1 integrase to catalyze the integration

of a plasmid via an *attB/attP* crossover (Figure 5.7). A strain of *P. falciparum*, Dd2 *attB*, contains an *attB* site at the *cg6* locus together with a drug selection marker, the *human dihydrofolate reductase* (*dhfr*) gene (Figure 5.7c). Hence, this strain of parasites was selected for using a selection drug, WR229210. A plasmid (pPLN-ffluc) containing the *attP* site, the *ffluc* cassette (consisting of the *ffluc* gene, 5' and 3' UTR of interest) and *blasticidin S deaminase (bsd)* selection marker (Figure 5.7a) was transfected into Dd2 *attB* parasites together with a plasmid containing the *bxb1 integrase* gene (Figure 5.7b, pINT). Recombination between the *attB* and *attP* sites which is facilitated by the Bxb1 integrase will allow the pPLN-ffluc plasmid to be integrated into the *cg6* locus of the Dd2 *attB* parasites (Figure 5.7d). This is an excellent platform as it allows the integration of the construct to be integrated at a specific genomic location. This will ensure that there are no confounding effects on gene transcription due to integration at different genomic locations.



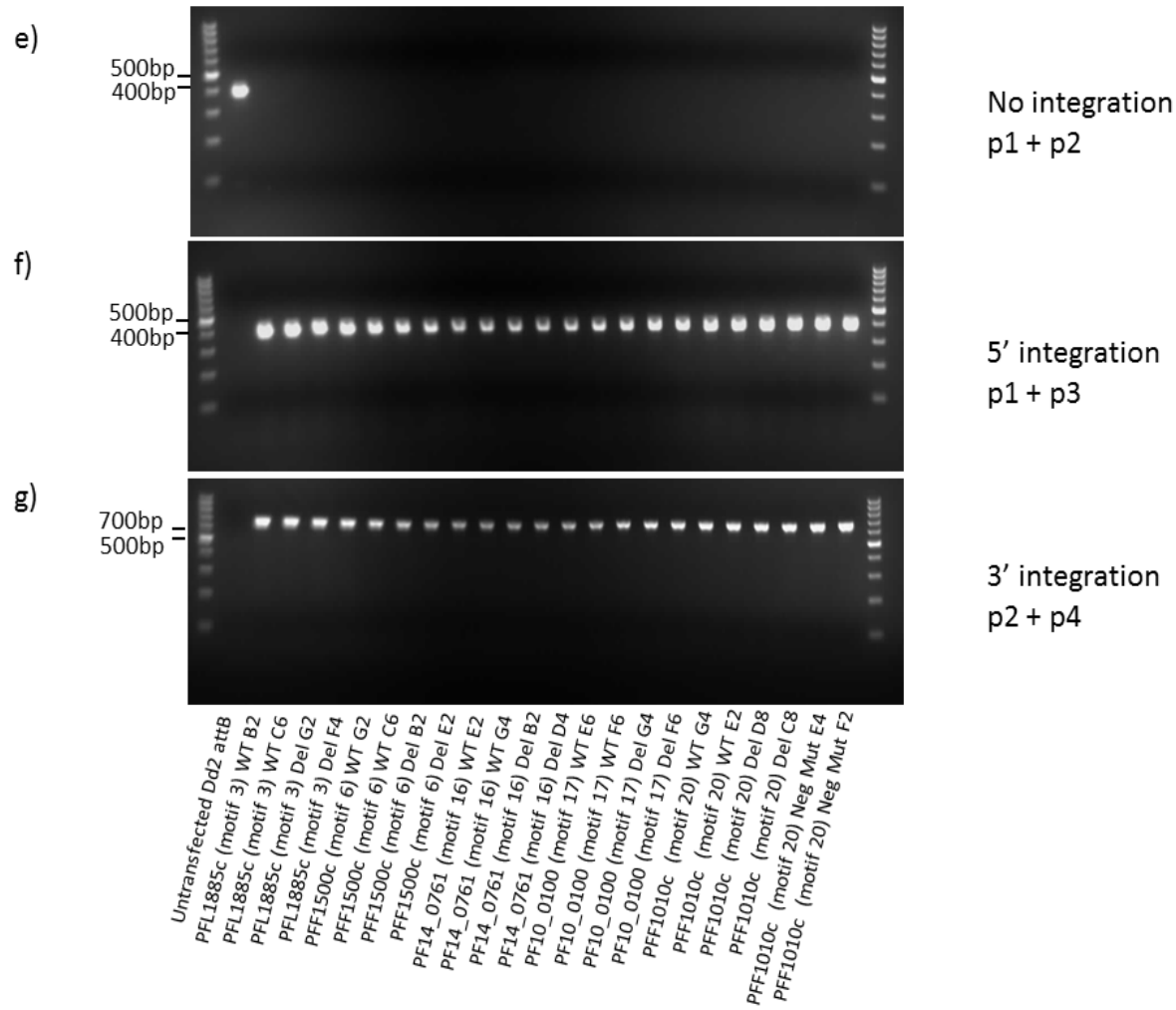


Figure 5.7 The Bxb1 integration system.

a) pPLN-ffluc plasmid. This plasmid was used for the integration of reporter gene (*ffluc*) into *cg6* locus in the Dd2 attB chromosome. The plasmid contains the *ffluc* gene which is flanked at the 5' end by 2 kb upstream (with respect to translational start site) of the translational start site of the gene of interest (*pfl1885c* for motif 3, *pff1500c* for motif 6, *pf14_0761* for motif 16, *pf10_0100* for motif 17 and *pff1010c* for motif 20). Together with the 5' upstream regions of the gene, the 3' end of the *ffluc* gene is flanked by the 1 kb downstream of the translational stop site of its respective gene (with respect to translational stop site). Constructs containing mutant 5' flanking sequences of the gene of interest which has the motif deleted were also created as detailed in Figure 4.8. Mutant constructs contained the same 3' flanking sequence as its corresponding wild type construct. pPLN-ffluc contains the *attP* sequence which is used for recombination with *attB* sequence on the *cg6* locus in the Dd2 *attB*. The plasmid also contains a *blasticidin-S-deaminase (bsd)* cassette which is used in the selection for parasites with successful integration. *Bsd* confers resistance to blasticidin S, a protein systhesis inhibitor. b) pINT plasmid. This construct contains the *bxb1 integrase* gene and Neomycin selection marker. It is selected for by drug G-418, another protein synthesis inhibitor. c) *cg6* locus of the Dd2 *attB*

parasite. In this strain of parasites, the *attB* sequence is integrated in the *cg6* locus. The *attB* sequence allows recombination with the pPLN-ffluc plasmid which contains the *attP* sequence. In addition, the human *dhfr* selection cassette is also present at the locus. This allows the Dd2 *attB* strain to be selected for using WR99210, an inhibitor against *P. falciparum* DHFR. Using primers p1 and p2 for PCR amplification on the Dd2 *attB* gDNA, a 400 bp PCR product will be amplified. p1 and p2 was used to confirm the presence of the *attB* site on the untransfected parasites by amplifying a 400 bp product. p1 and p2 was also used to confirm that the pPLN-ffluc plasmid is integrated at the attB site in the integrants by yielding no product. d) *cg6* locus after integration event. The sequences of the pPLN-ffluc were integrated into the cg6 locus. After recombination of *attB* and *attP* sequences, *attR* and *attL* flank the sequences of the pPLN-ffluc which is now integrated in the *cg6* locus. Primers p1 and p3 amplify a 400 bp product at the 5' end of the integrated junction. Primers p2 and p4 amplify a 700bp sequence on the 3' end. Primers p1 and p2 are sequences on the original locus. Primers p3 and p4 are sequences on the plasmid. Hence, amplification of products would indicate that an integration event has occurred. e, f, g) 50 µg of pPLN-ffluc plasmid and 50 µg of pINT plasmid was be co-transfected into Dd2 *attB* parasites. After drug selection using WR99210, blasticidin S and G-418 for 3 weeks and cycling on and off selection media twice for 2 weeks each after parasites were first detected. Each integrant line was then cloned by limiting dilution, 5 clones of each integrant line was selected and gDNA was isolated from each clone. A series of PCR amplifications were then done to verify the integration event and 2 clones of each line were selected for presentation in this figure. e) Amplification using p1 and p2. This primer pair amplified a 400 bp product on the original unintegrated *cg6* locus on the Dd2 *attB* gDNA. No product was observed for all the integrants. Strand extension time allowed for this PCR was 1 minute per cycle. f) Amplification using p1 and p3. This primer pair amplified a 400 bp product on the gDNA of all the integrants. No product was observed for the original unintegrated *cg6* locus on the Dd2 *attB* gDNA. Strand extension time allowed for this PCR was 1 minute per cycle. g) Amplification using p2 and p4. This primer pair amplified a 700 bp product on the gDNA of all the integrants. No product was observed for the original unintegrated *cg6* locus on the Dd2 *attB* gDNA. Strand extension time allowed for this PCR was 1 minute per cycle.

Using this platform, the strategy will be use to clone 2 kb 5' upstream regions of the genes used for the initial screening of motif 3, 6, 16, 17 and 20 (Figure 4.1b) together with 1 kb downstream of the respective gene's translational stop site as terminating sequences. After which, a construct containing a mutant form of the 5' upstream regions in which the motif of

interest is deleted will be created using a similar strategy as described in Figure 4.8. Constructs containing wild type and mutant promoters will be separately transfected into Dd2 *attB* parasites. After verifications on integration events and cloning by means of limiting dilution, RNA will be harvested at different time points for each cell line and quantified by real time qPCR.

After drug selection of 4 weeks, cycling on and off selection drug and cloning by limiting dilution, the selected parasites were checked by PCR to verify if the integration event had happened. Primers p1 and p2 which are on the endogenous locus amplified a 400 bp product only in the unintegrated Dd2 *attB* parasites gDNA (Figure 5.7e). However, no amplification was detected in all of the selected parasites (Figure 5.7e). This was because the sequences of p1 and p2 were approximately 9 kb apart on the integrated locus. This indicated that there is no locus in the population of integrants that was not integrated with pPLN-ffluc at the *cg6* locus. To check for integration at the 5' end of the plasmid, primers p1 and p3 was used (Figure 5.7d). The sequence on primer p3 is found on the plasmid and p1, on the endogenous locus (Figure 5.7d). Hence, there will be an amplified product of 400 bp only when the plasmid was integrated. The same strategy was used to check for integration at the 3' end of the plasmid using primers p4 and p2 (Figure 5.7d). A 700 bp product will be amplified only when the integration event occurred. 400 bp and 700 bp PCR products were amplified only in the drug selected parasites indicating of integration event occurred on the *cg6* locus (Figure 5.7f and g). On the other hand, when untransfected Dd2 *attB* strain gDNA was used for PCR with the p1/p3 and p2/p4 primer pairs, no products yielded (Figure 5.7f and g). This provides evidence of successful integration of the constructs into the *cg6* locus of the Dd2 *attB* parasites. In addition, the entire *ffluc* cassette (5', 3' flanking regions and the *ffluc* gene) was sequenced to ensure that there are no other mutations other than the deletion of the motif.

1 clone of each integrant line containing promoters with and without the motifs of interest were analyzed in pairs, wild type (WT) versus mutant with motif deleted (Del). In addition, 1 additional clone of each WT and Del for motif 6 containing the promoter of *pff1500c* was analyzed. RNA was harvested from the clones at 3 time points during the IDC: ring (8-12 hpi), trophozoite (20-24 hpi) and schizont (32-36 hpi). For each WT and Del integrant set, normalized *ffluc* transcripts of the Del integrant at each time point were presented as a fold change over the normalized *ffluc* transcripts of the WT integrant in each individual stage. After which, the fold change was normalized again to account for difference in copy number between WT and Del clones. A flow chart which summarizes the normalizing steps is in figure 5.8. Comparisons were done between WT and Del within individual stage. To be able to compare *ffluc* transcripts across 3 stages, a reference gene which is invariant across all 3 stages is required. Unfortunately such a gene is virtually impossible to find in *P. falciparum* as almost all genes exhibit a cyclic pattern of expression [19, 20].

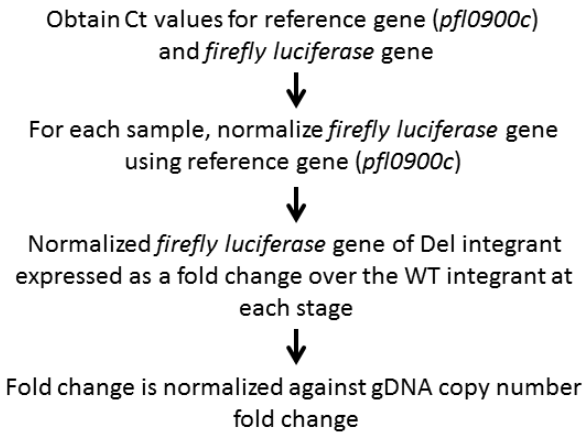
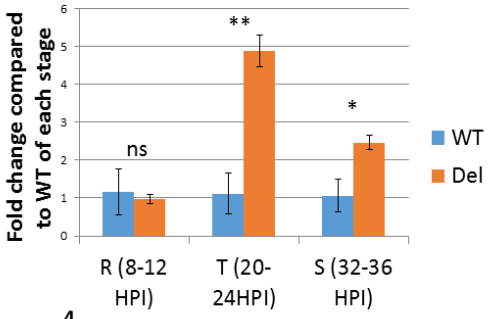


Figure 5.8 Flow chart of analysis steps in the analysis of luciferase transcripts quantifications for the stable reporter parasite lines.

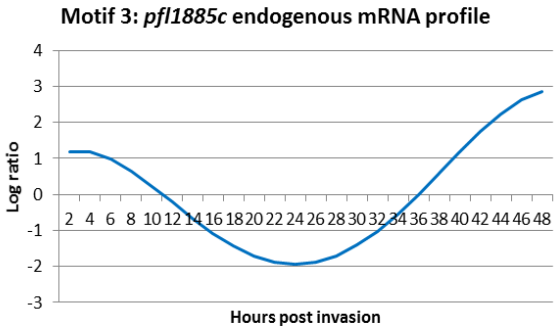
For the analysis of motif 3, 5' and 3' flanking sequences of *pfl1885c* were used. The endogenous transcript expression profile indicates that the gene is maximally expressed during the late schizont stage (Figure 5.9b). Transcript levels start to decline in the early ring stage until the minimum at the trophozoite stage, after which levels begin to increase until the peak at late schizont stages (Figure 5.9b). In the Del integrant where motif 3 was deleted, *ffluc* transcripts during the ring stage were in comparable levels with *ffluc* transcripts in the wild type integrant (Figure 5.9a). This indicates that the presence and absence of motif 3 did not affect transcript levels during ring stage. In contrast, an approximate 5 fold increase in transcripts was observed when comparing *ffluc* transcripts in the Del integrant to the wild type integrant during trophozoite stage (Figure 5.9a). As the parasites progressed through the IDC reaching the schizont stage (30-36 hpi), moderate increase in *ffluc* transcripts of 2.5 fold was observed in the Del integrant compared to the WT integrant (Figure 5.9a). This indicated that the deletion of motif 3 had led to increases in *ffluc* transcripts moderately during the schizont stage. Notably the most significant increase in *ffluc* transcripts occurred during the trophozoite stage (20-24 hpi), this is the time point where the endogenous *pfl1885c* transcript was the least abundant during the IDC (Figure 5.9b). In contrast, during ring and schizont stage where the endogenous *pfl1885c* transcripts were more abundant, no increase or moderate increase in *ffluc* transcripts were observed respectively (Figure 5.9a). This indicated that the repressive effect of motif 3 is the most pronounced during trophoozite stage which is the time point where the endogenous gene transcripts is the least abundant (Figure 5.9a and b).

a)

Motif 3

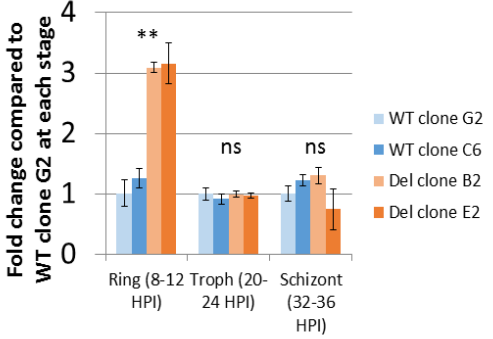


b)

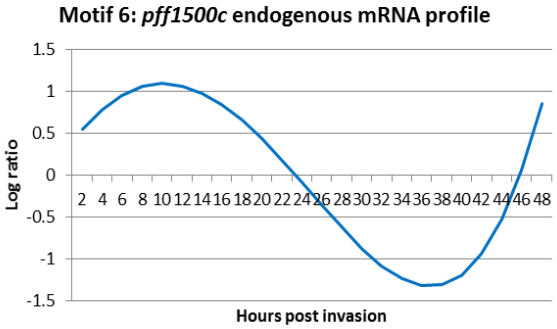


c)

Motif 6

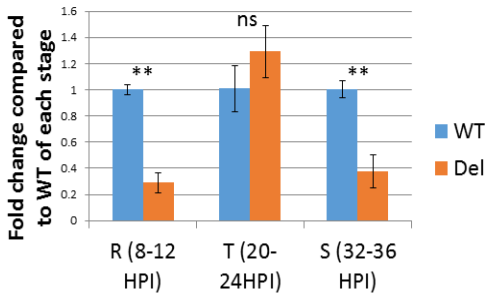


d)

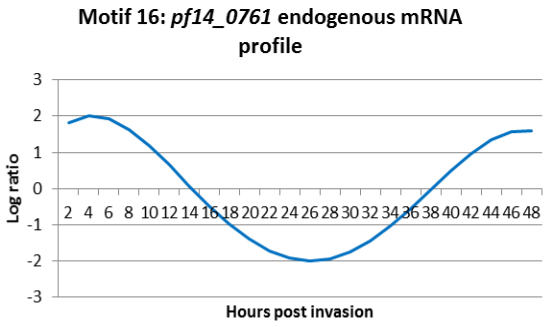


e)

Motif 16

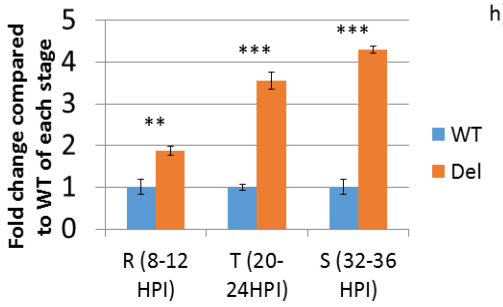


f)

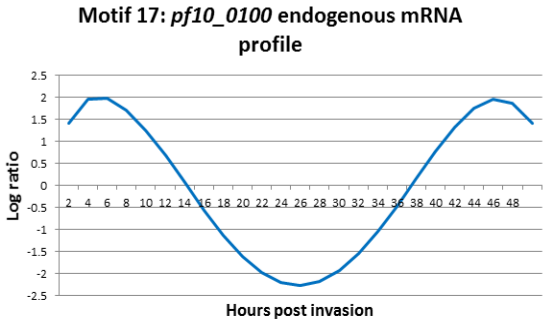


g)

Motif 17

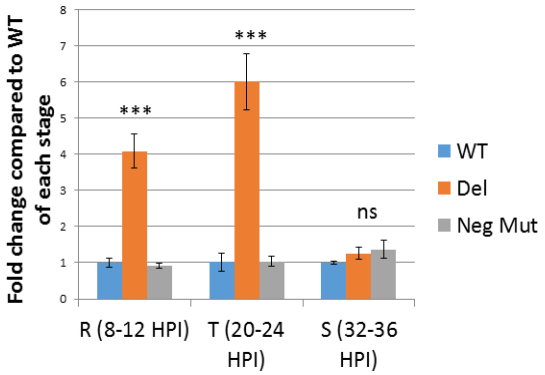


h)



i)

Motif 20



j)

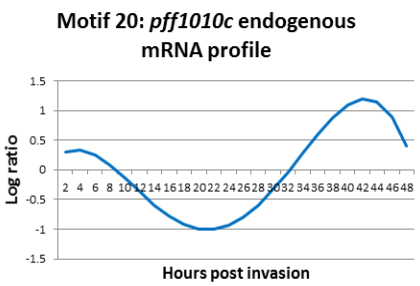


Figure 5.9 Analysis of *ffluc* transcripts from integrants and endogenous mRNA expression profiles of the genes of interest.

a, c, e, g, i) Analysis of ffluc transcripts relative to ffluc transcripts in WT integrant during rings stage. For each motif to be analyzed, a representative gene was chosen (*pfl1885c* for motif 3, *pff1500c* for motif 6, *pf14_0761* for motif 16, *pf10_0100* for motif 17, *pff1010c* for motif 20). For each gene/motif to be tested, 2 pPLN-ffluc constructs were made. 1 construct was with the 2 kb upstream of the gene of interest (with respect to the translation start site) driving the transcription of the *ffluc* reporter gene (WT). The other is with the mutant form of the 2 kb upstream of the gene of interest where the motif of interest was deleted (Del). In both WT and Del constructs, the *ffluc* gene is flanked at its 3' end with 1 kb downstream of the gene of interest with respect to the translational stop site. For motif 20 (i), an additional mutant of the 2 kb upstream of *pff1010c* was made where the "C" in the motif was mutated into a "T" (Neg mut). This mutant has been shown in transient transfection experiments to drive the similar levels of *ffluc* transcripts as the wild type putative promoter region. Hence, it is used as a negative control. Each construct was transfected into Dd2 *attB* parasites. Transfected parasites were selected in media containing WR99210, blasticidin S and G-418 for 3 weeks and after which, cycling on and off selection media twice for 2 weeks each. gDNA was isolated and verified for integration (Figure 5.7e, f, g). Total RNA and gDNA was harvested at 3 time points in the IDC: during rings (R) 8-12 hpi, during trophozoites (T) 20-24 pi and during schizonts (S) 32-36 hpi. cDNA was synthesized from the total RNA using primers specific for the *ffluc* gene and *pfl0900c*. *pfl0900c* will be used as a normalizing control in qPCR experiments. Real time qPCR was then carried out to obtain a Ct value for the *ffluc* gene and *pfl0900c* which is representative of the quantity of *ffluc* mRNA transcribed driven by the respective putative promoter (WT or Del) at R, T and S. The Ct for *ffluc* transcript in each sample from each integrant was normalized by its corresponding Ct for *pfl0900c* ($Ct_{(norm)} = Ct_{(ffluc)} - Ct_{(PFL0900c)}$). For each WT and Del integrant pair, the normalized *ffluc* transcripts of the Del integrant was expressed as a fold change over the normalized *ffluc* trancripts of the WT integrant at each individual stage. This was calculated by the ΔΔCt method, Fold change = $2^{-(Ct(T \text{ or } S \text{ norm}) - Ct(R \text{ WT norm}))}$. Using a similar method, relative copy numbers of the *ffluc* gene were calculated from real time qPCR of gDNA isolated from each integrant. The transcript fold change is normalized by the relative copy numbers of the *ffluc* gene. The statistical test used was the 2-tailed Student's T-test. ns, p>0.05, *, p<0.05; **, p<0.01; ***, p<0.001. Error bars represent samples harvested at 3 different lifecycles. a) 5' and 3' of *pfl1885c* flanked ffluc integrant for motif 3. For ring stage, *ffluc* transcripts were comparable in WT and Del integrants. For trophozoite stage, there were significantly more *ffluc* transcripts in the Del integrant compared to the WT integrant. During schizont stage, there was less increase in *ffluc* transcripts in the Del integrant than WT integrant. c) 5' and 3' of *pff1500c* flanked *ffluc* integrant for motif 6. During

trophozoite and schizont stages, *ffluc* transcript levels were similar in both WT and Del integrants. The *ffluc* transcript levels were much higher in the Del integrant when compared to WT integrant in the ring stage. This result was similar for the 2 clones of both WT and Del integrants. e) 5' and 3' of *pf14_0761* flanked ffluc integrant for motif 16. During ring and schizont stage, there were significantly less *ffluc* transcripts observed in the Del integrant compared to the WT integrant. During trophozoite stage, there were comparable levels of *ffluc* transcripts in WT and Del integrants. g) 5' and 3' of *pf10_0100* flanked *ffluc* integrant for motif 17. In all 3 stages, there were significantly more *ffluc* transcripts in the Del integrants compared to the WT integrants. Greater increases were observed for the trophozoite and schizont stage as compared to increases during the ring stage. i) 5' and 3' of *pff1010c* flanked ffluc integrant for motif 17. In all 3 integrants, there was a comparable level of ffluc transcripts during the schizont stage. In all 3 stages, *ffluc* transcripts in both WT and Neg Mut integrants were comparable. In the Del integrant, significantly higher levels of transcripts were observed when comparing to transcript levels in WT and Neg Del integrants during ring and trophozoite stage. b, d, f, h, j) Endogenous mRNA profiles of *pfl1885c*, *pff1500c*, *pf14_0761*, *pf10_0100* and *pff1010c*. mRNA profile of endogenous genes. Data was obtained from the published transcriptome of Dd2 [30]. Total RNA was harvested from the *P. falciparum* Dd2 strain in 2 hour intervals over the 48 hour IDC parasite life cycle. cDNA was synthesized and cDNA for each timepoint was labeled with Cy5. Pooled cDNA from all time points was labeled with Cy3. For each time point, Cy5 labeled cDNA was hybridized against the Cy3 labeled pool on a microarray chip containing oligonucleotides representing all the genes in *P. falciaprum*. The log ratio of Cy5 signals over Cy3 signals of the oligonucleotides representing the respective genes were plotted on the graph shown in this figure. b) Peak expression of *pfl1885c* occurs during the late schizont stages and minimum during the trophozoite stage. d) Peak expression of *pff1500c* occurs during the mid-late ring stages and minimum during the mid schizont stage. f) Peak expression of *pf14_0761* occurs during the late schizont/early ring stages and minimum during the trophozoite stage. h) Peak expression of *pf10_0100* occurs during the late schizont/early ring stages and minimum during the trophozoite stage. i) Peak expression of *pff1010c* occurs during the late schizont stages and minimum during the trophozoite stage.

5' and 3' flanking sequences of *pff1500c* was used for the analysis of motif 6. 2 clones of each WT and Del integrants after limiting dilution were selected for this analysis. *pff1500c* is a ring stage specific gene where the endogenous transcripts peak during the mid-ring stages (10-12 hpi) and subsequently decreases till the minimum at mid schizont stage (36-38 hpi)

(Figure 5.9d). Subsequently, *pff1500c* transcript levels begin to increase as the parasite progresses towards the late schizont stages. There was a significantly higher level of *ffluc* transcripts in the Del integrants than the wild type integrants by about 3 fold during the ring stage (Figure 5.9c). In contrast, *ffluc* transcript levels in the Del integrants were not significantly different from the wild type integrants during the trophozoite and schizont stages (Figure 5.9c). This indicated that the absence of the motif did not affect *ffluc* transcript levels during trophozoite and schizont stages but resulted in increased levels of transcripts during the ring stage. This provides evidence that the repressive nature of motif 6 which was first observed in the transient transfection assays is specific to the ring stage (Figure 5.9c). In contrast to motif 3, the results from the analysis of motif 6 appear to indicate that the repressive nature of motif 6 is during the time point where the endogenous *pff1500c* transcript levels were the most abundant (Figure 5.9b and c). This is an interesting observation because it may indicate that the repressive nature of the motifs is not just restricted to the time point where endogenous transcripts are the least abundant. It is likely that motifs on the same promoter work in concert with each other, acting at different time points in the IDC to achieve the wave-like transcript profile. In addition, results were similar in both clones of WT and Del integrants (Figure 5.9b). This indicated that the results were reproducible between different clones.

The next set of integrants to be analyzed contains the 5' and 3' flanking regions of *pf14_0761* for the testing of motif 16. Motif 16 is an activating motif where a reduction of *ffluc* activity was observed when deleted from the promoter in the transient transfections (Figure 5.3c). The endogenous *pf14_0761* transcript peaks during the early ring and late schizont stage, and is at its minimum during the trophozoite stage (Figure 5.9f). In the Del integrant, the *ffluc* transcript levels during ring and schizont stage were significantly lower than the *ffluc*

transcript in the wild type integrant (Figure 5.9e). In contrast, *ffluc* transcripts level was relatively similar during trophozoite stage in both WT and Del integrants (Figure 5.9e). Taken together, lowered levels of *ffluc* transcripts during ring and schizont stages observed in the Del integrant compared to the WT integrant coincided with the time point where the endogenous *pf14_0761* gene transcript was the most abundant in the IDC (Figure 5.9e and f). However, level of *ffluc* transcript was similar between WT and Del integrants in trophozoite stage where endogenous *pf14_0761* transcripts were least abundant (Figure 5.9e and f). This indicated the absence of motif 16 in the Del integrants has led to lowered *ffluc* transcript levels during ring and schizont stages, leading to the conclusion that the inductive effect of motif 16 occurs in ring and schizont stages where endogenous transcript abundance peaks (Figure 5.9e). This observation shows that the inductive effect of motif 16 was stage specific as well.

Motif 17 was analyzed by 5' and 3' flanking regions of *pf10_0100*. Motif 17 was found to be a repressive motif in the transient transfection assays (Figure 5.4c). The peak of *pf10_0100* endogenous mRNA abundance is also during the late schizont and early ring stage (Figure 5.9h). The minimum mRNA abundance of the endogenous gene is during the trophozoite stage. In this set of integrants, *ffluc* transcript levels were significantly higher in the Del integrants compared to the WT integrants at all 3 time points of the IDC (Figure 5.9g). Even though the repressive nature of motif 17 could be observed in all 3 stages, more pronounced increase in *ffluc* transcripts occurred during the later stages (trophozoite and schizont) (Figure 5.9g). Although deletion and mutation of the motif increased *ffluc* activity in transient transfection assays, it was also observed in additional transient transfection assays that increasing the copies of the motif did not enhance the repressive effect of the motif (Figure 5.4c). Taken together, these observations could indicate that motif 17 plays a general

repressive role in transcription in the IDC and additional regulatory mechanisms such as other DNA motifs or post-transcriptional regulation play more important roles in transcript regulation of *pf10_0100*.

Motif 20 is the motif with the strongest repressive effect from the transient transfection assays (Figure 4.10). 5' and 3' flanking regions of *pff1010c* was used in the investigation of stage specific effects of motif 20. The endogenous *pff1010c* mRNA expression pattern peaks during the mid-late schizont stages (Figure 5.9j). Subsequently, towards the end of schizont stage, *pff1010c* transcript begins to decline until the minimum at the trophozoite stage (Figure 5.9j). *ffluc* transcripts in the Del integrant were at similar levels as the wild type integrant during the schizont stage (Figure 5.9i). On the other hand, *ffluc* transcript level was at higher levels in the Del integrant where motif 20 was deleted than the wild type integrant during ring and trophozoite stage by 4 and 6 fold respectively (Figure 5.9i). This indicated that the presence of motif 20 in the WT integrant led to decreased *ffluc* transcript levels during ring and trophozoite stages (Figure 5.9i). The most pronounced effects of repression could be observed in trophozoite stage where the endogenous *pff1010c* transcript levels were least abundant (Figure 5.9i and j). In schizont stage, where endogenous transcript abundance was the highest, no repressive effect was observed (Figure 5.9i and j). As an additional control, another integrant was generated (Figure 5.9i, Neg Mut). In this integrant, a mutant form of the 5' flanking region of the *pff1010c* drives the transcription of *ffluc*. This mutant form of the promoter where motif 20 is mutated from "AAAACATGT" to "AAAATATGT" did not affect *ffluc* activity when compared to the wild type promoter in the transient transfection assays (Figure 5.5c). Hence, we hypothesized that *ffluc* transcript levels driven by this mutant form of the promoter will not be different from the *ffluc* transcript levels driven by the wild type promoter. When *ffluc* transcripts were analyzed from the Neg Mut integrant, the *ffluc*

transcript levels were the same as the wild type integrant at all 3 stages (Figure 5.9i, grey bars). This provides a good negative control that the differences in levels of *ffluc* transcript between the wild type and mutant promoter were not because of differences in integration or individual transfections.

In summary of the results in this section, the effects of motifs were stage specific (Figure 5.9). For motif 3 and 20, the repressive effects of the motif were most pronounced during the time point where the endogenous transcript abundances were the lowest (Figure 5.9a and i). Repressive effects of motif 6 were during the stage of the IDC where endogenous transcripts were the most abundant (Figure 5.9c). A general repressive effect of motif 17 was observed in all 3 stages, although the effect was the more pronounced during trophozoite and schizont stages (Figure 5.9g). Interestingly, a variety of mechanisms of transcriptional repression was observed. The inductive effect of motif 16 is also specific, for ring and schizont stages (Figure 5.9e).

5.5 Motifs have sequence-specific protein binding activities

In the regulation of gene transcription, transcriptional factors bind specific DNA sequences which then affect the transcription of a gene. Hence, the next step will be to identify whether the 5 motifs are able to bind to protein factors in the *P. falciparum* nuclear proteome. For this purpose, EMSAs were carried out to test if the motifs had specific protein binding ability.

In the following EMSA experiments, biotin labeled double stranded 50 bp probes containing the motif of interest in the center, flanked by the endogenous sequences in the respective promoters were tested for protein binding capabilities with 1 µg of *P. falciparum* nuclear extract. A shift will be observed when a protein-DNA complex is formed. The shifts observed

may, however, be due to unspecific binding to the entire length of the probe. Hence, more thorough investigations need to be done to understand whether the formation of complexes was motif sequence-specific. Additional investigations with unlabeled competitor probes with and without the motifs were done for motifs 3, 6, 16 and 20. Competitor probes with motif sequences contain the exact same sequence as the labeled probes with motif sequence but did not contained the biotin moiety. Competitor probes without motif sequences contain the same sequences flanking the motif as labeled probes with motif, but without the motif sequences and were not biotin labeled as well.

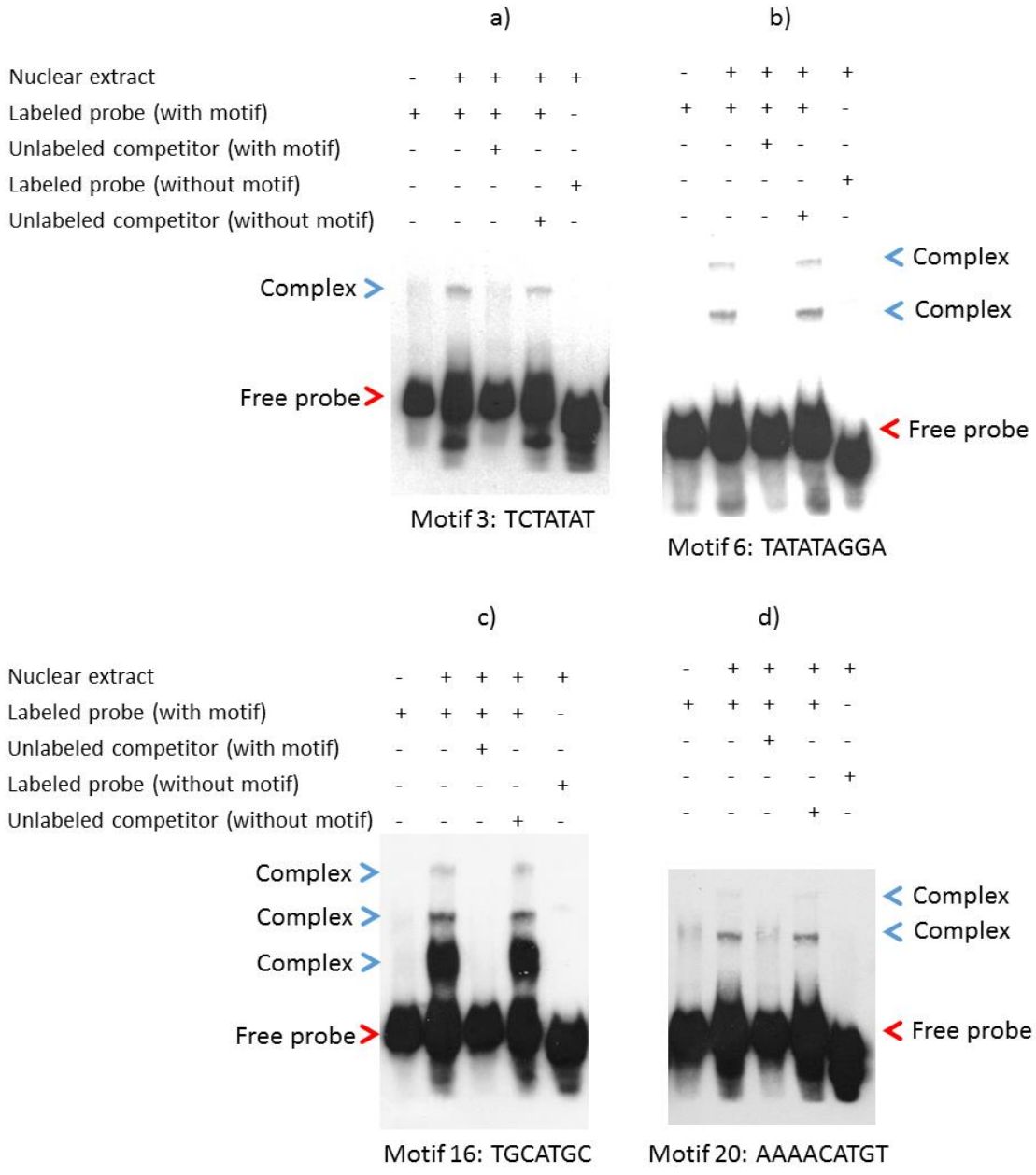


Figure 5.10 EMSA experiments testing motif 3, 6, 16 and 20.
Nuclear factors were extracted from schizont stage parasites and concentrated to 1 µg/µl. Double stranded probes were obtained by annealing single stranded oligonucleotides. Approximately 50 bp probes were designed with the motif of interest in the center and flanked on both sides by 20 bp of sequences which flank the motif in the endogenous loci (*pfl1885c* for motif 3, *pff1500c* for motif 6, *pf14_0761* for motif 16 and *pff1010c* for motif 20). Labeled probes were biotin labeled at the 5' ends of the oligonucleotides. Probes which contain only the 20 bp flanking sequences on each end (probes with no motif) were also synthesized. Unlabeled forms of the probes (with motif and without motif) were used as competitors in 200 fold in excess. The results from all 4 EMSA experiments were similar except for the number of bands shifted. No binding was observed when no nuclear extract was added to the labeled probe (20 fmol) which contains the motif (1st lane). Band shifts were observed when 10 µg of nuclear extract was allowed to bind to the labeled probe (20 fmol) which contains the motif (2nd lane). When competitor probe with the motif (4 pmol) was added 200 times in excess, the band shifts were competed away (3rd lane). However, when competitor probe without the motif (4 pmol) was added, the band shifts were not competed away (4th lane). In addition, when nuclear extract was allowed to bind to labeled probes without the motif (20fmol), no band shifts were observed (5th lane). a) Motif 3. 1 band shift observed. b) Motif 6. 2 band shifts observed. c) Motif 16. 3 band shifts observed. b) Motif 20. 2 band shifts observed.

For motif 3, no shift was observed when no nuclear extract was added to the labeled probe containing the motif (Figure 5.10a, lane 1). When nuclear extract was added to the labeled probe, a shift was observed (Figure 5.10a, lane 1). When unlabeled probe with motif 3 was added to the reaction, the shift was successfully competed away (Figure 5.10a, lane 3). However, an unlabeled probe where motif 3 was deleted did not out-compete the complex formed with the labeled probe with the motif (Figure 5.10a, lane 4). Furthermore, no complex was formed with the labeled probe without the motif (Figure 5.10a, lane 5). This provides evidence of the specificity of the DNA-protein complex formed between the motif and the protein element(s).

For motif 6, the labeled probe with motif 6 displayed no shift when nuclear extract was not added (Figure 5.10b, lane 1). Complexes indicated by 2 different sizes of shifts with the labeled probe with motif 6 were observed when nuclear extract was added to the reaction mixture (Figure 5.10b, lane 2). Again, the complexes were out competed by an unlabeled probe with motif 6, 200 times in excess (Figure 5.10b, lane 3). However, the complexes were not out competed by an unlabeled probe where motif 6 was deleted (Figure 5.10b, lane 4). In addition, no shifts were observed when binding reactions were carried out with a labeled probe where motif 6 was deleted (Figure 5.10b, lane 5). This revealed sequence-specific protein binding activity of motif 6 with the formation of 2 different complexes.

For motif 16, 3 separate complexes were formed with nuclear factors and the labeled probe containing motif 16 (Figure 5.10c, lane 2). 1 major band shift was observed and 2 other minor band shifts were detected as well. However, the band shifts were not observed in the binding reaction where no nuclear extract was added (Figure 5.10c, lane 1). In addition, the 3 complexes were out competed by an unlabeled probe containing motif 16 which was added 200 times in excess (Figure 5.10c, lane 3). To show that the binding observed from the band shifts were sequence-specific to motif 16, an unlabeled probe without motif 16 did not out compete the 3 complexes (Figure 5.10c, lane 4). In addition, shifts were not observed for binding reaction with the labeled probe without motif 16 (Figure 5.10c, lane 5).

For motif 20, 2 distinct band shifts were observed for binding of labeled probes with nuclear extract which were not observed in the binding reaction when nuclear extract was not added (Figure 5.10d, lane 1 and 2). Shifts were out competed by unlabeled probes with motif 20, but were not out competed by unlabeled probes which do not contain motif 20 (Figure 5.10d,

lane 3 and 4). Finally, no complexes were observed in binding reactions when nuclear extract was incubated with labeled probes which do not contain motif 20 (Figure 5.10d, lane 5).

In summary, 4 out of 5 motifs selected for further studies in this chapter had motif sequence-specific protein binding activities (Figure 5.10). The complexes formed represent nuclear protein factors binding to the DNA motifs which are likely to be transcription associated factors which carry out the functions by associating with the motifs.

5.6 Summary

Site-directed mutagenesis assays has validated the results obtained from the initial screening described in Chapter 4 for motifs 3, 6, 16, 17 and 20 (Figures 5.1 to 5.5). By using a variety of mutant forms of the motif in addition to deleting the motif, we have disrupted the function of the motif (Figures 5.1 to 5.5). In addition, site-directed mutagenesis assays has also identified important bases in the motif sequence (Figures 5.1 to 5.5). Next, we have also shown that the function of the motif is not only limited to the promoter that was used in the initial screening (Figure 5.6). Additional promoters containing the same motif were selected for deletion mutagenesis of the motif. We observed that the effect of the motif was conserved in a large proportion of the additional promoters tested (Figure 5.6). Importantly, we have also observed that the repressive and inductive effects of the motifs were stage specific (Figure 5.9). A plethora of repression mechanisms could also be deducted from observing the *ffluc* transcripts in the stable integrants (Figure 5.9). Finally, we have also observed that the motifs form protein-DNA complexes which protein elements from nuclear extracts (Figure 5.10). This result indicates that the motifs described in this chapter could bind to nuclear proteins which are potentially transcription factors.

6. Discussion

Although mechanisms of regulation of gene transcription in *P. falciparum* are generally viewed to be similar to the eukaryotic system, there are several notable differences between them. Specifically, the transcriptome during the *P. falciparum* IDC is characterized by a well regulated expression of genes where most genes (60-80%) are transcriptionally active in the IDC and display a sine wave-like transcript expression pattern [19, 20] (Figure 6.1a). However, the exact mechanism of how a gene is up regulated at a certain time point and down regulated in other time points is still unknown.

We have utilized FIRE which searches for motifs which are overrepresented in groups of genes which are co-expressed on an updated transcriptome (Figure 4.1) [30, 359]. By transient transfection and dual luciferase assays, 9 out of 16 motifs tested achieved statistical significance (Figure 4.10). 3 motifs induced transcriptional activity and 6 motifs repressed transcriptional activity (Figure 4.10a). Apart from the 9 motifs which achieved statistical significance, there were 3 motifs which displayed a slight repressive role but did not achieve statistical significance (Figure 4.10a). Interestingly, there appears to be an overrepresentation of repressive motifs out of the motifs tested in our assays.

Nonetheless, we have chosen to focus on 5 motifs which displayed significant effect in both dual luciferase assays and qPCR experiments for further investigations (Figure 4.10). Additional mutational assays on the motifs showed that the transcriptional differences observed in the initial screening were highly specific and not due to unspecific effects when alterations are made to the promoter (Figure 5.1 to 5.5). In addition, we have also shown that the motifs play a similar role on multiple promoters (Figure 5.6).

Interestingly, it was also observed the same motif had no effect or even opposite effect on a small number of promoters (Figure 5.6). This was not unexpected as it was observed that multiple motifs may occur in the same promoter region in *Plasmodium* regulatory motif prediction studies [357, 358]. This same observation was made in the motifs predicted in this thesis where most genes had an average of 5 different motifs associated with its promoter region (Supplementary Figure 3). It is also not uncommon to find multiple motifs on one promoter region in other eukaryotic organisms [203, 204, 208, 210]. Importantly, the function of these motifs have been shown to be affected by the context of the promoter it is located in and this could possibly be harnessed to fine tune gene expression using different combinations of motifs (Figure 5.6). In this thesis, only one motif was deleted in one promoter region, it would be important to explore deletions of multiple motifs in the same promoter to elucidate mechanisms in combinatorial functions of the motifs in the future experiments (Figures 4.10 and 5.6).

6.1 Repression as a mode of transcriptional regulation

An important observation made in the experiments described in this thesis is the demonstration of a plethora of repression mechanisms in the stable transfection experiments (Figure 5.9). Motif 3 and 20 displayed repression effects when the endogenous gene transcript abundance was at its minimum (Figure 5.9a and i). Motif 6 displayed repression effects when the endogenous gene transcript abundance was at its maximum (Figure 5.9c). Motif 17 displayed a general repressive effect in throughout the IDC (Figure 5.9g). When the motif 3 and 20 were deleted, most pronounced increases in levels of *ffluc* transcript were observed at the stage where endogenous transcript abundance was at the minimum (Figure 5.9a, b, i and j). At the time point where endogenous transcript abundance was at the

maximum, there was no difference in *ffluc* transcript levels. This suggests that sine wave-like transcript expression pattern of IDC transcripts occurs possibly by a repressive mechanism mediated through *cis*-regulatory motifs which exerts their roles at the point when transcript abundance is at the lowest (Figure 6.1c). However, this may not be the only repression mechanism as motif 6 and 17 exerts their roles at the time point where there is maximum endogenous transcript abundance and a throughout the IDC respectively (Figure 5.9c, d, g and h). It is likely that motif 6 and 17 functions in concert with other motifs not investigated in this thesis to achieve the transcript abundance patterns of its respective genes. Nonetheless, we have provided evidence that repression of transcripts plays a role in transcript regulation in *P. falciparum*, at least in part (Figure 5.9). This is in contrast with the mechanism where transcription of the gene is induced at the time point where the mRNA abundance level of the particular gene is at its maximum. Both mechanisms are not mutually exclusive and it is likely *P. falciparum* utilizes both mechanisms. However, an overrepresentation of repressive motifs observed from the screening of motifs in this thesis may indicate that repression of transcription may be a major mechanism of transcriptional regulation in *P. falciparum* (Figure 4.10).

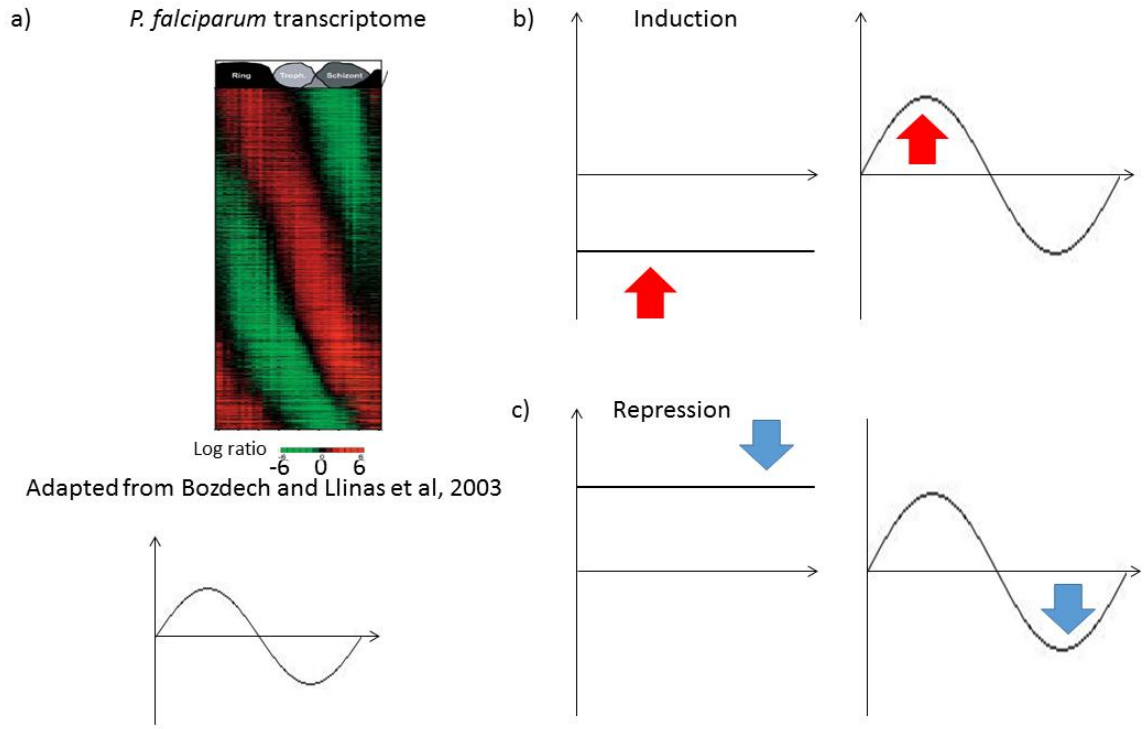


Figure 6.1 Repression as a mode of transcriptional regulation

a) *P. falciparum* transcriptome (adapted from Bozdech and Llinas at al., 2003 [19]). The *P. falciparum* transcriptome is characterized by a sine wave-like transcript expression pattern of each gene. 60-80% of the genes in the *P. falciparum* genome are transcriptionally active in the IDC. The transcript abundance of each gene peaks at a certain point in the IDC. b) Induction of transcription. The sine wave-like transcript expression pattern may be mediated by induction of transcription at the point where transcript abundance is at the maximum. c) Repression of transcription. On the other hand, the sine wave-like transcript expression pattern may also be mediated by repression of transcription at the point where transcript abundance is at the minimum.

There have been several lines of evidence previously reported in *P. falciparum* which supports the notion of repression as a mode of transcriptional regulation [55, 57, 98, 99, 303]. These studies support our hypothesis that gene promoters in *P. falciparum* are kept in a transcriptionally active state during the IDC and fine tuning of transcripts in the IDC temporally is then possibly regulated by the repression of transcription. These lines of evidence include (1) 'poised' assembly of basal transcriptional machinery over gene

promoters and (2) global permissive chromatin over the *P. falciparum* genome. These notions will be discussed in detail in the following sections.

6.2 Basal transcriptional machinery assembly over *P. falciparum* gene promoters

Almost all of the orthologs of the basal transcriptional machinery have been readily found in the *P. falciparum* genome [301]. However, a striking observation is that most of the TAFs in the TFIID complex are not found (Figure 1.1) [175]. TAFs have been shown to be regulating the selective recruitment of the PIC to specific core promoters in eukaryotic systems for spatiotemporal regulation of genes [184, 185]. However, without a variety of TAFs found in the *P. falciparum* genome, it could suggest that transcriptional regulation does not occur on the level of transcription initiation. Hence, it could be speculated that the PIC is poised over promoter regions in *P. falciparum*. This speculation is supported by another study which demonstrated that components of the PIC are assembled over both active and inactive promoters in the IDC.

Although a study on a small number of genes, PfTBP and PfTFIIB are demonstrated to bind to promoter regions of IDC transcribed genes regardless of stage specificity [303]. For example, PfTBP and PfTFIIB are assembled on promoter regions of ring specific genes in both ring and trophozoite stages [303]. However, this did not occur on the promoters of sporozoite and gametocyte specific genes. This may suggest that the PIC (represented here by TBP and TFIIB) may be assembled over promoter regions of IDC transcribed genes regardless whether they are maximally expressed at that stage or not. Again, this supports the notion that transcriptional initiation plays a minor role in *P. falciparum* transcriptional regulation.

However, it must be noted that this study was done on a limited group of genes, 2 ring specific genes, 2 trophozoite specific genes, 1 constitutively expressed gene, 1 sporozoite specific gene and 1 gametocyte specific gene [303]. Hence, this may or may not be a good representative of the whole genome. Next, authors did not directly show that RNAPII was also assembled in the same way as TBP and TFIIB. However, this study still provided indicative evidence that the PIC (or part of the PIC) is poised on promoters regions of IDC transcribed genes during the IDC. It would be important to characterize the binding of the PIC at a genome-wide scale in the future and elucidate whether RNAPII is associated with the PIC.

6.3 Permissive Chromatin

In a genome-wide study profiling the occupancy of H3K4me3 and H3K9ac histone modifications in the *P. falciparum* genome revealed that the 2 histone marks cover 91% of the *P. falciparum* genome [55]. In other eukaryotes, H3K4me3 and acetylation of histones are associated with actively transcribed genes and characterized as euchromatin or permissive chromatin histone marks [54, 411]. This indicates that most part of the *P. falciparum* genome is kept in euchromatic state which is transcriptionally permissive. In several other studies, it is reported that the presence of certain euchromatic marks correlate at varying degrees with mRNA abundance levels and are found to be highly abundant in the *P. falciparum* epigenome [56, 57]. However, it is still unclear if the histone marks precede transcription or the act of transcription causes changes in histone modifications.

The *P. falciparum* genome is characterized with high AT-richness, reaching 90% AT content in the intergenic regions [9]. It is reported that *P. falciparum* intergenic regions in general and

promoter regions are relatively nucleosome depleted [98, 99]. However, this observation requires further validation as there is a conflicting report [56]. In support of a nucleosome depleted intergenic region, studies from other eukaryotic systems have suggested that poly(dA:dT) tracts are rigid and intrinsically disfavor the formation of nucleosomes *in vitro* and nucleosomes are depleted over genomic regions which contain poly(dA:dT) tracts *in vivo* [101, 412, 413]. In contrast, it was also shown that moderately GC-rich sequences favor the formation of nucleosomes [414, 415]. Poly(dA:dT) tracts are highly frequent in the intergenic regions in the *P. falciparum* genome, in which tracts more than 10 bp make up 5% of the genome [378]. Amidst controversies, it may be likely that nucleosomes are depleted in the intergenic regions. As the intergenic regions are nucleosome depleted, *P. falciparum* promoter regions are highly accessible to transcription factors and the PIC. In further support, a study has shown a positive correlation between the AT-richness of a promoter and transcript levels of its associated gene [79].

Taken together, the epigenetic landscape seems to suggest that permissive chromatin covers the genome of *P. falciparum*. A highly permissive epigenome which allows DNA elements to be accessible to transcriptional machinery is in support that the *P. falciparum* genome is kept in a transcriptionally active state and stage specific expression of genes could be achieved through repression of transcription (Figure 6.1).

6.4 Post transcriptional regulation of transcripts

The stage specific expression of IDC genes from transcriptomic data is a measure of steady state mRNA levels [19]. However, steady state mRNA levels may not necessarily reflect transcriptional activity completely. mRNA transcript stability may play a role that is downstream of transcription and also affecting steady state mRNA levels. Transcriptional

activity could be more directly measured by the labeling and profiling of nascent transcripts [416]. At the time of preparation of this thesis, there are no studies on genome-wide nascent mRNA profiling in *P. falciparum* published. However, a study reported of global increases in mRNA half-life as the parasite progresses through the IDC from ring to schizont stages but did not report of mRNA decay rates fine tuning steady state mRNA levels [417]. In this study, mRNA stability of transcripts was measured by a genome-wide survey of transcript levels after RNAPII was inhibited by α-amanatin. It was reported that mRNA half-lives for all transcripts increase as the parasite progresses through the IDC. This is indicative of a global scale post-transcriptional regulation of transcripts rather than gene-specific regulation of transcripts.

In another study, authors measured transcriptional activity by nuclear run on experiments on about 20 gene transcripts [418]. Although this is a not a global study, comparisons between transcriptional activity and mRNA steady state abundance of the gene transcripts revealed that while there is a good proportion of genes where transcriptional activity correlated well with mRNA abundance, authors also reported a group of genes where transcriptional activity differed significantly from mRNA abundance which the authors proposed that post-transcriptional regulation plays a more important role in thesis genes [418]. Although, an unpublished result on the profiling of nascent mRNA in *P. falciparum* has revealed that transcriptional activity correlates well with steady state mRNA levels at least for the genes of the promoters that have been used in the stable transfection experiments (Painter et al., unpublished data, personal communication). Hence, regulation at both levels of transcription and mRNA stability may contribute to the mRNA abundance profiles in *P. falciparum*. It is still an open question whether either will take a center stage in *P. falciparum* gene regulation.

6.5 Other possible mechanisms of transcriptional regulation

A mode of transcriptional regulation that has emerged in the recent years is the regulation of transcriptional elongation by RNAPII proximal-promoter pausing [419]. In this mechanism, RNAPII is stalled at the promoter regions of genes and will be released for transcription upon receiving stimulus [294]. The release is mediated through the recruitment of release factors and also modifications to the RNAPII [420]. Although yet to be verified in a genome-wide scale and whether RNAPII is associated, the binding of PIC basal transcriptional machinery regardless of seems to indicate that the PIC is poised on gene promoters [303]. In addition, permissive chromatin allows transcription associated factors to access DNA easily [55, 99]. From the data presented in this thesis, we have observed that a large number of motifs are repressive in nature and the repressive effects are stage specific (Figure 4.10 and 5.9). A likely mechanism may be proposed that repressive motifs bind transcription factors which prevent the binding of RNAPII to the poised PIC on promoters and/or prevent elongation of RNAPII if RNAPII is already poised together with the PIC at the promoter regions to achieve stage specific regulation of genes in the IDC.

Although the field of long range enhancer-promoter interactions is still young in the *Plasmodium* species, a recent study using chromosome capture techniques coupling to NGS has revealed that apart from heterochromatin clusters in *var* gene regulation and rDNA loci cluster, there are few long range interactions detected [169]. Improvements to the current methods to detect 3-dimensional chromatin structures in *P. falciparum* has revealed strongest clustering of centromeres, telomeres and rDNA loci, but provided little information about other long range contacts [421]. Hence, more studies to identify long range interactions needs to be done for the *Plasmodium* parasites. In addition, most of the components in the eukaryotic Mediator complex, which are highly conserved in the eukaryotic cells, are not

found in the *P. falciparum* genome. Mediator, as detailed in the Introduction chapter, is responsible to relay signals from the enhancer elements to affect the actual gene transcription in other eukaryotic systems [422]. Although the lack of conservation of consensus sequence of Mediator components in the *P. falciparum* genome may be a factor, the lack of a Mediator complex could suggest that long range enhancer-promoter interactions may be a minor role in transcriptional regulation. In contrast, fine tuning gene transcription could be achieved by regulatory elements which are found in the target gene locality.

6.6 Future directions

An interesting observation made in this thesis and several other studies has revealed a combinatorial mode of transcription regulation in *P. falciparum* where multiple DNA motifs have been found on a single promoter (Supplementary Figure 3) [357, 358]. It is likely that the fine tuning of gene transcription is mediated by the binding of multiple transcription factors on multiple motifs in *P. falciparum* promoters. In this thesis, we have only managed to provide evidence of the manipulation of a single motif on a promoter (Figure 4.10, 5.1-5.6 and 5.9). Hence, it would be worthy to investigate the role of multiple motifs in the same promoter in future studies. This could be done through deleting or mutating multiple motifs on the same promoter and understanding the effects of the mutations on transcriptional activity of the promoter.

Through our EMSA experiments, we have detected that motifs 3, 6, 16 and 20 have the ability to bind protein elements (Figure 5.10). It is likely that these proteins are specific transcription factors which elicit the effects of the motifs in regulating transcriptional activities. Hence, it would be important to elucidate the identity of these factors. This could be done through the means of DNA affinity chromatography of nuclear proteins using a

fragment of the promoter which contains the motif of interest, followed by mass spectrometry analysis to identify the proteins which bind to the motif of interest. It would be interesting to investigate whether these proteins which have the potential to fine tune gene expression in the IDC are the ApiAP2 transcription factor family, which are the most well characterized group of transcription factors in *P. falciparum* currently, or they may be an entirely different class of transcription factors which is yet to be reported.

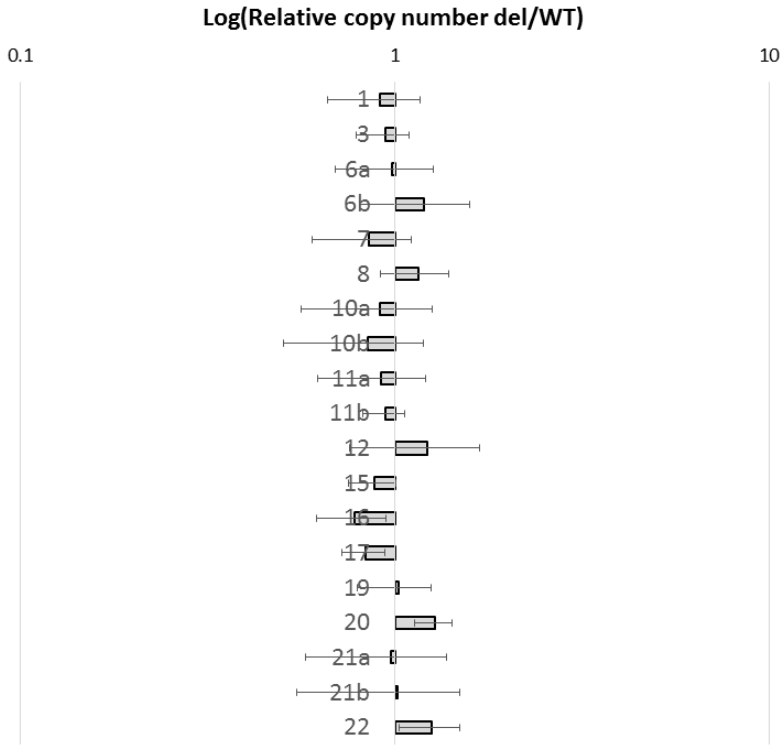
Epigenetics has been shown to be important in *P. falciparum* gene regulation in multiple studies as described in the first chapter of this thesis. It has also been shown in other eukaryotes that *cis*-regulatory elements correlate well with specific histone modifications [423]. Hence, it would be important to investigate the interplay between the DNA motifs newly described in this thesis and histone modifications in *P. falciparum*. Specifically, studies could be done to elucidate the changes in histone modifications in the vicinity of the DNA motif throughout the IDC. On the other hand, this could be also done through investigating the effects of the presence or absence of the DNA motif on the histone modification on the promoter. In addition, it has been shown in other eukaryotes that there is an association between transcription factor binding and nucleosome positioning at the transcription factor binding site [424, 425]. Hence, it would be also important to understand the effect in nucleosome positioning in the presence or absence of the newly described motifs in the vicinity of the DNA motifs throughout the IDC.

6.7 Conclusions

By deletion mutagenesis, we have characterized the *pfmrp2* promoter which is associated with differential regulation of an important drug resistant gene (Figure 3.3a). Through the experiments in this first part of the study, we have optimized a transfection and reporter assay

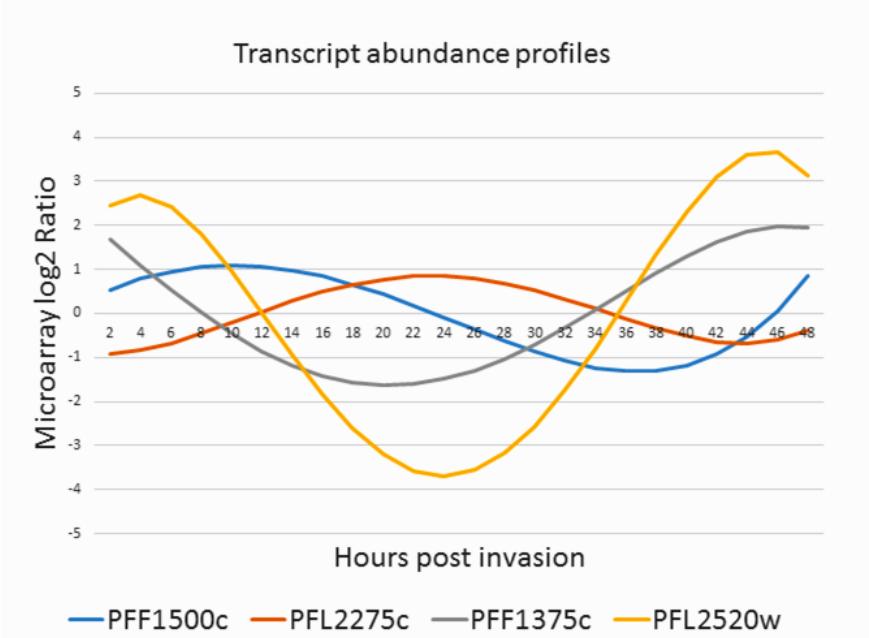
protocol which could be used to test for promoter activities (Section 2.1 and 2.2). In the next part of the study, we have utilized an algorithm to predict DNA motifs on 5' upstream regions of co-regulated genes (Figure 4.1). Using the protocol developed in the initial studies in this thesis, we screened the functional of the DNA motifs which are predicted *in silico* (Figure 4.10). We have observed an overrepresentation of motifs which play a repressive role in transcription (Figure 4.10). A series of site-direct mutagenesis assays has also verified the findings from the initial screening (Figure 5.1-5.5). In addition, it was observed that the same motif function similarly on multiple promoters (Figure 5.6). Importantly, we have also shown that the effects of the motifs were stage specific in the IDC and bind to potential transcription factors (Figure 5.9 and 5.10).

7. Appendix

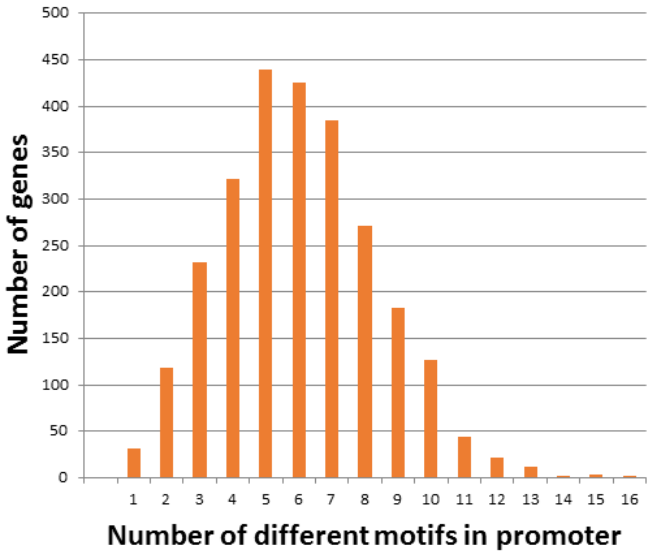


Supplementary Figure 1 Copy number differences between constructs containing the WT promoter and Del promoter (motif deleted).

Transient transfections were carried out as described in Chapter 2, 50 µg of pf86 (containing promoter of interest and *ffluc* gene) was co-transfected with pARL-5'3'-Re (containing *renilla luciferase* gene). After transfection, half of the parasites were lysed for dual luciferase assays. Total RNA and gDNA were harvested from the other half of the sample. Real time qPCR was then carried out to calculate copy number differences between constructs containing the WT promoter and Del promoter. A Ct value for the *ffluc* gene and *pfl0900c* which is representative of copy number from wild type and Del promoter constructs. The Ct for *ffluc* gene in each sample was normalized with its corresponding Ct for *pfl0900c* (Ct(norm) = Ct(*ffluc*) – Ct(*pfl0900c*)). The fold change of copy number differences obtained from mutant constructs relative to the transcripts obtained from wild type constructs was calculated by the ΔΔCt method, fold change = $2^{-(Ct(mutant\ norm) - Ct(wild\ type\ norm)}$). 2-tailed student's t-test was used to define whether the plasmid copy number from the mutant constructs were statistically different from the wild type constructs. Grey bars, statistically insignificant. Dark red/blue, $p<0.01$. Light red/blue, $p<0.05$. Error bars represent 3 independent transfections done on 3 separate days. None of the mutant constructs were found to be significantly different from the WT constructs in terms of copy number.



Supplementary Figure 2 The transcript abundance profiles obtained from previous microarray study [30].
The promoters of these genes were utilized to represent motif 6 and motif 21 in Figure 4.10.



Supplementary Figure 3 Number of different motifs predicted from this study present on a single *P. falciparum* promoter.
5-6 different motifs were found to be present on a single promoter for most of the *P. falciparum* promoters.

Primer name	Oligo sequence	Purpose
Chapter 3		
Mrp2_R(NcoI)	ATCG**CCATGG**AAAAACGGAATCTTTCTAACTTT	Reverse primer used for PCR amplification of *pfmrp2* promoter of all sizes
11C_2kb_F(XhoI)	ATCG**CTCGAG**AAATAGAACAATGGATAAAATCTCTTTG	Forward primer for 11C 2kb *pfmrp2* promoter
11C_1.8kb_F(XhoI)	ATCG**CTCGAG**GTGAATGTGCCTTATAAACAAC	Forward primer for 11C 1.8kb *pfmrp2* promoter
11C_1.5kb_F(XhoI)	ATCG**CTCGAG**GTAGTAAATATATATTCTGTTTTGTGCT	Forward primer for 11C 1.5kb *pfmrp2* promoter
11C_1.2kb_F(XhoI)	ATCG**CTCGAG**ATATGTATTTATTTGTATGAGCTGTG	Forward primer for 11C 1.2kb *pfmrp2* promoter
11C_0.9kb_F(XhoI)	ATCG**CTCGAG**GAAAAATATTTATTTTGTATTTAGAAGG	Forward primer for 11C 0.9kb *pfmrp2* promoter
11C_0.6kb_F(XhoI)	ATCG**CTCGAG**TTAACATCATAAATATTTATGAAATACG	Forward primer for 11C 0.6kb *pfmrp2* promoter
11C_0.5kb_F(XhoI)	ATCG**CTCGAG**TCTAAATATTAACATAGCAGTTTTGA	Forward primer for 11C 0.5kb *pfmrp2* promoter
11C_0.4kb_F(XhoI)	ATCG**CTCGAG**TTTAATTAGTCTCCTCTTTATTTATTTATA	Forward primer for 11C 0.4kb *pfmrp2* promoter
11C_0.3kb_F(XhoI)	ATCG**CTCGAG**AAGAAATACAGCACCTCCAG	Forward primer for 11C 0.3kb *pfmrp2* promoter
6A_2kb_F(XhoI)	ATCG**CTCGAG**ACACCTTTCTTTACATTTCGG	Forward primer for 6A 2kb *pfmrp2* promoter
6A_1.8kb_F(XhoI)	ATCG**CTCGAG**GCCAGATGAAAAGCTTCTG	Forward primer for 6A 1.8kb *pfmrp2* promoter
6A_1.5kb_F(XhoI)	ATCG**CTCGAG**CATACATCCGATGCGC	Forward primer for 6A 1.5kb *pfmrp2* promoter
6A_1.0kb_F(XhoI)	ATCG**CTCGAG**CATTGGAAGTGCATATTACTTT	Forward primer for 6A 1.0kb *pfmrp2* promoter
6A_0.8kb_F(XhoI)	ATCG**CTCGAG**GGAAGGTATAAATATTTTTCAAGAA	Forward primer for 6A 0.8kb *pfmrp2* promoter
6A_0.5kb_F(XhoI)	ATCG**CTCGAG**CTCATATGCCAAGAATTTCTTA	Forward primer for 6A 0.5kb *pfmrp2* promoter

pfmrp2-_F(NcoI)	ATCG**CCATGG**CATACATCCGATGCGC	Forward primer for inverted 6A 1.5kb *pfmrp2* promoter
pfmrp2-_R(XhoI)	ATCG**CTCGAG**AAAAACGGAATCTTTCTA ACTTT	Reverse primer for inverted 6A 1.5kb *pfmrp2* promoter
Chapter 4		
PFI0535w(wt)_F (XhoI)	ATCG**CTCGAG**ATGTGTAATATATTTTATT TATATTTACATGAT	Forward primer for *pfl0535w* promoter with motif
PFI0535w(del)_F (XhoI)	ATCG**CTCGAG**ATATATTTTATTTATATTT ACATG ATATATTCATAT	Forward primer for *pfl0535w* promoter without motif
PFI0535w_R(NcoI)	ATCG**CCATGG**CATTATTTTATCCGTTCAT GAA	Reverse primer for *pfl0535w* promoter
PFI0530c(wt)_F (XhoI)	ATCG**CTCGAG**TACACATAATTGAATGAG ATATTT ATTTTAAT	Forward primer for *pfl0530c* promoter with motif
PFI0530c(del)_F (XhoI)	ATCG**CTCGAG**AATTGAATGAGATATTTA TTTTAA TAAAGG	Forward primer for *pfl0530c* promoter without motif
PFI0530c_R(NcoI)	ATCG**CCATGG**CATTTTAGCACTTATTATA AGATA TCAAAA	Reverse primer for *pfl0530c* promoter
Hsp86_5'F(XhoI)	ATCG**CTCGAG**GCCTTGATATATTTTTAGA TATATGGATTA	Forward primer for *hsp86* promoter
Hsp86_3'R(NcoI)	ATCG**CCATGG**AGCCATGGCCGTTGAGAT TTTATTCGAAATGTGGG	Reverse primer for *hsp86* promoter
Hsp86_g-box-del_5'R	TTTTTTTTTTTTTTTAAAACACCTTC	Reverse primer to amplify 5' fragment of *hsp86* promoter without G-box motif
Hsp86_g-box-del_3'F	GGAAAGGGGCCATTGGATATATATTTAG TATTCC	Forward primer to amplify 3' fragment of *hsp86* promoter without G-box motif
Hsp86_1base-del_5'R	TTTTTTTTTTTTTTAAAACACCTTCA	Reverse primer to amplify 5' fragment of *hsp86* promoter with 1 base deleted
Hsp86_1base-del_3'F	AAAAAAGCCCCGCGGAAAGGGGCCATTG GATATATATTTAGTATTCC	Forward primer to amplify 3' fragment of *hsp86* promoter with 1 base deleted
Hsp86_6base-del_5'R	TTTTTTTTTAAAACACCTTCAATATA	Reverse primer to amplify 5' fragment of *hsp86* promoter with 1 bases deleted
Hsp86_6base-	AAAAAAGCCCCGCGGAAAGGGGCCATTG	Forward primer to

del_3'F	GATATATATTTAGTATTCC	amplify 3' fragment of *hsp86* promoter with 1 bases deleted
Motif1_PF14_0156_5'F(XhoI)	ATGC**CCATGG**ACATAACATATCATATAAAATTATGCTCTATT	Forward primer for *pf14_0156* promoter
Motif1_PF14_0156_3'R(NcoI)	ATGC**CTCGAG**GTGGATATGTATACATCGAAAATATATTAT	Reverse primer for *pf14_0156* promoter
Motif1_PF14_0156_5'R	AAAATTTTTAATTCCAAATATTTTCTG	Reverse primer to amplify 5' fragment of *pf14_0156* promoter without motif 1
Motif1_PF14_0156_3'F	TAAAATAAAAATATATATATATATATATATTATTTATATATATTTAATTTG	Forward primer to amplify 3' fragment of *pf14_0156* promoter without motif 1
Motif3_PFL1885c_5'F(XhoI)	ATGC**CCATGG**TTTTTTTATTTTTTCTTTTTAAACAATAA	Forward primer for *pfl1885c* promoter
Motif3_PFL1885c_3'R(NcoI)	ATGC**CTCGAG**ACATTTATTTCATTTTGCTTTTACTAC	Reverse primer for *pfl1885c* promoter
Motif3_PFL1885c_5'R	AATATATTTACTATTATCTATTTAGTTATATATATATATATAT	Reverse primer to amplify 5' fragment of *pfl1885c* promoter without motif 3
Motif3_PFL1885c_3'F	TATATATATCTTTACATTTTGATGACTT	Forward primer to amplify 3' fragment of *pfl1885c* promoter without motif 3
Motif6a_PFF1500c_5'F(XhoI)	ATGC**CCATGG**TCAAAAAAAAAAAAAAATAAATAAATATTAG	Forward primer for *pff1500c* promoter
Motif6a_PFF1500c_3'R(NcoI)	ATGC**CTCGAG**CCTTATGCATAATTTATATTCAGAAAC	Reverse primer for *pff1500c* promoter
Motif6a_PFF1500c_5'R	AGTATAAAATTATATTGCATATTTGAAATG	Reverse primer to amplify 5' fragment of *pff1500c* promoter without motif 6a
Motif6a_PFF1500c_3'F	AGAAATTAAAATGTAACGGTCAAA	Forward primer to amplify 3' fragment of *pff1500c* promoter without motif 6a
Motif6b_PFL2275c_5'F(XhoI)	ATGC**CCATGG**ATTATAAATAAATAGAAAAAAAAATGATAAATG	Forward primer for *pfl2275c* promoter
Motif6b_PFL2275c_3'R(NcoI)	ATGC**CTCGAG**CGTGTGCGATTAAGTGTATCTT	Reverse primer for *pfl2275c* promoter
Motif6b_PFL2275c_5'R	TTGTAACTATACACACATATATATATTATAAATATATAT	Reverse primer to amplify 5' fragment of *pfl2275c* promoter

		without motif 6b
Motif6b_PFL2275c_3'F	AAGGAAGAAAAAAAATAATTATGATAC	Forward primer to amplify 3' fragment of *pfl2275c* promoter without motif 6b
Motif7_PF14_0510_5'F(XhoI)	ATGC**CCATGG**GTTGTTTAAAATATAATTAAGTTTATAAAAAATAA	Forward primer for *pf14_0510* promoter
Motif7_PF14_0510_3'R(NcoI)	ATGC**CTCGAG**CATACGAAGCATGCTGAATT	Reverse primer for *pf14_0510* promoter
Motif7_PF14_0510_5'R	GTAATATATAAAATAAATTAAATATTGTGTTTTG	Reverse primer to amplify 5' fragment of *pf14_0510* promoter without motif 7
Motif7_PF14_0510_3'F	TTCTCAAATAGAAAAAAAAAATTTATT	Forward primer to amplify 3' fragment of *pf14_0510* promoter without motif 7
Motif8_PF14_0142_5'F(XhoI)	ATGC**CCATGG**AATAAATATTTACTTGAGCAAATTTATATTA	Forward primer for *pf14_0142* promoter
Motif8_PF14_0142_3'R(NcoI)	ATGC**CTCGAG**CATTTTAGTGTGTAGATTTTGTTGAT	Reverse primer for *pf14_0142* promoter
Motif8_PF14_0142_5'R	ATTTATTTATTTCTTTTTCGTTTTTC	Reverse primer to amplify 5' fragment of *pf14_0142* promoter without motif 8
Motif8_PF14_0142_3'F	TTGTTGTTAATATTACGTGAAATATAATC	Forward primer to amplify 3' fragment of *pf14_0142* promoter without motif 8
Motif10a_PFI1375w_5'F(XhoI)	ATGC**CCATGG**ATTTAAATGAAAATAATATAATACGTATATTTTATT	Forward primer for *pfl1375w* promoter
Motif10a_PFI1375w_3'R(NcoI)	ATGC**CTCGAG**GACCAATTTGGTAAGTGGTAAAG	Reverse primer for *pfl1375w* promoter
Motif10a_PFI1375w_5'R	TTCCCATTTAATGATGAATTATT	Reverse primer to amplify 5' fragment of *pfl1375w* promoter without motif 10a
Motif10a_PFI1375w_3'F	TAAAAAAAAAAAAATATATATATATATATATATTATATATTTA	Forward primer to amplify 3' fragment of *pfl1375w* promoter without motif 10a
Motif10b_PF14_0373_5'F(XhoI)	ATGC**CCATGG**TATATATGTAAATAAATAAATAAATAAATATAAATATAAATATAT	Forward primer for *pf14_0373* promoter
Motif10b_PF14_0373_3'R(NcoI)	ATGC**CTCGAG**GGGAATAACAATAGGAATAGAAATAG	Reverse primer for *pf14_0373* promoter
Motif10b_PF14_037	TATATATAATACATATATATATAATACAA	Reverse primer to

3_5'R	AACAATGAAA	amplify 5' fragment of *pf14_0373* promoter without motif 10b
Motif10b_PF14_0373_3'F	TTATTTTTTTTTAATATACATTTAATACAACG	Forward primer to amplify 3' fragment of *pf14_0373* promoter without motif 10b
Motif11a_PFB0310c_5'F(XhoI)	ATGC**CCATGG**TTTTATACAACTTATAATACATCTGAAAATATAA	Forward primer for *pfb0310c* promoter
Motif11a_PFB0310c_3'R(NcoI)	ATGC**CTCGAG**CTGTTCATAATAATATTTTATATAGCTAGAATAAA	Reverse primer for *pfb0310c* promoter
Motif11a_PFB0310c_5'R	TAATATATATATATATATATTATTTCGTA CATAATTATTT	Reverse primer to amplify 5' fragment of *pfb0310c* promoter without motif 11a
Motif11a_PFB0310c_3'F	GGTATTTTATTATATTTTTATTTTACACCTC	Forward primer to amplify 3' fragment of *pfb0310c* promoter without motif 11a
Motif11b_PFE1510c_5'F(XhoI)	ATGC**CCATGG**TTTTAATTATTTATATAACGATAATCCTTACAT	Forward primer for *pfe1510c* promoter
Motif11b_PFE1510c_3'R(NcoI)	ATGC**CTCGAG**AGAAAAATTTTATACCATTATTATAATAAAGAA	Reverse primer for *pfe1510c* promoter
Motif11b_PFE1510c_5'R	ATATAAATTAAAAAAAAATAATATATAAATGAATAA	Reverse primer to amplify 5' fragment of *pfe1510c* promoter without motif 11b
Motif11b_PFE1510c_3'F	ATAACTTTTATGAAATACACATATATATGAA	Forward primer to amplify 3' fragment of *pfe1510c* promoter without motif 11b
Motif12_MALP1.137_5'F(XhoI)	ATGC**CTCGAG**GAGAATTTGTCATTTATGAGAATATAGTAA	Forward primer for *malp1.137* promoter
Motif12_MALP1.137_3'R(NcoI)	ATGC**CCATGG**TTATATTAATTTGTGTAAATGCATATGATG	Reverse primer for *malp1.137* promoter
Motif12_MALP1.137_5'R	TGATAAAAAAGAGAAGAAAAATAATTATATA	Reverse primer to amplify 5' fragment of *malp1.137* promoter without motif 12
Motif12_MALP1.137_3'F	TTTGTATAATAATTTTATCTTGTAAAAAATG	Forward primer to amplify 3' fragment of *malp1.137* promoter without motif 12
Motif15_PFI0695c_5'F(XhoI)	ATGC**CCATGG**TTTTCTTTTTATCGGCTTCTCT	Forward primer for *pfl0695c* promoter
Motif15_PFI0695c_3'R(NcoI)	ATGC**CTCGAG**CATTTGTTGTTTTTTTCTAGGAATT	Reverse primer for *pfl0695c* promoter

Motif15_PFI0695c_5’R	TATATTTTTTCATTTTCTTTTTTTATTT	Reverse primer to amplify 5’ fragment of *pfl0695c* promoter without motif 15
Motif15_PFI0695c_3’F	GAGCTATTAATATTTATATAATACATAGATATATGTTT	Forward primer to amplify 3’ fragment of *pfl0695c* promoter without motif 15
Motif16_PF14_0761_5’F(XhoI)	ATGC**CCATGG**TTAAAAACAACAAAAATACGTTAATATTATATA	Forward primer for *pf14_0761* promoter
Motif16_PF14_0761_3’R(NcoI)	ATGC**CTCGAG**GGTATAATGTATATATTGCAATATATGTATGAT	Reverse primer for *pf14_0761* promoter
Motif16_PF14_0761_5’R	CGTACAATAAATTATTCTATGGTTATTGAAA T	Reverse primer to amplify 5’ fragment of *pf14_0761* promoter without motif 16
Motif16_PF14_0761_3’F	TATTTTTTATTTTTATTTTTAATTTTTTTTT	Forward primer to amplify 3’ fragment of *pf14_0761* promoter without motif 16
Motif17_PF10_0100_5’F(XhoI)	ATGC**CCATGG**AACAGTTAAAAAGAAATTGTTCTCC	Forward primer for *pf10_1010* promoter
Motif17_PF10_0100_3’R(NcoI)	ATCG**CCATGG**TTTTTTAATTTCTCTTTAAAATAATAAATAAA	Reverse primer for *pf10_1010* promoter
Motif17_PF10_0100_5’R	AACATGTAATTTTTTTTTTTAAATATAATATTAT	Reverse primer to amplify 5’ fragment of *pf10_1010* promoter without motif 17
Motif17_PF10_0100_3’F	TATTTTTTTTTTTTTTTTTTTATAATATATATTT	Forward primer to amplify 3’ fragment of *pf10_1010* promoter without motif 17
Motif19_PFI1335w_5’F(XhoI)	ATGC**CCATGG**TTTTTTCTTCTTTTTGATATAGAACAG	Forward primer for *pfl1335w* promoter
Motif19_PFI1335w_3’R(NcoI)	ATGC**CTCGAG**GTCTAGTTTTACTATCCTTAATTATTTTTCTT	Reverse primer for *pfl1335w* promoter
Motif19_PFI1335w_5’R	AACACATGAATATAGATGTATATATATATATATATAGG	Reverse primer to amplify 5’ fragment of *pfl1335w* promoter without motif 19
Motif19_PFI1335w_3’F	TATACATAATTTTTTTTTTTTTTTTT	Forward primer to amplify 3’ fragment of *pfl1335w* promoter without motif 19
Motif20_PFF1010c_5’F(XhoI)	ATGC**CCATGG**TGTGATATTTTATTAGTATGGTGGTAAT	Forward primer for *pff1010c* promoter

Motif20_PFF1010c_3'R(NcoI)	ATCG**CCATGG**TTTTGTAATTATATTAATATATATATATATATATATATATATGTGA	Reverse primer for *pff1010c* promoter
Motif20_PFF1010c_5'R	ATTTATTAAAGTTTTACAAAATTATTATTTTATG	Reverse primer to amplify 5' fragment of *pff1010c* promoter without motif 20
Motif20_PFF1010c_3'F	ATAAAGTGTACAATCTGTCAAAATAAAA	Forward primer to amplify 3' fragment of *pff1010c* promoter without motif 20
Motif21a_PFF1375c_5'F(XhoI)	ATGC**CTCGAG**GAGCATTGTAATTATTCTTATTCCTATTT	Forward primer for *pff1375c* promoter
Motif21a_PFF1375c_3'R(NcoI)	ATGC**CCATGG**TTTTATATATTATAAATTTAAAATACTTAAAAGAAAA	Reverse primer for *pff1375c* promoter
Motif21a_PFF1375c_5'R	ACACAATAAAAAATGAAAATTCAATAT	Reverse primer to amplify 5' fragment of *pff1375c* promoter without motif 21a
Motif21a_PFF1375c_3'F	GAAATAAAAAACAAATATATAATAAAAAAAAA	Forward primer to amplify 3' fragment of *pff1375c* promoter without motif 21a
Motif21b_PFL2520w_5'F(XhoI)	ATGC**CCATGG**AGTTCATATATTTTAAAGCATGCATT	Forward primer for *pfl2520w* promoter
Motif21b_PFL2520w_3'R(NcoI)	ATGC**CTCGAG**GTAATATTATATTCATTCCATATATCTTTTATTTAA	Reverse primer for *pfl2520w* promoter
Motif21b_PFL2520w_5'R	ATTCAAATCATCTCAAAATGTGTAA	Reverse primer to amplify 5' fragment of *pfl2520w* promoter without motif 21b
Motif21b_PFL2520w_3'F	ATATAAATAATAATATTAATATATTGTTAAGAAAAAAAA	Forward primer to amplify 3' fragment of *pfl2520w* promoter without motif 21b
Motif22_PF10_0281_5'F(XhoI)	ATGC**CCATGG**CTTTTTTCATTCTATTATTATAACATAACAAC	Forward primer for *pf10_0281* promoter
Motif22_PF10_0281_3'R(NcoI)	ATGC**CTCGAG**GAAGGGGAAATTGGAAAATT	Reverse primer for *pf10_0281* promoter
Motif22_PF10_0281_5'R	ATAGAAAAATTATATCACTATAAATGTTCATG	Reverse primer to amplify 5' fragment of *pf10_0281* promoter without motif 22
Motif22_PF10_0281_3'F	ATTATTTTTAATTTATATTTATATCCACATAAAATA	Forward primer to amplify 3' fragment of *pf10_0281* promoter without motif 22
Chapter 5		

Motif3_PFL1885c_3’F(mut1)	ACTGACTTCTATATTATATATATCTTTACATTTTGATGACTTAATAAT	Forward primer to amplify 3’ fragment of mutant *pfl1885c* promoter (mut1)
Motif3_PFL1885c_3’F(mut2)	ACTGACTTATATATATCTTTACATTTTGATGACTTAATAAT	Forward primer to amplify 3’ fragment of mutant *pfl1885c* promoter (mut2)
Motif3_PFL1885c_3’F(mut3)	TATATATTATATATATCTTTACATTTTGATGACTTAATAAT	Forward primer to amplify 3’ fragment of mutant *pfl1885c* promoter (mut3)
Motif6_PFF1500c_5’R(mut1)	TCAGTCAGTTCCTATATAAGTATAAAATTATATTGCATATTTGAAATG	Reverse primer to amplify 5’ fragment of mutant *pff1500c* promoter (mut1)
Motif6_PFF1500c_5’R(mut2)	TCAGTCAGTAGTATAAAATTATATTGCATATTTGAAATG	Reverse primer to amplify 5’ fragment of mutant *pff1500c* promoter (mut2)
Motif6_PFF1500c_5’R(mut3)	TTATATATAAGTATAAAATTATATTGCATATTTGAAATG	Reverse primer to amplify 5’ fragment of mutant *pff1500c* promoter (mut3)
Motif6_PFF1500c_5’R(mut2x)	TCCTATATATCCTATATAAGTATAAAATTATATTGCATATTTGAAATG	Reverse primer to amplify 5’ fragment of mutant *pff1500c* promoter (mut2x)
Motif6_PFF1500c_5’R(mut3x)	TCCTATATATCCTATATATCCTATATAAGTATAAAATTATATTGCATATTTGAAATG	Reverse primer to amplify 5’ fragment of mutant *pff1500c* promoter (mut3x)
Motif16_PF14_0761_5’R(mut1)	CAGTCAGTAGCATGCACGTACAATAAATTATTCTATGGTTATTGAAAT	Reverse primer to amplify 5’ fragment of mutant *pf14_0761* promoter (mut1)
Motif16_PF14_0761_5’R(mut2)	CAGTCAGTCGTACAATAAATTATTCTATGGTTATTGAA	Reverse primer to amplify 5’ fragment of mutant *pf14_0761* promoter (mut2)
Motif16_PF14_0761_5’R(mut3)	AGCATTTTCGTACAATAAATTATTCTATGGTTATTGAAAT	Reverse primer to amplify 5’ fragment of mutant *pf14_0761* promoter (mut3)
Motif16_PF14_0761_5’R(mut4)	ATTTTGCACGTACAATAAATTATTCTATGGTTATTGAAAT	Reverse primer to amplify 5’ fragment of mutant *pf14_0761* promoter (mut4)
Motif16_PF14_0761_5’R(mut5)	AGCATGCCCGTACAATAAATTATTCTATGGTTATTGAAAT	Reverse primer to amplify 5’ fragment of mutant *pf14_0761* promoter (mut5)

Motif16_PF14_0761 _5'R(mut2x)	AGCATGCAAGCATGCACGTACAATAAAT TATTCTATGGTTATTGAAAT	Reverse primer to amplify 5' fragment of mutant *pf14_0761* promoter (mut2x)
Motif16_PF14_0761 _5'R(mut3x)	AGCATGCAAGCATGCAAGCATGCACGTA CAATAAATTATTCTATGGTTATTGAAAT	Reverse primer to amplify 5' fragment of mutant *pf14_0761* promoter (mut3x)
Motif17_PF10_0100 _3'F(mut1)	ACTGACTGTATTTTTTTTTTTTTTTTTTTAT AATATATATTT	Forward primer to amplify 3' fragment of mutant *pfl10_0100* promoter (mut1)
Motif17_PF10_0100 _3'F(mut2x)	TTTACAACTTTACAACTATTTTTTTTTTTT TTTTTTTATAATATATATTT	Forward primer to amplify 3' fragment of mutant *pfl10_0100* promoter (mut2x)
Motif17_PF10_0100 _3'F(mut3x)	TTTACAACTTTACAACTTTACAACTATTTT TTTTTTTTTTTTTTTATAATATATATTT	Forward primer to amplify 3' fragment of mutant *pfl10_0100* promoter (mut3x)
Motif20_PFF1010c_ 5'R(mut1)	TCAGTCAGTATTTATTAAAGTTTTACAAA ATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut1)
Motif20_PFF1010c_ 5'R(mut2)	ACATGGGGGATTTATTAAAGTTTTACAAA ATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut2)
Motif20_PFF1010c_ 5'R(mut3)	AGGGGTTTTATTTATTAAAGTTTTACAAA ATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut3)
Motif20_PFF1010c_ 5'R(mut4)	ACATATTTTATTTATTAAAGTTTTACAAA ATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut4)
Motif20_PFF1010c_ 5'R(mut5)	ACAGGTTTTATTTATTAAAGTTTTACAAA ATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut5)
Motif20_PFF1010c_ 5'R(mut6)	ACGTGTTTTATTTATTAAAGTTTTACAAA ATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut6)
Motif20_PFF1010c_ 5'R(mut7)	AAATGTTTTATTTATTAAAGTTTTACAAA ATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut7)
Motif20_PFF1010c_ 5'R(mut2x)	ACATGTTTTACATGTTTTATTTATTAAAGT TTTACAAAATTATTATTTTA	Reverse primer to amplify 5' fragment of mutant *pff1010c* promoter (mut2x)
Motif20_PFF1010c_ 5'R(mut2x)	ACATGTTTTACATGTTTTACATGTTTTATT TATTAAAGTTTTACAAAATTATTATTTTA	Reverse primer to amplify 5' fragment of

		mutant *pff1010c* promoter (mut3x)
Motif3_PF08_0044_5'F(XhoI)	ATCG**CTCGAG**GTACGTATTTGTAAAAAGATATACATGATA	Forward primer for *pf08_0044* promoter
Motif3_PF08_0044_5'R	ATAAAAGACAATTAAACATTTTTAAATTTAT	Reverse primer for 5' fragment of *pf08_0044* promoter without motif 3
Motif3_PF08_0044_3'F	ATAAATATAAATTGCATAATTTAAAAGACTT	Forward primer for 3' fragment of *pf08_0044* promoter without motif 3
Motif3_PF08_0044_3'R(NcoI)	ATCG**CCATGG**CATGAAGAAAATCTGAAGGGTAA	Reverse primer for *pf08_0044* promoter
Motif3_PFF1145c_5'F(XhoI)	ATCG**CTCGAG**CTTCTAATGATGATATCATTTATTTTATG	Forward primer for *pff1145c* promoter
Motif3_PFF1145c_5'R	GTTAGAAATAAAATGTAAAATTAAAAATAAAA	Reverse primer for 5' fragment of *pff1145c* promoter without motif 3
Motif3_PFF1145c_3'F	AATTATCATACGAATATATATTTTAATGTG	Forward primer for 3' fragment of *pff1145c* promoter without motif 3
Motif3_PFF1145c_3'R(NcoI)	ATCG**CCATGG**TAAAAAAAAAAATGTATTTTATACTACTATATATATATAT	Reverse primer for *pff1145c* promoter
Motif3_MAL7P1.144_5'F(XhoI)	ATCG**CTCGAG**GGCGACATTATCAATAAATTAAATATA	Forward primer for *mal7p1.144* promoter
Motif3_MAL7P1.144_5'R	TTAATTTTTAAAATTATTTAAATAATAATAAAATAATAA	Reverse primer for 5' fragment of *mal7p1.144* promoter without motif 3
Motif3_MAL7P1.144_3'F	GTTATTTATTCATAATATATAATAATAATGTGTTCT	Forward primer for 3' fragment of *mal7p1.144* promoter without motif 3
Motif3_MAL7P1.144_3'R(NcoI)	ATCG**CCATGG**TATGATAAATGTAATGAAGGCGTT	Reverse primer for *mal7p1.144* promoter
Motif3_PFD1165w_5'F(XhoI)	ATCG**CTCGAG**GATACTTTATATAAATAAATATATGCATGTTTAATA	Forward primer for *pfd1165w* promoter
Motif3_PFD1165w_5'R	ATATTATTTCATATTATGGAAACGTGAA	Reverse primer for 5' fragment of *pfd1165w* promoter without motif 3
Motif3_PFD1165w_3'F	ATATATATATATATATATTTATAATATATATGATCATATATTTATTTTTGG	Forward primer for 3' fragment of *pfd1165w* promoter without motif

		3
Motif3_PFD1165w_3'R(NcoI)	ATCG**CCATGG**TTTTTTTTTTTTTTAATTTAGACTTTTTCTCTTTCGT	Reverse primer for *pfd1165w* promoter
Motif3_PF14_0423_5'F(XhoI)	ATCG**CTCGAG**CATCTTTTAAGAATAAATATAAATAAATACATACC	Forward primer for *pf14_0423* promoter
Motif3_PF14_0423_5'R	AATGAAATCCTTGGGGGACTTTT	Reverse primer for 5' fragment of *pf14_0423* promoter without motif 3
Motif3_PF14_0423_3'F	AGTTAAAAGTATTCTTGTATGTATAAAAATATATATTT	Forward primer for 3' fragment of *pf14_0423* promoter without motif 3
Motif3_PF14_0423_3'R(NcoI)	ATCG**CCATGG**TTTGTAATTCTCTTCATATAATTTATATATATATATA	Reverse primer for *pf14_0423* promoter
Motif16_PFB0440c_5'F(XhoI)	ATCG**CTCGAG**GTATTTATAAAACAAATTTTTCCTAGCA	Forward primer for *pfb0440c* promoter
Motif16_PFB0440c_5'R	CTTCAAAAAAAAAAAAAAATTTTTTTTTTCTGAAATTTCATATTAAAAAG	Reverse primer for 5' fragment of *pfb0440c* promoter without motif 16
Motif16_PFB0440c_3'F	TACGATGAAAAAAAAAAGTTAAAG	Forward primer for 3' fragment of *pfb0440c* promoter without motif 16
Motif16_PFB0440c_3'R(NcoI)	ATCG**CCATGG**TTCTAAATAATATACACAAAAAAATAAAACAC	Reverse primer for *pfb0440c* promoter
Motif16_PF14_0527_5'F(XhoI)	ATCG**CTCGAG**GGGTACATAAGAAAAAAACAGGA	Forward primer for *pf14_0527* promoter
Motif16_PF14_0527_5'R	AATATTTTATAACTAATAATTTTTATAAATACAACATC	Reverse primer for 5' fragment of *pf14_0527* promoter without motif 16
Motif16_PF14_0527_3'F	AAATATATTATCTATACAAATATACATTTTGTTAAATA	Forward primer for 3' fragment of *pf14_0527* promoter without motif 16
Motif16_PF14_0527_3'R(NcoI)	ATCG**CCATGG**ATACATATTTGTTGAGACAAACAAAAA	Reverse primer for *pf14_0527* promoter
Motif16_PF14_0013_5'F(XhoI)	ATCG**CTCGAG**GTGTTCTTACTAATTGATTAAAAAGTAATTTA	Forward primer for *pf14_0013* promoter
Motif16_PF14_0013_5'R	TTTTTATATATATTATATATATATAGATTTATTCTTTTATTAAATATATATAAATCATC	Reverse primer for 5' fragment of *pf14_0013* promoter without motif 16
Motif16_PF14_0013	AATTATTTTTTTTTTTTATATTTTGAATAA	Forward primer for 3'

_3'F	ACATAACAAATC	fragment of *pf14_0013* promoter without motif 16
Motif16_PF14_0013 _3'R(NcoI)	ATCG**CCATGG**ATTATCACCAATTTCTGTA GTGGA	Reverse primer for *pf14_0013* promoter
Motif16_PFD0085c_ 5'F(XhoI)	ATCG**CTCGAG**GGTGTATTATATATATAT GAAAAAATTGGAC	Forward primer for *pfd0085c* promoter
Motif16_PFD0085c_ 5'R	TAAATATTTCGATTAAATCTTTATATATA TATAATATATATATAAGCC	Reverse primer for 5' fragment of *pfd0085c* promoter without motif 16
Motif16_PFD0085c_ 3'F	ATAATGGTAGTGCATATAATAATATAATA AATTAT	Forward primer for 3' fragment of *pfd0085c* promoter without motif 16
Motif16_PFD0085c_ 3'R(NcoI)	ATCG**CCATGG**AATATAACAAAAAAATAA AATAAAATAAAAAATAAATTAAACC	Reverse primer for *pfd0085c* promoter
Motif16_PFA0125c_ 5'F(XhoI)	ATCG**CTCGAG**GAGGAAAAATTATGGAAA GGTATAG	Forward primer for *pfa0125c* promoter
Motif16_PFA0125c_ 5'R	TATATACACATATATATTATATTATATTA TTTTTATTTTATTTACTTTTACATAATCTT TCTTTC	Reverse primer for 5' fragment of *pfa0125c* promoter without motif 16
Motif16_PFA0125c_ 3'F	ATGTATTATTTAAAAAAAAAAATAAAAA ATAAAAAAATTGGATAATTCTC	Forward primer for 3' fragment of *pfa0125c* promoter without motif 16
Motif16_PFA0125c_ 3'R(NcoI)	ATCG**CCATGG**ATTATTCTACTATTATATT CTATATTATTTATAGTATACATACAAC	Reverse primer for *pfa0125c* promoter
Motif20_PF14_0333 _5'F(XhoI)	ATCG**CTCGAG**CAAACCATAGGAGAGTCA CAAA	Forward primer for *pf14_0333* promoter
Motif20_PF14_0333 _5'R	CCATATCAAAAATATATAAAATTAAAGC C	Reverse primer for 5' fragment of *pf14_0333* promoter without motif 20
Motif20_PF14_0333 _3'F	GTGGAAAAATATATAGATATTCTTGAATT AC	Forward primer for 3' fragment of *pf14_0333* promoter without motif 20
Motif20_PF14_0333 _3'R(NcoI)	ATCG**CCATGG**GTTATAAATTCCCTACAG GAAATAT	Reverse primer for *pf14_0333* promoter
Motif20_PFC0035w _5'F(XhoI)	ATCG**CTCGAG**GTTTTATATTGTCATGACT ATATAAAAGATTTC	Forward primer for *pfc0035w* promoter
Motif20_PFC0035w _5'R	AACGATATAGTAACAATAGAACTTATATT TTTTT	Reverse primer for 5' fragment of *pfc0035w* promoter without motif

		20
Motif20_PFC0035w_3'F	GAAAAATATATAAAAACATTTTAATTGTAATATAT	Forward primer for 3' fragment of *pfc0035w* promoter without motif 20
Motif20_PFC0035w_3'R(NcoI)	ATCG**CCATGG**CTTTATTGTGATACGCATATTATTTTAT	Reverse primer for *pfc0035w* promoter
Motif20_PFC0185w_5'F(XhoI)	ATCG**CTCGAG**CTTTTTTAAAATGCAACCTTGT	Forward primer for *pfc0185w* promoter
Motif20_PFC0185w_5'R	GTGTAAATTTTTTGCTATATAATTTGGTAT	Reverse primer for 5' fragment of *pfc0185w* promoter without motif 20
Motif20_PFC0185w_3'F	TTCTATACAAATGGAAAAAACGG	Forward primer for 3' fragment of *pfc0185w* promoter without motif 20
Motif20_PFC0185w_3'R(NcoI)	ATCG**CCATGG**TTTCTTGTACTAACGAAAGTGTGTC	Reverse primer for *pfc0185w* promoter
Motif20_MAL8P1.109_5'F(XhoI)	ATCG**CTCGAG**CTTATAATTGTTCCATAAGTATATACATGC	Forward primer for *mal8p1.109* promoter
Motif20_MAL8P1.109_5'R	AACCATCTTAACAAAAAATATAAAAAAT	Reverse primer for 5' fragment of *mal8p1.109* promoter without motif 20
Motif20_MAL8P1.109_3'F	TATTCCGTATAAATTGTTATTTTGTATG	Forward primer for 3' fragment of *mal8p1.109* promoter without motif 20
Motif20_MAL8P1.109_3'R(NcoI)	ATCG**CCATGG**GTCTTTATCTCCTTTTTCACAATAAA	Reverse primer for *mal8p1.109* promoter
Motif20_PFF0645c_5'F(XhoI)	ATCG**CTCGAG**GGTATATATATTTATAACATTTGTTAATTTTAAGAA	Forward primer for *pff0645c* promoter
Motif20_ PFF0645c_5'R	ATATATTTCAAATTATATATCACGTGTTAAAA	Reverse primer for 5' fragment of *pff0645c* promoter without motif 20
Motif20_ PFF0645c_3'F	ACAAAAACACACAAATGAATATAAAC	Forward primer for 3' fragment of *pff0645c* promoter without motif 20
Motif20_ PFF0645c_3'R(NcoI)	ATCG**CCATGG**TTTTTTTTTTTCTTTCTTCTTTTTTCTTCTTTTTTCGTCTTTTTTCTACTTTTCTCTTTTGGTTCACGTATATTC	Reverse primer for *pff0645c* promoter
Motif20_ PF14_0578_5'F(XhoI)	ATCG**CTCGAG**GATATATTATATCTAATCGTGCACAACTG	Forward primer for *pf14_0578* promoter
Motif20_ PF14_0578	TTCTATTTATATATATTATATATATATAAT	Reverse primer for 5'

_5'R	ATTTTTATTCTATC	fragment of *pf14_0578* promoter without motif 20
Motif20_ PF14_0578 _3'F	TTTTTTATTTTATTTTTTTTTTATTTATAATTTTGTAGAGATTATTTTTA	Forward primer for 3' fragment of *pf14_0578* promoter without motif 20
Motif20_ PF14_0578 _3'R(NcoI)	ATCG**CCATGG**CTTGAAAATTATTAATCATAAAAAAAAA3	Reverse primer for *pf14_0578* promoter
Stable_motif3_5'utr_ PFL1885C_F(PstI)	ATGC**CTGCAG**CCTTATGCATAATTTATATTCAGAAAC	Forward primer for promoter of *pfl1885c* for stable transfections
Stable_motif3_5'utr_ PFL1885C_R(BamHI)	ATGC**GGATCC**TCAAAAAAAAAAAAAAATAAATAAATATTAG	Reverse primer for promoter of *pfl1885c* for stable transfections
Stable_motif3_3'utr_ PFL1885C_F(XhoI)	ATGC**CTCGAG**GCTATGCCAAATAATGAATGG	Forward primer for 3'UTR of *pfl1885c* for stable transfections
Stable_motif3_3'utr_ PFL1885C_R(ApaI)	ATGC**GGGCCC**TGAATACTTTATAACGATTTAGCTATTATAG	Reverse primer for 3'UTR of *pfl1885c* for stable transfections
Stable_motif6_5'utr_ PFF1500C_F(PstI)	ATGC**CTGCAG**CCTTATGCATAATTTATATTCAGAAAC	Forward primer for promoter of *pff1500c* for stable transfections
Stable_motif6_5'utr_ PFF1500C_R(BamHI)	ATGC**GGATCC**TCAAAAAAAAAAAAAAATAAATAAATATTAG	Reverse primer for promoter of *pff1500c* for stable transfections
Stable_motif6_3'utr_ PFF1500C _F(XhoI)	ATGC**CTCGAG**ATAAAAACAATAAAAATATTATCATTATGTAGATATA	Forward primer for 3'UTR of *pff1500c* for stable transfections
Stable_motif6_3'utr_ PFF1500C _R(ApaI)	ATGC**GGGCCC**CATTAAAGTACGTACATTATAATAAATATTCTG	Reverse primer for 3'UTR of *pff1500c* for stable transfections
Stable_motif16_5'utr _PF14_0761_F(PstI)	ATGC**CTGCAG**GGTATAATGTATATATTGCAATATATGTATGAT	Forward primer for promoter of *pf14_0761* for stable transfections
Stable_motif16_5'utr _PF14_0761_R(BamHI)	ATGC**GGATCC**TTAAAAACAACAAAAATACGTTAATATTATATA	Reverse primer for promoter of *pf14_0761* for stable transfections
Stable_motif16_3'utr _ PF14_0761_F(XhoI)	ATGC**CTCGAG**AATGAAATAAAATCATATATTTATAATTATTATAATAA	Forward primer for 3'UTR of *pf14_0761* for stable transfections
Stable_motif16_3'utr _PF14_0761_R(ApaI)	ATGC**GGGCCC**ATAATCAAGGATAAGTTGAAAAAAATTA	Reverse primer for 3'UTR of *pf14_0761* for stable transfections
Stable_motif17_5'utr _PF10_0100_F(PstI)	ATGC**CTGCAG**GTACTACTATATACTTGTGGTTGCAAAA	Forward primer for promoter of *pf10_0100* for stable transfections
Stable_motif17_5'utr _PF10_0100_R(BamHI)	ATGC**GGATCC**TTTTTTAATTTCTCTTTAAAATAATAAATAAA	Reverse primer for promoter of *pf10_0100* for stable transfections
Stable_motif17_3'utr _	ATGC**CTCGAG**ATAAATAAATAAATGTATTAATATACAATTAATCATA	Forward primer for 3'UTR of *pf10_0100*

PF10_0100_F(XhoI)		for stable transfections
Stable_motif17_3'utr_PFF1010c_R(ApaI)	ATGC**GGGCCC**GTAAGAGTATATTAGAAGGTAATTGGGT	Reverse primer for 3'UTR of *pf10_0100* for stable transfections
Stable_motif20_5'utr_PFF1010c_F(PstI)	ATGC**CTGCAG**GATAATCCCTTATTTTATTTATTCATTC	Forward primer for promoter of *pff1010c* for stable transfections
Stable_motif20_5'utr_PFF1010c_R(BamHI)	ATGC**GGATCC**TTTTGTAATTATATTAATATATATATATATATATATATATATGTGATATTTTATTAGTATGGTGGTAAT	Reverse primer for promoter of *pff1010c* for stable transfections
Stable_motif20_3'utr_PFF1010c_F(XhoI)	ATGC**CTCGAG**AACAAAAAAAAAAAAAAAAAAAAAATTCATATGTATAAATGGG	Forward primer for 3'UTR of *pff1010c* for stable transfections
Stable_motif20_3'utr_PFF1010c_R(SpeI)	ATGC**ACTAGT**AATTGTGGAGTAATCTTATACATCATAA	Reverse primer for 3'UTR of *pff1010c* for stable transfections

Supplementary Table 1 List of primers used for cloning in the studies in this thesis.

This table summarizes all primers that have been used in the studies in this thesis. The primers have been listed by order of the chapters in this thesis. Restriction enzyme sites are in bold characters. Motif sequences are underlined. Mutations made to the motif sequences are in red. If 2 or 3 copies of the motif have been inserted, sequences are colored red and blue respectively.

8. References

1. WHO: **World Malaria Report**. *World Health Organization* 2013.
2. Francis SE, Sullivan DJ, Jr., Goldberg DE: **Hemoglobin metabolism in the malaria parasite Plasmodium falciparum**. *Annual review of microbiology* 1997, **51**:97-123.
3. Ginsburg H: **Should chloroquine be laid to rest?** *Acta tropica* 2005, **96**(1):16-23.
4. Vijaykadga S, Rojanawatsirivej C, Cholpol S, Phoungmanee D, Nakavej A, Wongsrichanalai C: **In vivo sensitivity monitoring of mefloquine monotherapy and artesunate-mefloquine combinations for the treatment of uncomplicated falciparum malaria in Thailand in 2003**. *Tropical medicine & international health : TM & IH* 2006, **11**(2):211-219.
5. Noedl H, Se Y, Schaecher K, Smith BL, Socheat D, Fukuda MM, Artemisinin Resistance in Cambodia 1 Study C: **Evidence of artemisinin-resistant malaria in western Cambodia**. *The New England journal of medicine* 2008, **359**(24):2619-2620.
6. Schwartz L, Brown GV, Genton B, Moorthy VS: **A review of malaria vaccine clinical projects based on the WHO rainbow table**. *Malaria journal* 2012, **11**:11.
7. Kirkman LA, Deitsch KW: **Antigenic variation and the generation of diversity in malaria parasites**. *Current opinion in microbiology* 2012, **15**(4):456-462.
8. Riley EM, Stewart VA: **Immune mechanisms in malaria: new insights in vaccine development**. *Nature medicine* 2013, **19**(2):168-178.
9. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S *et al*: **Genome sequence of the human malaria parasite Plasmodium falciparum**. *Nature* 2002, **419**(6906):498-511.
10. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL *et al*: **Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii**. *Nature* 2002, **419**(6906):512-519.
11. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P *et al*: **Comparative genomics of the neglected human malaria parasite Plasmodium vivax**. *Nature* 2008, **455**(7214):757-763.
12. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, Florens L, Janssen CS, Pain A, Christophides GK *et al*: **A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses**. *Science* 2005, **307**(5706):82-86.
13. Pain A, Bohme U, Berry AE, Mungall K, Finn RD, Jackson AP, Mourier T, Mistry J, Pasini EM, Aslett MA *et al*: **The genome of the simian and human malaria parasite Plasmodium knowlesi**. *Nature* 2008, **455**(7214):799-803.
14. Elgar G, Vavouri T: **Tuning in to the signals: noncoding sequence conservation in vertebrate genomes**. *Trends in genetics : TIG* 2008, **24**(7):344-352.
15. Volkman SK, Lozovsky E, Barry AE, Bedford T, Bethke L, Myrick A, Day KP, Hartl DL, Wirth DF, Sawyer SA: **Genomic heterogeneity in the density of noncoding single-nucleotide and microsatellite polymorphisms in Plasmodium falciparum**. *Gene* 2007, **387**(1-2):1-6.
16. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, Zhou Y, Johnson JR, Le Roch K, Sarr O, Ndir O *et al*: **A systematic map of genetic variation in Plasmodium falciparum**. *PLoS pathogens* 2006, **2**(6):e57.
17. Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S, Gujja S, Goldberg JM, Young S, Zeng Q *et al*: **The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum**. *Nature genetics* 2012, **44**(9):1046-1050.
18. Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, Kim S, Duru V, Bouchier C, Ma L *et al*: **A molecular marker of artemisinin-resistant Plasmodium falciparum malaria**. *Nature* 2014, **505**(7481):50-55.

19. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL: **The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum**. *PLoS Biol* 2003, **1**(1):E5.
20. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ *et al*: **Discovery of gene function by expression profiling of the malaria parasite life cycle**. *Science* 2003, **301**(5639):1503-1508.
21. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization**. *Mol Biol Cell* 1998, **9**(12):3273-3297.
22. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO *et al*: **Identification of genes periodically expressed in the human cell cycle and their expression in tumors**. *Mol Biol Cell* 2002, **13**(6):1977-2000.
23. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, Lemieux J, Barrell B, Pain A, Berriman M *et al*: **New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq**. *Mol Microbiol* 2010, **76**(1):12-24.
24. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: **Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains**. *Nucleic acids research* 2006, **34**(4):1166-1173.
25. Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast**. *Science* 2002, **296**(5568):752-755.
26. Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, Russell B, Ginsburg H, Nosten F, Day NP, White NJ *et al*: **The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites**. *Proc Natl Acad Sci U S A* 2008, **105**(42):16290-16295.
27. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL *et al*: **A proteomic view of the Plasmodium falciparum life cycle**. *Nature* 2002, **419**(6906):520-526.
28. Nirmalan N, Sims PF, Hyde JE: **Quantitative proteomics of the human malaria parasite Plasmodium falciparum and its application to studies of development and inhibition**. *Mol Microbiol* 2004, **52**(4):1187-1199.
29. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, Grainger M, Yan SF, Williamson KC, Holder AA, Carucci DJ *et al*: **Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle**. *Genome research* 2004, **14**(11):2308-2318.
30. Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, Bozdech Z: **Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum**. *Molecular & cellular proteomics : MCP* 2011, **10**(8):M110 006411.
31. Carvalho TG, Menard R: **Manipulating the Plasmodium genome**. *Current issues in molecular biology* 2005, **7**(1):39-55.
32. Balu B, Adams JH: **Advancements in transfection technologies for Plasmodium**. *International journal for parasitology* 2007, **37**(1):1-10.
33. Skinner-Adams TS, Lawrie PM, Hawthorne PL, Gardiner DL, Trenholme KR: **Comparison of Plasmodium falciparum transfection methods**. *Malaria journal* 2003, **2**:19.
34. Wu Y, Sifri CD, Lei HH, Su XZ, Wellems TE: **Transfection of Plasmodium falciparum within human red blood cells**. *Proc Natl Acad Sci U S A* 1995, **92**(4):973-977.
35. Fidock DA, Wellems TE: **Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil**. *Proc Natl Acad Sci U S A* 1997, **94**(20):10931-10936.
36. Deitsch K, Driskill C, Wellems T: **Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes**. *Nucleic acids research* 2001, **29**(3):850-853.

37. Hasenkamp S, Russell KT, Horrocks P: **Comparison of the absolute and relative efficiencies of electroporation-based transfection protocols for Plasmodium falciparum**. *Malaria journal* 2012, **11**:210.
38. Caro F, Miller MG, DeRisi JL: **Plate-based transfection and culturing technique for genetic manipulation of Plasmodium falciparum**. *Malaria journal* 2012, **11**:22.
39. Gopalakrishnan AM, Kundu AK, Mandal TK, Kumar N: **Novel nanosomes for gene delivery to Plasmodium falciparum-infected red blood cells**. *Scientific reports* 2013, **3**:1534.
40. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ: **Crystal structure of the nucleosome core particle at 2.8 A resolution**. *Nature* 1997, **389**(6648):251-260.
41. Jenuwein T, Allis CD: **Translating the histone code**. *Science* 2001, **293**(5532):1074-1080.
42. Jin J, Cai Y, Li B, Conaway RC, Workman JL, Conaway JW, Kusch T: **In and out: histone variant exchange in chromatin**. *Trends Biochem Sci* 2005, **30**(12):680-687.
43. Miao J, Fan Q, Cui L, Li J: **The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation**. *Gene* 2006, **369**:53-65.
44. Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG, Jensen ON: **Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum**. *J Proteome Res* 2009, **8**(7):3439-3450.
45. Issar N, Roux E, Mattei D, Scherf A: **Identification of a novel post-translational modification in Plasmodium falciparum: protein sumoylation in different cellular compartments**. *Cell Microbiol* 2008, **10**(10):1999-2011.
46. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, 3rd, Gingeras TR *et al*: **Genomic maps and comparative analysis of histone modifications in human and mouse**. *Cell* 2005, **120**(2):169-181.
47. Lee DY, Hayes JJ, Pruss D, Wolffe AP: **A positive role for histone acetylation in transcription factor access to nucleosomal DNA**. *Cell* 1993, **72**(1):73-84.
48. Loyola A, Almouzni G: **Bromodomains in living cells participate in deciphering the histone code**. *Trends Cell Biol* 2004, **14**(6):279-281.
49. Fan Q, An L, Cui L: **Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling**. *Eukaryot Cell* 2004, **3**(2):264-276.
50. Cui L, Miao J, Furuya T, Li X, Su XZ: **PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in Plasmodium falciparum**. *Eukaryot Cell* 2007, **6**(7):1219-1227.
51. Cui L, Miao J, Furuya T, Fan Q, Li X, Rathod PK, Su XZ: **Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro Plasmodium falciparum development**. *Eukaryot Cell* 2008, **7**(7):1200-1210.
52. Chaal BK, Gupta AP, Wastuwidyaningtyas BD, Luah YH, Bozdech Z: **Histone deacetylases play a major role in the transcriptional regulation of the Plasmodium falciparum life cycle**. *PLoS pathogens* 2010, **6**(1):e1000737.
53. Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T: **Histone H3 lysine 4 methylation patterns in higher eukaryotic genes**. *Nat Cell Biol* 2004, **6**(1):73-77.
54. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E *et al*: **Genome-wide map of nucleosome acetylation and methylation in yeast**. *Cell* 2005, **122**(4):517-527.
55. Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AM, Cohen AM, Janssen-Megens EM, van de Vegte-Bolmer M, Selzer RR, Iniguez AL *et al*: **Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum**. *Proc Natl Acad Sci U S A* 2009, **106**(24):9655-9660.
56. Bartfai R, Hoeijmakers WA, Salcedo-Amaya AM, Smits AH, Janssen-Megens E, Kaan A, Treeck M, Gilberger TW, Francoijs KJ, Stunnenberg HG: **H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3**. *PLoS pathogens* 2010, **6**(12):e1001223.

57. Gupta AP, Chin WH, Zhu L, Mok S, Luah YH, Lim EH, Bozdech Z: **Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite Plasmodium falciparum**. *PLoS pathogens* 2013, **9**(2):e1003170.
58. Vakoc CR, Mandat SA, Olenchock BA, Blobel GA: **Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin**. *Molecular cell* 2005, **19**(3):381-391.
59. Sanchez R, Zhou MM: **The role of human bromodomains in chromatin biology and gene transcription**. *Current opinion in drug discovery & development* 2009, **12**(5):659-665.
60. Eissenberg JC: **Structural biology of the chromodomain: form and function**. *Gene* 2012, **496**(2):69-78.
61. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, Ehlgen F, Ralph SA, Cowman AF, Bozdech Z *et al*: **Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors**. *PLoS pathogens* 2009, **5**(9):e1000569.
62. Lopez-Rubio JJ, Mancio-Silva L, Scherf A: **Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites**. *Cell Host Microbe* 2009, **5**(2):179-190.
63. Kyes SA, Kraemer SM, Smith JD: **Antigenic variation in Plasmodium falciparum: gene organization and regulation of the var multigene family**. *Eukaryot Cell* 2007, **6**(9):1511-1520.
64. Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, Gysin J, Lanzer M: **Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum**. *The EMBO journal* 1998, **17**(18):5418-5426.
65. Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, Scherf A: **5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites**. *Mol Microbiol* 2007, **66**(6):1296-1305.
66. Chookajorn T, Dzikowski R, Frank M, Li F, Jiwani AZ, Hartl DL, Deitsch KW: **Epigenetic memory at malaria virulence genes**. *Proc Natl Acad Sci U S A* 2007, **104**(3):899-902.
67. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, Scherf A, Crabb BS, Cowman AF: **Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum**. *Cell* 2005, **121**(1):13-24.
68. Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M, Duffy MF, Silva LM, Scherf A, Ivens A *et al*: **Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum**. *PLoS Biol* 2009, **7**(4):e84.
69. Perez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernandez-Cuevas NA, Delgadillo DM, Vargas M, Martinez-Calvillo S, Scherf A, Hernandez-Rivas R: **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 research* 2009, **37**(8):2596-2606.
70. Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, Yang W, Turner L, Lavstsen T, Theander TG *et al*: **PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum**. *Nature* 2013, **499**(7457):223-227.
71. Dolan SA, Miller LH, Wellems TE: **Evidence for a switching mechanism in the invasion of erythrocytes by Plasmodium falciparum**. *The Journal of clinical investigation* 1990, **86**(2):618-624.
72. Comeaux CA, Coleman BI, Bei AK, Whitehurst N, Duraisingh MT: **Functional analysis of epigenetic regulation of tandem RhopH1/clag genes reveals a role in Plasmodium falciparum growth**. *Mol Microbiol* 2011, **80**(2):378-390.
73. Jiang L, Lopez-Barragan MJ, Jiang H, Mu J, Gaur D, Zhao K, Felsenfeld G, Miller LH: **Epigenetic control of the variable expression of a Plasmodium falciparum receptor protein for erythrocyte invasion**. *Proc Natl Acad Sci U S A* 2010, **107**(5):2224-2229.

74. Crowley VM, Rovira-Graells N, Ribas de Pouplana L, Cortes A: **Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant Plasmodium falciparum genes linked to erythrocyte invasion**. *Mol Microbiol* 2011, **80**(2):391-406.
75. Hardy S, Robert F: **Random deposition of histone variants: A cellular mistake or a novel regulatory mechanism?** *Epigenetics : official journal of the DNA Methylation Society* 2010, **5**(5):368-372.
76. Thambirajah AA, Dryhurst D, Ishibashi T, Li A, Maffey AH, Ausio J: **H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation**. *J Biol Chem* 2006, **281**(29):20036-20044.
77. Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD: **Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin**. *Cell* 2005, **123**(2):233-248.
78. Petter M, Lee CC, Byrne TJ, Boysen KE, Volz J, Ralph SA, Cowman AF, Brown GV, Duffy MF: **Expression of P. falciparum var genes involves exchange of the histone variant H2A.Z at the promoter**. *PLoS pathogens* 2011, **7**(2):e1001292.
79. Hoeijmakers WA, Salcedo-Amaya AM, Smits AH, Francoijs KJ, Treeck M, Gilberger TW, Stunnenberg HG, Bartfai R: **H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the Plasmodium falciparum genome**. *Mol Microbiol* 2013, **87**(5):1061-1073.
80. Petter M, Selvarajah SA, Lee CC, Chin WH, Gupta AP, Bozdech Z, Brown GV, Duffy MF: **H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite Plasmodium falciparum**. *Mol Microbiol* 2013, **87**(6):1167-1182.
81. Lowell JE, Kaiser F, Janzen CJ, Cross GA: **Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in Trypanosoma brucei**. *Journal of cell science* 2005, **118**(Pt 24):5721-5730.
82. Dalmasso MC, Onyango DO, Naguleswaran A, Sullivan WJ, Jr., Angel SO: **Toxoplasma H2A variants reveal novel insights into nucleosome composition and functions for this histone family**. *J Mol Biol* 2009, **392**(1):33-47.
83. Hoeijmakers WA, Flueck C, Francoijs KJ, Smits AH, Wetzel J, Volz JC, Cowman AF, Voss T, Stunnenberg HG, Bartfai R: **Plasmodium falciparum centromeres display a unique epigenetic makeup and cluster prior to and during schizogony**. *Cell Microbiol* 2012, **14**(9):1391-1401.
84. Verma G, Surolia N: **Plasmodium falciparum CENH3 is able to functionally complement Cse4p and its, C-terminus is essential for centromere function**. *Molecular and biochemical parasitology* 2013, **192**(1-2):21-29.
85. Briquet S, Boschet C, Gissot M, Tissandie E, Sevilla E, Franetich JF, Thiery I, Hamid Z, Bourgouin C, Vaquero C: **High-mobility-group box nuclear factors of Plasmodium falciparum**. *Eukaryot Cell* 2006, **5**(4):672-682.
86. Nemeth A, Langst G: **Chromatin higher order structure: opening up chromatin for transcription**. *Briefings in functional genomics & proteomics* 2004, **2**(4):334-343.
87. Gissot M, Ting LM, Daly TM, Bergman LW, Sinnis P, Kim K: **High mobility group protein HMGB2 is a critical regulator of plasmodium oocyst development**. *J Biol Chem* 2008, **283**(25):17030-17038.
88. Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C: **A high-resolution atlas of nucleosome occupancy in yeast**. *Nature genetics* 2007, **39**(10):1235-1244.
89. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC *et al*: **Nucleosome organization in the Drosophila genome**. *Nature* 2008, **453**(7193):358-362.

90. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome**. *Cell* 2008, **132**(5):887-898.
91. Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD: **Evidence for nucleosome depletion at active regulatory regions genome-wide**. *Nature genetics* 2004, **36**(8):900-905.
92. Svaren J, Horz W: **Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast**. *Trends Biochem Sci* 1997, **22**(3):93-97.
93. Floer M, Wang X, Prabhu V, Berrozpe G, Narayan S, Spagna D, Alvarez D, Kendall J, Krasnitz A, Stepansky A *et al*: **A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding**. *Cell* 2010, **141**(3):407-418.
94. Barbaric S, Fascher KD, Horz W: **Activation of the weakly regulated PHO8 promoter in S. cerevisiae: chromatin transition and binding sites for the positive regulatory protein PHO4**. *Nucleic acids research* 1992, **20**(5):1031-1038.
95. Zawadzki KA, Morozov AV, Broach JR: **Chromatin-dependent transcription factor accessibility rather than nucleosome remodeling predominates during global transcriptional restructuring in Saccharomyces cerevisiae**. *Mol Biol Cell* 2009, **20**(15):3503-3513.
96. Angermayr M, Oechsner U, Gregor K, Schroth GP, Bandlow W: **Transcription initiation in vivo without classical transactivators: DNA kinks flanking the core promoter of the housekeeping yeast adenylate kinase gene, AKY2, position nucleosomes and constitutively activate transcription**. *Nucleic acids research* 2002, **30**(19):4199-4207.
97. Angermayr M, Schwerdffeger K, Bandlow W: **A nucleosome-free dG-dC-rich sequence element promotes constitutive transcription of the essential yeast RIO1 gene**. *Biol Chem* 2003, **384**(9):1287-1292.
98. Westenberger SJ, Cui L, Dharia N, Winzeler E: **Genome-wide nucleosome mapping of Plasmodium falciparum reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes**. *BMC Genomics* 2009, **10**:610.
99. Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG: **Nucleosome landscape and control of transcription in the human malaria parasite**. *Genome research* 2010, **20**(2):228-238.
100. Anderson JD, Widom J: **Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites**. *Mol Cell Biol* 2001, **21**(11):3830-3839.
101. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ: **Genome-scale identification of nucleosome positions in S. cerevisiae**. *Science* 2005, **309**(5734):626-630.
102. Ponts N, Harris EY, Lonardi S, Le Roch KG: **Nucleosome occupancy at transcription start sites in the human malaria parasite: a hard-wired evolution of virulence?** *Infect Genet Evol* 2011, **11**(4):716-724.
103. Watanabe J, Suzuki Y, Sasaki M, Sugano S: **Full-malaria 2004: an enlarged database for comparative studies of full-length cDNAs of malaria parasites, Plasmodium species**. *Nucleic acids research* 2004, **32**(Database issue):D334-338.
104. Brick K, Watanabe J, Pizzi E: **Core promoters are predicted by their distinct physicochemical properties in the genome of Plasmodium falciparum**. *Genome Biol* 2008, **9**(12):R178.
105. Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, Turner DJ, Macinnis B, Kwiatkowski DP, Swerdlow HP *et al*: **Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes**. *BMC Genomics* 2012, **13**:1.
106. Jones PA: **Functions of DNA methylation: islands, start sites, gene bodies and beyond**. *Nature reviews Genetics* 2012, **13**(7):484-492.
107. Venolia L, Gartler SM: **Comparison of transformation efficiency of human active and inactive X-chromosomal DNA**. *Nature* 1983, **302**(5903):82-83.
108. Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ *et al*: **DNA methylation status predicts cell type-specific enhancer activity**. *The EMBO journal* 2011, **30**(15):3028-3039.

109. Bell AC, Felsenfeld G: **Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene**. *Nature* 2000, **405**(6785):482-485.
110. Choi SW, Keyes MK, Horrocks P: **LC/ESI-MS demonstrates the absence of 5-methyl-2'-deoxycytosine in Plasmodium falciparum genomic DNA**. *Molecular and biochemical parasitology* 2006, **150**(2):350-352.
111. Pollack Y, Katzen AL, Spira DT, Golenser J: **The genome of Plasmodium falciparum. I: DNA base composition**. *Nucleic acids research* 1982, **10**(2):539-546.
112. Gissot M, Choi SW, Thompson RF, Greally JM, Kim K: **Toxoplasma gondii and Cryptosporidium parvum lack detectable DNA cytosine methylation**. *Eukaryot Cell* 2008, **7**(3):537-540.
113. Pollack Y, Kogan N, Golenser J: **Plasmodium falciparum: evidence for a DNA methylation pattern**. *Exp Parasitol* 1991, **72**(4):339-344.
114. Weber JL: **Analysis of sequences from the extremely A + T-rich genome of Plasmodium falciparum**. *Gene* 1987, **52**(1):103-109.
115. Ponts N, Fu L, Harris EY, Zhang J, Chung DW, Cervantes MC, Prudhomme J, Atanasova-Penichon V, Zehraoui E, Bunnik EM *et al*: **Genome-wide mapping of DNA methylation in the human malaria parasite Plasmodium falciparum**. *Cell Host Microbe* 2013, **14**(6):696-706.
116. Shearwin KE, Callen BP, Egan JB: **Transcriptional interference--a crash course**. *Trends in genetics : TIG* 2005, **21**(6):339-345.
117. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM *et al*: **Human DNA methylomes at base resolution show widespread epigenomic differences**. *Nature* 2009, **462**(7271):315-322.
118. Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM: **Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a**. *Molecular cell* 2010, **38**(5):662-674.
119. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C *et al*: **Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat**. *Nature* 2012, **491**(7424):454-457.
120. Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G, 3rd, Kenny PJ, Wahlestedt C: **Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase**. *Nature medicine* 2008, **14**(7):723-730.
121. Hildebrandt M, Nellen W: **Differential antisense transcription from the Dictyostelium EB4 gene locus: implications on antisense-mediated regulation of mRNA stability**. *Cell* 1992, **69**(1):197-204.
122. Kyes S, Christodoulou Z, Pinches R, Newbold C: **Stage-specific merozoite surface protein 2 antisense transcripts in Plasmodium falciparum**. *Molecular and biochemical parasitology* 2002, **123**(1):79-83.
123. Siegel TN, Hon CC, Zhang Q, Lopez-Rubio JJ, Scheidig-Benatar C, Martins RM, Sismeiro O, Coppee JY, Scherf A: **Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum**. *BMC Genomics* 2014, **15**(1):150.
124. Gardiner DL, Holt DC, Thomas EA, Kemp DJ, Trenholme KR: **Inhibition of Plasmodium falciparum clag9 gene function by antisense RNA**. *Molecular and biochemical parasitology* 2000, **110**(1):33-41.
125. Patankar S, Munasinghe A, Shoaibi A, Cummings LM, Wirth DF: **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* 2001, **12**(10):3114-3125.

126. Gunasekera AM, Patankar S, Schug J, Eisen G, Kissinger J, Roos D, Wirth DF: **Widespread distribution of antisense transcripts in the Plasmodium falciparum genome**. *Molecular and biochemical parasitology* 2004, **136**(1):35-42.
127. Militello KT, Patel V, Chessler AD, Fisher JK, Kasper JM, Gunasekera A, Wirth DF: **RNA polymerase II synthesizes antisense RNA in Plasmodium falciparum**. *Rna* 2005, **11**(4):365-370.
128. Lopez-Barragan MJ, Lemieux J, Quinones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K, Su XZ: **Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum**. *BMC Genomics* 2011, **12**:587.
129. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**(2):281-297.
130. Siau A, Toure FS, Ouwe-Missi-Oukem-Boyer O, Ciceron L, Mahmoudi N, Vaquero C, Froissard P, Bisvigou U, Bisser S, Coppee JY *et al*: **Whole-transcriptome analysis of Plasmodium falciparum field isolates: identification of new pathogenicity factors**. *The Journal of infectious diseases* 2007, **196**(11):1603-1612.
131. Tuteja R: **A method to inhibit the growth of Plasmodium falciparum by double-stranded RNA-mediated gene silencing of helicases**. *Methods in molecular biology* 2010, **587**:389-399.
132. Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, Waters AP, Cowman AF, Crabb BS, de Koning-Ward TF: **Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites**. *Nucleic acids research* 2009, **37**(11):3788-3798.
133. Kolev NG, Tschudi C, Ullu E: **RNA interference in protozoan parasites: achievements and challenges**. *Eukaryot Cell* 2011, **10**(9):1156-1163.
134. Xue X, Zhang Q, Huang Y, Feng L, Pan W: **No miRNA were found in Plasmodium and the ones identified in erythrocytes could not be correlated with infection**. *Malaria journal* 2008, **7**:47.
135. Rathjen T, Nicol C, McConkey G, Dalmay T: **Analysis of short RNAs in the malaria parasite and its red blood cell host**. *FEBS letters* 2006, **580**(22):5185-5188.
136. LaMonte G, Philip N, Reardon J, Lacsina JR, Majoros W, Chapman L, Thornburg CD, Telen MJ, Ohler U, Nicchitta CV *et al*: **Translocation of sickle cell erythrocyte microRNAs into Plasmodium falciparum inhibits parasite translation and contributes to malaria resistance**. *Cell Host Microbe* 2012, **12**(2):187-199.
137. Mourier T, Carret C, Kyes S, Christodoulou Z, Gardner PP, Jeffares DC, Pinches R, Barrell B, Berriman M, Griffiths-Jones S *et al*: **Genome-wide discovery and verification of novel structured RNAs in Plasmodium falciparum**. *Genome research* 2008, **18**(2):281-292.
138. Chakrabarti K, Pearson M, Grate L, Sterne-Weiler T, Deans J, Donohue JP, Ares M, Jr.: **Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis**. *Rna* 2007, **13**(11):1923-1939.
139. Mishra PC, Kumar A, Sharma A: **Analysis of small nucleolar RNAs reveals unique genetic features in malaria parasites**. *BMC Genomics* 2009, **10**:68.
140. Li F, Sonbuchner L, Kyes SA, Epp C, Deitsch KW: **Nuclear non-coding RNAs are transcribed from the centromeres of Plasmodium falciparum and are associated with centromeric chromatin**. *J Biol Chem* 2008, **283**(9):5692-5698.
141. Kung JT, Colognori D, Lee JT: **Long noncoding RNAs: past, present, and future**. *Genetics* 2013, **193**(3):651-669.
142. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E *et al*: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs**. *Cell* 2007, **129**(7):1311-1323.
143. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R: **Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription**. *Nature* 2008, **454**(7200):126-130.

144. Broadbent KM, Park D, Wolf AR, Van Tyne D, Sims JS, Ribacke U, Volkman S, Duraisingh M, Wirth D, Sabeti PC *et al*: **A global transcriptional analysis of Plasmodium falciparum malaria reveals a novel family of telomere-associated lncRNAs**. *Genome Biol* 2011, **12**(6):R56.
145. Sierra-Miranda M, Delgadillo DM, Mancio-Silva L, Vargas M, Villegas-Sepulveda N, Martinez-Calvillo S, Scherf A, Hernandez-Rivas R: **Two long non-coding RNAs generated from subtelomeric regions accumulate in a novel perinuclear compartment in Plasmodium falciparum**. *Molecular and biochemical parasitology* 2012, **185**(1):36-47.
146. Epp C, Li F, Howitt CA, Chookajorn T, Deitsch KW: **Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite Plasmodium falciparum**. *Rna* 2009, **15**(1):116-127.
147. Liao Q, Shen J, Liu J, Sun X, Zhao G, Chang Y, Xu L, Li X, Zhao Y, Zheng H *et al*: **Genome-wide identification and functional annotation of Plasmodium falciparum long noncoding RNAs from RNA-seq data**. *Parasitology research* 2014, **113**(4):1269-1281.
148. Brown JM, Green J, das Neves RP, Wallace HA, Smith AJ, Hughes J, Gray N, Taylor S, Wood WG, Higgs DR *et al*: **Association between active genes occurs at nuclear speckles and is modulated by chromatin environment**. *The Journal of cell biology* 2008, **182**(6):1083-1097.
149. Morey C, Kress C, Bickmore WA: **Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression**. *Genome research* 2009, **19**(7):1184-1194.
150. Iborra FJ, Pombo A, Jackson DA, Cook PR: **Active RNA polymerases are localized within discrete transcription "factories' in human nuclei**. *Journal of cell science* 1996, **109 (Pt 6)**:1427-1436.
151. Jackson DA, Iborra FJ, Manders EM, Cook PR: **Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei**. *Mol Biol Cell* 1998, **9**(6):1523-1536.
152. Faro-Trindade I, Cook PR: **A conserved organization of transcription during embryonic stem cell differentiation and in cells with high C value**. *Mol Biol Cell* 2006, **17**(7):2910-2920.
153. Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R: **SR proteins function in coupling RNAP II transcription to pre-mRNA splicing**. *Molecular cell* 2007, **26**(6):867-881.
154. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd, Frank J, Manley JL: **Molecular architecture of the human pre-mRNA 3' processing complex**. *Molecular cell* 2009, **33**(3):365-376.
155. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS *et al*: **Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells**. *Nature genetics* 2010, **42**(1):53-61.
156. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W *et al*: **Active genes dynamically colocalize to shared sites of ongoing transcription**. *Nature genetics* 2004, **36**(10):1065-1071.
157. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH *et al*: **An oestrogen-receptor-alpha-bound human chromatin interactome**. *Nature* 2009, **462**(7269):58-64.
158. Papantonis A, Larkin JD, Wada Y, Ohta Y, Ihara S, Kodama T, Cook PR: **Active RNA polymerases: mobile or immobile molecular machines?** *PLoS Biol* 2010, **8**(7):e1000419.
159. Moraes CB, Dorval T, Contreras-Dominguez M, Dossin Fde M, Hansen MA, Genovesio A, Freitas-Junior LH: **Transcription sites are developmentally regulated during the asexual cycle of Plasmodium falciparum**. *PLoS One* 2013, **8**(2):e55539.
160. Eskiw CH, Fraser P: **Ultrastructural study of transcription factories in mouse erythroblasts**. *Journal of cell science* 2011, **124**(Pt 21):3676-3683.
161. Mancio-Silva L, Zhang Q, Scheidig-Benatar C, Scherf A: **Clustering of dispersed ribosomal DNA and its role in gene regulation and chromosome-end associations in malaria parasites**. *Proc Natl Acad Sci U S A* 2010, **107**(34):15117-15122.

162. Issar N, Ralph SA, Mancio-Silva L, Keeling C, Scherf A: **Differential sub-nuclear localisation of repressive and activating histone methyl modifications in P. falciparum**. *Microbes and infection / Institut Pasteur* 2009, **11**(3):403-407.
163. Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, Tonkin CJ, Langer C, Crabb BS, Cowman AF: **Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in Plasmodium falciparum**. *International journal for parasitology* 2010, **40**(1):109-121.
164. Lopez-Rubio JJ, Mancio-Silva L, Scherf A: **Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites**. *Cell Host Microbe* 2009, **5**(2):179-190.
165. Ralph SA, Scheidig-Benatar C, Scherf A: **Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations**. *Proc Natl Acad Sci U S A* 2005, **102**(15):5414-5419.
166. Guizetti J, Martins RM, Guadagnini S, Claes A, Scherf A: **Nuclear pores and perinuclear expression sites of var and ribosomal DNA genes correspond to physically distinct regions in Plasmodium falciparum**. *Eukaryot Cell* 2013, **12**(5):697-702.
167. Coleman BI, Ribacke U, Manary M, Bei AK, Winzeler EA, Wirth DF, Duraisingh MT: **Nuclear repositioning precedes promoter accessibility and is linked to the switching frequency of a Plasmodium falciparum invasion gene**. *Cell Host Microbe* 2012, **12**(6):739-750.
168. Rodley CD, Bertels F, Jones B, O'Sullivan JM: **Global identification of yeast chromosome interactions using Genome conformation capture**. *Fungal genetics and biology : FG & B* 2009, **46**(11):879-886.
169. Lemieux JE, Kyes SA, Otto TD, Feller AI, Eastman RT, Pinches RA, Berriman M, Su XZ, Newbold CI: **Genome-wide profiling of chromosome interactions in Plasmodium falciparum characterizes nuclear architecture and reconfigurations associated with antigenic variation**. *Mol Microbiol* 2013, **90**(3):519-537.
170. Ong CT, Corces VG: **Enhancer function: new insights into the regulation of tissue-specific gene expression**. *Nature reviews Genetics* 2011, **12**(4):283-293.
171. Gruenbaum Y, Goldman RD, Meyuhas R, Mills E, Margalit A, Fridkin A, Dayani Y, Prokocimer M, Enosh A: **The nuclear lamina and its functions in the nucleus**. *International review of cytology* 2003, **226**:1-62.
172. Akhtar A, Gasser SM: **The nuclear envelope and transcriptional control**. *Nature reviews Genetics* 2007, **8**(7):507-517.
173. Weiner A, Dahan-Pasternak N, Shimoni E, Shinder V, von Huth P, Elbaum M, Dzikowski R: **3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite Plasmodium falciparum**. *Cell Microbiol* 2011, **13**(7):967-977.
174. Thomas MC, Chiang CM: **The general transcription machinery and general cofactors**. *Crit Rev Biochem Mol Biol* 2006, **41**(3):105-178.
175. Bischoff E, Vaquero C: **In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum**. *BMC Genomics* 2010, **11**:34.
176. Butler JE, Kadonaga JT: **The RNA polymerase II core promoter: a key component in the regulation of gene expression**. *Genes & development* 2002, **16**(20):2583-2592.
177. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT: **The RNA polymerase II core promoter - the gateway to transcription**. *Current opinion in cell biology* 2008, **20**(3):253-259.
178. Thomas MC, Chiang CM: **The general transcription machinery and general cofactors**. *Critical reviews in biochemistry and molecular biology* 2006, **41**(3):105-178.
179. Smale ST, Kadonaga JT: **The RNA polymerase II core promoter**. *Annu Rev Biochem* 2003, **72**:449-479.
180. FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C: **Clustering of DNA sequences in human promoters**. *Genome research* 2004, **14**(8):1562-1574.

181. Basehoar AD, Zanton SJ, Pugh BF: **Identification and distinct regulation of yeast TATA box-containing genes**. *Cell* 2004, **116**(5):699-709.
182. Kim J, Iyer VR: **Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles**. *Mol Cell Biol* 2004, **24**(18):8104-8112.
183. Paulson M, Press C, Smith E, Tanese N, Levy DE: **IFN-Stimulated transcription through a TBP-free acetyltransferase complex escapes viral shutoff**. *Nat Cell Biol* 2002, **4**(2):140-147.
184. Wright KJ, Marr MT, 2nd, Tjian R: **TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter**. *Proc Natl Acad Sci U S A* 2006, **103**(33):12347-12352.
185. Demeny MA, Soutoglou E, Nagy Z, Scheer E, Janoshazi A, Richardot M, Argentini M, Kessler P, Tora L: **Identification of a small TAF complex and its role in the assembly of TAF-containing complexes**. *PLoS One* 2007, **2**(3):e316.
186. Reina JH, Hernandez N: **On a roll for new TRF targets**. *Genes & development* 2007, **21**(22):2855-2860.
187. Mohan WS, Jr., Scheer E, Wendling O, Metzger D, Tora L: **TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice**. *Mol Cell Biol* 2003, **23**(12):4307-4318.
188. Hiller M, Chen X, Pringle MJ, Suchorolski M, Sancak Y, Viswanathan S, Bolival B, Lin TY, Marino S, Fuller MT: **Testis-specific TAF homologs collaborate to control a tissue-specific transcription program**. *Development* 2004, **131**(21):5297-5308.
189. Deato MD, Tjian R: **Switching of the core transcription machinery during myogenesis**. *Genes & development* 2007, **21**(17):2137-2149.
190. Hart DO, Raha T, Lawson ND, Green MR: **Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3**. *Nature* 2007, **450**(7172):1082-1085.
191. Bajic VB, Tan SL, Suzuki Y, Sugano S: **Promoter prediction analysis on the whole human genome**. *Nature biotechnology* 2004, **22**(11):1467-1473.
192. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM: **Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome**. *Genome research* 2006, **16**(1):1-10.
193. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC *et al*: **Genome-wide analysis of mammalian promoter architecture and evolution**. *Nature genetics* 2006, **38**(6):626-635.
194. Latinkic BV, Umbhauer M, Neal KA, Lerchner W, Smith JC, Cunliffe V: **The Xenopus Brachyury promoter is activated by FGF and low concentrations of activin and suppressed by high concentrations of activin and by paired-type homeodomain proteins**. *Genes & development* 1997, **11**(23):3265-3276.
195. Quitschke WW, DePonti-Zilli L, Lin ZY, Paterson BM: **Identification of two nuclear factor-binding domains on the chicken cardiac actin promoter: implications for regulation of the gene**. *Mol Cell Biol* 1989, **9**(8):3218-3230.
196. Chalkley GE, Verrijzer CP: **DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator**. *Embo J* 1999, **18**(17):4835-4845.
197. Smale ST, Baltimore D: **The "initiator" as a transcription control element**. *Cell* 1989, **57**(1):103-113.
198. Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E: **Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters**. *Gene* 2007, **389**(1):52-65.
199. FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C: **Comparative genomics of Drosophila and human core promoters**. *Genome Biol* 2006, **7**(7):R53.
200. Burke TW, Kadonaga JT: **The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila**. *Genes & development* 1997, **11**(22):3020-3031.

201. Gershenzon NI, Trifonov EN, Ioshikhes IP: **The features of Drosophila core promoters revealed by statistical analysis**. *BMC Genomics* 2006, **7**:161.
202. Kadonaga JT: **The DPE, a core promoter element for transcription by RNA polymerase II**. *Exp Mol Med* 2002, **34**(4):259-264.
203. Juven-Gershon T, Hsu JY, Kadonaga JT: **Caudal, a key developmental regulator, is a DPE-specific transcriptional factor**. *Genes & development* 2008, **22**(20):2823-2830.
204. Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT: **The MTE, a new core promoter element for transcription by RNA polymerase II**. *Genes & development* 2004, **18**(13):1606-1617.
205. Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A: **A code for transcription initiation in mammalian genomes**. *Genome research* 2008, **18**(1):1-12.
206. Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH: **New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB**. *Genes & development* 1998, **12**(1):34-44.
207. Deng W, Roberts SG: **A core promoter element downstream of the TATA box that is recognized by TFIIB**. *Genes & development* 2005, **19**(20):2418-2423.
208. Evans R, Fairley JA, Roberts SG: **Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB**. *Genes & development* 2001, **15**(22):2945-2949.
209. Lewis BA, Kim TK, Orkin SH: **A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts**. *Proc Natl Acad Sci U S A* 2000, **97**(13):7172-7177.
210. Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA: **Functional characterization of core promoter elements: the downstream core element is recognized by TAF1**. *Mol Cell Biol* 2005, **25**(21):9674-9686.
211. Tokusumi Y, Ma Y, Song X, Jacobson RH, Takada S: **The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters**. *Mol Cell Biol* 2007, **27**(5):1844-1858.
212. Ohtsuki S, Levine M, Cai HN: **Different core promoters possess distinct regulatory activities in the Drosophila embryo**. *Genes & development* 1998, **12**(4):547-556.
213. Emami KH, Navarre WW, Smale ST: **Core promoter specificities of the Sp1 and VP16 transcriptional activation domains**. *Mol Cell Biol* 1995, **15**(11):5906-5916.
214. Maston GA, Evans SK, Green MR: **Transcriptional regulatory elements in the human genome**. *Annu Rev Genomics Hum Genet* 2006, **7**:29-59.
215. Su W, Jackson S, Tjian R, Echols H: **DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1**. *Genes & development* 1991, **5**(5):820-826.
216. Calhoun VC, Stathopoulos A, Levine M: **Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex**. *Proc Natl Acad Sci U S A* 2002, **99**(14):9243-9247.
217. de Laat W, Grosveld F: **Spatial organization of gene expression: the active chromatin hub**. *Chromosome Res* 2003, **11**(5):447-459.
218. Zhao H, Dean A: **Organizing the genome: enhancers and insulators**. *Biochem Cell Biol* 2005, **83**(4):516-524.
219. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE *et al*: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project**. *Nature* 2007, **447**(7146):799-816.
220. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA: **Interchromosomal associations between alternatively expressed loci**. *Nature* 2005, **435**(7042):637-645.

221. Maston GA, Evans SK, Green MR: **Transcriptional regulatory elements in the human genome**. *Annu Rev Genomics Hum Genet* 2006, **7**:29-59.
222. Dekker J, Rippe K, Dekker M, Kleckner N: **Capturing chromosome conformation**. *Science* 2002, **295**(5558):1306-1311.
223. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W: **Looping and interaction between hypersensitive sites in the active beta-globin locus**. *Molecular cell* 2002, **10**(6):1453-1465.
224. Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W: **The active spatial organization of the beta-globin locus requires the transcription factor EKLF**. *Genes & development* 2004, **18**(20):2485-2490.
225. Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA: **Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1**. *Molecular cell* 2005, **17**(3):453-462.
226. Jing H, Vakoc CR, Ying L, Mandat S, Wang H, Zheng X, Blobel GA: **Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus**. *Molecular cell* 2008, **29**(2):232-242.
227. Murrell A, Heeson S, Reik W: **Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops**. *Nature genetics* 2004, **36**(8):889-893.
228. Spilianakis CG, Flavell RA: **Long-range intrachromosomal interactions in the T helper type 2 cytokine locus**. *Nature immunology* 2004, **5**(10):1017-1027.
229. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W: **CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus**. *Genes & development* 2006, **20**(17):2349-2354.
230. Mishiro T, Ishihara K, Hino S, Tsutsumi S, Aburatani H, Shirahige K, Kinoshita Y, Nakao M: **Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster**. *The EMBO journal* 2009, **28**(9):1234-1245.
231. Bedford DC, Kasper LH, Fukuyama T, Brindle PK: **Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases**. *Epigenetics : official journal of the DNA Methylation Society* 2010, **5**(1):9-15.
232. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA *et al*: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome**. *Nature genetics* 2007, **39**(3):311-318.
233. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F *et al*: **ChIP-seq accurately predicts tissue-specific activity of enhancers**. *Nature* 2009, **457**(7231):854-858.
234. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW *et al*: **Histone modifications at human enhancers reflect global cell-type-specific gene expression**. *Nature* 2009, **459**(7243):108-112.
235. Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J *et al*: **A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate**. *Nature immunology* 2010, **11**(7):635-643.
236. Caikovski M, Yokthongwattana C, Habu Y, Nishimura T, Mathieu O, Paszkowski J: **Divergent evolution of CHD3 proteins resulted in MOM1 refining epigenetic control in vascular plants**. *PLoS genetics* 2008, **4**(8):e1000165.
237. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M: **FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription**. *Cell* 2008, **132**(6):958-970.
238. Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G: **H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions**. *Nature genetics* 2009, **41**(8):941-945.

239. He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M *et al*: **Nucleosome dynamics define transcriptional enhancers**. *Nature genetics* 2010, **42**(4):343-347.
240. Fakhouri TH, Stevenson J, Chisholm AD, Mango SE: **Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA**. *PLoS genetics* 2010, **6**(8).
241. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL *et al*: **Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages**. *Immunity* 2010, **32**(3):317-328.
242. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities**. *Molecular cell* 2010, **38**(4):576-589.
243. Xu J, Watts JA, Pope SD, Gadue P, Kamps M, Plath K, Zaret KS, Smale ST: **Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells**. *Genes & development* 2009, **23**(24):2824-2838.
244. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S *et al*: **Widespread transcription at neuronal activity-regulated enhancers**. *Nature* 2010, **465**(7295):182-187.
245. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA *et al*: **A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression**. *Nature* 2011, **472**(7341):120-124.
246. Ogbourne S, Antalis TM: **Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes**. *Biochem J* 1998, **331 (Pt 1)**:1-14.
247. Privalsky ML: **The role of corepressors in transcriptional regulation by nuclear hormone receptors**. *Annu Rev Physiol* 2004, **66**:315-360.
248. Harris MB, Mostecki J, Rothman PB: **Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function**. *J Biol Chem* 2005, **280**(13):13114-13121.
249. Li L, He S, Sun JM, Davie JR: **Gene regulation by Sp1 and Sp3**. *Biochem Cell Biol* 2004, **82**(4):460-471.
250. Srinivasan L, Atchison ML: **YY1 DNA binding and PcG recruitment requires CtBP**. *Genes & development* 2004, **18**(21):2596-2601.
251. Chen L, Widom J: **Mechanism of transcriptional silencing in yeast**. *Cell* 2005, **120**(1):37-48.
252. Feinberg AP: **Phenotypic plasticity and the epigenetics of human disease**. *Nature* 2007, **447**(7143):433-440.
253. Gaszner M, Felsenfeld G: **Insulators: exploiting transcriptional and epigenetic mechanisms**. *Nature reviews Genetics* 2006, **7**(9):703-713.
254. Sasaki H, Ishihara K, Kato R: **Mechanisms of Igf2/H19 imprinting: DNA methylation, chromatin and long-distance gene regulation**. *Journal of biochemistry* 2000, **127**(5):711-715.
255. Kanduri C, Holmgren C, Pilartz M, Franklin G, Kanduri M, Liu L, Ginjala V, Ulleras E, Mattsson R, Ohlsson R: **The 5' flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication**. *Curr Biol* 2000, **10**(8):449-457.
256. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G: **CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species**. *Molecular cell* 2004, **13**(2):291-298.
257. Gerasimova TI, Byrd K, Corces VG: **A chromatin insulator determines the nuclear localization of DNA**. *Molecular cell* 2000, **6**(5):1025-1035.
258. Byrd K, Corces VG: **Visualization of chromatin domains created by the gypsy insulator of Drosophila**. *The Journal of cell biology* 2003, **162**(4):565-574.
259. Maeda RK, Karch F: **Making connections: boundaries and insulators in Drosophila**. *Current opinion in genetics & development* 2007, **17**(5):394-399.

260. Nolis IK, McKay DJ, Mantouvalou E, Lomvardas S, Merika M, Thanos D: **Transcription factors mediate long-range enhancer-promoter interactions**. *Proc Natl Acad Sci U S A* 2009, **106**(48):20222-20227.
261. Valenzuela L, Kamakaka RT: **Chromatin insulators**. *Annual review of genetics* 2006, **40**:107-138.
262. Grewal SI, Jia S: **Heterochromatin revisited**. *Nature reviews Genetics* 2007, **8**(1):35-46.
263. Litt MD, Simpson M, Recillas-Targa F, Prioleau MN, Felsenfeld G: **Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci**. *The EMBO journal* 2001, **20**(9):2224-2235.
264. Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G: **Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus**. *Science* 2001, **293**(5539):2453-2455.
265. West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G: **Recruitment of histone modifications by USF proteins at a vertebrate barrier element**. *Molecular cell* 2004, **16**(3):453-463.
266. Noma K, Cam HP, Maraia RJ, Grewal SI: **A role for TFIIIC transcription factor complex in genome organization**. *Cell* 2006, **125**(5):859-872.
267. Roeder RG: **Transcriptional regulation and the role of diverse coactivators in animal cells**. *FEBS letters* 2005, **579**(4):909-915.
268. Kornberg RD: **Mediator and the mechanism of transcriptional activation**. *Trends Biochem Sci* 2005, **30**(5):235-239.
269. Malik S, Roeder RG: **Dynamic regulation of pol II transcription by the mammalian Mediator complex**. *Trends Biochem Sci* 2005, **30**(5):256-263.
270. Backstrom S, Elfving N, Nilsson R, Wingsle G, Bjorklund S: **Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit**. *Molecular cell* 2007, **26**(5):717-729.
271. Bourbon HM: **Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex**. *Nucleic acids research* 2008, **36**(12):3993-4008.
272. Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FC, Werner M: **A high resolution protein interaction map of the yeast Mediator complex**. *Nucleic acids research* 2004, **32**(18):5379-5391.
273. Elmlund H, Baraznenok V, Lindahl M, Samuelsen CO, Koeck PJ, Holmberg S, Hebert H, Gustafsson CM: **The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II**. *Proc Natl Acad Sci U S A* 2006, **103**(43):15788-15793.
274. Cai G, Imasaki T, Takagi Y, Asturias FJ: **Mediator structural conservation and implications for the regulation mechanism**. *Structure* 2009, **17**(4):559-567.
275. Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M: **Mediator-dependent recruitment of TFIIH modules in preinitiation complex**. *Molecular cell* 2008, **31**(3):337-346.
276. Baek HJ, Kang YK, Roeder RG: **Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly**. *J Biol Chem* 2006, **281**(22):15172-15181.
277. Black JC, Choi JE, Lombardo SR, Carey M: **A mechanism for coordinating chromatin modification and preinitiation complex assembly**. *Molecular cell* 2006, **23**(6):809-818.
278. Wallberg AE, Yamamura S, Malik S, Spiegelman BM, Roeder RG: **Coordination of p300-mediated chromatin remodeling and TRAP/mediator function through coactivator PGC-1alpha**. *Molecular cell* 2003, **12**(5):1137-1149.
279. Ding N, Zhou H, Esteve PO, Chin HG, Kim S, Xu X, Joseph SM, Friez MJ, Schwartz CE, Pradhan S *et al*: **Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation**. *Molecular cell* 2008, **31**(3):347-359.

280. Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT: **Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells**. *Genes & development* 2011, **25**(7):742-754.
281. Malik S, Barrero MJ, Jones T: **Identification of a regulator of transcription elongation as an accessory factor for the human Mediator coactivator**. *Proc Natl Acad Sci U S A* 2007, **104**(15):6182-6187.
282. Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, Zhou Q: **Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4**. *Molecular cell* 2005, **19**(4):535-545.
283. Malagon F, Tong AH, Shafer BK, Strathern JN: **Genetic interactions of DST1 in Saccharomyces cerevisiae suggest a role of TFIIS in the initiation-elongation transition**. *Genetics* 2004, **166**(3):1215-1227.
284. Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C *et al*: **Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II**. *Mol Cell Biol* 2003, **23**(12):4207-4218.
285. Yudkovsky N, Ranish JA, Hahn S: **A transcription reinitiation intermediate that is stabilized by activator**. *Nature* 2000, **408**(6809):225-229.
286. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS *et al*: **Mediator and cohesin connect gene expression and chromatin architecture**. *Nature* 2010, **467**(7314):430-435.
287. Fraser NW, Sehgal PB, Darnell JE: **DRB-induced premature termination of late adenovirus transcription**. *Nature* 1978, **272**(5654):590-593.
288. Gariglio P, Bellard M, Chambon P: **Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes**. *Nucleic acids research* 1981, **9**(11):2589-2598.
289. Rougvie AE, Lis JT: **The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged**. *Cell* 1988, **54**(6):795-804.
290. Rasmussen EB, Lis JT: **In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes**. *Proc Natl Acad Sci U S A* 1993, **90**(17):7923-7927.
291. Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA: **Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase**. *Molecular cell* 2006, **24**(5):735-746.
292. Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS: **NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila**. *Mol Cell Biol* 2008, **28**(10):3290-3300.
293. Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA: **RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo**. *Nature genetics* 2007, **39**(12):1512-1516.
294. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA: **A chromatin landmark and transcription initiation at most promoters in human cells**. *Cell* 2007, **130**(1):77-88.
295. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters**. *Science* 2008, **322**(5909):1845-1848.
296. Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI: **X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila**. *Nature* 2011, **471**(7336):115-118.
297. Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, Adelman K: **Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation**. *Cell* 2010, **143**(4):540-551.
298. Gilchrist DA, Nechaev S, Lee C, Ghosh SK, Collins JB, Li L, Gilmour DS, Adelman K: **NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly**. *Genes & development* 2008, **22**(14):1921-1933.

299. Lanzer M, de Bruin D, Ravetch JV: **Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates**. *The EMBO journal* 1992, **11**(5):1949-1955.
300. Lanzer M, de Bruin D, Ravetch JV: **A sequence element associated with the Plasmodium falciparum KAHRP gene is the site of developmentally regulated protein-DNA interactions**. *Nucleic acids research* 1992, **20**(12):3051-3056.
301. Coulson RM, Hall N, Ouzounis CA: **Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum**. *Genome research* 2004, **14**(8):1548-1554.
302. Ruvalcaba-Salazar OK, del Carmen Ramirez-Estudillo M, Montiel-Condado D, Recillas-Targa F, Vargas M, Hernandez-Rivas R: **Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions**. *Molecular and biochemical parasitology* 2005, **140**(2):183-196.
303. Gopalakrishnan AM, Nyindodo LA, Ross Fergus M, Lopez-Estrano C: **Plasmodium falciparum: Preinitiation complex occupancy of active and inactive promoters during erythrocytic stage**. *Exp Parasitol* 2009, **121**(1):46-54.
304. Templeton TJ, Iyer LM, Anantharaman V, Enomoto S, Abrahante JE, Subramanian GM, Hoffman SL, Abrahamsen MS, Aravind L: **Comparative analysis of apicomplexa and genomic diversity in eukaryotes**. *Genome research* 2004, **14**(9):1686-1695.
305. Balaji S, Babu MM, Iyer LM, Aravind L: **Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains**. *Nucleic acids research* 2005, **33**(13):3994-4006.
306. Wessler SR: **Homing into the origin of the AP2 DNA binding domain**. *Trends in plant science* 2005, **10**(2):54-56.
307. Coulson RM, Hall N, Ouzounis CA: **Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum**. *Genome Res* 2004, **14**(8):1548-1554.
308. Callebaut I, Prat K, Meurice E, Mornon JP, Tomavo S: **Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes**. *BMC genomics* 2005, **6**:100.
309. Gissot M, Briquet S, Refour P, Boschet C, Vaquero C: **PfMyb1, a Plasmodium falciparum transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation**. *J Mol Biol* 2005, **346**(1):29-42.
310. Hahn S, Young ET: **Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators**. *Genetics* 2011, **189**(3):705-736.
311. Fogel GB, Weekes DG, Varga G, Dow ER, Craven AM, Harlow HB, Su EW, Onyia JE, Su C: **A statistical analysis of the TRANSFAC database**. *Bio Systems* 2005, **81**(2):137-154.
312. Boschet C, Gissot M, Briquet S, Hamid Z, Claudel-Renard C, Vaquero C: **Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 Plasmodium falciparum clones**. *Molecular and biochemical parasitology* 2004, **138**(1):159-163.
313. Komaki-Yasuda K, Okuwaki M, Kano S, Nagata K, Kawazu S: **5' sequence- and chromatin modification-dependent gene expression in Plasmodium falciparum erythrocytic stage**. *Molecular and biochemical parasitology* 2008, **162**(1):40-51.
314. Yano K, Komaki-Yasuda K, Kobayashi T, Takemae H, Kita K, Kano S, Kawazu S: **Expression of mRNAs and proteins for peroxiredoxins in Plasmodium falciparum erythrocytic stage**. *Parasitology international* 2005, **54**(1):35-41.
315. Komaki-Yasuda K, Okuwaki M, Nagata K, Kawazu S, Kano S: **Identification of a novel and unique transcription factor in the intraerythrocytic stage of Plasmodium falciparum**. *PLoS One* 2013, **8**(9):e74701.
316. Valverde R, Edwards L, Regan L: **Structure and function of KH domains**. *The FEBS journal* 2008, **275**(11):2712-2726.

317. Goyal M, Alam A, Iqbal MS, Dey S, Bindu S, Pal C, Banerjee A, Chakrabarti S, Bandyopadhyay U: **Identification and molecular characterization of an Alba-family protein from human malaria parasite Plasmodium falciparum**. *Nucleic acids research* 2012, **40**(3):1174-1190.
318. Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinas M: **Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite**. *PLoS pathogens* 2010, **6**(10):e1001165.
319. De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, Bulyk ML, Llinas M: **Specific DNA-binding by apicomplexan AP2 transcription factors**. *Proc Natl Acad Sci U S A* 2008, **105**(24):8393-8398.
320. Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, Waters AP, Kato T, Kaneko I: **Identification of a transcription factor in the mosquito-invasive stage of malaria parasites**. *Mol Microbiol* 2009, **71**(6):1402-1414.
321. Yuda M, Iwanaga S, Shigenobu S, Kato T, Kaneko I: **Transcription factor AP2-Sp and its target genes in malarial sporozoites**. *Mol Microbiol* 2010, **75**(4):854-863.
322. Iwanaga S, Kaneko I, Kato T, Yuda M: **Identification of an AP2-family protein that is critical for malaria liver stage development**. *PLoS One* 2012, **7**(11):e47557.
323. Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BT, Moes S, Bozdech Z, Jenoe P, Stunnenberg HG, Voss TS: **A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology**. *PLoS pathogens* 2010, **6**(2):e1000784.
324. Voss TS, Kaestli M, Vogel D, Bopp S, Beck HP: **Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters**. *Mol Microbiol* 2003, **48**(6):1593-1607.
325. Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA *et al*: **A transcriptional switch underlies commitment to sexual development in malaria parasites**. *Nature* 2014, **507**(7491):248-252.
326. Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, Religa AA, Bushell E, Graham AL, Cameron R *et al*: **A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium**. *Nature* 2014, **507**(7491):253-257.
327. Russell K, Hasenkamp S, Emes R, Horrocks P: **Analysis of the spatial and temporal arrangement of transcripts over intergenic regions in the human malarial parasite Plasmodium falciparum**. *BMC Genomics* 2013, **14**:267.
328. Lopez-Estrano C, Gopalakrishnan AM, Semblat JP, Fergus MR, Mazier D, Haldar K: **An enhancer-like region regulates hrp3 promoter stage-specific gene expression in the human malaria parasite Plasmodium falciparum**. *Biochim Biophys Acta* 2007, **1769**(7-8):506-513.
329. Militello KT, Dodge M, Bethke L, Wirth DF: **Identification of regulatory elements in the Plasmodium falciparum genome**. *Molecular and biochemical parasitology* 2004, **134**(1):75-88.
330. Crabb BS, Cowman AF: **Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodium falciparum**. *Proc Natl Acad Sci U S A* 1996, **93**(14):7289-7294.
331. Horrocks P, Kilbey BJ: **Physical and functional mapping of the transcriptional start sites of Plasmodium falciparum proliferating cell nuclear antigen**. *Molecular and biochemical parasitology* 1996, **82**(2):207-215.
332. Dechering KJ, Kaan AM, Mbacham W, Wirth DF, Eling W, Konings RN, Stunnenberg HG: **Isolation and functional characterization of two distinct sexual-stage-specific promoters of the human malaria parasite Plasmodium falciparum**. *Mol Cell Biol* 1999, **19**(2):967-978.
333. Osta M, Gannoun-Zaki L, Bonnefoy S, Roy C, Vial HJ: **A 24 bp cis-acting element essential for the transcriptional activity of Plasmodium falciparum CDP-diacylglycerol synthase gene promoter**. *Molecular and biochemical parasitology* 2002, **121**(1):87-98.

334. Sunil S, Chauhan VS, Malhotra P: **Distinct and stage specific nuclear factors regulate the expression of falcipains, Plasmodium falciparum cysteine proteases**. *BMC Mol Biol* 2008, **9**:47.
335. Young JA, Johnson JR, Benner C, Yan SF, Chen K, Le Roch KG, Zhou Y, Winzeler EA: **In silico discovery of transcription regulatory elements in Plasmodium falciparum**. *BMC Genomics* 2008, **9**:70.
336. Horrocks P, Lanzer M: **Mutational analysis identifies a five base pair cis-acting sequence essential for GBP130 promoter activity in Plasmodium falciparum**. *Molecular and biochemical parasitology* 1999, **99**(1):77-87.
337. Porter ME: **Positive and negative effects of deletions and mutations within the 5' flanking sequences of Plasmodium falciparum DNA polymerase delta**. *Molecular and biochemical parasitology* 2002, **122**(1):9-19.
338. Wickham ME, Thompson JK, Cowman AF: **Characterisation of the merozoite surface protein-2 promoter using stable and transient transfection in Plasmodium falciparum**. *Molecular and biochemical parasitology* 2003, **129**(2):147-156.
339. Azevedo MF, del Portillo HA: **Promoter regions of Plasmodium vivax are poorly or not recognized by Plasmodium falciparum**. *Malaria journal* 2007, **6**:20.
340. Olivieri A, Silvestrini F, Sanchez M, Alano P: **A 140-bp AT-rich sequence mediates positive and negative transcriptional control of a Plasmodium falciparum developmentally regulated promoter**. *International journal for parasitology* 2008, **38**(3-4):299-312.
341. Helm S, Lehmann C, Nagel A, Stanway RR, Horstmann S, Llinas M, Heussler VT: **Identification and characterization of a liver stage-specific promoter region of the malaria parasite Plasmodium**. *PLoS One* 2010, **5**(10):e13653.
342. Wong EH, Hasenkamp S, Horrocks P: **Analysis of the molecular mechanisms governing the stage-specific expression of a prototypical housekeeping gene during intraerythrocytic development of P. falciparum**. *J Mol Biol* 2011, **408**(2):205-221.
343. Militello KT, Dodge M, Bethke L, Wirth DF: **Identification of regulatory elements in the Plasmodium falciparum genome**. *Mol Biochem Parasitol* 2004, **134**(1):75-88.
344. Gunasekera AM, Myrick A, Militello KT, Sims JS, Dong CK, Gierahn T, Le Roch K, Winzeler E, Wirth DF: **Regulatory motifs uncovered among gene expression clusters in Plasmodium falciparum**. *Mol Biochem Parasitol* 2007, **153**(1):19-30.
345. Voss TS, Kaestli M, Vogel D, Bopp S, Beck HP: **Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters**. *Mol Microbiol* 2003, **48**(6):1593-1607.
346. Horrocks P, Dechering K, Lanzer M: **Control of gene expression in Plasmodium falciparum**. *Mol Biochem Parasitol* 1998, **95**(2):171-181.
347. Osta M, Gannoun-Zaki L, Bonnefoy S, Roy C, Vial HJ: **A 24 bp cis-acting element essential for the transcriptional activity of Plasmodium falciparum CDP-diacylglycerol synthase gene promoter**. *Mol Biochem Parasitol* 2002, **121**(1):87-98.
348. Olivieri A, Silvestrini F, Sanchez M, Alano P: **A 140-bp AT-rich sequence mediates positive and negative transcriptional control of a Plasmodium falciparum developmentally regulated promoter**. *Int J Parasitol* 2008, **38**(3-4):299-312.
349. Polson HE, Blackman MJ: **A role for poly(dA)poly(dT) tracts in directing activity of the Plasmodium falciparum calmodulin gene promoter**. *Mol Biochem Parasitol* 2005, **141**(2):179-189.
350. Wickham ME, Thompson JK, Cowman AF: **Characterisation of the merozoite surface protein-2 promoter using stable and transient transfection in Plasmodium falciparum**. *Mol Biochem Parasitol* 2003, **129**(2):147-156.
351. Tham WH, Payne PD, Brown GV, Rogerson SJ: **Identification of basic transcriptional elements required for rif gene transcription**. *Int J Parasitol* 2007, **37**(6):605-615.

352. Gopalakrishnan AM, Lopez-Estrano C: **Role of cis-regulatory elements on the ring-specific hrp3 promoter in the human parasite Plasmodium falciparum**. *Parasitol Res* 2010, **106**(4):833-845.
353. Bussemaker HJ, Li H, Siggia ED: **Regulatory element detection using correlation with expression**. *Nature genetics* 2001, **27**(2):167-171.
354. Elemento O, Slonim N, Tavazoie S: **A universal framework for regulatory element discovery across all genomes and data types**. *Mol Cell* 2007, **28**(2):337-350.
355. Iengar P, Joshi NV: **Identification of putative regulatory motifs in the upstream regions of co-expressed functional groups of genes in Plasmodium falciparum**. *BMC Genomics* 2009, **10**:18.
356. Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinas M: **Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite**. *PLoS Pathog* 2010, **6**(10):e1001165.
357. van Noort V, Huynen MA: **Combinatorial gene regulation in Plasmodium falciparum**. *Trends in genetics : TIG* 2006, **22**(2):73-78.
358. Wu J, Sieglaff DH, Gervin J, Xie XS: **Discovering regulatory motifs in the Plasmodium genome using comparative genomics**. *Bioinformatics* 2008, **24**(17):1843-1849.
359. Elemento O, Slonim N, Tavazoie S: **A universal framework for regulatory element discovery across all genomes and data types**. *Molecular cell* 2007, **28**(2):337-350.
360. Mok S, Liong KY, Lim EH, Huang X, Zhu L, Preiser PR, Bozdech Z: **Structural polymorphism in the promoter of pfmrp2 confers Plasmodium falciparum tolerance to quinoline drugs**. *Mol Microbiol* 2014, **91**(5):918-934.
361. Trager W, Jensen JB: **Human malaria parasites in continuous culture**. *Science* 1976, **193**(4254):673-675.
362. Dassa E, Bouige P: **The ABC of ABCS: a phylogenetic and functional classification of ABC systems in living organisms**. *Res Microbiol* 2001, **152**(3-4):211-229.
363. Dean M, Rzhetsky A, Allikmets R: **The human ATP-binding cassette (ABC) transporter superfamily**. *Genome research* 2001, **11**(7):1156-1166.
364. Djimde A, Doumbo OK, Cortese JF, Kayentao K, Doumbo S, Diourte Y, Coulibaly D, Dicko A, Su XZ, Nomura T *et al*: **A molecular marker for chloroquine-resistant falciparum malaria**. *The New England journal of medicine* 2001, **344**(4):257-263.
365. Jones PM, George AM: **Multidrug resistance in parasites: ABC transporters, P-glycoproteins and molecular modelling**. *International journal for parasitology* 2005, **35**(5):555-566.
366. Raj DK, Mu J, Jiang H, Kabat J, Singh S, Sullivan M, Fay MP, McCutchan TF, Su XZ: **Disruption of a Plasmodium falciparum multidrug resistance-associated protein (PfMRP) alters its fitness and transport of antimalarial drugs and glutathione**. *J Biol Chem* 2009, **284**(12):7687-7696.
367. Dahlstrom S, Ferreira PE, Veiga MI, Sedighi N, Wiklund L, Martensson A, Farnert A, Sisowath C, Osorio L, Darban H *et al*: **Plasmodium falciparum multidrug resistance protein 1 and artemisinin-based combination therapy in Africa**. *The Journal of infectious diseases* 2009, **200**(9):1456-1464.
368. Dahlstrom S, Veiga MI, Martensson A, Bjorkman A, Gil JP: **Polymorphism in PfMRP1 (Plasmodium falciparum multidrug resistance protein 1) amino acid 1466 associated with resistance to sulfadoxine-pyrimethamine treatment**. *Antimicrob Agents Chemother* 2009, **53**(6):2553-2556.
369. Veiga MI, Ferreira PE, Jornhagen L, Malmberg M, Kone A, Schmidt BA, Petzold M, Bjorkman A, Nosten F, Gil JP: **Novel polymorphisms in Plasmodium falciparum ABC transporter genes are associated with major ACT antimalarial drug resistance**. *PLoS One* 2011, **6**(5):e20212.
370. Mu J, Ferdig MT, Feng X, Joy DA, Duan J, Furuya T, Subramanian G, Aravind L, Cooper RA, Wootton JC *et al*: **Multiple transporters associated with malaria parasite responses to chloroquine and quinine**. *Mol Microbiol* 2003, **49**(4):977-989.

371. Sullivan DJ, Jr., Matile H, Ridley RG, Goldberg DE: **A common mechanism for blockade of heme polymerization by antimalarial quinolines**. *J Biol Chem* 1998, **273**(47):31103-31107.
372. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, Ribas de Pouplana L, Preiser PR, Bozdech Z, Cortes A: **Transcriptional variation in the malaria parasite Plasmodium falciparum**. *Genome research* 2012, **22**(5):925-938.
373. Paumi CM, Chuk M, Snider J, Stagljar I, Michaelis S: **ABC transporters in Saccharomyces cerevisiae and their interactors: new technology advances the biology of the ABCC (MRP) subfamily**. *Microbiology and molecular biology reviews : MMBR* 2009, **73**(4):577-593.
374. Mok S, Liong KY, Lim EH, Huang X, Zhu L, Preiser PR, Bozdech Z: **Structural polymorphism in the promoter of pfmrp2 confers Plasmodium falciparum tolerance to quinoline drugs**. *Mol Microbiol* 2014, **91**(5):918-934.
375. Mok S, Imwong M, Mackinnon MJ, Sim J, Ramadoss R, Yi P, Mayxay M, Chotivanich K, Liong KY, Russell B *et al*: **Artemisinin resistance in Plasmodium falciparum is associated with an altered temporal pattern of transcription**. *BMC Genomics* 2011, **12**:391.
376. Milner DA, Jr., Pochet N, Krupka M, Williams C, Seydel K, Taylor TE, Van de Peer Y, Regev A, Wirth D, Daily JP *et al*: **Transcriptional profiling of Plasmodium falciparum parasites from patients with severe malaria identifies distinct low vs. high parasitemic clusters**. *PLoS One* 2012, **7**(7):e40739.
377. Holmes EC, Burch SS: **The causes and consequences of genetic variation in dengue virus**. *Trends Microbiol* 2000, **8**(2):74-77.
378. Dechering KJ, Cuelenaere K, Konings RN, Leunissen JA: **Distinct frequency-distributions of homopolymeric DNA tracts in different genomes**. *Nucleic acids research* 1998, **26**(17):4056-4062.
379. Levinson G, Gutman GA: **Slipped-strand mispairing: a major mechanism for DNA sequence evolution**. *Mol Biol Evol* 1987, **4**(3):203-221.
380. Streisinger G, Owen J: **Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4**. *Genetics* 1985, **109**(4):633-659.
381. Canceill D, Viguera E, Ehrlich SD: **Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency**. *J Biol Chem* 1999, **274**(39):27481-27490.
382. Farabaugh PJ, Schmeissner U, Hofer M, Miller JH: **Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the lacI gene of Escherichia coli**. *J Mol Biol* 1978, **126**(4):847-857.
383. Albertini AM, Hofer M, Calos MP, Tlsty TD, Miller JH: **Analysis of spontaneous deletions and gene amplification in the lac region of Escherichia coli**. *Cold Spring Harbor symposia on quantitative biology* 1983, **47 Pt 2**:841-850.
384. van Belkum A, van Leeuwen W, Scherer S, Verbrugh H: **Occurrence and structure-function relationship of pentameric short sequence repeats in microbial genomes**. *Res Microbiol* 1999, **150**(9-10):617-626.
385. Tan JC, Tan A, Checkley L, Honsa CM, Ferdig MT: **Variable numbers of tandem repeats in Plasmodium falciparum genes**. *J Mol Evol* 2010, **71**(4):268-278.
386. Zilversmit MM, Volkman SK, DePristo MA, Wirth DF, Awadalla P, Hartl DL: **Low-complexity regions in Plasmodium falciparum: missing links in the evolution of an extreme genome**. *Mol Biol Evol* 2010, **27**(9):2198-2209.
387. Haerty W, Golding GB: **Increased polymorphism near low-complexity sequences across the genomes of Plasmodium falciparum isolates**. *Genome Biol Evol* 2011, **3**:539-550.
388. Guler JL, Freeman DL, Ahyong V, Patrapuvich R, White J, Gujjar R, Phillips MA, DeRisi J, Rathod PK: **Asexual populations of the human malaria parasite, Plasmodium falciparum, use a two-step genomic strategy to acquire accurate, beneficial DNA amplifications**. *PLoS pathogens* 2013, **9**(5):e1003375.

389. Jurgelenaite R, Dijkstra TM, Kocken CH, Heskes T: **Gene regulation in the intraerythrocytic cycle of Plasmodium falciparum**. *Bioinformatics* 2009, **25**(12):1484-1491.
390. Hu G, Llinas M, Li J, Preiser PR, Bozdech Z: **Selection of long oligonucleotides for gene expression microarrays using weighted rank-sum strategy**. *BMC bioinformatics* 2007, **8**:350.
391. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU: **Molecular chaperone functions in protein folding and proteostasis**. *Annual review of biochemistry* 2013, **82**:323-355.
392. Banumathy G, Singh V, Pavithra SR, Tatu U: **Heat shock protein 90 function is essential for Plasmodium falciparum growth in human erythrocytes**. *J Biol Chem* 2003, **278**(20):18336-18345.
393. Pavithra SR, Banumathy G, Joy O, Singh V, Tatu U: **Recurrent fever promotes Plasmodium falciparum development in human erythrocytes**. *J Biol Chem* 2004, **279**(45):46692-46699.
394. Watanabe J: **Cloning and characterization of heat shock protein DnaJ homologues from Plasmodium falciparum and comparison with ring infected erythrocyte surface antigen**. *Molecular and biochemical parasitology* 1997, **88**(1-2):253-258.
395. Nicoll WS, Botha M, McNamara C, Schlange M, Pesce ER, Boshoff A, Ludewig MH, Zimmermann R, Cheetham ME, Chapple JP *et al*: **Cytosolic and ER J-domains of mammalian and parasitic origin can functionally interact with DnaK**. *The international journal of biochemistry & cell biology* 2007, **39**(4):736-751.
396. Botha M, Chiang AN, Needham PG, Stephens LL, Hoppe HC, Kulzer S, Przyborski JM, Lingelbach K, Wipf P, Brodsky JL *et al*: **Plasmodium falciparum encodes a single cytosolic type I Hsp40 that functionally interacts with Hsp70 and is upregulated by heat shock**. *Cell stress & chaperones* 2011, **16**(4):389-401.
397. Vanmuylder N, Werry-Huet A, Rooze M, Louryan S: **Heat shock protein HSP86 expression during mouse embryo development, especially in the germ-line**. *Anatomy and embryology* 2002, **205**(4):301-306.
398. Bopp SE, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, McCormack S, Plouffe D, McNamara CW, Walker JR *et al*: **Mitotic evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen families**. *PLoS genetics* 2013, **9**(2):e1003293.
399. Clarke LA, Rebelo CS, Goncalves J, Boavida MG, Jordan P: **PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences**. *Molecular pathology : MP* 2001, **54**(5):351-353.
400. Walsh PS, Fildes NJ, Reynolds R: **Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA**. *Nucleic acids research* 1996, **24**(14):2807-2812.
401. Klintschar M, Wiegand P: **Polymerase slippage in relation to the uniformity of tetrameric repeat stretches**. *Forensic science international* 2003, **135**(2):163-166.
402. Su XZ, Wu Y, Sifri CD, Wellems TE: **Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA**. *Nucleic acids research* 1996, **24**(8):1574-1575.
403. Lopez-Barragan MJ, Quinones M, Cui K, Lemieux J, Zhao K, Su XZ: **Effect of PCR extension temperature on high-throughput sequencing**. *Molecular and biochemical parasitology* 2011, **176**(1):64-67.
404. Chevet E, Lemaitre G, Katinka MD: **Low concentrations of tetramethylammonium chloride increase yield and specificity of PCR**. *Nucleic acids research* 1995, **23**(16):3343-3344.
405. Fazekas A, Steeves R, Newmaster S: **Improving sequencing quality from PCR products containing long mononucleotide repeats**. *BioTechniques* 2010, **48**(4):277-285.
406. Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, Milner DA, Jr., Daily JP, Sarr O, Ndiaye D, Ndir O *et al*: **A genome-wide map of diversity in Plasmodium falciparum**. *Nature genetics* 2007, **39**(1):113-119.
407. Mu J, Awadalla P, Duan J, McGee KM, Keebler J, Seydel K, McVean GA, Su XZ: **Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome**. *Nature genetics* 2007, **39**(1):126-130.

408. Walliker D, Quakyi IA, Wellems TE, McCutchan TF, Szarfman A, London WT, Corcoran LM, Burkot TR, Carter R: **Genetic analysis of the human malaria parasite Plasmodium falciparum**. *Science* 1987, **236**(4809):1661-1666.
409. Ponnudurai T, Leeuwenberg AD, Meuwissen JH: **Chloroquine sensitivity of isolates of Plasmodium falciparum adapted to in vitro culture**. *Tropical and geographical medicine* 1981, **33**(1):50-54.
410. Nkrumah LJ, Muhle RA, Moura PA, Ghosh P, Hatfull GF, Jacobs WR, Jr., Fidock DA: **Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase**. *Nature methods* 2006, **3**(8):615-621.
411. Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ: **Single-nucleosome mapping of histone modifications in S. cerevisiae**. *PLoS Biol* 2005, **3**(10):e328.
412. Iyer V, Struhl K: **Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure**. *The EMBO journal* 1995, **14**(11):2570-2579.
413. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J *et al*: **The DNA-encoded nucleosome organization of a eukaryotic genome**. *Nature* 2009, **458**(7236):362-366.
414. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A: **Determinants of nucleosome organization in primary human cells**. *Nature* 2011, **474**(7352):516-520.
415. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST: **A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling**. *Cell* 2009, **138**(1):114-128.
416. Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dolken L *et al*: **Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast**. *Molecular systems biology* 2011, **7**:458.
417. Shock JL, Fischer KF, DeRisi JL: **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* 2007, **8**(7):R134.
418. Sims JS, Militello KT, Sims PA, Patel VP, Kasper JM, Wirth DF: **Patterns of gene-specific and total transcriptional activity during the Plasmodium falciparum intraerythrocytic developmental cycle**. *Eukaryot Cell* 2009, **8**(3):327-338.
419. Adelman K, Lis JT: **Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans**. *Nature reviews Genetics* 2012, **13**(10):720-731.
420. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: **c-Myc regulates transcriptional pause release**. *Cell* 2010, **141**(3):432-445.
421. Ay F, Bunnik EM, Varoquaux N, Bol SM, Prudhomme J, Vert JP, Noble WS, Le Roch KG: **Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression**. *Genome research* 2014, **24**(6):974-988.
422. Malik S, Roeder RG: **The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation**. *Nature reviews Genetics* 2010, **11**(11):761-772.
423. Hon GC, Hawkins RD, Ren B: **Predictive chromatin signatures in the mammalian genome**. *Human molecular genetics* 2009, **18**(R2):R195-201.
424. Zaugg JB, Luscombe NM: **A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast**. *Genome research* 2012, **22**(1):84-94.
425. Hu G, Schones DE, Cui K, Ybarra R, Northrup D, Tang Q, Gattinoni L, Restifo NP, Huang S, Zhao K: **Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1**. *Genome research* 2011, **21**(10):1650-1658.