Epigenetic regulation of clonally variant gene expression in *P. falciparum*

Valerie Margarita Crowley

Aquesta tesi ha estat realitzada al Laboratori de Traducció Genètica, del programa del Biologia Cel·lular i del Desenvolupament al Institut de Recerca Biomèdica (IRB) Barcelona, sota la direcció del Dr. Alfred Cortés i el Dr. LluÍs Ribas de Pouplana, des de octubre de 2007 al agost de 2011.

TESI DOCTORAL UPF / 2011

Dr. Alfred Cortés
Assistant Research Professor
Barcelona Centre for
International Health Research
(CRESIB)

Dr. Lluís Ribas de Pouplana
ICREA Research Professor
Gene Translation Laboratory
Cell & Developmental Biology
IRB Barcelona

To Mom, Dad and Brian
It‘s in English. No really, it is…

Abstract

Clonally variant gene expression (CVGE) is a common survival strategy used by many pathogens, including the malaria parasite *Plasmodium falciparum*. Among the genes that show CVGE are several members of the *clag* and *eba* families. The active or repressed state of *clag3.1*, *clag3.2* and *eba-140* is transmitted epigenetically and is controlled by the euchromatic or heterochromatic state of bistable chromatin domains, characterised by H3K9ac and H3K9me respectively. Both of these modifications are maintained throughout the asexual cycle to transmit epigenetic memory. We have characterised CVGE on a genome-wide level and found that it is widespread. Using *clag3.2* in proof-of-principle experiments, we demonstrate that CVGE for genes other than *var* can be reproduced in an episomal system and provide preliminary evidence that CVGE is regulated by the spontaneous formation of facultative heterochromatin. The mutually exclusive expression of *clag3.1* and *clag3.2* is also addressed, including the characterisation of a neighbouring noncoding RNA.

Resum

La variació clonal en l'expressió gènica (CVGE) és una estratègia de supervivència comú utilitzada per molts patògens, incloent el paràsit de la malària *Plasmodium falciparum*. Entre els gens que mostren CVGE hi ha diversos membres de les famílies *clag* i *eba*. L'estat actiu o inactiu de *clag3.1*, *clag3.2* i *eba-140* es transmet epigenèticament i és controlat per l'estat eucromatínic o heterocromatínic de dominis de cromatina biestables, caracteritzats respectivament per H3K9ac i H3K9me. Ambdues modificacions es mantenen durant tot el cicle asexual per transmetre la informació epigenètica. També hem caracteritzat la CVGE a nivell de tot el genoma, i hem vist que afecta un gran nombre de gens. Experiments preliminars amb *clag3.2* ens demostren que la CVGE de gens diferents que *var* es pot reproduir en un sistema episomal i proporcionen evidència preliminar de què la CVGE està regulada per la formació espontània d'heterocromatina facultativa. També s'ha estudiat l'expressió mútuament exclusiva de *clag3.1* i *clag3.2*, incloent la caracterització d'un ARN no-codificant veí.

Index

Abstract v
Abbreviations xiii

1. Introduction 1
1.1 An introduction to *Plasmodium falciparum* 3
1.1.1 An introduction to *P. falciparum* malaria..... 3
1.1.2 The *P. falciparum* life cycle..... 4
1.1.3 *P. falciparum* genome features and organization..... 6

1.2 An introduction to eukaryotic transcription 7
1.2.1 General aspects of eukaryotic transcription 7
1.2.2 General aspects of *P. falciparum* transcription..... 10
1.2.3 Transcription along the intraerythrocytic cycle (IEC)..... 13
1.2.4 Specific transcription factors (STFs) are underrepresented in *P. falciparum* 14
1.2.5 Clonally variant gene expression (CVGE) 16

1.3 Chromatin structure and dynamics..... 18
1.3.1 Histones and nucleosomes, the building blocks of chromatin 18
1.3.2 Higher order chromatin structures 20
1.3.3 Post-translational histone modifications (PTMs) and gene regulation 21
1.3.4 Histone modifying enzymes 23
1.3.5 The histone code hypothesis..... 24
1.3.6 Histones and PTMs in *P. falciparum*..... 25
1.3.7 Histone modifying enzymes and chromatin effectors in *P. falciparum* 27

1.4 An introduction to epigenetics..... 30
1.4.1 Epigenetics and criteria defining *bona fide* epigenetic signals 30
1.4.2 Factors that contribute to the maintenance of epigenetic pathways in *P. falciparum*..... 32
1.4.3 Transmission of epigenetic memory by PTMs 33

1.5 Clonally variant gene expression (CVGE) of *var* genes..... 39
1.5.1 *var* genes, the paradigm of antigenic variation in *P. falciparum* 39

1.5.3 Epigenetic control of *var* genes 44
1.5.4 *P. falciparum* nuclear organization and architecture 46

1.6 Clonally variant gene expression (CVGE) beyond *var* genes 49
1.6.1 CVGE in other gene families apart from *var* 49
1.6.2 Cytoadherence-linked asexual genes, *clag3.1* and *clag3.2* 50
1.6.3 Biochemical function of *clag3.1* and *clag3.2* 52
1.6.4 Erythrocyte binding antigen-140 (*eba*-140) 54

1.7 Introduction to monoallelic and mutually exclusive gene expression 57
1.7.1 Non-random and random monoallelic gene expression 58
1.7.2 Mutually exclusive gene expression of mammalian odorant receptors and human trichromatic colour vision 61
1.7.3 Mutually exclusive expression of antigenic variation in pathogenic organisms 66

1.8 *clag3.1, clag3.2 and eba-140* as models to study the epigenetic control of CVGE in *P. falciparum* 71

2. Objectives 73

3. Materials and Methods 77

3.1 *P. falciparum culture* 79
3.1.1 Media 79
3.1.2 Basic maintenance of *P. falciparum* cultures 79
3.1.3 Parasite lines and invasion rates 79
3.1.4 Washing of erythrocytes 80
3.1.5 Preparation of Giemsa-stained thin smears 81
3.1.6 Creation of *P. falciparum* stocks 81
3.1.7 Thawing of *P. falciparum* stocks 82
3.1.8 Saponin lysis 82
3.1.9 Determining parasitemias 83
3.1.9.1 Microscopy 83
3.1.9.2 Flow cytometry 83

3.2 *P. falciparum* synchronisation 84
3.2.1 Ring stage synchronisation by sorbitol lysis 84
3.2.2 Magnetic purification of late forms 84
3.2.3 70% Percoll™ density purification of late forms 85
3.2.4 Tight synchronisation 85

3.2.5 Estimation of *P. falciparum* culture age 86

3.3 Nucleic acid isolation 87
3.3.1 Purification of genomic DNA, phenol-chloroform extraction, ethanol precipitation and quantification of nucleic acids 87
3.3.2 Total RNA extraction 88
3.3.3 cDNA synthesis 89
3.3.3.1 Dnase I treatment and RNA purification 89
3.3.3.2 Reverse transcription and cDNA purification 89

3.4 General Molecular Biology Techniques 91
3.4.1 Standard PCR 91
3.4.2 Difficult PCR 91
3.4.3 Quantitative Real-Time PCR (qPCR) 91
3.4.4 DNA sequencing 92

3.5 Luciferase assays in transient transfections 94
3.5.1 Plasmid preparation 94
3.5.2 Erythrocyte preparation 94
3.5.3 Electroporation of erythrocytes 95
3.5.4 Parasite preparation 95
3.5.5 Luciferase assay 96

3.6 Generation of 10Gluc transgenic parasite line 97
3.6.1 HLHprom3.2bsd plasmid 97
3.6.2 Transfection of 10G with HLHprom3.2bsdR 97
3.6.3 Subcloning of 10Gluc by limiting dilution 97
3.6.4 Luciferase assays on stable transfectants 98

3.7 Quantitative Chromatin Immunoprecipitation (qChIP) 99
3.7.1 Fixed cell immunoflorescence (IFA) 99
3.7.1.1 Acetone-methanol fixation 99
3.7.1.2 Formaldehyde fixation 100
3.7.1.3 Antibody preparation and microscopy 100
3.7.2 Harvesting of parasite material 101
3.7.3 Formaldehyde cross-linking, homogenization and chromatin sonication 101
3.7.4 Confirmation of fragmentation size and DNA quantification 102
3.7.5 Preclearing and immunoprecipitation 103
3.7.6 Washes, elution and reversal of cross-links 104
3.7.7 Protein digestion, DNA extraction and qPCR 104

3.8 Digestion of nuclei by restriction enzymes for chromosome conformation capture (3C) and circular chromosome conformation capture (4C) 106
3.8.1 Parasite material and formaldehyde cross-linking of nuclei 106
3.8.2 Restriction enzyme digestion of cross-linked nuclei 106
3.8.3 Preparation of samples to check efficiency of restriction enzyme digestion by Southern blot 107

3.9 Southern blot 109
3.9.1 Amplification, preparation and labelling of DNA probes 109
3.9.2 Southern blot digestion and agarose gel electrophoresis 110
3.9.3 DNA transfer 110
3.9.4 Washing of the membrane 111
3.9.5 Phosphor imaging and X-ray film exposure 112

3.10 Optimised conditions for the restriction enzyme digestion of cross-linked nuclei for 3C and 4C 113

4. Article 1 115

4.1 Introduction to Article 1 117

4.2 Article 1 118

4.3 Additional results 147
4.3.1 Verification that commercial antibodies for qChIP cross-react with *P. falciparum* histones 147
4.3.2 qChIP data using antibodies against core histone H3 were highly variable 148

5. Article 2 151

5.1 Introduction to Article 2 153

5.2 Article 2 157

5.3 Additional results 201
5.3.1 Generation of a parasite line that stably maintains an episome with the *clag3.2* promoter 201
5.3.2 10Gluc is a population of individual parasites that differentially express luciferase 201

5.3.3 Tightly synchronised 10Gluc-derived subclones differentially express luciferase as determined by gene reporter assays and confirmed by qPCR 204
5.3.4 Differences in luciferase expression observed between 10Gluc-derived subclones does not increase 206
5.3.5 Differences in luciferase expression between 10Gluc-derived subclones are not due to an episomal recombination..... 209
5.3.6 The episome that is expressed is enriched in H3K9ac, whereas the episome that is repressed is enriched in H3K9me3 210
5.3.7 In transient transfection assays, HLHprom3.2bsdR appears to be repressed 212

6. Mutually exclusive gene expression in *P. falciparum* 215

6.1 Hypotheses regarding *clag3* mutual exclusion 217

6.2 Results 225
6.2.1 Mutually exclusive expression of *clag3.1* and *clag3.2* 225
6.2.1.1 10Gluc and 10Gluc-derived subclones have switched *clag3* expression 225
6.2.1.2 The ncRNA PF03TR002 is correlated with the expression of *clag3.1* and the repression of *clag3.2* 226
6.2.1.3 The differential expression of the *clag3* genes in 10G and 1.2B parasite lines is not the result of nucleotide differences in their 5' intergenic regions 230
6.2.1.4 *clag3.1* and *clag3.2* are mutually exclusive in parasite lines other than 3D7 233
6.2.2 Chromosome conformation capture (3C) and circular chromosome conformation capture (4C) 235
6.2.2.1 3C strategy for *clag3.2* and 4C strategy for *clag3.2* and the *var* gene PF07_0048 235
6.2.2.2 Design of Southern blot experiments for the analysis of restriction enzyme digestion efficiency in cross-linked nuclei 238
6.2.2.3 With optimization, cross-linked nuclei can be efficiently digested by *Bcl*I, *Bgl*II and *Dpn*II 240

7. General Discussion 245

8. Conclusions 269

9. Bibliography 273
Acknowledgments 301

Abbreviations

2TM	two transmembrane
3C	chromosome conformation capture
4C	circular chromosome conformation capture
5mC	5-methyl-2'-deoxycytosine
6mA	6-methyl-2'-deoxyadenine
A	adenine
ATP	adenosine triphosphate
bp	basepair
BRE	TFIIB-recognition element
BSA	bovine serum albumin
CD	chromodomain
CGH	comparative genome hybridisation
clag	cytoadherence-linked asexual gene
CSD	chromo shadow domain
CTCF	CCCTC-binding factor
CVGE	clonally variant gene expression
DAPI	4'-6-Diamidino-2-phenylindole dihydrochloride
DHS	DNA hypersensitive site
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
DNMT	DNA methyltransferase
dNTP	deoxyribonucleotide triphosphate
DPE	downstream promoter element
DTT	dithiothreitol
eba	erythrocyte binding antigen
EBL	erythrocyte-binding-like
EDTA	ethylenediamine tetra-acetic acid
EGTA	ethylene glycol tetra-acetic acid
EMP1	erythrocyte membrane protein
ES	expression site
EtBr	ethidium bromide
GO	gene ontology
GTF	general transcription factor
HAT	histone acetyl transferase

HDAC	histone deacetylase
HDM	histone demethylase
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid
HKMT	histone lysine methyl transferase
HMGB	high mobility group box
HMT	histone methyl transferase
HP1	heterochromatin protein 1
hpi	hours post invasion
ICE	imprint control element
IEC	intraerythrocytic cycle
IFA	immunofluorescence assay
Inr	initiator
JHDM	Jumonji-C domain-containing histone demethylase
kb	kilobase
KMT	lysine methyl transferase
LCR	locus control region
LiCl	lithium chloride
lncRNA	long noncoding RNA
LSD	lysine specific demethylases
MB	mega base pairs
MTE	motif ten element
ncRNA	noncoding RNA
nm	nanometer
NPP	new permeability pathway
ORF	open reading frame
PBS	phosphate buffered saline
PcG	polycomb group
PCR	polymerase chain reacton
PERC	perinuclear repressive centres
PEV	position effect variegation
PF	positive feedback
Pf	*Plasmodium falciparum*
PIC	preinitiation complex
PRC	polycomb repressive complex
PSAC	plasmodial surface anion channel

PTM	covalent post-translational modification
PV	parasitophorous vacuole
qChIP	quantitative chromatin immunoprecipitation
qPCR	quantitative PCR
RBC	red blood cell
RBL	reticulocyte-binding-like
RE	restriction enzyme
RLU	relative light unit
RNA	ribonucleic acid
RNAi	RNA interference
RNaseA	ribonuclease A
rpm	revolutions per minute
SCGDC	sex chromosome gene dosage compensation
SDS	sodium dodecyl suphate
siRNA	short-interfering RNA
STF	specific transcription factor
T	thymine
TAP	transcription associated protein
TARE	telomere-associated repeat element
TBE	Tris-borate-EDTA buffer
TBP	TATA-binding protein
TE	tris-EDTA buffer
TF	transcription factor
TRIS	Tris-(hydroxymethyl) aminomethane
tRNA	transfer RNA
trxG	trithorax group
TSS	transcription start site
U	unit
uORF	upstream open reading frame
UTR	untranslated region
VSG	variable surface glycoprotein
VSP	variable surface protein
XCI	X chromosome inactivation
Xic	X inactivation center

Histone modifications

H3K27me2	dimethylation of histone H3 lysine 27
H3K27me3	trimethylation of histone H3 lysine 27
H3K4me2	dimethylation of histone H3 lysine 4
H3K4me3	trimethylation of histone H3 lysine 4
H3K9ac	acetylation of histone H3 lysine 9
H3K9me2	dimethylation of histone H3 lysine 4
H3K9me3	trimethylation of histone H3 lysine 9
H4ac	global histone H4 acetylation lysines 5,8,12,16
H4K20me3	trimethylation of histone H4 lysine 20

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

1. Introduction

1.1 An introduction to *Plasmodium falciparum*

1.1.1 An introduction to *P. falciparum* malaria

Apicomplexan protozoans form a phylum of obligately intracellular parasites that include important human pathogens, including *Plasmodium* sp., the causative agent of malaria. Malaria is a mosquito-borne infectious disease. Approximately half of the world's population is at risk of contracting malaria; however most malaria cases and deaths occur in sub-Saharan Africa (Figure 1.1).

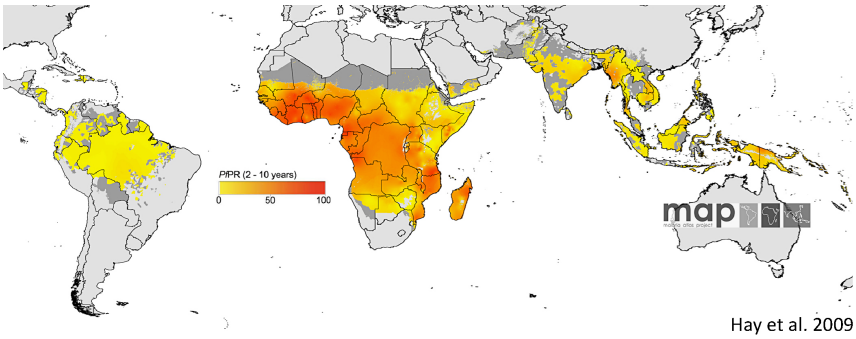


Figure 1.1 The spatial distribution of *P. falciparum* endemicity. The data mapped are the annual mean *P. falciparum* parasite rate (PfPR) which refers to the proportion those sampled showing detectable parasites in blood samples for the endemicity class of children between 2 and 10 years of age. Figure reproduced from Hay et al 2009.

In 2009, there were 225 million cases of clinical malaria and approximately one million deaths, mainly in African children (WHO 2010). Most of these deaths are caused by *Plasmodium falciparum*, one of the four distinct *Plasmodium* species that parasitize humans. Clinical malaria caused by *P. falciparum* results in a wide range of flulike symptoms including headache, nausea and cycles of fever. Complications of *P. falciparum* malaria may lead to severe malaria that could result in

symptoms such as coma, respiratory distress, severe anaemia and death. Children, primigravid women and their unborn children and naive adults are particularly susceptible to *P. falciparum* infection. Several important features of *P. falciparum* biology that make it a successful parasite are its high asexual multiplication rate, its ability to invade erythrocytes of all ages due to redundancy in erythrocyte invasion pathways, and its efficient evasion of host immunity by sequestration in peripheral circulation and antigenic variation (Mackinnon & Marsh 2010). Despite the substantial impact of this pathogen on human health, much of its basic biology is poorly understood (Callaway 2007).

1.1.2 The *P. falciparum* life cycle

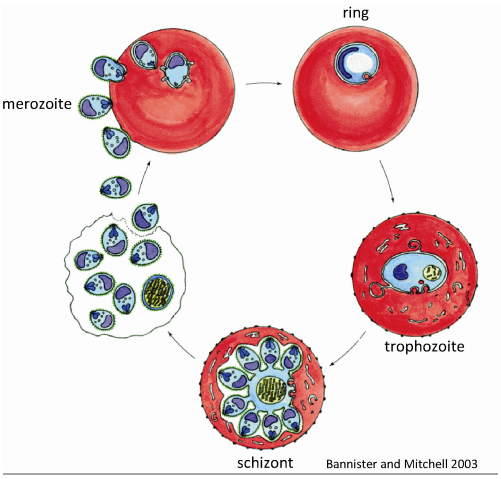


Figure 1.2 The intraerythrocytic cycle (IEC) of *P. falciparum*. Merozoites invade erythrocytes, which progress through morphologically distinct ring, trophozoite and schizont stages. Schizonts burst, releasing merozoites that invade erythrocytes and the cycle continues. See text for further details. Figure reproduced from Bannister & Mitchell 2003.

P. falciparum has a complex life cycle involving several differentiated forms and two different hosts. During a blood meal, a parasitized female *Anopheles* mosquito releases sporozoites that migrate through the skin, into the circulation and then to the liver where they invade hepatocytes. Each invading sporozoite differentiates and mitotically divides into thousands of merozoites that are released into the circulation where they invade erythrocytes. This begins the asexual

intraerythrocytic cycle (IEC) that lasts approximately 48 h (Figure 1.2). All clinical symptoms of malaria are associated with the IEC, which begin approximately 7-15 days after initial infection, and is the target for many antimalarial drug and vaccine strategies.

Merozoite invasion is a multistep process that involves the molecular interaction of numerous host receptors and parasite ligands. Parasite molecules implicated in invasion are usually discharged from its apical secretory organelles (micronemes, rhoptries and dense granules). Merozoite invasion begins with a low affinity attachment of the parasite to the erythrocyte which is followed by its reorientation to appose its apical end with the erythrocyte membrane. A tight junction is formed between the parasite and the erythrocyte membrane. This junction moves from the apical to the posterior end of the merozoite as the parasite propels itself into the erythrocyte by the action of its actin-myosin motor. The merozoite invades in such a way that the internalized parasite resides within a parasitophorous vacuole (PV), a membrane composed of lipids that are derived both from the erythrocyte and the merozoite, in which the parasite proliferates (Zuccala & Baum 2011). Once inside the erythrocyte, the merozoite develops through morphologically distinct ring, trophozoite and schizont stages. IEC progression is accompanied by extensive remodelling of the host erythrocyte where some parasite proteins are secreted and exposed on the surface of the infected erythrocyte. In the schizont, many mitotic divisions followed by nuclear envelope division and segmentation of the cytoplasm occurs, resulting in approximately 16 developing merozoites. Schizont rupture releases merozoites into circulation that are able to invade other erythrocytes and the IEC continues. Parasites in the IEC are

haploid, the only diploid forms of the parasite occur in the mosquito. A small fraction of merozoites develop into gametocytes which are taken up by mosquitoes upon a blood meal. The male gametocyte divides into 8 flagellated microgametes that escape from the erythrocyte, and once it has fertilized a female macrogamete, forms a motile ookinete that moves from the stomach to form an oocyst. Oocysts undergo clonal expansion to form thousands of developing sporozoites. Once the oocyst bursts, the sporozoites migrate to the salivary glands and are transmitted to a new host upon mosquito feeding (Miller et al 2002).

1.1.3 *P. falciparum* genome features and organization

Sequencing of the *P. falciparum* genome was completed in 2002 (Gardner et al 2002). It is 23.3 Mb in size with a karyotype of 14 chromosomes. In addition to the nuclear genome the parasite has a 6 kb linear mitochondrial and a 35 kb circular apicoplastic genomes. *P. falciparum* has the most (A+T) rich genome sequenced to date, with an overall average (A+T) content of 80.6%, which can surpass 90% at intergenic and intronic regions. Genes associated with antigenic variation are concentrated in the subtelomeres and the genome lacks DNA transponsons and retrotransposons (Wickstead et al 2003). At the time of genome sequencing, approximately 60% of the predicted protein coding genes lacked sufficient sequence homology with proteins from other organisms and thus precluded their functional assignment (Gardner et al 2002). An abundance of antisense transcripts and non protein coding RNAs (ncRNAs) also transcribed during the IEC (Gunasekera et al 2004; Otto et al 2010; Broadbent et al 2011).

1.2 An introduction to eukaryotic transcription

1.2.1 General aspects of eukaryotic transcription

To survive, all organisms require the proper temporal expression of the appropriate genes. The expression of protein coding genes can be regulated at many different levels. However, most regulation is believed to occur at the level of transcription. Transcription is a complex process that requires the coordinated recruitment of multiple DNA-binding proteins, activators, coactivators and chromatin modifying enzymes (collectively referred to as transcription associated proteins, TAPs) to the specific DNA sequence that is to be transcribed (Figure 1.3). Typical eukaryotic genes contain two *cis*-acting DNA regulatory elements, the core promoter with its proximal regulatory elements (collectively referred to as the promoter) and additional, distally located regulatory elements (Figure 1.4)

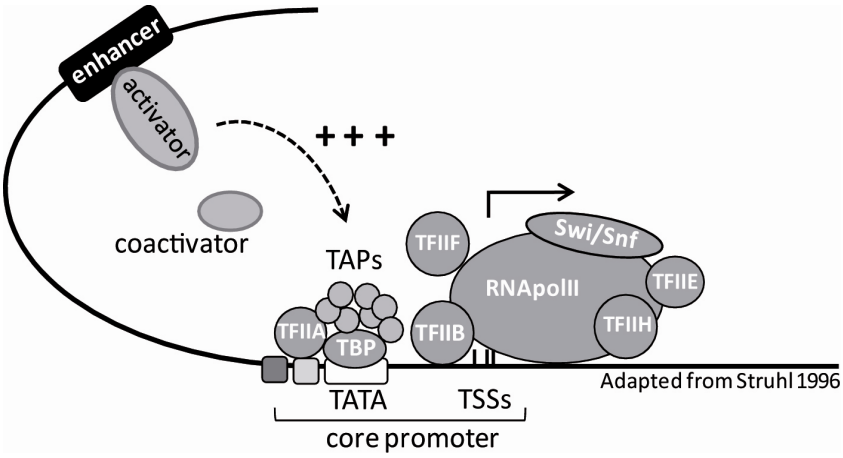


Figure 1.3 Typical eukaryotic transcriptional machinery. Activators bind to enhancer elements that stimulate basal transcription via protein-protein interactions with various components of the general transcription machinery. See text for more information. Figure adapted from Struhl 1996 .

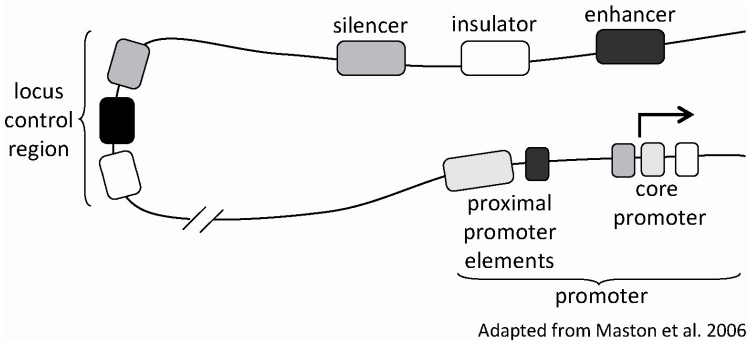


Figure 1.4 Schematic of a typical eukaryotic gene regulatory region. Regulatory regions consist of the promoter and distally located elements that may come in contact with the promoter via looping out of the intervening DNA. See text for further details. Figure adapted from Maston et al. 2006.

These *cis*-acting regions contain sequences for *trans*-acting transcription factor (TF) binding which can activate or repress transcription. Eukaryotic core promoters are composed of a number of *cis*-acting sequence elements that may include well-defined elements such as the TATA-box, initiator (Inr) element, downstream promoter element (DPE), TFIIB-recognition element (BRE) and the motif ten element (MTE) (Hampsey 1998; Maston et al 2006). Transcription begins with the sequential assembly of several general transcription factors (GTFs, which themselves are multiprotein complexes) over the core promoter to form the preinitiation complex (PIC). The PIC is required to direct RNA polymerase II to the transcription start site (TSS). These components are only sufficient for basal transcription, and the activity of this complex is modulated by the interaction of another class of DNA-binding proteins referred to as activators (such as specific transcription factors, STFs) (Kornberg & Lorch 1999). Since STFs recognize specific DNA sequences, they are important molecules in conferring specificity to the general transcriptional machinery.

Distal DNA regulatory regions can be broadly classified into several categories that include enhancers, silencers and insulators. Enhancers are regulatory elements that increase transcription. These elements typically influence gene expression over long distances and are usually distance and orientation independent relative to the promoter. Silencers share many of the properties described for enhancers, however, they negatively regulate transcription. Insulators are genetic elements that prevent genes from being influenced by the transcriptional activity of their neighbours and thus partition the genome into transcriptional domains. Insulators are divided into two categories based on their mode of action. Enhancer-blocking insulators are *cis*-acting elements that prevent the interaction of an enhancer with a promoter when physically placed between the two. Barrier insulators are *cis*-acting elements that prevent against the spread of heterochromatin-mediated silencing (Gaszner & Felsenfeld 2006). The *cis*-acting elements involved in boundary formation are largely unknown, but CTCF (CCCTC-binding factor), an 11-zinc-finger protein, has been implicated as an enhancer-blocking insulator in vertebrates and tRNA loci have been implicated in preventing the spread of pericentromeric heterochromatin in *S. pombe* (Bell et al 1999; Cam et al 2005). However, some insulators have been shown to posses both enhancer blocking and barrier activity (Donze & Kamakaka 2001; Ishii et al 2002).

In eukaryotes, DNA is wrapped around histone proteins and organised into chromatin. Not only does chromatin serve to efficiently package DNA into the nucleus, but chromatin structure also plays an important role in the regulation of gene expression. Chromatin and its role in regulating gene expression will be discussed further in section 1.3. It is all

of these aforementioned factors that collectively contribute to the appropriate regulation of gene expression.

1.2.2 General aspects of *P. falciparum* transcription

Transcription in *P. falciparum* shares many of the characteristics that are common to eukaryotic transcription. Functional analysis has demonstrated that parasite promoters are of a typical bipartite structure and often contain multiple TSSs that usually encode monocistronic transcripts (Watanabe et al 2002; Lanzer et al 1992). The core promoter elements of a TATA box-like and Inr-like sequences have been functionally characterized in the *kahrp* and a in *var* gene respectively (Ruvalcaba-Salazar et al 2005; Calderwood et al 2003). DNA elements with enhancer and repressor activities have been identified (Horrocks et al 1998; Lopez-Estrano et al 2007; Olivieri et al 2008; Osta et al 2002). However these elements do not share similarities to known regulatory regions in other eukaryotes and no common regulatory features have emerged in the regulation of gene expression (Militello et al 2004). No insulator elements have been identified in *P. falciparum* and a CTCF homolog has not yet been described.

Two transcriptional components to be functionally characterized in *P. falciparum* have been the large subunit of RNA pol II (Li et al 1989) and the TATA binding protein (TBP), part of the TFIID complex (McAndrew et al 1993). Regarding *trans*-acting factors, extensive bioinformatic analyses from several laboratories has lead to the identification of many TAPs (Aravind et al 2003; Coulson et al 2004; Callebaut et al 2005; Bischoff & Vaquero 2010). All 12 subunits of the RNA pol II complex are accounted for, as well as 44 additional GTFs (Coulson et al 2004;

Callebaut et al 2005; Bischoff & Vaquero 2010). Although functional analysis for the vast majority of these GTFs is still lacking, their total number and relative conservation with other eukaryotes suggest that they control transcriptional initiation and elongation as in other eukaryotic organisms (Bischoff & Vaquero 2010).

There is conflicting data regarding the transcriptional responsiveness of the *P. falciparum* genome to mount relevant responses to environmental perturbations or drug exposure that has lead to polarised opinions in the field. Numerous groups have detected large changes in gene expression in response to environmental challenges such as febrile temperature and to various chemical perturbations with drugs and growth inhibiting compounds, including but not limited to, artenusate, artemisinin, quinine, chloroquine and apidicine (Oakley et al 2007; Natalang et al 2008; Hu et al 2010; Tamez et al 2008). However, other groups have controversially proposed that *P. falciparum* does not mount relevant transcriptional responses (Ganesan et al 2008). This conclusion was been based on observations that when treated with a lethal antifolate drug, the gene expression of the main target of the drug did not increase, nor were there any consistently reliable increases in RNA levels for any genes across the entire genome (Ganesan et al 2008). A similar conclusion was reached by another study that was undertaken to elucidate the target of a potent antimalarial compound, T4, a choline analog (Le Roch et al 2008). There were no transcriptional changes observed in the genes believed to be involved in phosphatidylcholine (PC) biosynthesis, and only after proteomic and metabolomic studies was the down regulation of the enzyme involved in the final step of PC synthesis identified (Le Roch et al 2008). Consequently, it was concluded by these authors that

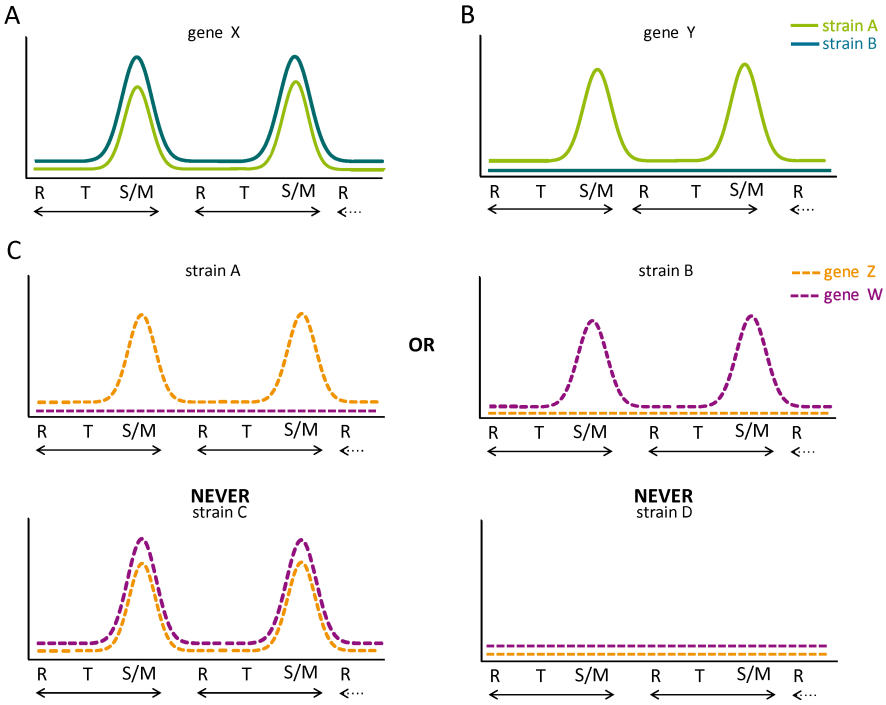


Figure 1.5 Three types of transcriptional silencing observed along the IEC in *P. falciparum*. A. Silencing associated with life cycle progression. B. Clonally variant gene expression (CVGE). C. Mutually exclusive gene expression. A. The transcript abundance for many asexual genes exhibit one expression peak during the IEC. Gene X is representative of a gene expressed in schizonts. Two full asexual cycles are shown. Solid lines represent parasite strains. **B.** Clonally variant gene expression (CVGE) refers to the phenomenon where some genes can be found in different transcriptional states in genetically identical parasites at the same stage of the IEC. In this example, gene Y is expressed in strain A and repressed in strain B. **C.** Mutually exclusive gene expression. In this case, genes Z and W (discontinuous lines) are members of the same gene family, but are never expressed at the same time in a given parasite strain. In this example, gene Z is only expressed in strain A, whereas gene W is only expressed in strain B. Mutual exclusion does not permit for the examples shown in hypothetical strains C and D where both genes are expressed or repressed in the same parasite strain. In all cases, strains A and B are genetically identical.

the transcriptome of *P. falciparum* appears to respond in a general as opposed to a specific or targeted manner when faced with a drug challenge (Ganesan et al 2008; Le Roch et al 2008). Whether *P. falciparum* is capable of directed transcriptional responses is still a matter of debate.

1.2.3 Transcription along the intraerythrocytic cycle (IEC)

The IEC encompasses four morphologically distinct stages: ring, trophozoite, schizont and merozoite and this progression is largely controlled at the level of transcription. There is a striking periodicity in mRNA expression profiles that parallels progression along the IEC. The majority of genes follow what has been described as a "just-in-time" pattern of expression, such that the time of gene expression correlates with the stage when the protein is required, and repressed for the rest of the cycle (Figure 1.5A, Bozdech et al 2003; Le Roch et al 2003). It was additionally observed that over 80% of the genome is expressed during the IEC and that there are no large co-regulated chromosomal regions, suggesting that genes are each individually regulated. However, mRNA steady state levels do not correlate well with translational activity, as recent proteomic data has shown that protein abundances also exhibit one expression peak during the IEC, but that this peak occurs considerably later (a mean of 11 h) than is observed for the corresponding transcript, implicating post-transcriptional processes in progression along the IEC (Foth et al 2011).

1.2.4 Specific transcription factors (STFs) are underrepresented in *P. falciparum*

At the time the *P. falciparum* genome was sequenced, gene ontology (GO) terms were assigned to the 40% of proteins with sufficient sequence homology and it was observed that several GO categories were underrepresented in the genome when compared to *S. cerevisiae*, a free-living eukaryote with a similar number of genes. Importantly, it was noted that the number of proteins designated as TAPs, GTFs and STFs were underrepresented (Gardner et al 2002).

Soon after, initial comparative genomics analyses by other groups identified a large number of GTFs and TAPs, but also noted the lack of STFs, and thus it was hypothesized that transcription was predominantly regulated by epigenetic and post-translational mechanisms (Callebaut et al 2005; Coulson et al 2004).

An expanded family of 27 STFs was later identified by bioinformatic analyses and this family is emerging as potential “master regulators” of *P. falciparum* transcription (Balaji et al 2005; De Silva et al 2008; Campbell et al 2010; Flueck et al 2010). These TFs represent a novel family of proteins that posses a modified version of the AP2 (Apetala2)-integrase DNA binding domain which is present in many plant TFs, and have thus been termed ApiAP2s (Balaji et al 2005). This globular AP2 domain is approximately 60 amino acids and can be found in various architectures from single to triple arrangements within a single protein. The ApiAP2s show a lineage specific distribution not only within other *Plasmodium* spp. but are present in all other apicomplexans. The Llinás laboratory demonstrated that protein-DNA specificity is conserved between orthologs of phylogenetically distant apicomplexan species, when a

Crytosporidium paravum protein bound the same DNA sequence as the *P. falciparum* homolog, although the individual ApiAP2s are predicted to regulate different sets of genes in the different species (De Silva et al 2008).

Using protein binding microarrays, the Llinás laboratory has also determined that in addition to high affinity primary sequence binding motifs, 14 ApiAP2s are able to bind secondary sequences, increasing the complexity of ApiAP2 regulatory networks (Campbell et al 2010). The ApiAP2s also show a cyclical pattern of expression along the IEC, and it has been suggested that the ApiAP2s could regulate each other along the IEC and collectively contribute to the cascade of gene expression observed along the IEC (Bozdech et al 2003; Balaji et al 2005). Other evidence implicating the ApiAP2s as master regulators of *P. falciparum* transcription are based on yeast two hybrid studies that found physical interactions of various ApiAP2s with putative chromatin modifying enzymes and with other ApiAP2s (LaCount et al 2005). An ApiAP2, PfSIP2, has been found to localize exclusively to subtelomeric heterochromatin and to be associated with various proteins associated with DNA metabolism (Flueck et al 2010 , discussed further in section 1.5.2).

Two ApiAP2 homologs in *P. berghei* have been shown to be involved in controlling gene expression in stages apart from the IEC; AP2-O was reported to regulate the genes involved in midgut invasion of the ookinete, and the disruption of Ap2-Sp lead to a total loss of sporozoite formation (Yuda et al 2009; Yuda et al 2010). Additionally, putative

binding sites for ApiAP2s were identified for a gene expressed exclusively during liver stages in *P. berghei* (Helm et al 2010).

Additional bioinformatic analysis has identified another large group of zinc finger-containing proteins (37 in total) and one cold-shock STF, none of which have been functionally characterized. Eight helix-turn-helix containing proteins have been identified and one (PfMyb1), has been subject to some biochemical characterization, however, the role of these other STFs in transcriptional regulation is currently unclear (Bischoff & Vaquero 2010; Gissot et al 2005).

Although the total number of STFs identified in this parasite has increased (73 to date), this group of proteins remains underrepresented in the genome as the number is still low when compared to other eukaryotic organisms with a comparable number of protein coding genes, such as *S. pombe* and *S. cerevisiae* (Bischoff & Vaquero 2010).

1.2.5 Clonally variant gene expression (CVGE)

Clonally variant gene expression (CVGE) refers to the phenomenon where some genes can be found in different transcriptional states in genetically identical parasites at the same stage of life cycle progression while grown under homogenous conditions (Figure 1.5B). In *P. falciparum*, populations of genetically identical parasites are transcriptionally heterogeneous (Cortés et al 2007) consequently, CVGE is often studied in subclones. Subclones refer to parasite populations that originate from a single clonal parasite. When subclones are cultured for the minimum number of cycles necessary to amass sufficient material for transcriptional analysis, for genes that show stable, clonally

transmitted expression patterns, it is akin to studying a single parasite from within the clonal population from which the subclone was derived (Cortés et al 2007).

Antigenic variation, a form of CVGE, refers to the expression of alternate forms of a particular surface antigen in order to evade host immunity. CVGE is used by many pathogens in order to proliferate and spread to new hosts by evading mechanical clearance, recognition and destruction (Deitsch 2009). In *P. falciparum*, the *var* genes are the paradigm of antigenic variation and are discussed in section 1.5.

1.3 Chromatin structure and dynamics

1.3.1 Histones and nucleosomes, the building blocks of chromatin

In eukaryotic cells, the packaging of DNA into chromatin is a major mechanism by which the access to genomic DNA is restricted and regulated (Kornberg & Lorch 1999). The basic unit of chromatin is the nucleosome and is comprised of 147 bp of DNA wrapped around an octamer of the four core histone proteins: H2A, H2B, H3 and H4. The unstructured amino-terminal domains of each histone protrude from the globular nucleosome and are referred to as histone tails (Kornberg 1977). Both histone tails and globular domains are subject to numerous covalent post-translational modifications (PTMs) by reactions that are enzymatically reversible (Kouzarides 2007). The amino acid sequence of the core histones is remarkably similar, even between phylogenetically distinct species (Figure 1.6). Canonical histones can be replaced by histone variants, which are similar to canonical histones but differ by a small number of amino acid changes. Histone variants can alter the functional properties of chromatin by changing stability, DNA wrapping, regulating access to DNA and by creating structurally and functionally distinct chromatin domains (Talbert & Henikoff 2010).

Chromatin can be classified into three different conformational varieties: euchromatin, constitutive heterochromatin and facultative heterochromatin. Generally, euchromatin is uncondensed, accessible and the transcriptionally active form of chromatin, whereas heterochromatin is condensed, inaccessible and transcriptionally inactive. Facultative chromatin is able to convert between euchromatic and heterochromatic states (Trojer & Reinberg 2007). However, euchromatin is not devoid of

```
A
Pf.H3    MARTKQTARKSTGGKAPRKQLASKAARKSAPVSTGIKKPHRYRPGTVALREIRKFQKSTD 60
Hs.H3    MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
Sc.H3    MARTKQTARKSTGGKAPRKQLASKAARKSAPSTGGVKKPHRYKPGTVALREIRRFQKSTE 60
         **********************:******** : *:******:**********::****:

Pf.H3    LLIRKLPFQRLVREIAQEYKTDLRFQSQAVLALQEAAEAYLVGLFEDTNLCAIHAKRVTI 120
Hs.H3    LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEACEAYLVGLFEDTNLCAIHAKRVTI 120
Sc.H3    LLIRKLPFQRLVREIAQDFKTDLRFQSSAIGALQESVEAYLVSLFEDTNLAAIHAKRVTI 120
         *****************::********.*: ****: *****.*******.*********

Pf.H3    MPKDIQLARRIRGERS 136
Hs.H3    MPKDIQLARRIRGEA- 135
Sc.H3    QKKDIKLARRLRGE-- 134
         ***:****:***

B
Pf.H4    MSGRGKGGKGLGKGGAKRH-RKILRDNIQGITKPAIRRLARRGGVKRISGLIYEEIRGVL 59
Hs.H4    MSGRGKGGKGLGKGGAKRH-RKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVL 59
Tt.H4    MAG-GKGGKGMGKVGAKRHSRKSNKASIEGITKPAIRRLARRGGVKRISSFIYDDSRQVL 59
         *:* ******:** ***** **  : .*:********************.:**:: * **

Pf.H4    KVFLENVIKDSIMYTEHAKRKTVTAMDIVYSLKRQGRTLYGFGG 103
Hs.H4    KVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG 103
Tt.H4    KSFLENVVRDAVTYTEHARRKTVTAMDVVYALKRQGRTLYGFGG 103
         * *****::*:: *****:********:**:*************
```

Figure 1.6 Multiple sequence alignment of *P. falciparum* core histones H3 (A) and H4 (B) with other phylogenetically distant species. Histone proteins are well conserved through evolutionary history. Acetyl- or methyl- modified residues highlighted in bold were the target for antibodies used in ChIP experiments (discussed in Chapter 4). * indicates a conserved residue, : indicates a conservative substitution and . indicates a semi-conservative substitution. Hs, Homo sapiens, Sc, *Saccharomyces cerevisiae*, Tt *Tetrahymena thermophila*, Pf *Plasmodium falciparum*. Sequences were aligned using ClustalW (Larkin et al 2007). Accession numbers: Pf.H3 (PFF0510w), Hs.H3 (AAA52651.1), Sc.H3 (AAS64341.1), Pf.H4 (PF11_0061), Hs.H4 (NP_003529), Tt.H4 (XP_001016593).

nucleosomes, and in order for transcription to proceed, nucleosomes need to be repositioned and replaced. Consequently, chromatin conformation and nucleosome positioning need to be highly dynamic processes depending on the transcriptional needs of a cell at a particular time. In a process called ATP-dependent nucleosome remodelling, protein complexes are able to alter DNA-nucleosome interactions to transiently allow the access of the transcriptional apparatus to DNA (Becker & Horz 2002). Some of the complexes involved in these processes are the SWI/SNF, IWSI, Mi-2 and NURD complexes (Becker & Horz 2002). Through the energy released from the hydrolysis of ATP,

these complexes are able to slide, displace and space histone octamers as well as exchange canonical histones for histone variants. Apart from ATP-dependent nucleosome remodelling, PTMs, particularly on histones tails, are able to dynamically recruit a vast array of enzymes and structural proteins that alter chromatin structure in order to either permit or prevent transcription and are responsible for regulating higher order chromatin structure (Kouzarides 2007).

1.3.2 Higher order chromatin structures

Eukaryotic chromatin is successively organized into different layers and each layer is implicated in various aspects of gene regulation. Single nucleosomal arrays, often referred to as “beads-on-a-string”, are 11 nm in diameter and represents the first level of chromatin organization. The incorporation of linker histone H1 organizes the 11 nm fibre into what is referred to as the 30 mn fibre, which constitutes the second level of chromatin organization. Two different topologies have been proposed for the organization of the 30 nm fibre, referred to as the zig-zag and solenoid models (Li & Reinberg 2011). All chromatin conformations that involve self association of 30 nm fibres with itself or that involve inter-fibre contacts are referred to as higher order chromatin structures. Transcription is frequently regulated through the communication of distal *cis*-regulatory elements that on the linear chromosome may be separated by large distances, and how these interactions occur is not entirely clear. Long-range chromatin interactions, such as the formation of chromatin loops have been proposed as a way to bring distal regulatory elements in contact with their promoters (Li & Reinberg 2011). The interaction of the transcriptional repressor CTCF with cohesin complexes has been proposed as a way to mediate long distance

interactions between enhancers and promoters within and between chromatin fibres (Gondor & Ohlsson 2008). It is also becoming clear that spatial organization of chromosomes within the nucleus and specific chromosomal positions within defined chromosomal territories play a role in genetic regulation (Li & Reinberg 2011; Zhu & Reinberg 2011).

1.3.3 Post-translational histone modifications (PTMs) and gene regulation

Individual histone residues are selectively modified and the same residue may be the target of antagonistic PTMs. Many different PTMs exist on histones and they can be classified into eight different categories (Table 1). These various PTMs are associated with different functions such as DNA repair, replication, chromosome condensation and gene expression (Kouzarides 2007). Two classes of PTMs with a clear role in the regulation

Table 1 Categories of PTMs and their associated functions

Chromatin Modifications	Residues Modified	Functions Regulated
acetylation	K-ac	transcription, DNA repair, DNA replication, chromosome condensation
methylation (lysine)	K-me1 K-me2 K-me3	transcription, DNA repair
methylation (arginine)	R-me1 R-me2a R-me2s	transcription
phosphorylation	S-ph T-ph	transcription, DNA repair, chromosome condensation
ubiquitination	K-ub	transcription, repair
sumoylation	K-su	transcription
ADP ribosylation	E-ar	transcription
deimination	R --> Cit	transcription
proline isomerization	P-cis --> P-trans	transcription

Kouzarides 2007

of gene expression are acetylation and methylation of lysine residues. There are two described mechanisms for the way that PTMs function: by disrupting chromatin contacts between adjacent nucleosomes and/or between nucleosomes and DNA; by affecting the recruitment of nonhistone proteins to chromatin, both of which are not mutually exclusive (Bannister & Kouzarides 2011).

Acetylation is generally associated with gene activation. Acetylation is believed to reduce chromatin compaction by neutralizing the basic charges of the ε-amino groups on lysines (Wolffe & Pruss 1996), affect higher order chromatin structures by preventing self-association and the formation of tertiary chromatin structures (Fletcher & Hansen 1996), recruitment of nonhistone proteins that contain bromodomains that function as transcriptional coactivators (Jacobson et al 2000) and by directly competing against the deposition of repressive marks on the same histone residue.

Unlike acetylation, methylation does not change the charge of the histone and it is unlikely that the chemical moiety itself alters chromatin compaction (Bannister & Kouzarides 2011). Methylation can be associated with either activation or silencing depending on which histone residue is methylated, whether the methylated position corresponds to the promoter or coding region of a gene, and is further complicated by the fact that lysines may be mono-, di- or tri-methylated (Kouzarides 2002). Although exceptions exist and methylation marks are difficult to classify, methylation of H3K4, H3K36 and H3K79 are generally associated with gene activation, whereas the methylation of H3K9, H3K27 and H4K20 are generally related to transcriptional silencing. At the molecular level, euchromatin is associated with hyperacetylation and H3K4me3, constitutive heterochromatin is associated with H3K9me3, H4K20me3 and hypoacetylation, whereas facultative heterochromatin is generally hypoacetylated and has H3K27me2 or H3K27me3 and/or H3K9me2 (Trojer & Reinberg 2007). The recruitment of nonhistone proteins to chromatin, often called chromatin effectors, occurs via the specific interaction between a PTM and particular protein domains.

The interaction of enzymes with their target PTMs can be highly specific and context dependent. Chromo-, Tudor, PhD, PWWP and MBT domains recognize methylated residues, while acetylation is recognized by bromodomains. The proteins containing these domains may associate with other proteins and lead to the formation of large multiprotein complexes that act in concert to alter chromatin structure in a variety of ways. One such example is the recruitment of heterochromatin protein 1 (HP1), a repressive binding effector, to H3K9me3. HP1 contains two binding domains, the chromodomain (CD) and the chromo shadow domain (CSD). The CD domain binds to H3K9me3, while the CDS domain is involved in HP1 dimerization and protein-protein interactions. Two well characterized interactors are an H3K9 methyltransferase, SUV39H1 (KMT1) and an H3K9 deacetylase, HDAC1 (Yamamoto & Sonoda 2003; Schotta et al 2002; Czermin et al 2001). Since HP1 interacts with both KMT1 and HDAC1, a model of heterochromatin spreading involving these components was proposed, where HDAC1 removes acetylation on H3K9, permitting for the methylation of the same residue by KMT1 which allows for HP1 binding, that in turn is able to recruit more KMT1 and HDAC1 in a self-propagating process (Grewal & Elgin 2002).

1.3.4 Histone modifying enzymes

The steady state levels of PTMs is the result of a delicate balance between their addition and removal. Histone acetylation is regulated by the opposing actions of histone acetyltransferases (HATs) and histone deacetylases (HDACs) that respectively add and remove acetyl groups from histones. HATs are classified into many families based on conserved structural motifs, three of which include the GNAT, MYST and p300/CBP families. In general, these enzymes have broad specificities, as they

acetylate more than one lysine residue (Kouzarides 2007). HDACs are grouped into four different classes (I, II, III, IV) based on similarities to yeast proteins and each class may be further divided into different subgroups. These enzymes generally do not show specificity for a specific residue but may show preference for a particular histone (Kouzarides 2007). Sirtuins are class III NAD^+-dependent HDACs that deacetylate various acetylated lysine residues (with some preference for H4K16ac) but which also have other nonhistone targets (Vaquero 2009).

Histone methylation is also regulated by the antagonistic enzymatic activities of histone lysine methyltransferases (HKMTs) and histone lysine demethylases (HDMs). HKMTs show marked specificities, typically methylating a particular lysine on a particular histone. All HKMTs catalyze the transfer of a methyl group from S-adenosylmethionine to a lysine's ε-amino group (Zhang & Reinberg 2001). Two families of histone demethylases have been identified, lysine specific demethylases 1 (LSD1) and Jumonji-C domain-containing histone demethylases (JHDMs) with lysine and methyl state specificities (Klose et al 2006).

1.3.5 The histone code hypothesis

Different PTMs are able to recruit enzymes that alter local chromatin conformations and different patterns of histone modifications are associated with distinct transcriptional states. It is also believed that a single PTM does not lead to large alterations in chromatin structure nor transcription. The Allis laboratory hypothesized that different combinations of PTMs on different histone tails could constitute an information-rich code. This histone code could be "written" and "read" by different enzymes to orchestrate specific DNA-based tasks for

particular biological outcomes (Jenuwein & Allis 2001). Considering the vast number of chemical moieties that can be found on histones, all their potential combinations, and the added complexity of potential cross-talk between them, this could result in an almost infinite amount of stored information. This hypothesis assumes that specific combinations of each PTM leads to a defined biological consequence. Currently, the identity of many PTMs are known, but there is no clear picture of how all these modifications, and in what combinations, orchestrate specific biological functions, and whether they constitute a code in the strictest sense of the word is yet to be determined.

1.3.6 Histones and PTMs in *P. falciparum*

P. falciparum posses the four canonical core histones and the histone variants H2A.Z, H2Bv, H3.3 and H3cen but is lacking histone H2AX (Longhurst & Holder 1997; Miao et al 2006; Trelle et al 2009). Unlike in vertebrates, histone transcripts are polyadenylated (Miao et al. 2006). A recent study has found PfH2A.Z to be constantly and ubiquitously localized to euchromatic intergenic regions and strongly correlated with H3K9ac and H3K4me3 (Bartfai et al 2010). There is no annotated linker histone H1 in the *P. falciparum* genome, nor has it been identified in any other Apicomplexan sequenced thus far (Gardner et al 2002; Sullivan et al 2006). This could have implications in the formation of higher order chromatin structures in Apicomplexa, and especially *P. falciparum* as (A+T) rich sequences are less efficiently packaged into nucleosomes (Nelson et al 1987). This is reflected in the fact that the chromosomes of *P. falciparum* do not condense during metaphase (Gerald et al 2011 and sources therein).

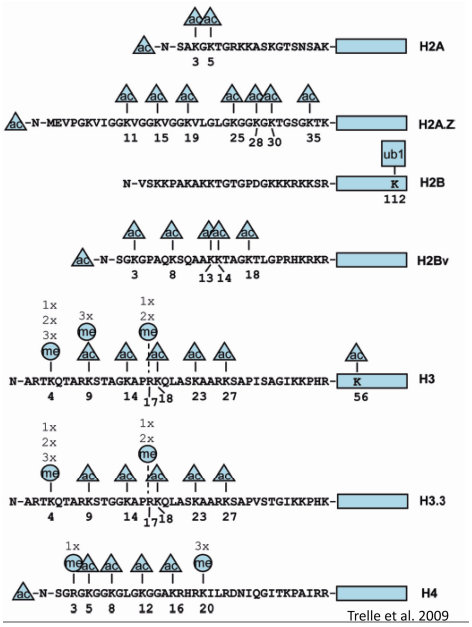


Figure 1.7 PTMs in *P. falciparum*. A summary of the PTMs identified by mass spectrometry indicating the identity and position of each modification. ac, acetylation, me, methylation, 1x, mono, 2x, di, 3x, tri, ub1, ubiquitylation. Figure reproduced from Trelle et al 2009.

Mass spectrometry analysis has identified a total of 44 different PTMs thus far on all *P. falciparum* histones, and multiple modifications are found on singe histone tails (Figure 1.7, Miao et al 2006; Trelle et al 2009). The histone tail of H2A.Z is longer in *P. falciparum* than in mammals and is associated with high levels of acetylation that distinguishes this *P. falciparum* histone from its human host (Trelle et al 2009). Importantly, there are an abundance of marks associated with gene activation and a lack of marks associated with gene repression; notably methylation of H3K27, H3K36 and H3K79 are missing (Trelle et al 2009). However, there is some conflicting data, as the Scherf group has identified H3K79me3 by both western blot and immunofluorescence assays (Issar et al 2009).

Genome wide ChIP-on-chip mapping of H3K9me3 from the Scherf lab and later confirmed by the Stunnenberg lab, revealed that this mark is specifically localized to gene families that are associated with host-parasite interactions that are generally subtelomerically located, as well as a few internal chromosome clusters (Lopez-Rubio et al 2009).

H3K9me3 was not found to be a general mark of silencing in *P. falciparum*, as it was not found on inactive genes outside of the heterochromatin islands in either study. This was in contrast to the other histone marks H4K20me3, H3K4me3 and H3K9ac that showed a broad chromosomal distribution. Using RNA-seq, the Stunnenberg lab demonstrated that H3K9ac at promoter regions is correlated with transcriptional status, whereas H3K4me3 shows cycle-regulated deposition, peaking at the trophozoite stage, but it does not correlate well with changes in gene expression (Bartfai et al. 2010).

1.3.7 Histone modifying enzymes and chromatin effectors in *P. falciparum*

In terms of HATs, *P. falciparum* has PfGCN5 and PfMYST, members of the GNAT and MYST family of HATs respectively, as well as an additional six putative HATs (Bischoff & Vaquero 2010). PfGCN5 has been shown to acetylate H3K9 and H3K14 *in vitro* and has been shown to interact with the coactivator PfADA2, suggesting that they may interact *in vivo* to form large multiprotein HAT complexes as described in yeast (Fan et al 2004).

Regarding HDACs, *P. falciparum* has PfHDAC1, a zinc-dependent class I HDAC (Joshi et al 1999) and two sirtuins. Sirtuins are NAD^+-dependent HDACs. PfSir2A is most similar to group I sirtuins and also has ADP-ribosyltransferase activity, whereas PfSir2B is most similar to group IV sirtuins (Freitas-Junior et al 2005; Merrick & Duraisingh 2007; Tonkin et al 2009). There are also an additional three HDAC homologs that have yet to be characterized (Chaal et al 2010).

Different bioinformatic analyses have identified either six or nine putative HKMTs in *P. falciparum* and two of these proteins expressed in a eukaryotic system showed HKMT activity. PfSET8 was able to methylate H4K20me1/2/3 and PfSET2 methylated histone H3 *in vitro* (Cui et al 2008a; Bischoff & Vaquero 2010). Through immunolocalization, PfKMT1 was shown to be located at the periphery of *P. falciparum* nuclei (Lopez-Rubio et al 2009). This was confirmed by the Cowman lab, who in addition to PfKMT1, also localised an additional 7 putative HMTs to the nucleus (Volz et al 2010).

Two putative JHDMs have also been identified by bioinformatic analysis, PfJmjC1 and PfJmjC2 (Cui et al 2008a). An LSD1 homolog identified by bioinformatic analysis has also been observed to localize to the nucleus (Iyer et al 2008; Volz et al 2010).

Drugs that interfere with histone modifying activities have been described as antimalarials *in vitro*. The inhibitory effects of apicidin, a type I HDAC inhibitor was reported for *P. berghei* and another study found that *P. falciparum* growth was inhibited by anacardic acid, a compound that inhibits the HAT PfGCN5 (Darkin-Rattray et al 1996; Cui et al 2008b). The Bozdech lab showed that treatment of *P. falciparum* with apicidin led to *P. falciparum* death and to genome-wide hyperacetylation which resulted in a deregulation of transcription along the IEC (Chaal et al 2010).

The sole chromatin effector to be characterized in *P. falciparum* thus far has been HP1. Studies found that PfHP1 specifically binds di- and tri-methylated H3K9 but not to other methylated or repressive marks, the

protein dimerizes, and is found in discrete foci localized to chromosomal end clusters. PfHP1 was found to be localized to silent *var* genes and was highly correlated with previously mapped H3K9me3 locations (Perez-Toledo et al 2009; Flueck et al 2009).

High mobility group box (HMGB) proteins are proteins that can either bind in sequence dependent or independent fashions to regulate transcription and they have also have been implicated in chromatin remodeling. Four HMGB homologs have been identified in *P. falciparum*, and two (PfHMGB1 and PfHMGB2) have been shown to bind DNA *in vitro* (Briquet et al 2006; Kumar et al 2008). Knockouts of HMGB2 in the rodent malaria *P. yoelii* resulted in a reduced capacity to form oocysts (Gissot et al 2005).

1.4 An introduction to epigenetics

1.4.1 Epigenetics and criteria defining *bona fide* epigenetic signals

Epigenetics refers to alternative states of gene expression that are stable over time, heritable and are not associated with changes in DNA sequence (Campos & Reinberg 2009). It is clear that many factors contribute to the maintenance of an epigenetic phenotype, including DNA methylation, PTMs, histone variants and various types of ncRNAs. However, there is a tendency to classify any molecular factor, such as every ncRNA or histone variant as an "epigenetic signal". Since no singular mechanism contributes to the maintenance of an epigenetic phenotype, an attempt to integrate these factors has resulted in what is referred to as "epigenetic pathways" (Berger et al 2009). This permits for the inclusion of all the factors that contribute to the maintenance of an epigenetic phenotype but permits us to distinguish between which molecular signals are actually transmitting epigenetic memory.

For a molecular signal to be considered truly epigenetic, it is required to i) self perpetuate ii) be heritable iii) have an effect on gene expression and iv) be reversible (Bonasio et al 2010). Of the many factors that participate in an epigenetic pathway, DNA methylation is one of the best characterized that meets all of the above stated criteria to be considered a true epigenetic signal that transmits information (Bonasio et al 2010). The semi-conservative mode of DNA replication results in two hemi-methylated double strands that are fully methylated by DNA methyltransferases that specifically target hemi-methylated DNA (Goll & Bestor 2005) ensuring its propagation through cell divisions and DNA methylation patterns are robustly stable through many cell divisions

(Wigler et al 1981). DNA methylation negatively regulates transcription (Bird & Wolffe 1999), and DNA methylation is readily reversible by DNA demethylases (Goll & Bestor 2005). However, epigenetic inheritance is observed in several species that have little to no DNA methylation, as in *S. pombe* , *S. cerevisiae*, *C. elegans* and *D. melanogaster* (Lavrov & Kibanov 2007 and sources therein). This strongly suggesting that at least in these organisms, other factors apart from DNA methylation function as epigenetic signals.

In contrast to DNA methylation, it is clear that various types of ncRNAs are important in the maintenance of epigenetic phenotypes (Margueron & Reinberg 2010), but many are not themselves epigenetic signals. For example, long ncRNAs such as the *Kcnq1ot1* and *Xist* transcripts have been implicated in the genetic imprinting of specific loci (Koerner et al 2009, discussed further in section 1.7.1). It has been suggested that these ncRNAs participate in epigenetic pathways by recruiting, in a sequence-specific manner, the deposition of epigenetic signals, such as DNA methylation and PTMs (Nagano et al 2008; Pandey et al 2008). Similarly, short RNAs, derived from RNAi pathways, have also been demonstrated to interact with DNA methyltransferases (Matzke et al 2007). However, the ncRNAs themselves do not transmit cellular memory and thus are not epigenetic signals, but can be considered effectors involved in maintaining an epigenetic phenotype.

Similarly, histone variants may assist in maintaining an epigenetic phenotype but do not function as epigenetic signals. For example, nucleosomes containing histone H3.3 are unstable and this variant is often deposited at transcribed genes, promoters and gene regulatory

elements (Talbert & Henikoff 2010). A euchromatic epigenetic phenotype might be associated with histone H3.3 but the histone itself may not transmit cellular information, but may be required in the epigenetic pathway to facilitate transcription. Whether PTMs function as true epigenetic signals is currently debated and will be discussed below in section 1.4.3.

1.4.2 Factors that contribute to the maintenance of epigenetic pathways in *P. falciparum*

As stated above, many factors including DNA methylation, PTMs, histone variants and various types of ncRNAs contribute to the maintenance of an epigenetic phenotype. However, this list is somewhat attenuated in *P. falciparum*, as several of these factors are missing. There is no 5-methyl-2'-deoxycytosine (5mC) DNA methylation, nor has a gene encoding the cognate methyltransferase (DNMTI) been detected in *P. falciparum* (Choi et al 2006; Ponger & Li 2005). Although based on bioinformatic analysis, an enzyme capable of methylating 6-methyl-2'-deoxyadenine (6mA), is present in the *P. falciparum* genome, this methylated DNA residue has been detected at low levels using methylation-sensitive restriction enzymes, but the methyltransferase, when expressed in both *E. coli* and yeast did not display enzymatic activity (Hattman 2005; Pollack et al 1991). Therefore, *P. falciparum* likely has little to no DNA methylation.

There is no regulation of gene expression via short-interfering RNAs (siRNAs), as *P. falciparum* does not have a functional RNAi pathway, and transient transfection assays expressing antisense RNA have not resulted in gene silencing (Baum et al 2009). Gene silencing by antisense transcripts could be possible in certain contexts by base pairing with

cognate sense transcripts, (Militello et al 2004). Other ncRNAs could be involved in the maintenance of epigenetic pathways; a recent study identified 107 novel ncRNA transcripts using RNA-seq, the majority of which are shared with other sequenced *Plasmodium* species (Otto et al 2010). Some ncRNAs have been found to be transcribed from *P. falciparum* centromeres and to associate with centromeric chromatin (Li et al 2008), and a recent study has identified a large number of telomere associated long ncRNAs (Broadbent et al 2011).

Regarding histone variants, two recent publications have focused their attention on H2A.Z, a variant that has been implicated in numerous diverse functions in other organisms, ranging from heterochromatin formation, chromosome stability and transcriptional activation (Talbert & Henikoff 2010). The Stunnenberg lab analysed the genome-wide distribution of PfH2A.Z and they suggest that this variant defines euchromatic intergenic regions and propose that it could serve as a platform for the assembly of chromatin effectors (Bartfai et al 2010).
Other work on H2A.Z from the M. Duffy lab has shown that the deposition of this variant was correlated with *var* gene expression but not with the activity of other transcribed genes (Petter et al 2011).

PTMs have been implicated in the transmission of epigenetic memory in *P. falciparum* of clonally variant genes, discussed further in section 1.5.3.

1.4.3 Transmission of epigenetic memory by PTMs

It is perhaps intuitive to implicate PTMs in epigenetic processes as they are closely associated with DNA and many have direct functional consequences that affect transcription. However, not all PTMs transmit

cellular memory, and it is debatable whether PTMs can fulfil any of the criteria of an epigenetic signal as described above (section 1.4.1).

A mechanism is required to initiate and establish the PTMs for the desired epigenetic phenotype, once established they need to be propagated, mechanisms must exist to transmit PTMs through cell divisions and the PTMs must be reversible. Data supporting PTMs as epigenetic signals is scant, however, several compelling models have emerged that support the hypothesis that PTMs transmit epigenetic memory.

STFs are the leading candidates to orchestrate the deposition of PTMs for the establishment of a specific phenotype, as they have been show to interact with enzyme complexes that possess chromatin modifying activities (Bonasio et al 2010). An important example is in the Polycomb/Trithorax system of transcriptional regulation that requires Polycomb group (PcG) and Trithoriax group (trxG) proteins to respectively silence and activate genes. PcG and trxG group proteins mediate these processes by the deposition of either H3K27me3, in the case of repression, or H3K4me3 in the case of activation (Simon & Kingston 2009). It has been shown that sequence-specific DNA binding proteins are what recruit these chromatin modifying complexes to particular sequences and it is hypothesized that these sequence specific DNA binding proteins serve as platforms to recruit PcG and trxG complex binding. These DNA-binding proteins include Pleiohomeotic, Pleiohomeotic-like and Trithorax (Schuettengruber et al 2009). STFs can also recruit DNA methyltransferases in a sequence-specific fashion (Hervouet et al 2009).

The replication of chromatin requires DNA synthesis and nucleosome assembly to be carried out simultaneously. Nucleosomes ahead of the replication fork need to be disrupted and reassembled behind the fork on both the leading and lagging strands and replication experiments carried out *in vitro* have shown that chromatin is faithfully reassembled close behind the replication fork (Lucchini et al 2001). How parental nucleosomes are deposited and segregated during DNA replication is at the root of epigenetic transmission by PTMs.

Nucleosomes on replicating DNA are derived both from parental histones and new, naive histones; but only parental histones contain any pre-replication information. It is accepted that parental nucleosomes, carrying the original PTMs, are deposited to both strands of newly replicating DNA (Pospelov et al 1982; Jackson & Chalkley 1985). Thereby, both copies of newly replicated DNA would contain information present in the original template. However, the manner in which parental histones are segregated remains unclear. Several models of histone segregation have been proposed including a random model and two semi-conservative models of histone segregation based on whether histones are segregated as dimers or tetramers. As the name implies, the random model of histone segregation suggests that histones are randomly segregated between the leading and lagging strands and whether the histones segregate as tetramers or dimers is inconsequential. The semi-conservative models proposes that histones are equally distributed between the leading and lagging strands, and only differ in whether the H3-H4 tetramer is divided or remains intact during replication (Annunziato 2005). If the H3-H4 tetramer is divided and the dimers are

segregated onto the new DNA strands, the parental histones associate with naive dimers to reconstitute the tetramer. This would imply that all the newly formed nucleosomes contain H3-H4 tetramers that are parental-daughter hybrids. If the tetramer remains intact, then the newly replicated chromatin would contain two distinct populations of nucleosomes, one that contains exclusively parental derived H3-H4 tetramers and another that are composed solely of new H3-H4 tetramers. It is still not entirely clear which is the method, but *in vitro* replication studies have shown that the segregation of parental H3-H4 tetramers and histone octamers, not dimers or half nucleosomes, are deposited onto newly replicated DNA (Annunziato 2005 and sources therein). Regardless of the mode of histone segregation during replication, only part of the newly formed chromatin contains the original pre-replication information. This implies that mechanisms must exist to propagate the message carried by the original PTMs.

Based on the PTMs that characterise mating-type switches in *S. pombe*, the Thon lab has proposed a theoretical model of heritable bistability for the transmission of PTM-mediated epigenetic memory, whereby PTMs function as epigenetic signals (Figure 1.8, Dodd et al 2007).

The model is based on heritable bistability that is maintained by positive feedback loops and cooperativity in PTMs. Bistability refers to networks in which two states are equally stable under the same conditions, and in this case, refers to heritable patterns of gene expression and/or gene

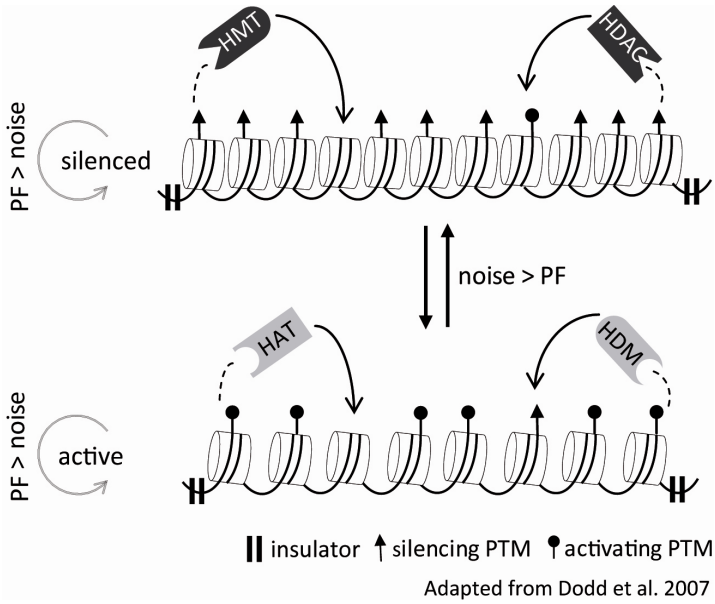


Figure 1.8 Model for the transmission of epigenetic memory by PTMs. PF and PTM refer to positive feedback and post-translational histone modification respectively. HMT, HDAC, HAT and HDM refer to histone methyltransferases, histone deacetylases, histone acetyl transferases and histone demethylases respectively. Circular black arrows indicate an enzymatic action, discontinuous black lines indicate cooperativity. Grey arrows indicate stable transmission when PF>noise, and straight black arrows indicate switches from one stable state to the other when noise > PF. See text for further details. Figure adapted from Dodd et al 2007.

repression. Inherent to bistable systems are stochastic, low frequency switch-like transitions from the one state to the other. Positive feedback in their model refers to recruitment-based interconversions which refers to the ability of a nucleosome that carries a particular PTM to directly or indirectly recruit enzyme activities that either place the same PTM or removes antagonistic PTMs on adjacent nucleosomes. Cooperativity refers to the ability of modified nucleosomes not only to affect nearby nucleosomes but also to non-adjacent nucleosomes. Another parameter in their model is that of noise, which refers to recruitment-independent conversions that occur due to the random gain or loss of PTMs. Their

model is able to produce stable and heritable patterns of gene expression depending on the positive feedback to noise ratios in PTMs, even if the PTMs are highly dynamic and even if nucleosomes are randomly partitioned during replication. It is the relationship between positive feedback and noise that contributes either the maintenance of a particular epigenetic state (positive feedback > noise) or a switch to the opposite state (positive feedback < noise) (Dodd et al 2007).

Although the classification of PTMs as molecular signals is largely based on theoretical models, it is clear that PTMs at least act as effectors in epigenetic pathways.

1.5 Clonally variant gene expression (CVGE) of *var* genes

1.5.1 *var* genes, the paradigm of antigenic variation in *P. falciparum*

var genes are a large family of ≈ 60 members per haploid genome that code for different forms of erythrocyte membrane protein 1 (PfEMP1) (Gardner et al 2002). PfEMP1 is a parasite-derived, clonally variant adhesion molecule that is exported to the surface of infected erythrocytes and is an important factor in *P. falciparum* pathogenicity (Baruch et al 1995; Su et al 1995; Scherf et al 1998). Different PfEMP1s bind to different host molecules present on the endothelial lining of host capillaries. These host molecules include CD36, CD31/PECAM1, CD54/ICAM1 and CSA (Kraemer and Smith 2006). This binding of infected erythrocytes to host molecules effectively sequesters the parasite within the host vasculature away from circulation and thus the parasite avoids clearance by the spleen (Miller et al 1994). PfEMP1 also mediates a process called rosetting, the spontaneous binding of infected erythrocytes to uninfected erythrocytes, via an interaction with several different receptors on uninfected erythrocytes, a phenotype that may be associated with severe malaria (Rowe et al 1995; Rowe et al 1997; Juillerat et al 2011).

Importantly, PfEMP1 is also a key antigenic molecule. The expression of PfEMP1 is mutually exclusive such that an individual parasite only expresses a single form of PfEMP1 at a time, a phenomenon that involves the simultaneous silencing of all other family members (Figure 1.5C, Scherf et al 1998). PfEMP1 is exposed on the surface of the infected erythrocyte and is exposed to the host immune system which readily

makes antibodies against the expressed PfEMP1 (Bull et al 1998). However, the *var* family of genes undergoes low level switching such that before an infection has been successfully cleared, a subpopulation emerges that has switched to express a distinct PfEMP1 that is no longer recognized by host immunity (Roberts et al. 1992). This antigenic variation mediated by PfEMP1 leads to waves of parasitemia and persistent infections despite antibody-mediated immune pressure in non-naive individuals (Miller et al 1994; Molineaux et al 2001). The mutually exclusive expression program coupled with low level switching allows the parasite to guard its antigenic repertoire. However, with clinical parasite loads that can exceed 10^{10} parasites and only ≈60 *var* genes, it is unclear how the parasite prevents the squandering of its limited antigens to establish chronic infections (Recker et al 2011). Based on long term culturing of clones and mathematical modeling, the Newbold lab has recently suggested that *var* gene switching is non-random and highly structured, and that each *var* gene may possess an intrinsic switching rate, with both slow and fast switching phenotypes (Recker et al 2011). They suggest that this switching mode, termed single-many-single (*sms*), permits the parasite to protect its repertoire and simultaneously establish chronic infections in non-naive individuals that already have immunity to some *var* variants and show temporary cross-reactive responses. It is known that antigenic switching is accomplished without DNA rearrangements, and involves epigenetic mechanisms (Scherf et al 1998). Yet, the molecular mechanisms underlying this process are not completely understood.

var genes have a conserved genetic structure composed of two exons separated by an intron . The first exon is hypervariable and codes for the

extracellular and single transmembrane domains of PfEMP1. The second exon is conserved and codes for a region that anchors the protein within a knob structure at the erythrocyte membrane (Smith et al 2000). *var* genes are disinguished into different groups based on their 5' promoter types, which are classified by chromosomal location and sequence homology (Kraemer et al 2007). *var* genes have an additional promoter with bidirectional promoter activity originating from within the intron that results in the production of sense and antisense transcripts (Su et al 1995; Calderwood et al 2003; Epp et al 2009). These transcripts are not polyadenylated and are transcribed when a particular *var* gene is active or silent, and is thus independent of 5' promoter activity (Epp et al 2009). The two promoters are active at different times: the upstream promoter is active in ring stages (8-20 h post invasion), whereas the promoter within the intron is active 24-48 h post invasion (pigmented trophozoites and schizonts, Kyes et al 2003; Kyes et al 2007; Scherf et al 1998). It is know that the expression of *var* genes is regulated at the level of transcription initiation (Scherf et al 1998; Kyes et al 2003).

1.5.2 Genetic elements involved in the mutually exclusive expression of *var* genes

The *var* gene family shows mutually exclusive expression, and the unknown molecular mechanism that maintains this transcriptional phenomenon is often referred to as a counting mechanism (Scherf et al 2008). Several studies have implicated various genetic elements that somehow establish or cooperate with the counting mechanism to keep only one *var* gene active and the remaining *var* genes silent. It was demonstrated that when removed from their chromosomal context, *var* genes placed on an episome are transcriptionally active, implicating other elements in their silencing (Deitsch et al 1999). It was later

demonstrated that the silencing of episomal *var* promoters requires not only passage through S phase (the phase of the cell cycle where chromatin is assembled), but also the activity of both *var* promoters, that within the 5' UTR and also from the intron (Deitsch et al 2001). This was supported by another study that reported that the spontaneous deletion of elements within the intron resulted in the activation of an episomal *var* gene (Gannoun-Zaki et al 2005). This was supported by the previous identification of var_{COMMON}, a *var* gene found in several field isolates taken from infected children in Gabon, that is constitutively transcribed and lacks both the intron and exon 2 domains (Winter et al 2003). It was later shown that *var* promoters become active when they are unpaired from their intron (Frank & Deitsch 2006). However, the necessity of the intron in *var* gene silencing was debated by other groups, when it was shown that not all silent *var* genes were accompanied by the transcription of the sterile transcript (Ralph et al 2005).

The Cowman group provided evidence that a *var* promoter itself is sufficient to enter into the mutually exclusive counting mechanism, suggesting that the intron was not required (Voss et al 2006), however, it was later observed that any active promoter could pair with a 5' *var* promoter and silence a *var* gene after passage through S-phase (Dzikowski et al. 2007). This was evidence to suggest that neither the *var* intron nor its sterile transcript are required for *var* gene silencing, but rather, it is the transcriptional activity of the intron that mediates silencing. This observation could account for the silencing of a *var* promoter in the absence of the *var* intron, as in these experiments described by the Cowman lab, used a plasmid that contained an additional promoter, that of *hsp86*. It was also demonstrated that *var*

gene silencing is not dependent on production of the PfEMP1 protein, suggesting that the 5' UTR and the intron contain all the necessary elements to become included in the counting mechanism, and the silent state is not maintained by negative feedback loops (Dzikowski & Deitsch 2006). *var* genes are also known to be silenced at the level of individual genes as opposed to general heterochromatin spreading (Duraisingh et al 2005; Lopez-Rubio et al 2009; Flueck et al 2009; Swamy et al 2011). The Deitsch group also showed that a titratable nuclear factor was involved in *var* gene activation (Dzikowski & Deitsch 2008). How the phenomenon of promoter pairing involves *var* genes in the counting mechanism is for the moment unclear.

Genetic elements found in the upstream regions of different subsets of *var* genes have been shown to be bound by different protein activities (Voss et al 2003). These conserved motifs include SPE1 and SPE2, common to *ups*B promoter types and CPE, common to *ups*C promoter types. Recognition of SPE2 by a protein only occurrs during S-phase. Further work on these elements revealed that the SPE2 element is bound by an ApiAP2 transcription factor that has been named PfSIP2 (Flueck et al 2010). The consensus sequence of PfSIP2 was biochemically mapped and PfSIP2 binding sites were almost exclusively found in subtelomeric heterochromatin in the upstream regions of *ups*B *var* genes. Purification of native PfSIP2 resulted in the co-purification of 18 other proteins, which included two other ApiAP2 transcription factors, and several other proteins involved in DNA replication/repair and chromatin remodelling (Flueck et al 2010).

The Deitsch lab has shown that both sense and antisense transcripts that originate from the *var* intron localize to distinct spots within the nucleus and that they are associated with chromatin, however their expression does not correlate with *var* gene expression status and whether they localize to *var* DNA sequences in the nucleus is unknown (Epp et al 2009).

1.5.3 Epigenetic control of *var* genes

Although the transcriptional regulation of the *var* gene family is not fully understood, it is known to involve epigenetic mechanisms. It was known that the active *var* gene is clonally inherited after many rounds of invasion, suggestive of some form of cellular memory. One of the first indications of epigenetic control was the observation that silent *var* genes are activated *in situ* and the activation does not involve DNA recombination (Scherf et al 1998). Two studies published concurrently in 2005 strengthened this hypothesis when together they showed that *var* gene silencing is reversible and mediated by heterochromatin formation (Duraisingh et al 2005; Freitas-Junior et al 2005). The authors analysed PfSir2, an HDAC, and observed that it spreads along the telomere and is associated with silent *var* genes. It was observed that this HDAC is absent from the promoter of active *var* genes where acetylation occurs and that the activation of a *var* gene involves repositioning of the activated locus to a subdomain in the heterochromatic nuclear periphery. They also showed that its disruption lead to a partial loss of mutual exclusion (Duraisingh et al 2005; Freitas-Junior et al 2005). Further work on the two *P. falciparum* Sir2 homologs, PfSir2A and PfSir2B showed that both proteins are involved in the maintenance of mutually exclusive *var* gene expression (Tonkin et al 2009). The Cowman lab demonstrated that a *var* promoter itself contains sufficient information to nucleate the formation

of a silent chromatin state on an episome that is stably inherited (Voss et al 2006). The Deitsch group suggests that active transcription is required to establish the epigenetic memory in the *var* gene family but is independent from antigen production (Dzikowski & Deitsch 2008). They silenced all the endogenous *var* genes by introducing unpaired *var* promoters on an episome. When the episomes were removed, the parasite resumed the mutually exclusive expression of a single *var* gene, however, *var* genes were activated at random and were not the *var* genes that had previously been expressed. They also noted that unpaired *var* promoters (i.e. episomes containing a *var* promoter without any other additional promoter activities) could not be epigenetically silenced.

Two studies published in 2007 showed that certain PTMs were associated with *var* cellular memory (Chookajorn et al 2007; Lopez-Rubio et al 2007a). Lopez-Rubio and colleagues characterised the PTMs that were associated with a *var* gene that was expressed, bookmarked for expression (poised) or repressed. The poised state refers to a *var* gene that is transcriptionally inactive (i.e. in trophozoite and schizont stages) but that will be activated during the ring stage in the next cycle. It was observed that transcriptionally silent *var* genes were marked with H3K9me3 throughout the IEC. The active *var* gene was associated with H3K4me3, H3K4me2 and H3K9ac, whereas the poised state of a *var* gene was associated with H3K4me2 (Figure 1.9, Lopez-Rubio et al 2007a).

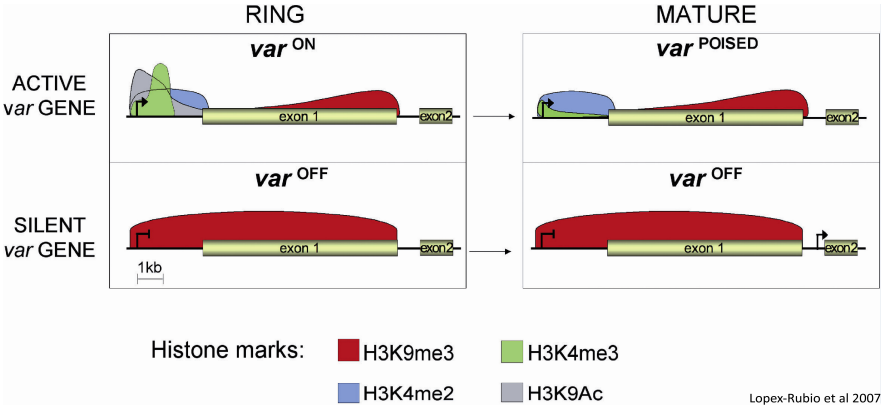


Figure 1.9 A summary of histone H3 PTMs associated with the silent, bookmarked for activation (poised) or active states of *var* genes. See text for details. Reproduced from Lopez-Rubio et al. 2007.

1.5.4 *P. falciparum* nuclear organization and architecture

Almost all of what is known about the nuclear architecture in *P. falciparum* has been acquired by the study of *var* genes. *P. falciparum* subtelomeric ends share a common overall organisation. Telomeric repeats are followed by six blocks of noncoding repeats called telomere-associated repeat elements (TARES 1-6, Gardner et al 2002). These are followed by subtelomeric *var* genes that are often interspersed with *rifin* genes, although some *var* genes are found in internal chromosomal clusters. Sequences found in the subtelomeres are responsible for the physical clustering of chromosome ends, which are in turn tethered to the nuclear periphery (Figueiredo et al 2002). One of the TAREs, TARE6 (also called Rep20) was shown to play a direct role in the physical association of chromosome ends, presumably mediated by Rep20 binding proteins (O'Donnell et al 2002). A protein PfOrc1 was shown to bind several TAREs, and PfOrc1 in combination with PfSir2 contributed to the stability of chromosomal end clusters that perhaps involves the subtelomeres folding back on themselves (Mancio-Silva et al 2008). The

clustering of chromosome ends results in high levels of ectopic recombination between members of antigenically important family and explains why there is minimal overlap in *var* gene repertoires between different strains (Freitas-Junior et al 2000).

Several studies have demonstrated that *var* activation is accompanied with locus repositioning and have demonstrated that silent subtelomeric *var* genes are associated with telomere clusters at the nuclear periphery (Duraisingh et al. 2005, Ralph et al. 2005 PNAS, Voss et al. 2006). There is debate as to the localisation of active central *var* genes, as one group reports that they move away from a telomeric cluster, whereas another reports an active central *var* gene at a cluster, both studies used DNA-FISH (Ralph et al. 2005 PNAS; Voss et al. 2006). A study using RNA-FISH provided evidence that active *var* genes, regardless of chromosomal location, were all localised to the nuclear periphery in areas distinct from telomeric clusters (Lopez-Rubio et al. 2009). However, as the nuclear periphery itself contains both condensed and noncondensed chromatin (Ralph et al. 2005 PNAS), and whether active *var* genes move towards or away from a telomeric cluster, these studies support the hypothesis that the nucleus is divided into distinct microenvironments that either permit or restrict transcription.

It has been suggested that mutually exclusive *var* gene expression is due to limited space in a *var*-specific expression site in the nuclear periphery, although this theory has been challenged by the observation that two active *var* promoters can be simultaneously accommodated in an expression site (Dzikowski & Deitsch 2008; Dzikowski et al 2007) and with the isolation of a single parasite line that expressed two different PfEMP1

on the surface of an infected erythrocyte (Joergensen et al 2010). The Scherf lab found other genes apart from *var* genes that are marked with H3K9me3 also cluster at the nuclear periphery. They have proposed a model where genes that are epigenetically silenced converge on what they have termed perinuclear repressive centres (PERCs) via the formation of small chromatin loops, the formation of which could be mediated by H3K9me3 marks themselves (Lopez-Rubio et al 2009). These PERCs are located in the nuclear periphery and are hypothesized to contain the required enzyme activities to promote the deposition of repressive marks and the removal of active marks, such as PfKMT1, PfHP1 and PfSir2. An immunolocalization survey of several histone methyl PTMs showed their enrichments in distinct sub nuclear zones, again suggestive of transcriptional active and repressive compartments (Issar et al 2009). A recent study has also identified a novel set of telomere-associated long ncRNAs, terms lncRNA-TAREs. The function of these ncRNA-TAREs is currently unknown but they are coordinately expressed after S-phase and each one is adjacent to an *ups*B type *var* gene (Broadbent et al 2011).

1.6 Clonally variant gene expression (CVGE) beyond *var* genes

1.6.1 CVGE in other gene families apart from *var*

In large part because of their role in pathogenesis, much attention has focused on *var* genes. However, it has emerged that other genes exported to the erythrocyte membrane also undergo expression switches, as CVGE (Figure 1.5B) has also been observed in the *rifin*, *stevor* and *Pfmc-2TM* families, that contain approximately 150, 35, and 14 members respectively (Kyes et al 1999; Lavazec et al 2007; Niang et al 2009). These three gene families are related as they constitute the two-transmembrane (2TM) superfamily. However, the biological function of the 2TM superfamily and how their expression is regulated remains largely unknown. CVGE had also been observed in the small gene family, *PfRh*, which consists of six members (Taylor et al 2001; Duraisingh et al 2003).

However, CVGE is not limited to large multigene families, *PfRh*, nor to gene families that are exposed for long durations of time to host immunity, as it was demonstrated that a wide variety of small gene families (with 4-6 members) implicated in erythrocyte invasion also showed CVGE (Cortés et al 2007). These authors showed that parasite populations are mixtures of isogenic parasites that are transcriptionally different and they observed CVGE for members of the *eba*, *clag*, and *acbp* multigene families, and confirmed the clonally variant expression of *PfRh*. They also found that two members of the *clag* gene family, *clag3.1* and *clag3.2,* display mutually exclusive expression (Figure 1.5C), the only other known example of mutual exclusion in *P. falciparum* apart from the *var* genes.

Other evidence indicating that epigenetic mechanisms were involved in the transmission of cellular memory beyond *var* genes was shown by Cortés and colleagues (2007) when they showed that expression status of *eba*-140, *clag3.1* and *clag3.2* was clonally transmitted, promoter mediated and occurred without detectable DNA rearrangements. The Miller lab demonstrated that *PfRh4*, a clonally variant gene involved in erythrocyte invasion, when transcriptionally silent was marked by H3K9me3, and the active state of the gene was associated with H3ac, H4ac, and H3K4me3, suggesting that PTMs may be involved in the transmission of epigenetic memory in other families apart from *var* genes (Jiang et al 2010).

1.6.2 Cytoadherence-linked asexual genes, *clag3.1* and *clag3.2*

The cytoadherence-linked asexual gene (*clag*) family is comprised of five members, *clag*2, *3.1*, *3.2*, *8* and *9* that encode for RhopH1 (Kaneko et al 2001). RhopH1 is one member of the high molecular weight rhoptry (RhopH) complex which also consists of RhopH2 and RhopH3, both of which are encoded by single copy genes (Holder et al 1985; Cooper et al 1988). The inability to genetically disrupt *rhopH*2 or *rhopH*3 suggests that the function of the complex is essential (Comeaux et al 2011 and sources therein). The *clag* genes show lineage specific expansion as they are found only in plasmodia and in all plasmodia species (Iriko et al 2008). Phylogenetic analysis has shown that the *clag* family is split into two distinct lineages, Type-AP *clags* that consist solely of *clag*9 and Type-A *clags* that consist of the other four members. Type-A *clags* are highly polymorphic and the polymorphisms of *clag*8 and *9* sequences are

suggestive of diversifying selection (Kaneko et al 2005; Iriko et al 2008). *clag*3.1 and *clag3.2* are subtelomerically located on the left arm of chromosome 3 and are only separated by a *var* pseudogene and a ncRNA (Figure 1.10).

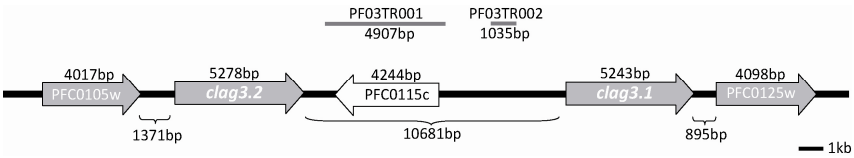


Figure 1.10 Genomic context of *clag3.1* and *clag3.2*. *clag3.1* (PFC0120w) and *clag3.2* (PFC0110w) are located in tandem in a head to tail arrangement on the left arm of chromosome 3. *clag3.1* and *clag3.2* are separated by a *var* pseudogene, PFC0115c (whose transcript is referred to as PF03TR001) and a ncRNA, PF03TR002. Upstream of *clag3.2* is PFC0105w, a putative serine/threonine protein kinase. *clag3.1* is followed by PFC0125w, a putative ABC transporter. The size of the genes and their intergenic regions are indicated in base pairs. For exon-intron structure of *clag3.1* and *clag3.2*, see Article 1, Figure 5A.

At the nucleotide level, their coding sequences are 95% identical, although their 5' regions are more divergent. *clag3.1* was recently identified as one of eight genes that are differentially upregulated in children compared to pregnant women in field isolates from Tanzania, suggesting that it may have a role in childhood malaria pathogenesis (Vignali et al 2011). *clag3*.1 and *3.2* are exclusively and highly expressed in schizonts, which correlates with the time of RhopH1 production (Kaneko et al 2005). Previous work has shown that *clag3.1*, *3.2* and *2* show CVGE, as different subclones that express one or the other can be isolated. It is also known that the silencing of these genes is clonally transmitted and stable for at least 25-30 asexual generations and does not involve detectable changes in DNA sequence. Due to these observations, the silencing of these genes was hypothesized to be

mediated by epigenetic mechanisms (Cortés et al 2007). These authors also found the expression of *clag3.1* and *clag3.2* to be mutually exclusive, an observation that was later confirmed (Comeaux et al 2011).

The authors who characterized PfSIP2 tested the ability of this ApiAP2 transcription factor to bind SPE2-like consensus sequences that are present in the upstream regions of other genes, including the *clags* and *rhopH*3 (Young et al 2008; Flueck et al 2010). Although the consensus sequences are quite similar, the sequences found in *clag3.1* and *clag9* were not bound by PfSIP2, whereas the sequence found in *rhopH*3 was recognized by PfSIP2, although chromatin immunoprecipitations (ChIP) experiments did not confirm their interaction (Flueck et al 2010).

1.6.3 Biochemical function of *clag3.1* and *clag3.2*

The RhopH complex was first identified as a three member complex found in the rhoptries (Holder et al 1985) composed of three distinct peptides (Cooper et al 1988). It was hypothesized that the complex was involved either in invasion or the establishment/maintenance of the PV (Cooper et al 1988). The genes that coded for the proteins in the complex were identified and RhopH1 was found to be encoded by a member of the *clag* family (Kaneko et al 2001). The *clag* family was named as such because the absence or disruption of the first *clag* gene to be identified, *clag9*, in addition to other phenotypes, resulted in reduced adhesion of infected erythrocytes to CD36, a glycoprotein that mediates numerous adhesive processes (Day et al 1993; Chaiyaroj et al 1994; Holt et al 1999; Trenholme et al 2000; Gardiner et al 2004). Hydrophobicity analysis of CLAG9 predicts that the protein contains several transmembrane domains (Trenholme et al 2000).

Subcellular localization experiments localized the RhopH complex to rhoptry bulbs in schizonts (Gardiner et al 2004) and to the apical end of merozoites (Ling et al 2004; Kaneko et al 2005). The complex persists in ring (Lustigman et al 1988; Ling et al 2004) and is distributed into the host erythrocyte membrane and into the membrane of the parasitophorous vacuole (Hiller et al 2003; Ling et al 2003). The complex was found to be composed of a 1:1:1 ratio of the three components and since RhopH1 is encoded by five different genes, it was hypothesized that the different gene products could confer some specificity to the complex (Kaneko et al 2005). It has also been shown that antibodies against components of the RhopH complex partially inhibit the growth of *P. falciparum in vivo* and *in vitro* (Kaneko 2007 and sources therein). Cumulatively, these data supported the original hypothesis, implicating the RhopH complex in invasion or in PV establishment for many years.

However recently, and unexpectedly, the products of *clag3.1* and *3.2* have been identified by the Desai laboratory as either a main component or a regulator of the plasmodium surface anion channel (PSAC) involved in the new permeability pathways (NPP) of infected erythrocytes (Nguitragool et al 2011). These NPP refers to the induction of new pathways of solute permeability some hours after merozoite invasion in the infected erythrocyte membrane upon the proliferation of *P. falciparum*. NPP are necessary for the uptake of nutrients and the disposal of parasite metabolic waste, they involve the translocation of diverse organic and inorganic solutes and could involve several ion-channel types (Huber et al 2005; Hill & Desai 2010). The Desai laboratory screened two *P. falciparum* strains, HB3 and Dd2 with a chemical library

to identify compounds that inhibited PSAC-associated lysis. They identified two such compounds that differentially inhibited PSAC activity in the two strains and they took advantage of a genetic cross that had been performed between HB3 and Dd2 to genetically map and eventually identify and confirm the association of *clag3.1* and *clag3.2* with PSAC activity. They convincingly correlated *clag3* with PSAC activity by allelic exchange experiments, were able to use the two different chemical inhibitors to select for *clag3.1* or *clag3.2* expressing populations and they demonstrated that part of the CLAG3.1 and CLAG3.2 proteins are exposed on the surface of the parasitized erythrocyte.

In light of these recent findings, many questions regarding *clag3* involvement in PSAC activity arise; CLAG3 is one member of a three member protein complex, what is the involvement of the RhopH complex in PSAC function? as the CLAG3 products are transmembrane proteins, how are they transported from the rhoptry bulb to the surface of the infected erythrocyte? if the gene is transcribed and the protein translated in late schizonts, but PSAC activity does not begin until ≈20 h after merozoite invasion, what is the role of CLAG3 during this intervening time? why is *clag3* mutually exclusively expressed if it encodes an ion channel? These questions can now be addressed experimentally and are sure to produce interesting results in the near future.

1.6.4 Erythrocyte binding antigen-140 (*eba*-140)

Two multigene families that encode proteins that bind erythrocytes and determine the invasion pathways utilized by the parasite have been described in *P. falciparum*, the erythrocyte-binding-like (EBL) and

reticulocyte-binding-like (RBL) families (Gaur et al 2004). These two families participate in erythrocyte invasion, a complex process that involves several redundant molecular interactions between parasite ligands and host receptors that appear to function in a hierarchical fashion (Baum et al 2005; Stubbs et al 2005). *eba-140* is part of the EBL family, whose other members include *eba-175*, *eba-181*, *ebl-1*, and *eba-165* (a pseudogene). *eba-140* is subtelomerically located, expressed during late schizont stages and also shows CVGE (Cortés et al 2007). The genomic context of *eba-140* can be found in Figure 1.11. *eba-140* shows CVGE which is stably transmitted and does not involve detectable changes in DNA sequence, implicating an epigenetic signal in the maintenance of *eba-140* expression status (Cortés et al 2007). *eba*-140 expression is clinically relevant as it was determined to be a dominant invasion ligand expressed in clinical isolates from The Gambia that were cultured *ex vivo* (Gomez-Escobar et al 2010).

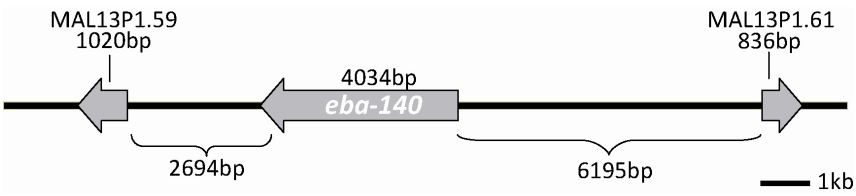


Figure 1.11 Genomic context of *eba-140*. *eba-140* is found on the left arm of chromosome 13. Upstream of *eba-140* is MAL13P1.61, a member of the *hyp8* family of genes that encode for exported proteins of unknown function. Downstream of *eba-140* is MAL13P1.59, a member of the *phista* family of genes that encode for exported proteins of unknown function. The size of the genes and their intergenic regions are indicated in base pairs. For exon-intron structure of *eba-140*, see Article 1, Figure 5A.

The biochemical function of *eba-140* is well described. EBA-140 resides in the micronemes and is discharged upon erythrocyte invasion. In concert

with other proteins, EBA-140 assists in tight junction formation by binding glycophorin C present on erythrocyte membranes via sialic acid residues (Lobo et al 2003). High levels of antibodies against the EBA-140, 175 and 181 in children from Papua New Guinea were strongly associated with protection from symptomatic malaria and high density parasitemia, indicating that EBLs are important target of acquired immunity (Richards et al 2010).

1.7 Introduction to monoallelic and mutually exclusive gene expression

Monoallelic gene expression (allelic exclusion) refers to the random or predetermined transcription of a single allele in diploid organisms. Although the expression of a single *var* gene is often referred to as monoallelic, *P. falciparum* possesses a haploid genome in the human host, and accordingly, there is no allelic selection. For clarification, the expression of one member of a multigene family and the simultaneous silencing of all other members of that family will be termed mutually exclusive gene expression (Figure 1.5C). Mutually exclusive gene expression has been documented solely for the ≈60 member *var* gene family and two members, *clag3.1* and *clag3.2* of the five member *clag* family (Scherf et al 1998; Cortés et al 2007). These two genes are collectively referred to as *clag3*.

Antigenic variation is a common form of CVGE used by diverse pathogens that often involves the mutually exclusive expression and periodic switching of the exposed antigen in order to evade host immunity (Deitsch 2009). In diploid organisms, mutually exclusive expression is not associated with antigenic variation, but is often used when expression of more than one gene would confound a certain physiological process, such as in the case of mammalian olfaction, colour vision, or generation of antibody diversity in the immune system. Mammalian olfaction and colour vision are discussed in section 1.7.2.

Although monoallelic gene expression and mutually exclusive gene expression are mechanistically different processes, they are analogous in the sense that the end result is the same; only one member is selectively

expressed while the other member(s) are repressed. Even though the molecular mechanisms by which a single allele or one gene of a multigene family is expressed have not been fully elucidated, several recurrent themes have emerged in their regulation, including asymmetric DNA methylation, PTMs, chromatin looping and ncRNAs. An overview of the different forms of monoallelic and mutually exclusive gene expression in diploid organisms and pathogens follows below.

1.7.1 Non-random and random monoallelic gene expression

Genetic imprinting (non-random monoallelic expression)

Monoallelically expressed genes fall into two major categories: imprinted and randomly monoallelic. Imprinted alleles are expressed exclusively from the maternally or paternally inherited chromosome in a parent-of-origin dependent manner, thus imprinted loci are functionally haploid. This is in contrast to random monoallelic gene expression where alleles can be expressed from either the maternal or paternal chromosome.

Imprinted genes often occur in clusters of 2-12 genes and one of these genes often encodes a ncRNA (Koerner et al 2009). Within imprinted clusters, imprinted gene expression is regulated by the imprint control element (ICE), a DNA sequence that controls the expression of multiple genes in *cis*. The ICE shows differential CpG methylation, where one parental chromosome is methylated (imprinted) and the other unmethylated, the later ultimately controlling the expression of the entire cluster. Two major mechanisms have been described to regulate

imprinted gene clusters; transcription of ncRNAs and allele-specific insulation mediated by CTCF and cohesin.

The regulation of different imprinted clusters involves ncRNAs that interact with various molecular components to orchestrate imprinted gene expression. Once such example is the maternally imprinted *Kcnq1* cluster in mice and its cognate ncRNA, *Kcnq1ot1*. The *Kcnq1* cluster is a large imprinted cluster spanning 800 kb and contains nine imprinted mRNAs that lie up- and downstream from where *Kcnq1ot1* is transcribed. *Kcnq1ot1* is transcribed from a promoter within the unmethylated ICE on the paternal allele (Mancini-Dinardo et al 2006). Truncation of the *Kcnq1ot1* transcript from the paternal allele results in the loss of repression of all the protein-coding genes in the imprinted cluster, demonstrating that this ncRNA, not only the transcriptional activity associated with its expression, is required for imprinted gene expression (Mancini-Dinardo et al 2006). *Kcnq1ot1* was shown to physically associate with several silent genes located hundreds of kb away from the site of its transcription (Pandey et al 2008). It was also shown that *Kcnq1ot1* physically interacts with polycomb repressive complex 2 (PRC2) and with the HMT G9A, responsible for the deposition of the repressive PTMs of H3K27me3 and H3K9me2 respectively (Pandey et al 2008). Furthermore, it was shown that the ncRNA also physically interacts with the DNA methyltransferase Dnmt1 (Mohammad et al 2010). Taken together, *Kcnq1ot1* acts to target HMTs and DNA methyltransferases to the entire imprinted cluster in a lineage-specific manner.

CTCF can act as a methylation-sensitive insulator element to control imprinted gene clusters (Koerner et al 2009). One such example of CTCF

mediation of imprinted gene expression is in the mammalian paternally imprinted *Igf2* cluster and its cognate ncRNA, *H19*. It has been shown that H19 is not necessary for the imprinted expression of the *Igf2* cluster. Rather, CTCF binds the unmethylated maternal allele and prevents the interaction of enhancers with the *H19* promoter, whereas CTCF is unable to bind the methylated paternal allele and the enhancers are able to interact with the *H19* promoter (Bell & Felsenfeld 2000).

Random monoallelic expression of sex-linked genes

In contrast, genes showing random monoallelic expression can be expressed from either the maternal or paternal chromosomes, in a stochastic manner. The best documented example of random monoallelic gene expression is in sex chromosome gene dosage compensation (SCGDC). The strategies of SCGDC are diverse in different organisms, however, in all known examples, compensation is targeted exclusively to the X chromosome of one sex to modulate gene expression solely in that sex (Lavrov & Kibanov 2007). In placental female mammals, a hallmark of X-chromosome inactivation (XCI) is the expression of numerous ncRNAs that regulate each other from the X-inactivation centre (Xic), a region of approximately 500 kb. The X chromosome that is to be silenced is randomly established during embryogenesis yet the silencing is stably maintained through many cell divisions. To date, at least seven distinct ncRNA genes have been identified to have a regulatory role in XCI (Lee 2009 and sources therein). The ncRNA first identified to be involved in this process was the X-inactive specific transcript (*Xist*), a long 17 kb ncRNA that is transcribed solely from the X chromosome that will be inactivated that accumulates and physically coats the silent X (Brockdorff et al 1992). *Xist* itself is regulated by

another ncRNA, *Tsix*, a transcript of approximately 40 kb that is the antisense counterpart that antagonizes *Xist* and is required for the establishment of the active X-chromosome. *Tsix* is in turn regulated by *Xite*, another long ncRNA transcript, as well as by two shorter transcripts that can have positive and negatively influence *Tsix* that arise from the bidirectional promoter activity at the *DXPas34* microsatellite repeat element. In addition, the 1.6 kb *RepA* transcript and several other very small (21-42 nt) RNAs are also implicated in XCI (Lee 2009 and sources therein). Both *Tsix* and *RepA* transcripts have been shown to bind PRC2 (Zhao et al 2008). To summarize a proposed model for the spread of silencing in an allele-specific manner along the entire X chromosome, the initial heterochromatin seeding of the silent X is believed to involve the recruitment of the PRC2 by *RepA* that deposits H3K27me3 at the 5' end of *Xist*. As *Xist* coats the entire chromosome, it is believed to then recruit and distribute PRC2 which deposits repressive marks along the entire length of the chromosome, whereas *Tsix* blocks these activities on the active X (Lee 2009). Other PTMs and histone variants distinguish the silent X chromosome from its active counterpart, including hypoacetylation of histones H3 and H4 and enrichment of macro-H2A.1 (Koerner et al 2009).

1.7.2 Mutually exclusive gene expression of mammalian odorant receptors and human trichromatic colour vision

Mammalian odorant receptors

The strict control of mutually exclusive gene expression is exemplified by the mammalian olfactory system which relies on the expression of a single gene from a large family of ≈1400 odorant receptor (OR) genes

that are located in tandem arrays across the genome. The effective expression of only one OR gene and the repression of all others is crucial for the system, as transcription of the unselected ORs would lead to sensory confusion. The exact mechanism of this process has not been elucidated but is known to have several levels of regulation.

It is known that the activation of a single OR gene is not associated with DNA rearrangements (Shykind et al 2004). *cis* elements located within the proximal promoter are required for the mutually exclusive expression of OR genes (Rothman et al 2005). The analysis of mutants that resulted in full length mRNA transcripts but truncated peptides, resulted in the activation of other OR genes, which suggested that the OR protein was required to initiate a regulatory feedback process that ensured the expression of a single OR gene (Serizawa et al 2003; Lewcock & Reed 2004; Shykind et al 2004). It was also observed that the episomal expression of an OR by different exogenous promoters lead to the silencing of the endogenous OR, reinforced the feedback loop hypothesis and also indicated that the coding sequence contains information for the regulation of mutually exclusive OR expression (Nguyen et al 2007).

The identification of a conserved, distant 124 bp enhancer sequence termed the *H* element, was found to regulate the expression of an OR gene cluster (Serizawa et al 2003; Lewcock & Reed 2004). It was hypothesized that chromatin-remodelling complexes, together with the transcriptional apparatus, assemble on the *H* element to activate one OR gene at a time by the physical association with only one OR gene (Serizawa et al 2004; Shykind et al 2004). Another group observed that the *H* sequence could interact in *trans* with OR promoters on

nonhomologous chromosomes (Lomvardas et al 2006). This lead to the hypothesis that this *H* element was a limiting enhancer that mediated the genome-wide regulation of OR gene expression by acting as a "pan-enhancer".

However, two groups showed that deletion of the *H* element abolished the expression of three proximal OR genes located within a cluster, but did not have an effect on OR genes located further away within the same cluster, nor on genes located outside of the cluster (Fuss et al 2007; Nishizumi et al 2007). This indicated that this single *H* element was not an enhancer for all OR genes in the genome and that multiple *cis*-acting elements reside within each cluster (Fuss et al 2007).

Recently, it was shown that the nonselected OR genes are silenced by heterochromatin formation mediated by H3K9me3 and H4K20me3, whereas the active OR is enriched in H3K4me3 (Magklara et al 2011). The heterochromatin-mediated silencing does not require OR expression, suggesting that heterochromatinization it is not the result of an OR-elicited feedback signal. The authors suggest that a limited enzymatic activity removes these repressive marks from a stochastically selected allele permitting its transcription (Magklara et al 2011).

Trichromatic colour vision in humans

In humans, the expression of a single photopigment in each cone cell is fundamental to trichromatic colour vision. The duplication and sequence divergence of a gene on the X chromosome resulted in what are now red and green cone pigment genes (Smallwood et al 2002). Human red and

green cone pigment genes lie in tandem in a head to tail arrangement. A LCR lies between 3.1 and 3.7 kb upstream of the red cone pigment TSS and is highly conserved between diverse mammals and is not duplicated in the human genome (Smallwood et al 2002 and sources therein). Deletion of the LCR abolishes the expression of both red and green pigments (Carroll et al 2010). There is dramatic variation in the red:green pigment ratios in normal colour vision in humans and it was proposed that differences in the nucleotide sequence in the promoter regions contributed to the difference in pigment ratios (Smallwood et al 2002). However it was later shown that the degree of variation was not associated with nucleotide differences in red and green promoters between individuals with different red:green ratios (McMahon et al 2004). Evidence also exists that both pigment genes are recognized by the same transcription factor (Deeb et al 2010).

A stochastic model of mutually exclusive pigment expression was proposed whereby the two pigment promoters compete to interact with the single LCR (Smallwood et al 2002). This was supported by evidence that switching between red and green pigment expression could occur after cell division, as single cells expressing a single pigment would result in clones that contained a mixture of red or green expressing cells (Deeb 2006). The authors suggest that the LCR-promoter interaction dissociates during DNA replication, and upon reestablishment (mediated by protein complexes) in daughter cells, occurs stochastically to generate a mixed population of clones.

The resultant clonal population showed a slight preference (2.7:1) to express red pigment, potentially because it is the pigment gene located

closest to the LCR, a distance of 3 kb versus 42 kb (Deeb 2006; McMahon et al 2008). However, calculations based on the linear genetic distance between the LCR and the two promoters, a ratio of 1000:1 (red:green) would be predicted and the authors suggested that some mechanism must exist for the observed ratio that is much nearer to equal than predicted (McMahon et al 2008). Although experimental data is lacking, the authors propose that this genomic region is organized into looped chromatin domains, as packing of chromatin loops would be a mechanism whereby the LCR could be brought into nearly equal physical proximity to the promoters of both genes (McMahon et al 2008).

Chromatin looping has been shown to play a critical role for the regulation of genes within clusters that resulted from gene duplication events (Sproul et al 2005). One such example is the regulation of the murine β-globin locus. This β-globin locus is composed of five genes spread along a 200 kb region that are not mutually exclusively expressed but show differential expression throughout development. The expression of these genes is controlled by a single LCR that is located approximately 40 kb from the most proximal globin gene. The 5' end of the cluster contains numerous DNA hypersensitive sites (DHS), and the 3' end also contains a DHS. DHS are sites that are hypersensitive to cleavage by nucleases as they are nucleosome depleted regions that are presumably TF binding sites. Using chromosome conformation capture (3C, described in section 6.1, Figure 6.1), the de Laat group showed that the active globin genes come in close physical proximity to the LCR, with the intervening chromatin forming a loop that contains the inactive globin genes. This interaction is mediated by two DHS sites that are separated by a linear distance of 130 kb. Additionally, this physical

clustering in the nucleus was solely observed in globin-expressing tissues, such as erythroid cells and not in non-expressing cells such as in brain tissue (Tolhuis et al 2002). Similar roles for chromatin looping in the regulation of duplicated genes has also been observed in homeobox (HOX) and major histocompatibility complex (MHC) gene families (Sproul et al 2005 and sources therein).

1.7.3 Mutually exclusive expression of antigenic variation in pathogenic organisms

The mutually exclusive expression of immunogenic antigens is a common survival strategy that is achieved in a number of different ways in various pathogenic organisms. The mechanisms involved in the mutually exclusive expression of *var* genes involves genetic and epigenetic mechanisms that were previously described in sections 1.5.2 and 1.5.3 respectively. However, the means by which mutually exclusive antigen expression is accomplished in different pathogens can be broadly classified into three categories: those that involve changes in the underlying DNA sequence, those that involved post-transcriptional mechanisms and those that involve transcriptional mechanisms all of which could involve layers of epigenetic regulation.

Mutually exclusive expression of antigens as a result of changes in DNA sequence

Large and small scale changes to DNA sequence can lead to the mutually exclusive expression of variant antigens. Large scale changes in DNA sequence can arise by various forms of DNA recombination, including nonhomologous recombination, inversions, deletions and conversions. One well characterized mechanism of mutually exclusive antigenic

variation is gene conversion, a mechanism found in several pathogenic organisms, but is typified by *Trypanosoma brucei*. Evasion of host immunity by *T. brucei* is mediated by the large (≈1000) variant surface glycoprotein (*vsg*) family of genes that are subtelomerically located. *vsg* transcripts are polycistronically transcribed in a mutually exclusive manner from numerous telomeric bloodstream form *vsg* expression sites (ES) (Pay 2005). The active ES is located in an extranucleolar body that contains the appropriate enzymatic activities for transcription and RNA processing necessary for *vsg* expression. *T. brucei* can undergo two different forms of VSG switching, referred to as transcriptional and recombination switching (Horn & McCulloch 2010). Both types of switching refer to the coordinated silencing and activation of different ESs from one telomere to another. However, transcriptional switching refers to the expression of pre-existing *vsg* genes, whereas recombination switching refers to the expression of a new, chimeric *vsg* that arose from frequent segmental gene conversions that occur in this parasite. These frequent recombination events play an important role in the generation of new *vsg* variants that can evade acquired host immunity (Marcello & Barry 2007). However, the mechanism used to co-ordinately silence and activate different ESs is currently unclear. It is known to be a rapid process that occurs stochastically at a low frequency and does not involve DNA rearrangements. An ISWI homolog (a chromatin remodelling protein), and DOT1B, an HMT responsible for the deposition of H3K76me3 (equivalent to H3K79me in yeast and mammals) were shown to be involved in the derepression of silent ESs (Hughes et al 2007; Figueiredo et al 2008).

Smaller scale differences on the order of a few or even single nucleotides can also lead to changes in gene expression. These types of mutations can occur during DNA replication from slipped-strand mispairing (Deitsch 2009). This type of regulation has been documented for the *trp* genes in *Treponema pallidum*, the causative agent of syphilis, whose gene products are predicted to be exposed surface antigens. This family of 12 paralogs contain homopolymeric guanosine (G) repeats of varying lengths (7-12 bp) located immediately upstream of their TSS within the promoters of some of these genes. The transcription of some of these *trp* operons with 8 or fewer Gs are transcribed, whereas the transcription of those with longer poly-G tracts was dramatically reduced (Giacani et al 2007).

Mutually exclusive expression of antigens by post-transcriptionall regulation

The *var2csa* gene in *P. falciparum* is a unique *var* variant that is associated with sequestration of the parasite by binding to the proteoglycan chondroitin sulphate (CSA) in the placenta of pregnant women (Fried & Duffy 1996). *var2csa* has a unique promoter type, *upsE*, that distinguishes it from other *var* gene promoters (Kraemer & Smith 2006). The *var2csa* gene is transcribed into an mRNA that contains an additional small upstream ORF (uORF) that is able to repress translation (Amulic et al 2009). The authors demonstrate that the peptide of this uORF is not required in the translational repression of the gene. The selection of cultured parasites to express *var2csa* suggests that an environmental cue is not necessary to induce expression of the second ORF, rather that it might happen stochastically within distinct subpopulations. It is hypothesized that the uORF is transcribed in men,

children and non-pregnant women and that the ribosome dissociates from the mRNA before reaching the *var2csa* ORF. However, the second ORF is only transcribed in pregnant women where the encoded PfEMP1 functions in the presence of a placenta. As other *var* genes, *var2csa* is also marked by H3K9me3 when transcriptionally silent (Lopez-Rubio et al 2009).

In contrast, antigenic variation in the parasite *Giardia lamblia* is also achieved by post-transcriptional regulation, however, not by a uORF, but via RNAi (Prucca et al 2008). *G. lamblia* expresses one variant-specific surface (VSP) protein from a repertoire of ≈ 190 genes. By nuclear run-on analysis, these authors observed that most *vsp* genes were simultaneously transcribed and their transcription was accompanied by a low abundance of antisense transcripts for the expressed VSP and an abundance of antisense transcriptions for the VSPs that are transcribed but not translated. Deletion of key components in the RNAi pathway resulted in the expression of multiple VSPs in individual parasites (Prucca et al 2008). VSP expression changes by spontaneous switches to different variants that occurs by an unknown mechanism, but is believed to depend on the chromatin state of VSG genes (Kulakova et al 2006).

Mutually exclusive expression of antigens by transcriptional mechanisms

The *var* genes in *P. falciparum* represent a family of antigens that show mutually exclusive expression that is controlled at the level of transcription. The mutually exclusive regulation of this gene family involves both genetic and epigenetic mechanisms that were previously described in sections 1.5.2 and 1.5.3 respectively.

1.8 *clag3.1, clag3.2 and eba-140* as models to study the epigenetic control of CVGE in *P. falciparum*

Much of the attention on deciphering the molecular mechanisms of CVGE have focused on *var* genes, in large part because *var* was the first gene family where variegated expression was detected and partly because of their major role in pathogenesis. However, the discovery that other small gene families involved in unrelated processes also show CVGE that also has an epigenetic basis, is suggestive that CVGE is not solely a phenomenon associated with *var* genes or antigenic variation and therefore, the epigenetic regulation of CVGE may be central to *P. falciparum* biology. The mutually exclusive expression of the *clag3* genes also presents an opportunity to investigate the mechanisms of mutually exclusive expression in *P. falciparum*. The *var* genes represent a peculiar family of genes; it is a large family where the majority of the members are repressed. Therefore, it is unknown whether the same mechanisms that regulate CVGE in other families, where the individual members are often expressed, will apply.

The genes *clag3.1*, *clag3.2* and *eba-140* are good models to study the epigenetic control of CVGE and mutually exclusive expression (in the case of *clag3.1* and *clag3.2*) in *P. falciparum* for several reasons. We have distinct subcloned populations of parasites that are genetically identical but have these genes either expressed or repressed. These parasite lines have not been genetically manipulated and thus we can compare their activation or repression as it occurs endogenously. By looking at these three genes, at different life stages and in parasite lines where the genes are differentially expressed, we can discriminate between the

mechanisms that are involved in silencing associated with life cycle progression, clonally variant and mutually exclusive gene expression (Figure 1.5). Collectively, this should give us a comprehensive understanding of the control of transcriptional variation in *P. falciparum*.

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

2. Objectives

The general objective of this thesis is to investigate the epigenetic regulation of clonally variant gene expression in *P. falciparum* using the genes *clag3.1*, *clag3.2* and *eba-140* as models. More specifically,

1. Identify the promoter regions of *clag3.1*, *clag3.2* and *eba-140,* and functionally characterise their activity in isogenic parasite lines that differentially express these three genes (Chapter 4).

2. Characterise the relevant histone modifications associated with the clonally variant expression and those associated with life cycle dependent silencing of *clag3.1*, *clag3.2* and *eba-140* using isogenic parasite lines that differentially express these genes (Chapter 4).

3. Characterise the level of transcriptional variation in *P. falciparum* on a genome-wide scale (Chapter 5).

4. Mimic the clonally variant gene expression of the *clag* gene family in an episomal system using *clag3.2* as a model (Chapter 5).

5. Investigate the mechanism of mutually exclusive gene expression in *P. falciparum* using *clag3.1* and *clag3.2* as models. This includes characterising the expression of a proximal ncRNA and optimising the initial steps of chromosome conformation capture experiments (Chapter 6).

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

3. Materials and Methods

3.1 *P. falciparum culture*

3.1.1 Media

P. falciparum complete medium is custom made (Invitrogen, 041 91762 A NIMR) and contains Albumax II and gentamycin (0.01 mg/ml). Complete medium is supplemented with L-glutamine (Invitrogen, 25030024) to a final concentration of 1 mM. Incomplete media (washing media) contains 10.43 g/l RPMI 1640 (Gibco ® 51800-019, with L-glutamine and without $NaHCO_3$), 5.95 g/l HEPES (free acid, Sigma H3375), pH adjusted to 7.4.

3.1.2 Basic maintenance of *P. falciparum* cultures

P. falciparum is cultured under sterile conditions according to standard procedure in media without human sera as described by Trager and Jensen (1976). To maintain sterile conditions, all laboratory prepared solutions are filter sterilised. Cultures are maintained at 3% hematocrit. Schizont parasitemias are not to exceed 8%. Media is changed every second day, but with high schizont parasitemias (≥ 3%), the media is changed daily. Cultures are grown in Falcon ® brand Petri dishes in Billups-Rothenberg modular hypoxia incubation chambers (MIC-101). Cultures are grown under low oxygen conditions, typically 94 %N_2, 5% CO_2, 3% O_2 at 37°C in incubators supplemented with 5% CO_2.

3.1.3 Parasite lines and invasion rates

The parasite lines used in this study and their corresponding invasion rates can be found in Table 3.1. Invasion rate refers to the number of

merozoites that successfully invade and infect erythrocytes upon schizont bursting and is used to calculate expected parasitemias in the next IEC. 3D7-A is a clonal parasite line originating from 3D7 that has been maintained in culture for many years (Rosario 1981). 10G and 1.2B are subclones of 3D7-A. 10Gluc and 10Gluc-derived subclones are described in section 3.6.

Table 3.1. List of parasite lines used in these studies and their corresponding invasion rates

Parasite line	Invasion Rate (IR)
3D7-A	7.5
10G	7.5
1.2B	7.5
10Gluc	≈4
10Gluc-derived subclones	≈4

3.1.4 Washing of erythrocytes

Total blood of the B^+ blood group is purchased from the Banc de Sang i Teixits (Generalitat de Catalunya, Departament de Salut). Full blood is washed to remove plasma and buffy coat. Blood is centrifuged for 5 min at 1217 x *g* at room temperature. The supernatant, containing plasma and buffy coat, is removed by aspiration and the erythrocytes are resuspended in washing media. The erythrocytes are centrifuged again for 10 min at 1217 x *g* at room temperature. The supernatant is removed and the erythrocytes are resuspended in washing media, and centrifuged at 1217 x *g*. During this final centrifugation step, the rate of rotor deceleration is lowered. The supernatant is removed and the washed erythrocytes are resuspended in one pellet volume of complete media and stored at 4^oC.

3.1.5 Preparation of Giemsa-stained thin smears

Giemsa specifically stains for phosphate groups of DNA, preferentially at AT rich regions and permits for the morphological discrimination of the different stages of the asexual IEC via microscopy. The erythrocyte is stained pink, the chromatin content of the nuclei are stained dark purple, the cytoplasm is stained blue, and crystals of hemozoin in mature parasites appear yellow-brown. Approximately 0.2 ml of culture are centrifuged for 30 s at 500 x *g*. The supernatant is removed and approximately 3 µL of erythrocytes are used for the smear. The smear is air dried and briefly fixed with methanol. Excess methanol is decanted and while the slide is still wet, slides are stained with freshly prepared Giemsa (Merck 109201), diluted 1:10 in Sorenson's buffer (3 g/l Na_2HPO_4, 0.6 g/l KH_2PO_4, pH adjusted to 7.2, stored at room temperature) for approximately 10 min, and then rinsed abundantly with water. Slides are air or tissue dried.

3.1.6 Creation of *P. falciparum* stocks

Only ring stages survive freezing and thawing. Accordingly, the culture must contain rings, ideally 5% or higher, for stabilate preparation. Cultures are centrifuged for 5 min at 405 x *g*. The supernatant is removed by aspiration and the erythrocyte pellet is slowly resuspended by the gradual addition (drop by drop) of two pellet volumes of freezing solution (for 50 ml, 0.324 g NaCl, 1.512 g D-sorbitol [Sigma S3889], 14 ml glycerol [Sigma], 36 ml H_2O, stored at -20°C). Cells are thoroughly mixed, incubated at room temperature for 5 min and frozen directly in liquid N_2.

3.1.7 Thawing of *P. falciparum* stocks

Parasite stocks are removed from liquid N_2 and warmed in your hand. The contents are transferred to a tube and the volume is estimated. All thawing solutions are added gradually, drop by drop, with constant swirling. To the stabilate, 0.2 ml per ml of stabilate of 12% NaCl (w/v) in PBS is added. This is followed by the addition of 10 ml per ml of stabiliate of 1.6% NaCl (w/v) in PBS. To this, 10 ml per ml of stabilate of 0.9% (w/v) NaCl + 0.2% (w/v) glucose in PBS is added. The sample is centrifuged for 4 min at 259 x *g*, the supernatant removed by aspiration and the infected erythrocytes are washed with the gradual addition of 10 ml washing medium. This is followed by centrifugation at 259 x *g* for 4 min, the supernatant is removed by aspiration and the erythrocyte pellet is slowly resuspended in complete medium. The appropriate amount of washed erythrocytes are added to establish the culture.

3.1.8 Saponin lysis

Saponin lysis is used to liberate parasites from the erythrocyte whilst maintaining the parasitophorous vacuole intact. Parasite cultures are centrifuged for 6 min at 461 x *g*. The supernatant is removed by aspiration and the RBC pellet is resuspended in 2 pellet volumes of 0.15% cold saponin (prepared in PBS and filter sterilised) and maintained on ice for 5 min. This is followed by centrifugation at 1411 x *g* for 6 min at room temperature and the supernatant is removed by aspiration. The parasite pellet is washed twice with cold PBS with centrifugations of 6 min at 1411 x *g*. The resultant parasites are ready to be used in various downstream applications.

3.1.9 Determining parasitemias

3.1.9.1 Microscopy

Parasite growth is monitored by the microscopic examination of Giemsa-stained thin smears, using an Olympus BX41 ® optical microscope.

3.1.9.2 Flow cytometry

Parasitemias are also determined via flow cytometry, using either Beckman Coulter FC500-MPL or Epics XL flow cytometers to read either 96 well plates (200 µl volumes), or volumes of 1 ml respectively. The dye used for parasite detection is the green-fluorescent nucleic acid stain SYTO®11 (Invitrogen, Molecular Probes, S7573). SYTO®11-stained DNA absorbs at 508 nm and emits at 527 nm. Stocks of STYO®11 are 5 mM prepared in DMSO which are kept in the dark at 4°C. For analysis, parasite cultures are diluted 1/200 in room temperature PBS, to which 1 µL of STYO®11 stock solution is added and the sample is mixed well prior to cytometry. Data is analysed with Summit software, v4.3.

3.2 *P. falciparum* synchronisation

3.2.1 Ring stage synchronisation by sorbitol lysis

Sorbitol is used to synchronise cultures to the ring stage, as sorbitol lyses infected erythrocytes containing parasites >20 h. Cultures are centrifuged at 405 x *g* for 5 min. The supernatant is removed by aspiration and the erythrocyte pellet is resuspended in 7 pellet volumes of 5% sorbitol (D-sorbitol Sigma, S3889). Samples are incubated for 7 min at 37°C, and subsequently centrifuged for 4 min at 405 x *g*. The supernatant is discarded and the infected erythrocytes are washed with 10 ml of washing medium. The samples are centrifuged for 5 min at 405 x *g* and complete media is added to re-establish the culture.

3.2.2 Magnetic purification of late forms

Late *P. falciparum* forms contain an abundance of hemozoin. Hemozoin is the by-product of haemoglobin degradation and contains iron in the Fe^{3+} state. Consequently, schizonts can be purified by depletion using magnetic bead columns. CS columns (Miltenyi Biotech, 130-041-305) are placed in the VarioMACS^{TM} Separator (Miltenyi Biotech). To the column, a 21 gauge needle (Becton Dickinson, 304432) is affixed to restrict the flow rate. Parasite cultures are pooled and 15 ml of culture are applied to the magnetic column and allowed to pass through the column by gravity flow. Each 15 ml application of culture at 3% haematocrit, is washed with 30 ml of washing medium. To elute the bound schizonts, the column is removed from the VarioMACS^{TM} Separator, the flow restrictor is removed and the schizonts are eluted with washing medium. After the last elution, the remaining schizonts are flushed out of the column with a syringe. The schizonts are centrifuged for 6 min at 583 x *g* and the appropriate amount of washed erythrocytes

are added to re-establish the culture. Columns are reused by cleaning with Milli-Q H_2O, NaOH, Milli-Q H_2O again, ethanol, and are stored at 4^oC in ethanol. The level of purity obtained is typically >80%.

3.2.3 70% Percoll™ density purification of late forms

Parasitized erythrocytes are less dense than uninfected erythrocytes and thus can be purified on density gradients. Parasites cultures are resuspended and centrifuged for 5 min at 405 x *g*. Most of the supernatant is removed by aspiration and with the remaining supernatant, the erythrocyte pellet is resuspended and slowly applied on top of a tube containing 10 ml of 70% Percoll™ (in actuality, the Percoll™ solution is 63%. For 142.86 ml, 10 ml 10X PBS [2 tablets Sigma P4417-50TAB in 40 ml H_2O], 90 ml Percoll ™ [GE Healthcare, 17-0891-02], 42.86 ml washing media, store at 4^oC). The samples are centrifuged for 11 min at 1411 x *g* and the rate of rotor deceleration is lowered. The dark uppermost layer contains purified schizonts. The schizonts are removed, washed with washing medium and centrifuged for 6 min at 583 x *g*. The supernatant is discarded and the schizonts are added to an appropriate amount of washed erythrocytes to re-establish the culture. The level of purity obtained is typically >90%.

3.2.4 Tight synchronisation

The analysis of gene expression in *P. falciparum* is complicated by the periodic nature of gene expression associated with life cycle progression. As a result, small differences in the stage of parasites between samples can lead to artefactual differences in gene expression levels. Tight synchronization allows us to i) separate the phenomenon of life cycle dependent silencing from clonal variant gene expression ii) compare

different parasite strains. Tight synchronisation refers to the synchronisation of a parasite culture first to the ring stage, followed by a synchronisation to the schizont stage (either by magnetic separation or density gradient) either during the same IEC or the following IEC and typically when the ratio of schizonts to rings is at least 1:1. A culture is re-established with these schizonts and typically 5 h later, the culture is synchronised to the ring stage. This ensures that the resultant parasites are all within a 5 h window of each other in terms of progression along the IEC.

3.2.5 Estimation of *P. falciparum* culture age

To ensure that tightly synchronised schizont samples are indeed at a similar stage, the following calculation is used to estimate the age of the culture. The age of tightly synchronised parasite cultures is defined as the percentage of bursted schizonts. The number of rings and schizonts is determined by microscopy and with the invasion rate of the parasite, the following calculation is applied: % bursted schizonts = {(#rings/invasion rate) / [#late + (#rings/invasion rate)]} *100

3.3 Nucleic acid isolation

3.3.1 Purification of genomic DNA, phenol-chloroform extraction, ethanol precipitation and quantification of nucleic acids

Cultures are centrifuged at 405 x *g* for 5 min in polypropylene tubes. The supernatant is removed by aspiration and the erythrocyte pellet is resuspended in 4 pellet volumes of Buffer A (50 mM sodium acetate, pH 5.2, 100 mM NaCl, 1 mM EDTA pH 8) at room temperature. Sample is mixed well and the erythrocytes and parasites are lysed by the addition of 1 pellet volume of 18% SDS. The sample is mixed by inversion and left at room temperature for 2 min. An equal volume of room temperature phenol/chloroform/isoamyl alcohol pH 8 (25:24:1, Amresco, 0883) is added and the sample is mixed by inversion and centrifuged for 10 min at 720 x *g* at room temperature. The upper phase is recovered and aliquoted into 1.5 ml microcentrifuge tubes, to which 1/10 volumes of 3 M sodium acetate pH 5.2 and 2.5 volumes of cold ethanol are added. The samples are mixed well and the DNA is left to precipitate overnight at -20°C . Microcentrifuge tubes are centrifuged for 15 min at 16 100 x *g* at 4°C , the supernatant is discarded and the pellet is resuspended in 500 µL of TE. One volume of phenol/chloroform/isoamyl alcohol is added as before, the tubes are mixed well by inversion and are centrifuged in a microcentrifuge for 5 min at room temperature at 16 100 x *g*. The supernatant is transferred to a new microcentrifuge tube and to it, 1/10 volume of 3 M sodium acetate pH 5.2 and 2.5 volumes of cold ethanol are added. The samples are mixed well and the DNA is left to precipitate overnight at -20°C. The samples are centrifuged for 30 min at 16 100 x *g* at 4°C, the supernatant discarded and the DNA pellet is washed with 600 µL of cold 80% ethanol. The samples are centrifuged at 16 100 x *g* for 5

min at 4 ^{o}C, the supernatant is discarded and the DNA pellet is air-dried at room temperature for 5 min. DNA is passively redissolved by a 10 min incubation at room temperature in either TE or Milli-Q H_2O. The DNA is ready for use in various downstream applications.

Genomic DNA extracted in this manner contains a large amount of RNA. If RNA-free genomic DNA is required, a sample of genomic DNA is treated with RNaseA (Fermentas, EN0531) for 30 min at 37 ^{o}C. The DNA is subsequently purified via a phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation as described above. When the DNA yield is expected to be low, the ethanol precipitation is supplemented with 40 µg of glycogen (Sigma, G0885).

Small volumes (1-2 µL) of nucleic acids are quantified directly on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific), using 40 µg/ml per $1AU_{260nm}$ and 50 µg/ml per $1AU_{260nm}$ conversion factors for RNA and dsDNA respectively.

3.3.2 Total RNA extraction

Parasites at the appropriate stage are resuspended and centrifuged for 6 min at 461 x *g*. The supernatant is removed by aspiration and the erythrocyte pellet is resuspended in 20 volumes of TRIzol® Reagent (Invitrogen, 15596) prewarmed to 37°C. The sample is thoroughly homogenized by extensive pipetting and incubated at 37°C for 5 min and then stored at -80°C. To extract RNA, TRIzol extracts are passively thawed to room temperature and 0.2X volumes of chloroform per ml of TRIzol used are added. The samples are mixed well by inversion and subsequently centrifuged for 30 min at 3007 x *g* at room temperature. At

least 0.5X the initial TRIzol volume of the aqueous phase is recovered, and to it, 5/6 of the volume of the aqueous phase of isopropanol is added. The sample is aliquoted into microcentrifuge tubes and allowed to precipitate overnight at -20°C. The following day, samples are centrifuged for 30 min at 16 x 100 *g* at 4°C. The supernatant is discarded and the RNA pellet is washed with 0.5X per ml of initial TRIzol volume used of 75% cold ethanol. Samples are centrifuged at 4°C for 5 min at 16 100 x *g*, the supernatant is discarded and the pellet is air dried at room temperature for 10 min. The RNA pellet is resuspended in an appropriate amount of RNase-free H_2O and incubated at 63°C for 10 min to facilitate resuspension of the pellet. RNA is stored at -80°C. Prior to cDNA synthesis, the integrity of the RNA is determined via agarose gel electrophoresis and quantified as described in section 3.3.1.

3.3.3 cDNA synthesis

3.3.3.1 Dnase I treatment and RNA purification

Total RNA containing traces of genomic DNA, is treated with Dnase I in solution using the RNase-Free DNase Set (Qiagen, 7954) as per the manufacturer's instructions, but with a 30 min incubation at room temperature and the reaction is not stopped by the addition of EDTA. RNA is purified on spin columns (RNeasy minElute Cleanup kit, Qiagen, 74204) and resuspended in 10 µL of nuclease-free H_2O. RNA is quantified as described in section 3.3.1.

3.3.3.2 Reverse transcription and cDNA purification

A minimum of 0.5 µg of DNA-free, total RNA is reverse transcribed using the Reverse Transcription System (Promga-A3500) according to the manufacturer's instructions but with the following modifications. Samples are primed either with oligo$(dT)_{15}$, random primers, or with a

mixture of both (provided with the kit), depending on the downstream application. No reverse transcriptase controls are always included. Once the reactions have been set up, the samples are incubated at room temperature for 10 min, followed by a 1 h incubation at 42°C. The samples are then kept on ice for 5 min and diluted appropriately with nuclease-free H_2O. If the cDNA needs to be quantified, unincorporated nucleotides are removed from the cDNA by spin column purification and the cDNA is quantified as described in section 3.3.1.

3.4 General Molecular Biology Techniques

3.4.1 Standard PCR

PCR is performed in an Eppendorf Mastercycler gradient thermal cycler using recombinant Taq polymerase (Invitrogen, 10342-020). Standard cycling conditions are a denaturation step of 94°C 2 min followed by 32 cycles of a 30 s denaturation step at 94°C, a 1 min annealing step at the appropriate primer annealing temperature (typically 50-60°C), and 1 min per kb of amplified product extension step at 68°C. The 32 cycles are followed by a final elongation step of 5 min at 68°C.

3.4.2 Difficult PCR

PCR amplification from difficult templates, such as the amplification of large AT rich fragments, or fragments with long homopolymeric repeats, are performed using TaKaRa LA Taq™ polymerase (Takara Bio Inc., RR002A). Thermal cycler conditions are a denaturation step of 92°C 1 min followed by 30 cycles of a 20 s denaturation step at 92°C, a 1 min annealing step at the appropriate primer annealing temperature (typically 50-60°C), and 1 min per kb of amplified product extension step at 60°C. The 30 cycles are followed by a final elongation step of 5 min at 68°C.

3.4.3 Quantitative Real-Time PCR (qPCR)

qPCR was performed using Power SYBR Green master mix (Applied Biosystems, 4368702) in 96 well MicroAmp ® Fast Optical PCR plates (Applied Biosystems, 4346906) in a StepOnePlus™ Real Time PCR System (Applied Biosystems). Plates were sealed with MicroAmp® Optical Adhesive film (Applied Biosystems, 4311971) and final reaction volumes are 20 µL. Prior to placement in the thermal cycler, samples are mixed in

an Eppendorf Mixmate for 30 sec at 1000 rpm and centrifuged for 1 min at 402 x *g*. Thermal cycler conditions are 95°C 10 min, 40 cycles of 95°C 15 sec (denaturation), 57°C 30 sec (annealing) , 60°C 30 sec (extension) and melt curve analysis consisting of 95 °C 15 sec, 60°C 1 min with an increment of 0.3°C/min up to 95°C, 95°C 15 sec. Melt curves were analysed for the presence of multiple peaks or the presence of primer dimers. Fluorescence measurements are taken during the extension phase. Data was analyzed using delta Ct or relative standard curve methods. Primers used for qPCR analyses are listed in Table 3.2.

Table 3.2. List of primers used in qPCR analysis

Name in text	Primer name	PlasmoDB ID	Sequence 5' to 3'	Final Concentration (nM)
seryl tRNA synthetase	PF07_0073(seryl)_F	PF07_0073	aagtagcaggtcatcgtggtt	500
	PF07_0073(seryl)_R		ttcggcacattcttccataa	500
luciferase or luc	luc_qPCR_F	-	aactgcctgcgtcagattctc	200
	luc_qPCR_R		cggaatgatttgattgccaaa	200
bsd	bsdF_qPCR	-	tctctgaagactacagcgtc	200
	bsdR_qPCR		acgatacaagtcaggttgcc	200
lucA	lucA_qPCR_F	-	ccggcgccattctatcct	200
	lucA_qPCR_R		tcttcatagccttatgcagttgct	200
TR1 P2	3TR001_2F	PF03TR001	tgcaataattatctgcccattca	200
	3TR001_2R		tgatcccacacccgaagact	200
TR2 PA	3TR002_AF	PF03TR002	cttatgcgcaaattttattttaag	500
	3TR002_AR		ataaccgtatttctatgtttgac	1000
TR2 P2	3TR002_2F	PF03TR002	gacacctaaaatgttttgtgatt	200
	3TR002_2R		tttcttttggtgttctaatgaag	200
TR2 P3	3TR002_3F	PF03TR002	attatcacctgatatatataccca	200
	3TR002_3R		tcatgtttctttttaatcatcttc	200
TR2 PC	3TR002_Cf	PF03TR002	gtgcattaattataattctttttga	200
	3TR002_Cr		aatattttgatatattgtcttttcg	200

3.4.4 DNA sequencing

Plasmids were sequenced via single primer extension by Macrogen Inc. to confirm the presence of the correct inserted sequence, proper orientation, and to ensure no additional mutations have been introduced during PCR amplification. The 5' flanking regions of *clag3.1* and *clag3.2* PCR amplified from 10G and 1.2B were also sequenced in this fashion, and the primers used are listed in Table 3.3. Due to the length and the high number of low complexity regions, the sequencing of the *clag3.1* 5'

flanking region could only be achieved by sequencing with a number of different primers.

Table 3.3. Primers used to sequence *clag3.1* and *clag3.2* from 10G and 1.2B parasites. The primers marked with asterisks were used to amplify the gene by PCR for the sequencing

	Primer Name	Sequence 5' to 3'
clag3.1	prom3.1F1kbAsp*	tggtgttggtaccataatgcactcattaataattttaaaccgtt
	3.1_qRT_P3F	cataaacataagagggtgattta
	3.1seq1fwd	gaagaaaaaagatgatatttaata
	3.1seq3rev	ttttctttttttacagtttattctc
	3.1_qRT_P2r	gaagtgagataaaaatacatattc
	3.1_3.2_ qRT_P5r*	catggattttaattgttcaatattg
	3.1seq3fwd	gagaataaactgtaaaaaagaaaa
	3.1seq2rev	attatgcacttaattcttgttgac
clag3.2	Prom3.2fAsp*	tggtgttggtacctagaatatatttatttatagtcaaaatttgc
	3.2seq2fwd	ttcacggttacgtttctgttttt
	3.1_3.2_ qRT_P5r*	catggattttaattgttcaatattg
	3.2seq1rev	tgatcggagtcttaaaaaatgaa

3.5 Luciferase assays in transient transfections

3.5.1 Plasmid preparation

The plasmids used in the transient transfection assays are described in the experimental procedures of Article 1. A mixture of 50 µg of test plasmid (and 10 µg of control plasmid if required) in a final volume of 270 µl is precipitated with 750 µl of cold absolute ethanol and 30 µl of 3M sodium acetate (pH 5.2) overnight at 20°C. Samples are centrifuged in a microcentrifuge precooled to 4 ^{o}C for 30 min at 16 100 x *g*. All subsequent steps are carried out under sterile conditions. The supernatant is removed and the DNA pellet is washed with 900 µl of cold 70% sterile ethanol. Samples are centrifuged at 4°C for 10 min and supernatant is removed. The samples are briefly centrifuged again to remove the residual ethanol and the pellet is air dried for 8 min. DNA is passively resuspended with gentle tapping in 22.5 µl sterile TE at room temperature for 10 min.

3.5.2 Erythrocyte preparation

200 µl of erythrocytes are needed per transfection. The appropriate volume of washed erythrocytes is centrifuged for 5 min at 583 x *g*. The supernatant is removed by aspiration and 4 pellet volumes of cold cytomix (120 mM KCl, 0.15 mM $CaCl_2$, 2 mM EGTA, 5 mM $MgCl_2$, 10 mM K_2HPO_4 pH 7.6, 25 mM HEPES pH 7.6, filter sterilised). Untransfected erythrocytes are included in every experiment as a control. These controls are prepared in the same way, but are resuspended in complete medium instead of cytomix. Samples are thoroughly mixed by pipetting and 1 ml of this mixture is aliquoted into 15 ml tubes and centrifuged at room temperature for 3 min at 405 x *g*. The supernatant is removed by aspiration.

3.5.3 Electroporation of erythrocytes

To introduce the plasmids into the erythrocytes, under sterile conditions, 277.5 µl of cytomix is added to the corresponding DNA and this 300 µl is added to 200 µl of erythrocytes in cytomix. This is mixed, transferred to a 0.2 cm gap electroporation cuvette (Gene Pulser ® cuvette, BioRad, 165-2086) and immediately electroporated in a GenePulser X Cell electroporator (Bio-Rad) at 310 V, ∞ ohmns, 950 µF and high capacitance. Time constants are expected to be between 7 and 15. Electroporated material is transferred to a 15 ml tube with a plastic Pasteur pipette and cuvettes are rinsed twice with complete media. Samples are centrifuged for 3 min at 405 x *g*, the supernatant is removed by aspiration and plasmid-loaded erythrocytes are resuspended to 5 ml with complete medium.

3.5.4 Parasite preparation

Percoll™-purified parasites (as described in section 3.2.3) are resuspended in 1 ml complete medium. A 1:100 dilution of parasites is prepared, 10 µl of this dilution are mixed with 10 µl of Trypan blue and parasites are counted with a Countess® cell counter (Invitrogen) and the required volume of parasites to add to each culture dish is determined using the following formula:

$$V\ (\mu l) = 10^7 \times \frac{\text{Vol. dish (in ml)} \times \text{Parasitemia wanted (in \%)} \times \text{Hematocrit (in \%)}}{\text{Reading from countess}}$$

For these experiments, 10 ml plates at 0.8% parasitemia and 2% haematocrit are required. Parasites are cultured for 24, 48 or 50 h (depending on the desired stage and on the promoter).

3.5.5 Luciferase assay

Cultures are centrifuged for 5 min at 405 x *g* and smears are taken. The supernatant is removed by aspiration and the parasites are released by saponin lysis as described in section 3.1.8. The supernatant is removed by aspiration and the parasites are resuspended in 900 µl PBS and transferred to a microcentrifuge tube. Parasites are centrifuged in a microcentrifuge precooled to 4°C for 3 min at 1500 x *g*. The supernatant is removed and parasites are lysed with 85 µl of 1X *Renilla* lysis buffer from the *Renilla* luciferase assay system (Promega, E2810) for 15 min at room temperature. Firefly reagent (Promega, E1500) (and if necessary, *Renilla* reagent, Promega, E2810) are passively warmed to room temperature and maintained in a water bath at room temperature until use. Parasites are centrifuged in a microcentrifuge for 1 min at 16 100 *g* at 4°C and the supernatant is transferred to a new tube. Gloves are removed to prevent static and luminescence is measure in a Lumat LB 9507 luminometer for 10 s by adding 20 µl of sample to 100 µl of either firefly or *Renilla* assay reagent.

3.6 Generation of 10Gluc transgenic parasite line

3.6.1 HLHprom3.2bsd plasmid

The HLHprom3.2bsdR plasmid was prepared by Núria Rovira-Graells. This plasmid was created by the insertion of the blasticidin cassette and Rep20 sequence from the plasmid pHBupsC^{R} into 3.2-LH-1371. 3.2-LH-1371 is described in Article 1, experimental procedures. This fragment was digested out of pHBupsC^{R} by *Sma*I/*Bgl*II and was subcloned into *Sma*I/*BamH*1 digested 3.2-LH-1371.

3.6.2 Transfection of 10G with HLHprom3.2bsdR

The stable 10Gluc parasite line was created by Núria Rovira-Graells by the transfection of 100 µg of HLHprom3.2bsdR into 200 µl RBC of ring infected 10G parasites (≈ 7% parasitemia). Electroporation conditions were as described in section 3.5.3. Approximately 20 h after transfection, blasticidin was added to the transfected culture to a final concentration of 2.5 µg/ml. To prevent plasmid loss, this parasite line and its subclones are continually cultured with 2.5 µg/ml of blasticidin. Media was changed three times a week and the culture diluted in half once a week until parasites were detected by microscopy, approximately 4 weeks post transfection.

3.6.3 Subcloning of 10Gluc by limiting dilution

It is ring stages that are subcloned and the cycle prior to subcloning, cultures are gently shaken (Minitron, 40 rpm, 37°C) to reduce the number of double-infected erythrocytes. Parasitemias were determined via flow cytometry (as described in section 3.1.9.2) and cultures were diluted to 0.4 parasites/well (experimental wells) with uninfected erythrocytes in a 96-well plate. Control wells containing 10000 and 100

parasites /well are also included. The culture media was changed in the experimental wells every second day and these wells were diluted in half once a week. Control wells were checked regularly via microscopy. Experimental wells were initially screened for positives via microscopy on day 16, followed by a second screen one week later.

3.6.4 Luciferase assays on stable transfectants

The initial luciferase assay screen on 10Gluc and 10Gluc-derived subclones at the schizont stage was performed as described in section 3.5.5, however, cultures were maintained in 24-well plates (volume of 1 ml at 2% haematocrit per well) and ring parasitemias were determined by flow cytometry as described in section 3.1.9.2. Parasites were resuspended in 50 µl of *Renilla* lysis buffer and luciferase readings were normalised by ring parasitemias.

3.7 Quantitative Chromatin Immunoprecipitation (qChIP)

3.7.1 Fixed cell immunoflorescence (IFA)

Prior to performing qChIP, the antibodies were tested for cross-reactivity with *P. falciparum* samples via fixed cell IFA. The antibodies tested are listed in Table 3.4. Approximately 200 µl of culture are centrifuged for 30 s at 500 x *g*. The supernatant is removed and the erythrocyte pellet is resuspended with 20 µl PBS and approximately, 4 µl are used per smear. Dried smears are ready for fixation.

Table 3.4. List of antibodies used in qChIP analysis and their corresponding host species, isotype and immunogen information

Antibody	Company/Reference Number	Host species/Isotype	Immunogen
anti-H3	Abcam / ab1791	rabbit polyclonal / IgG	synthetic peptide derived from within residues 100 to the C-terminus of human histone H3
anti-H3K9me3	Millipore / 07-442	rabbit polyclonal / IgG	2X-branched synthetic peptide of human histone H3 and trimethylated on K9
anti-H3K9ac	Millipore / 07-352	rabbit polyclonal / IgG	synthetic peptide corresponding to amino acids 4-14 of yeast histone H3 with a C-terminal cysteine added for conjugation purposes and acetylated on K9
anti-H3K4me3	Millipore / 04-745 clone MC315	mouse monoclonal / IgG	synthetic peptide derived from human histone H3 trimethylated on K4
anti-H4ac	Millipore / 06-866	rabbit polyclonal / IgG	peptide corresponding to a.a. 2-19 of *Tetrahymena* histone H4, acetylated on lysines 5, 8, 12 and 16

3.7.1.1 Acetone-methanol fixation

The smear is fixed in 45 ml of freshly prepared 90% acetone-10% methanol for 2 min on ice. The acetone-methanol solution is not pre-cooled. The slide is dried vertically and hydrophobic circles are drawn using a hydrophobic marker (Dako, S2002). The slide is washed in PBS for 10 min.

3.7.1.2 Formaldehyde fixation

Hydrophobic smears are drawn on the smear. The smear is fixed in 1% formaldehyde in PBS for 5 min and rinsed twice in PBS. The cells are permeabilised with 0.1% Triton X-100 in PBS for 5 min and subsequently rinsed twice in PBS.

3.7.1.3 Antibody preparation and microscopy

Primary antibodies are diluted 1:100 in filtered 0.75% BSA in PBS. To avoid over-drying, the slide is dipped in 0.75% BSA in PBS and the excess liquid is dried with a tissue. Once the hydrophobic circles are dried, diluted antibody is applied to each spot. The primary antibodies are incubated at room temperature for 1 h at room temperature or overnight at 4°C in a humid chamber. The slides are washed 3 times in PBS for 5 min each. The secondary antibody is diluted in 0.75% BSA in PBS + DAPI (Sigma, D9542, 1:500 from a 0.5 mg/ml stock). The secondary antibody used was AlexFluor 555 goat anti-rabbit IgG (Invitrogen, Molecular Probles) which absorbs at 555 nm and emits at 565 nm. The slide is prepared and incubated as described for the primary antibody but protected from light. The slides are washed 3 times in PBS for 5 min each. Excess liquid is dried and to each circle, Mowiol 4-88 mounting media (Merck, 475904) prepared as per the manufacturer's instructions, is applied and spots are covered with a coverslip. The slides are kept at 4°C for at least a few hours before microscopy with a Nikon E600 fluorescence microscope and images were obtained by analySIS image acquisition software.

3.7.2 Harvesting of parasite material

Parasites synchronised to the appropriate stage are released by saponin lysis (as previously described in section 3.1.8) and resuspended in room temperature PBS to a concentration of approximately $1x10^5$ parasites/µl.

3.7.3 Formaldehyde cross-linking, homogenization and chromatin sonication

Parasites are cross-linked with a final formaldehyde (Sigma, F8775) concentration of 1% and incubated in a water bath at 37^oC for 5 min with frequent agitation. Formaldehyde is quenched by the addition of glycine to a final concentration of 125 mM and left on ice for 5 min with occasional agitation. Samples are centrifuged for 5 min at 461 x *g*, the supernatant removed by aspiration and pellets are washed 3 times with 10 ml of cold PBS. The cross-linked parasite pellets are resuspended in cold lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA pH 8, 0.1 mM EGTA pH 8, 1 mM DTT, 1X complete EDTA-free protease inhibitor cocktail [Roche, 11873580001]) and left on ice for 30 min. Sample are maintained on ice from this point forward. Samples are transferred to a precooled glass dounce homogenizer (Kimble-Kontes 885300-0002, 2 ml capacity), NP-40 (Roche 1133247300) is added to a final concentration of 0.25% and samples are homogenized with 200 strokes with tight pestle “B” (clearance, 0.0005-0.0025 inches) to release cross-linked nuclei. Homogenized samples are pooled, transferred to microcentrifuge tubes and cross-linked nuclei are pelleted by centrifugation in a precooled microcentrifuge for 10 min at 16 100 x *g*. Supernatants are discarded and pellets of cross-linked nuclei are resuspended to a concentration of approximately $1.38x10^5$ parasites/µl in cold SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1)

and split into 300 µl aliquots in microcentrifuge tubes. Samples are sonicated in a Diagenode Bioruptor ® sonicating water bath at 4°C supplemented with ice, on high power using the following parameters: 20 sec ON, 2 min OFF, 20 sec ON, 2 min OFF, 30 sec ON. Samples are centrifuged in a precooled microcentrifuge for 5 min at 16 100 x *g*. Chromatin, now in the aqueous phase, is pooled and split into 100 µl single-use aliquots in microcentrifuge tubes and stored at -80°C.

3.7.4 Confirmation of fragmentation size and DNA quantification

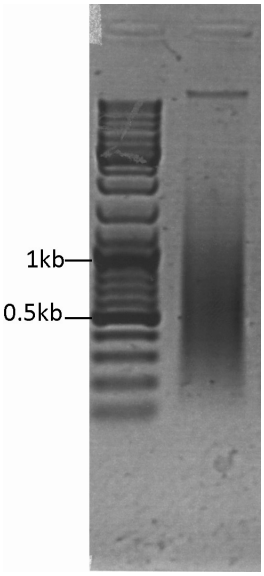


Figure 3.1 Chromatin fragmentation by sonication. Representative chromatin smear centred around 500 bp suitable for qChIP analysis.

To ensure that the sonicated chromatin fragments are of the appropriate size and to determine the concentration of DNA in the chromatin sample, 30 µl of chromatin are analyzed by agarose gel electrophoresis and spectrometry. To reverse cross-links, NaCl is added to a final concentration of 0.22M and Milli-Q H_2O is added to the chromatin for a final volume of 900 µl with Milli-Q H_2O and incubated for at least 4 h at 65°C. RNA is degraded by the addition of 50 µg of RNaseA (Fermentas, EN0531) and samples are incubated at 37°C for 30 min. Subsequently, proteins are digested by the addition of 50 µg of proteinase K (Sigma, P2308), 10 mM EDTA pH 8, 40 mM Tris HCl pH 7.4 and incubated at 45°C for 2 h. DNA is extracted by a phenol-chloroform extraction followed by an ethanol precipitation (as previously described 3.3.1, except the samples are precipitated at -80°C for 1 h) and resuspended in 10 µl of Milli-Q H_2O.

The average size of the DNA fragments is determined by running 7 µl on a 1% agarose-EtBr gel at 90 V, 61 mA and 5 W for 45 min – 1 h. The appropriate fragmentation size for proceeding with the ChIP protocol is a smear centred around 500 bp. A representative example of appropriate chromatin fragmentation is shown in Figure 3.1. DNA concentration is quantified as described in 3.3.1.

3.7.5 Preclearing and immunoprecipitation

Frozen chromatin aliquots are thawed on ice. Approximately 1 µg of DNA is used per ChIP and no antibody control, and approximately 0.2 µg of DNA will used for the INPUT sample. All ChIP samples are to be precleared at the same time in the same tube and aliquoted for separate ChIPs after preclearing. Consequently, the appropriate amount of chromatin is diluted in cold dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA pH 8, 16.7 mM Tris HCl pH 8.1, 167 mM NaCl, 1X protease inhibitor cocktail) to a concentration of 0.002 µg DNA/µl . Samples are precleared by the addition of 1/12 volume of the total diluted chromatin volume of 50% protein G agarose/salmon sperm DNA (Millipore 16-201, i.e. 3.6mL diluted chromatin is precleared with 300 µl of 50% beads) and rotated for 1 h at 4°C. Agarose is pelleted by centrifugation at 4000 rpm for 1 min in a precooled centrifuge. The precleared supernatant is removed, 100 µl is retained as the INPUT sample and is frozen at -20°C, and 500 µl aliquots are prepared for the no antibody control and each immunoprecipitation. To each ChIP, the appropriate antibody is added for a final dilution of 1:100. ChIP and the no antibody control are rotated overnight at 4°C.

3.7.6 Washes, elution and reversal of cross-links

To bind the immune complex, 40 µl of 50% of protein G agarose is added to each ChIP and to the no antibody controls and samples are rotated for 1 h at 4°C. Agarose beads are pelleted by centrifugation in a microcentrifuge at 1500 x *g* for 1 min. The supernatant is discarded and beads are washed once in low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, Tris-HCl pH 8.1, 150 mM NaCl), once in high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, Tris-HCl pH 8.1, 500 mM NaCl), once in LiCl wash buffer (250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris HCl pH 8) and twice in TE (10 mM Tris-HCl pH 8). Each wash is performed with 900 µl of freshly prepared, cold buffer and consists of a 5 min rotation at 4 ^{o}C with centrifugations of 1 min at 1500 x *g* in a precooled microcentrifuge. After the final wash in TE, remove all TE with gel loading tips (Costar 4884). DNA-protein complexes are eluted by incubating the agarose beads with 100 µl of freshly prepared elution buffer (1% SDS, 0.1 M $NaHCO_3$), samples are vortexed, briefly centrifuged for 2 s at 100 x *g* and incubated for 15 min at 65 ^{o}C. Samples are vortexed again, centrifuged for 1 min at 1500 x *g* and the supernatant containing DNA-protein complexes is retained. The agarose beads are extracted a second time as described with another 100 µl of elution buffer, except incubated at 65°C for 5 min. The supernatant is extracted with gel loading tips and pooled with the previous eluant. 100 µl of elution buffer is added to the thawed INPUTS. The cross-links of all samples (ChIPs, no antibody controls and INPUTS) are reversed by incubating overnight at 65°C.

3.7.7 Protein digestion, DNA extraction and qPCR

Proteins are digested by the addition of 1 µg of proteinase K for 2 h at 45 ^{o}C, and the DNA is extracted via a phenol-chloroform extraction and

precipitated with ethanol overnight at -20°C as described in section 3.3.1, with the addition of 40 µg of glycogen and resuspended in 300 µl of Milli-Q H_2O. qPCR was performed as described in section 3.4.3 using 2 µl of DNA per reaction and results were calculated using the relative standard curve method against sonicated, RNaseA-treated gDNA. The sequences of the primers used in qChIP analysis are listed in Article 1, Supplementary Table S1, and in Table 3.2.

3.8 Digestion of nuclei by restriction enzymes for chromosome conformation capture (3C) and circular chromosome conformation capture (4C)

3.8.1 Parasite material and formaldehyde cross-linking of nuclei

These methods have been adapted from Hagege et al (2007) and Zhao et al (2006) and reflect standard 3C and 4C procedures. Parasites are isolated and cross-linked with formaldehyde as previously described for ChIP in section 3.7.1 and 3.7.2 until after the homogenization step. After homogenization, the samples are pooled, transferred to microcentrifuge tubes and centrifuged at 4°C for 10 min at 16 100 x *g*. The supernatant is discarded and the cross-linked nuclei are resuspended in PBS and aliquoted into in individual aliquots at ≈ 1.8x10^{6} parasites per aliquot (which corresponds to ≈2.95x10^{7} nuclei) and stored at -80°C.

3.8.2 Restriction enzyme digestion of cross-linked nuclei

A single aliquot of cross-linked nuclei are used per condition tested. Each experiment also contains an undigested control that is treated in parallel without the addition of restriction enzymes. The restriction enzymes for 3C analysis are *Bcl*I and *Bgl*II (Roche, 40 U/µl, 11 097 059 001 and 10 899 224 001, respectively). The restriction enzyme for 4C analysis is *Dpn*II (NEB, 50 U/µl, R0543M). Frozen cross-linked nuclei are thawed on ice and pelleted by centrifugation at 4°C in a microcentrifuge at 2300 x *g* for 5 min. The supernatant is discarded and pellets are resuspended in 400 µl of 1X of appropriate restriction enzyme digestion buffer (Roche SureCut buffer M for 3C samples and NEBuffer *Dpn*II for 4C samples). Samples are centrifuged at 4°C for 5 min at 2300 x *g*, the supernatant is removed and pellets are resuspended in 627 µl of the appropriate

restriction enzyme digestion buffer. To perforate the nuclear membrane and to open chromatin for restriction enzyme digestion, SDS is added to a final concentration of 0.2%. To avoid the formation of nuclear aggregates, samples are gently but thoroughly resuspended by pipetting up and down and are subsequently incubated at 37°C with gentle agitation for 30 min (Minitron incubator, 95 rmp). To sequester the SDS, Triton X-100 is added to a final concentration of 1% and samples are incubated for 30 min at 37°C with gentle agitation (Minitron incubator, 95 rmp). 3C samples are not subject to a simultaneous double digestion because the optimal digestion temperatures for *Bcl*I (50°C) and *Bgl*II (37°C) are incompatible, and are first digested with 500 U *Bgl*II. 4C samples are digested with 500 U of *Dpn*II. The samples are incubated at 37°C overnight with gentle agitation in a Minitron incubator at 95 rpm. 4C samples are subsequently kept at 4°C until the following day. 3C samples are supplemented with 1 mM DTT, and 500 U of *Bcl*I are added. These samples are incubated at 50 ^{o}C overnight with gentle agitation in a Minitron incubator at 95 rpm.

3.8.3 Preparation of samples to check efficiency of restriction enzyme digestion by Southern blot

The following day, to inactivate the restriction enzymes and to release digested chromatin, SDS is added to all samples (undigested controls, 3C and 4C samples) to a final concentration of 2%. To reverse formaldehyde cross-links, samples are incubated at 65°C overnight. Samples are subsequently centrifuged at 16 100 x *g* at room temperature in a microcentrifuge for 10 min, and the supernatant is transferred to another microcentrifuge tube. To the supernatant, 380 µl TE buffer and 100 µg RNaseA are added and the samples are incubated at 37°C for 30 min. 100 µg of proteinase K are added and samples are incubated for 2 h

at 45°C. The DNA is purified according to the phenol-chloroform extraction and ethanol precipitation protocol previously described in section 3.3.1 with the addition of 40 µg of glycogen and DNA pellets are washed 3 times with 70% ethanol and resuspended in Milli-Q H_2O. DNA is quantified as described in section 3.3.1 and approximately 300 ng of DNA are run on a 1% agarose gel for visual verification of digestion prior to proceeding with Southern Blot analysis.

3.9 Southern blot

3.9.1 Amplification, preparation and labelling of DNA probes

Specific DNA probes were amplified by PCR from 3D7-A genomic DNA and their sizes were verified by agarose gel electrophoresis. The primers used to amplify the probes are listed in Table 3.5. Probes are excised from agarose gels and purified on spin columns. 30 ng of purified probes are radiolabelled with α^{32}-dATP (PerkinElmer, NEG012H250UC) using the NEBlot kit (New England Biolabs, 15005) as per the manufacturer's instructions. Nucleotides are removed from the probe using Probe-Quant G-50 microcolumns (GE Healthcare, 28-9034-08) as per the manufacturer's instructions and the labelled probe is subsequently kept on ice. Just prior to hybridization, the probe is incubated at 95°C for 5 min, kept on ice for 5 min and then added to the hybridization solution at 62°C.

Table 3.5 List of primers used to amplify the probes used for southern blot analysis

Probe	Primer name	Sequence 5' to 3'
3C BAIT *clag3.2*	SBAITfwd	acaacgataacgaaaaacaacaa
	SBAITrev	gtgtttttggtccatattgttat
3C *clag3.2* fragment 2	3CS2fwd	cacaactacaaaaactatactca
	3CS2rev	ttgtggtgaagataatagagtaa
3C *clag3.2* fragment 7	3CS1fwd	gtagtacaccaaaaatatgactt
	3CS1rev	aaggtgatacgttttatttggta
4C BAIT *clag3.2*	3.2seq2fwd	ttcacggttacgtttctgttttt
	3.2_qRT_P4r	actaatactttattcacgtgtac
4C BAIT *var*	4CVBS_F	aataatgtgcatgcatgtcagtggt
	4CVBS_R	tcggtgccagttgtttccgt

3.9.2 Southern blot digestion and agarose gel electrophoresis

0.5-2 µg of genomic DNA is digested with the appropriate restriction enzymes in a 37°C incubator for 4 h. The details of the digestions for each Southern blot analysis are outlined in Figures 6.9-6.13. Digested samples are run on large (20 x 20 cm) 1% agarose gels, prepared in 1X TBE, supplemented with 0.5 µg/ul ethidium bromide. The gels are run in 1X TBE at 30 – 60 V overnight (depending on the expected fragment sizes). The following day, a picture of the gel, with a fluorescent ruler is taken. The gel is depurinized by incubation for 10 min in 0.25 M HCl and then rinsed twice with Milli-Q H_2O. To denature the double-stranded DNA, the gel is incubated for 30 min in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) and then rinsed with Milli-Q H_2O. The alkaline gel is neutralized with two, 30 min incubations in neutralizing buffer (1.5 M NaCl, 0.5 M Tris pH7.4, 1 mM EDTA pH=8).

3.9.3 DNA transfer

The transfer is performed in the same electrophoresis tank in which the gels were run. To the tank, 10X SSC is added (20X SSC is 0.3 M tri-sodium citrate dehydrate, 3 M NaCl). A saline bridge is made with Whatmann 3MM paper. 10X SSC is added on top of the saline bridge. Six Whatmann 3MM papers cut to the appropriate size and wet with 10X SCC are placed on top of the saline bridge. The gel is placed next and Parafilm is placed around the gel if necessary to prevent contact between papers above and below the gel. A positively charged Nylon membrane, cut to the appropriate size (GE Healthcare, Amersham Hybond^{TM} N, RPN203B) is wet with Milli-Q H_2O then incubated in 10X SSC for 5 min and is placed on top of the gel. Six Whatmann 3MM papers, wet with 10X SSC are placed on top of the membrane. 90 pieces of dry bench paper and 15

pieces of dry, folded paper towels, cut to the appropriate size are placed on top of the wet Whatmann 3MM papers. Even pressure is applied to ensure good contact and the DNA is transferred to the membrane by capillary action overnight. The following morning, the uppermost wet papers are removed and replaced with dry layers. The transfer is dismantled, the outline of the gel is traced onto the membrane in pencil. The membrane is placed on Whatmann 3MM paper and the DNA is fixed in a Stratagene UV Stratalinker ® using the auto cross-link cycle, twice, changing the position of the membrane in between cycles. Fixed membranes are incubated in 2X SSC-0.1% SDS. Membranes are placed into a hybridization tube and prehybridized for 1 h at 62°C with prehybridization solution (6X SSC, 5X Denhardt's solution, 0.5% SDS, 200 µg/ml sonicated salmon sperm DNA. 100X Denhardt's solution is 2% w/v Ficoll^{TM} PM400, 2% w/v Polyvinylpyrrolidone, 2% w/v BSA Fraction V) whilst rolling. Prehybridization solution is then discarded and hybridization solution (same composition as prehybridization solution but with the addition of the radiolabelled probe) is added to the membrane and the membrane is left rolling overnight at 62 ^{o}C.

3.9.4 Washing of the membrane

Hybridization solution is decanted and stored at -20 ^{o}C and membranes are washed twice with 6X SSC-0.1% SDS and twice with 2X SSC-0.1% SDS. Washes are for 15 min, rolling, at 62°C. Membranes are checked with a Geiger counter, and if the radioactivity is too strong, two additional washes are performed with 1X SSC-0.1% SDS.

3.9.5 Phosphor imaging and X-ray film exposure

The washed membranes are placed in an exposure cassette (Molecular dynamics) between layers of plastic wrap and covered with a Hyperscreen™ intensifying storage phosphor screen (GE Healthcare, Amersham, previously treated in the Image Eraser 810-UNV, Molecular Dynamics) for a minimum of 4 h. The phosphor image signal from the Hyperscreen™ is digitised in a Typhoon™ 8600 Scanner (Molecular Dynamics). The membrane is subsequently exposed on X-ray film (Super RX Fuji medical X-ray film) in a Hypercassette™ (GE Healthcare, Amersham), at -80°C overnight (at least). The film is developed in a Hyperprocessor™ Automatic Film Processor (GE Healthcare, Amersham).

3.10 Optimised conditions for the restriction enzyme digestion of cross-linked nuclei for 3C and 4C

Southern blot analysis revealed that under the standard digestion conditions described in section 3.8.2, *Bcl*I digested the DNA in formaldehyde cross-linked nuclei efficiently, whereas *Bgl*II and *Dpn*II did not. Consequently the 3C/4C restriction enzyme digestion protocol is performed as described in section 3.8.2, however, with the following modifications:

-formaldehyde cross-linked nuclei are used immediately upon preparation (i.e. they are not frozen)

-for 3C, samples are first digested with *Bcl*I at 50°C overnight, followed by *Bgl*II digestion at 37°C

-for 4C, samples are incubated at 50°C overnight and subsequently digested with *Dpn*II the following day i

-the units of restriction enzyme per digestion was increased from 500 U to 5960 U in the case of *Bgl*II and from 500 U to 7500 U in the case of *Dpn*II

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

4. Article 1

4.1 Introduction to Article 1

The genes *clag3.1*, *clag3.2* and *eba-140* are good models to study epigenetic regulation in *P. falciparum*. These three genes show life cycle dependent silencing and CVGE (Figure 1.5B and C). *clag*3.1, *clag3.2* and *eba*-140 are never expressed outside of the schizont stage, and the clonal transmission of their variant expression status implies that in rings and trophozoites, these genes are either bookmarked for expression or bookmarked for repression.

By comparing these three genes at different life stages and in isogenic parasite lines where the genes are either transcriptionally active or repressed, or bookmarked for activation or repression, we intend to determine the mechanisms involved in silencing associated with progression along the IEC and with CVGE, and if different mechanisms are involved, attempt to discriminate between them.

Towards this aim, we characterised the upstream sequences required for promoter activity, determined nucleosome positions and degree of chromatin compaction and characterised several relevant histone modifications at different life stages in parasite lines where the genes are differentially expressed.

4.2 Article 1

Crowley VM, Rovira-Graells N, Ribas de Pouplana L and A Cortés. Heterochromatin Formation in Bistable Chromatin Domains Controls the Epigenetic Repression of Clonally Variant *Plasmodium falciparum* Genes Linked to Erythrocyte Invasion. Molecular Microbiology.2011;80(2):391-406.

SUPPORTING INFORMATION

Heterochromatin Formation in Bistable Chromatin Domains Control the Epigenetic Repression of Clonally Variant *Plasmodium falciparum* Genes

Valerie M. Crowley[1], Núria Rovira-Graells[1], Lluís Ribas de Pouplana[1,2] and Alfred Cortés[1,2,*]

[1]Institute for Research in Biomedicine (IRB), 08028 Barcelona, Catalonia, Spain

[2]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain

*To whom correspondence should be addressed. Tel. +34 934034867; Fax: +34 934037109; Email: alfred.cortes@irbbarcelona.org

A

clag3.1

ATTTCCATAATTTATTTTCTAATAATTATCTTCTTTTCTTTCCATTGCACCGATATTAAAAACAAAAA
CAAAAAAAAAAAAACCAAAAAACACCATTCATATTTATATTGAATATGTATTTTTATCTCACTTCTTT
TTATTTACCT**A**$^{-454}$AGTAAAT**A**$^{-446}$TATATATA ATCATATAACAAATAACATGTTATGCAAAAATATG
CATAAACATAAGAGGGTGATTTAAAATAAAAATAAAAATATACTTTACAATATGTTC**A-**^{345}AAATATA
ATTCGT **A**$^{-331}$ TATTCATAT**A**$^{-321}$TTCTCGAATAGTTTTATTTCCTTTTATATTAC ATAATGTACAC
TTT**A**$^{-273}$AAAC**A**$^{-269}$AAATGT**G**$^{-262}$GAC**A**$^{-258}$AATATGATTTTGTACTAAAAGAAAAAAAAAAAAAAGT
A^{-220}CAACATTAAAAAAATTCTTTT TGATATATATATATATGTTT**A**$^{-178}$AACAT**A**$^{-172}$TATTTGTGTA
CTCCTTTTTGTACATGGTTTTTGATTTTGTTATATTTATGTACTATTCAATTCTGTACTTTTTTTTAA
TAATAACATAATAATAATTACTTATATATTTGTGCAATATATCAAAGTGTACATGCAATAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAA**ATG**

clag3.2

TTTTATTTTATTTTTTTATTTTTTGTTTTTATGGAGTTGCACCGACATTAAAAAGAAACGAAACGATA
TATACTTTTTTATAATCAACGTTCACGGTTACGTTTCTGTTTTTCTTTCTTTTTTTTTAATTATTTAT
TTATGTGTATTAACTTAAATATTTGGAAATATGCAAAATTATGCATAATACTGAAAATAAAGACTAAA
CAAATGGAAAATCAACATATATTTTATGTCTATTATATATGTCATTTCGATATTGTTATTGTTTATTA
ACAAATGGGTAAA**A**$^{-315}$TAATAAAAATGGTCATCATCGGTATTGTACAAAAAT**G**$^{-278}$ TGAAAATTAAAA
AATATGTACAAAAGTAAAAATATGTTATTCAATTATATATATATATATATA**T**$^{-216}$ATATATAATATTTC
A^{-201}C**A**$^{-199}$TACACAAACATATATATAACATATAATTGT**G**$^{-168}$TACTCCTTTTTGAATACGGTTTTTA
TTTTTGTTATTGTTTTTTCTTTTTATTCTGGGTTATTTTCTTCTGTACTTTTTTAAATATATATCGTA
ATAATAGTTGAGTACGCACTAATATGTCAATTTGAAAAAAAATATTGTACACGTGAATAAAGTATTAG
TAAAAA**ATG**

eba-140

ATATATAT**A**$^{-592}$TATTTAT**A**$^{-584}$T**T**$^{-582}$TTTTTTTTCCAT**A**$^{-569}$CTCTAGAAC**A**$^{-559}$ATTCTATAT**A**$^{-549}$
TTTAAATATTTCTTTCAATATGTAGTAATTATATATATATATATATATAATATATATTTTTTTTT
TATTTATTTATTCATAAAAAATGTATGTTAATTTTTATTATTAAAAAGGTTTTAAAATAAATGGGAAA
AATTATAAAAAAAATATAACAAATGAGAAATAAAAAGAGAGAGTAAGTTAAGAATTAAAAAAAGAATT
AACTTCAATAATTAAATGGTTAATATTTCTACTGCTTAAATTAATTAATTAAATATAATTCAAATTAA
TTTTGAAATGTGTTTTATAATTAATTTTTAATTCAATTTTTGGGTTGTAATTTTATATGGTTGTACAT
ATATATATATATATATATATATATATATATATATATATGTATATTTAATTAAATTTTAATATTTAATTTT
TTTTTTTTTTTACTACCAAATATATTTATATATAGTACTTTTATAATATCGTTTTCTTATGAGCATTA
TATCATATAAATAATTAATTAAATAAATATATATATATATATATAATTATAATTAGACCAATAAATTATA
TATA**ATG**

B

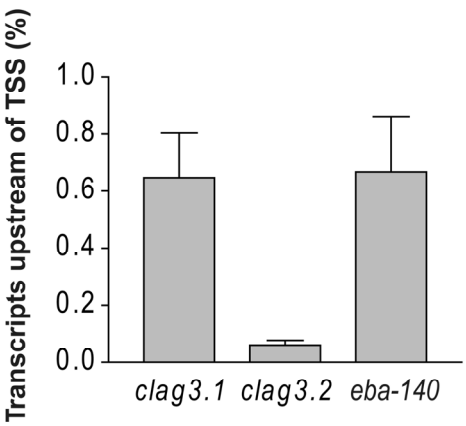


Fig. S1. Transcription start sites (TSSs) in *clag3.1*, *clag3.2* and *eba-140*. **(A)** Nucleotide positions of TSSs as determined by 5' RACE. The nucleotides corresponding to the TSSs of *clag3.1*, *clag3.2* and *eba-140* are indicated in bold and are underlined. The numbers in superscript adjacent to the TSSs indicate the position

relative to the translation start codon, which is indicated in bold. The positions of the primers used for nested PCR are underlined. **(B)** Percentage of transcripts initiated upstream of the mapped TSSs, as determined by qPCR. cDNA was generated by reverse transcription with oligo $(dT)_{15}$ and analysed by qPCR with primer pairs located upstream and downstream of the mapped TSSs (Table S1). Levels of cDNA for each primer pair region were calculated relative to RNAse-treated genomic DNA using the $2^{\Delta Ct}$ method. Experiments with cDNA generated with random primers gave identical results.

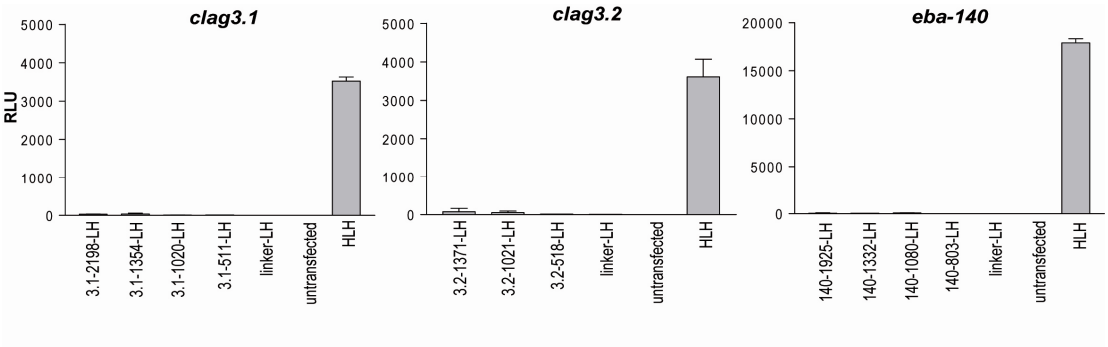
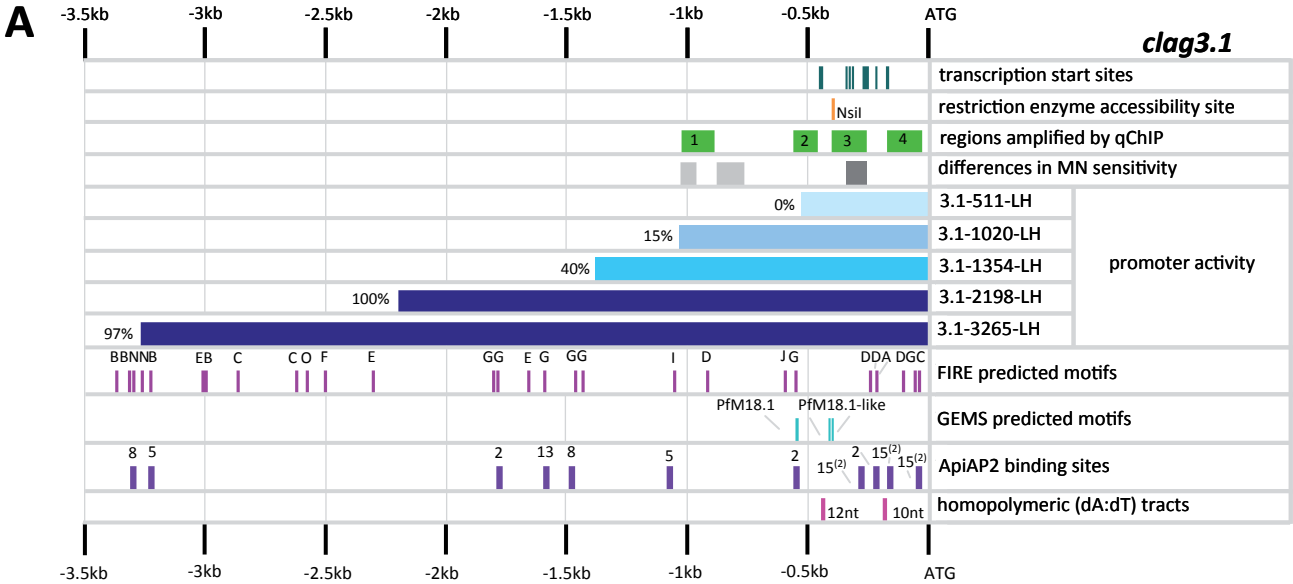
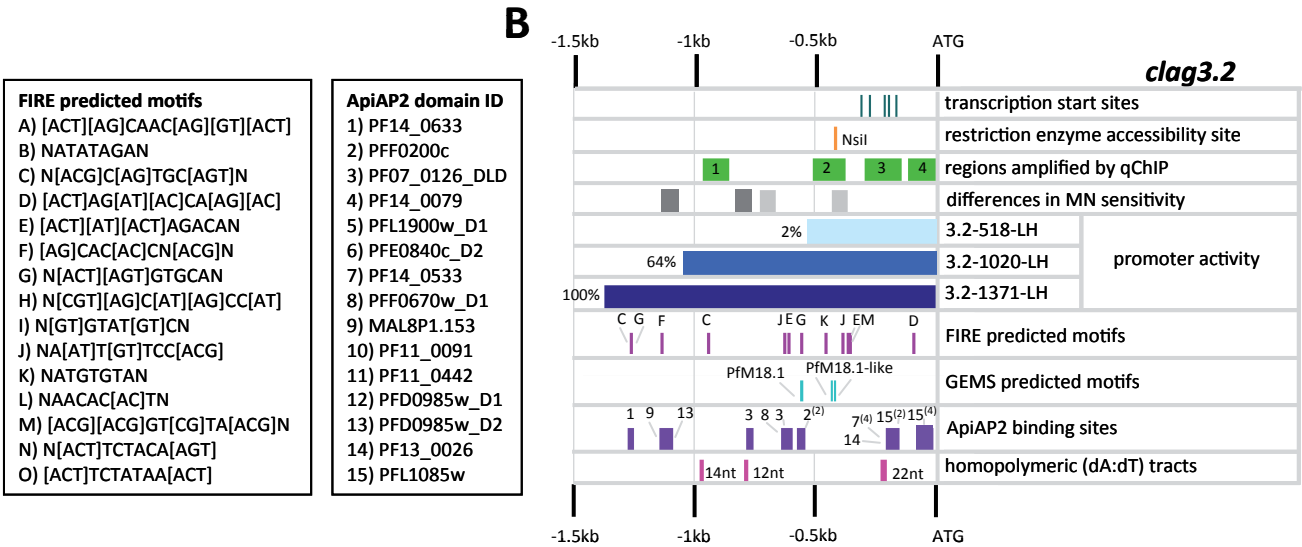


Fig. S2. Promoter activity in parasites at non-schizont stages. Luciferase assays were performed with the indicated plasmids in 3D7-A parasites at late ring/early trophozoite stage. The plasmids HLH (*hrp3* promoter) and linker-LH (no promoter) were used as positive and negative controls, respectively. Each construct was tested in duplicate. Values correspond to average luciferase activity (in relative light units, RLU), with standard deviation.





FIRE predicted motifs

A) [ACT][AG]CAAC[AG][GT][ACT]
B) NATATAGAN
C) N[ACG]C[AG]TGC[AGT]N
D) [ACT]AG[AT][AC]CA[AG][AC]
E) [ACT][AT][ACT]AGACAN
F) [AG]CAC[AC]CN[ACG]N
G) N[ACT][AGT]GTGCAN
H) N[CGT][AG]C[AT][AG]CC[AT]
I) N[GT]GTAT[GT]CN
J) NA[AT]T[GT]TCC[ACG]
K) NATGTGTAN
L) NAACAC[AC]TN
M) [ACG][ACG]GT[CG]TA[ACG]N
N) N[ACT]TCTACA[AGT]
O) [ACT]TCTATAA[ACT]

ApiAP2 domain ID

1) PF14_0633
2) PFF0200c
3) PF07_0126_DLD
4) PF14_0079
5) PFL1900w_D1
6) PFE0840c_D2
7) PF14_0533
8) PFF0670w_D1
9) MAL8P1.153
10) PF11_0091
11) PF11_0442
12) PFD0985w_D1
13) PFD0985w_D2
14) PF13_0026
15) PFL1085w

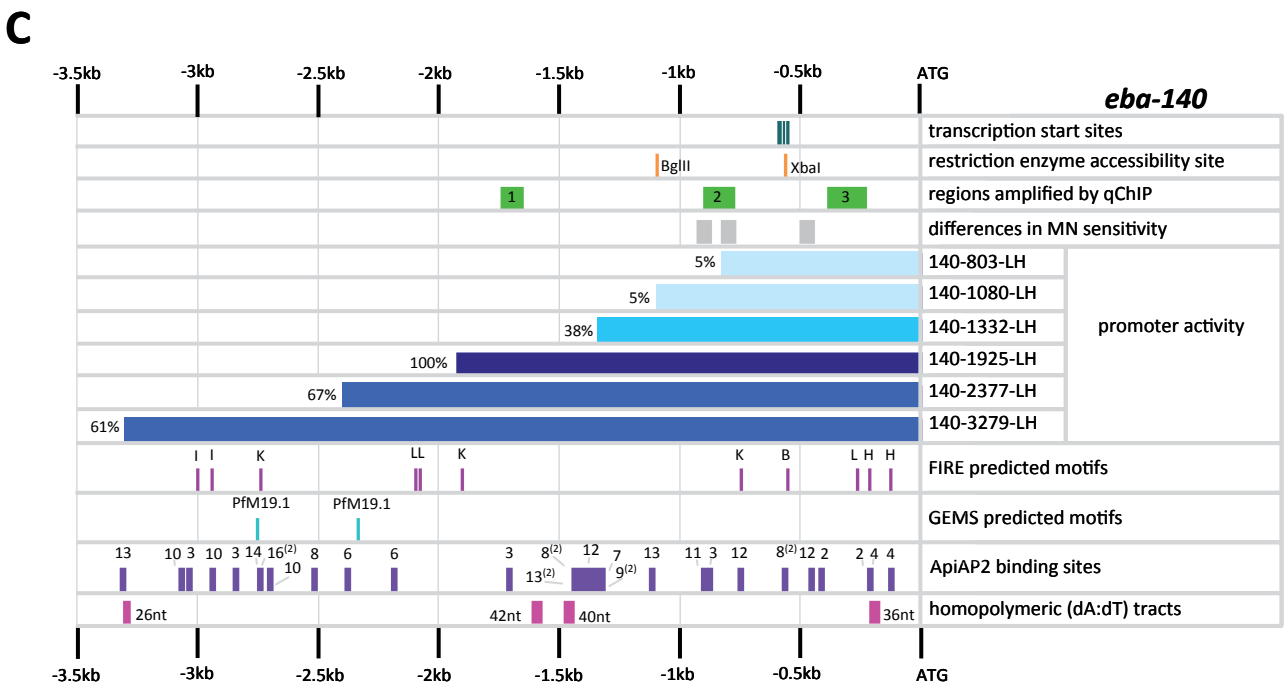


Fig. S3. Alignment of experimentally derived data with predicted regulatory motifs in the upstream regions of *clag3.1* **(A)**, *clag3.2* **(B)**, and *eba-140* **(C)**. All positions are relative to the translation start site. Transcription start sites, as determined by 5' RLM-RACE, are as in Fig. 1A. The position of restriction enzymes sites used in accessibility assays (Fig. 4) are indicated in orange. Regions amplified for qCHIP analysis are shown in green and are numbered as in Fig. 5 and S6. Regions where major differences in micrococcal nuclease sensitivity were observed (arrows in Fig. 3) are indicated as dark grey boxes, whereas regions with minor differences are indicated as light grey boxes (arrowheads and asterisks in Fig. 3). 5' upstream fragments used for gene reporter experiments (Fig. 1B) are shown in blue, with darker colours representing higher promoter activity. The percentage of promoter activity relative to the construct with highest activity is indicated. Motifs enriched in the upstream region of genes expressed in schizonts, as predicted by the FIRE algorithm (Elemento *et al.*, 2007), are indicated in magenta. The consensus sequence of these motifs is shown in the inset. *cis*-regulatory elements predicted by the GEMS algorithm (Young *et al.*, 2008) are indicated in turquoise. PfM18.1 (GGTGCA), which is similar to FIRE motif G, is a motif common to genes that encode proteins localised to the rhoptries. Two copies of a PfM18.1-like motif, separated by six nucleotides, are present in *clag3.1* and *clag3.2*, a pattern common to other rhoptry genes (Young *et al.*, 2008). PfM19.1 (GTTGTATA), found in *eba-140* upstream region, is a motif common to genes involved in cell invasion. Motifs recognised by ApiAP2 transcriptional regulators whose expression pattern suggests that they may act as activators of genes expressed in schizonts (Campbell *et al.*, 2010) are shown in purple. The positions of the motifs were obtained from PlasmoDB (www.plasmodb.org). The numbers in brackets refer to the number of binding sites for the indicated ApiAP2, and the ApiAP2 domain IDs are listed in the inset (Campbell *et al.*, 2010). All homopolymeric (dA:dT) tracts greater than four nucleotides in length are shown in pink together with the number of nucleotides (nt) encompassing the tract. A potential role of these tracts in regulating transcription in *P. falciparum* has been suggested (Polson and Blackman, 2005).

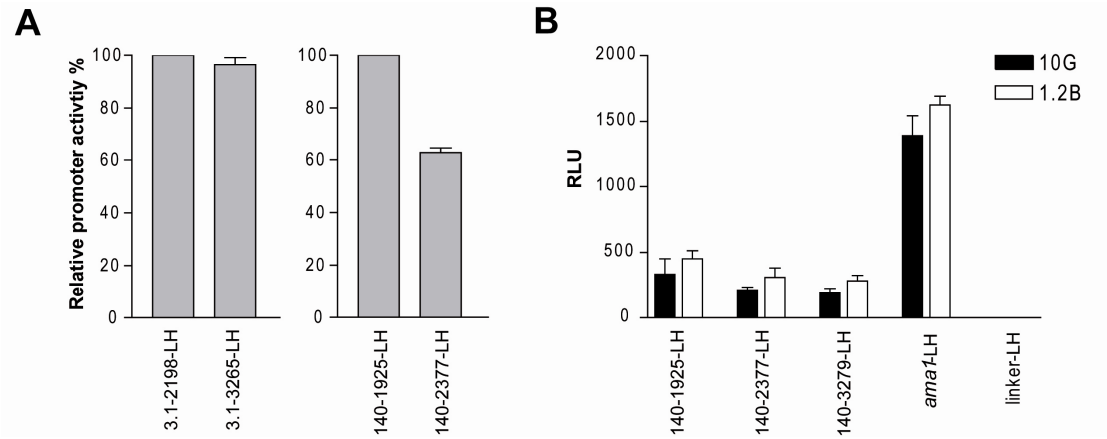


Fig. S4. Luciferase assays with constructs encompassing longer fragments of *clag3.1* and *eba-140* upstream regions. Luciferase activity is tested at the schizont stage. The length of the upstream sequences tested is included in the name of the constructs, as in Fig. 1B. **(A)** Luciferase activity in 3D7-A parasites. Values are % luciferase activity relative to 3.1-2198-LH or 140-1925-LH, as in Fig. 1B, and correspond to the average of two independent experiments (each performed in duplicate), with standard deviation. **(B)** Luciferase activity in 10G and 1.2B parasites. Values correspond to average luciferase activity (in relative light units, RLU), with standard deviation, normalised to 1% parasitemia, as in Fig. 2D.

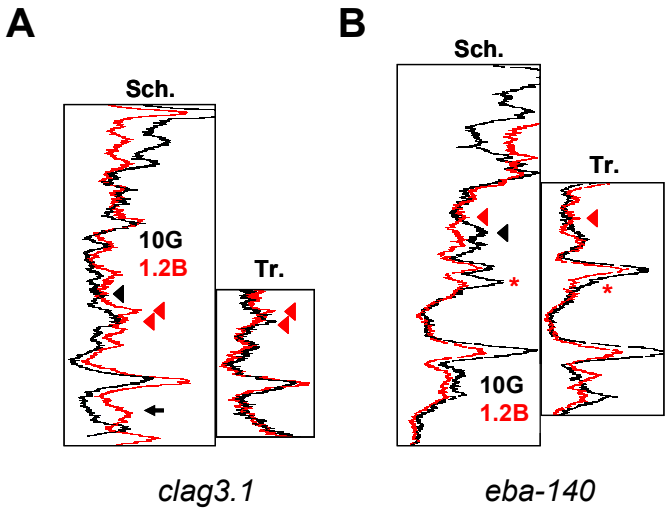


Fig. S5. Phosphor imager traces of the micrococcal nuclease digestion experiments in Fig. 3, for *clag3.1* **(A)** or *eba-140* **(B)**. Traces correspond to 0.5U schizont lanes (Sch. panels) or 0.5U trophozoites lanes (Tr. panels). Arrows, arrowheads, asterisk and all other symbols are as in Figure 3. It is important to note that it is impossible to adjust precisely the amount of radioactivity and the degree of digestion between different samples. Hence, differences between 10G and 1.2B in peak intensity over a large

region are not considered authentic differences in nucleosome positions. Only major changes between 10G and 1.2B in peak intensities relative to neighbouring peaks, absence of specific peaks or changes in peak position, observed both in digestion with 0.5 and 1.5 MN units, are considered to reflect differences in nucleosome positions.

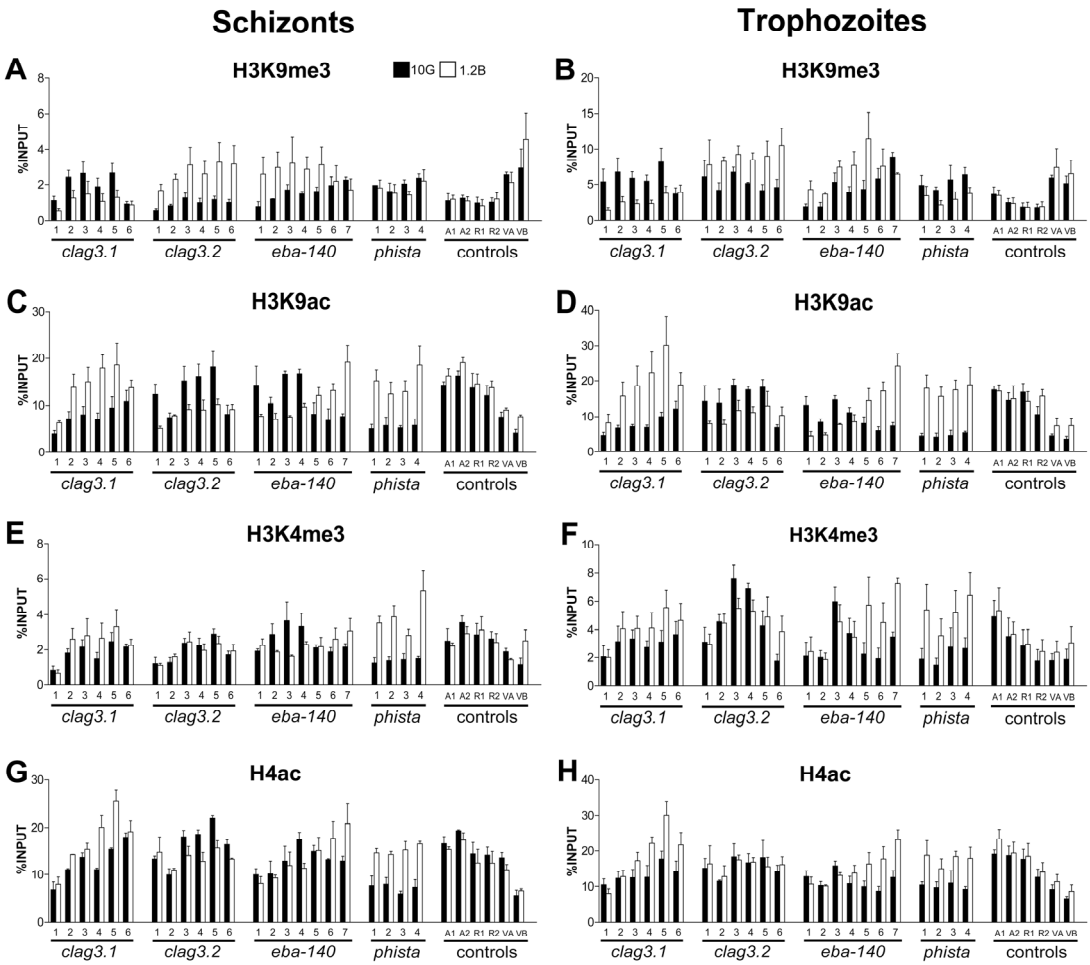


Fig. S6. Raw qChIP data for the analysis of histone modifications associated with the variant expression of *clag3.1, clag3.2*, *eba-140* and MAL13P1.59 in 10G and 1.2B parasites. ChIP data is expressed as % input, defined as the percentage of immunoprecipitated chromatin relative to the input sample. Chromatin from schizonts (A,C,E,G) or trophozoites (B,D,F,H) was immunoprecipitated with antibodies against the histone modifications indicated. Values are the average of three independent experiments, with standard error. Two silent *var* genes (VA: MAL7P1.55 ORF, VB: PFL1950w upstream region), and two single copy genes that do not display variant expression, *ama1* (A1: upstream region, A2: ORF) and *rhopH2* (R1: upstream region, R2: ORF) served as controls.

Table S1.

Primer Name	Sequence 5' – 3'
5' RLM-RACE	
3.1RACEin	TTTGT**GGATCC**GCATGTACACTTTGATATATTGC
3.1RACEout	TCATGTTGGTTGACATTTTGAC
3.2RACEin	TTTGT**GGATCC**CAAATTGACATATTAGTGCGTAC
3.2RACEout	GCATTTTCGTTTTTATTTTCGCC
eba140RACEin	TTTGT**GGATCC**GATATAATGCTCATAAGAAAACG
eba140RACEout	TGTCTATTATAAAGTGTTCTACC

Primers used to generate luciferase plasmids

Plasmid	Primer Name	Sequence 5' – 3'
3.1-511-LH	prom3.1F0.5kbAsp	TGGTGTT**GGTACC**ACACCATTCATATTTATATTGAATATGTA
	prom3.1r1Nsi	TGGAG**ATGCAT**TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTGC ATGTACACTTTGATATATTGC
3.1-1020-LH	prom3.1F1kbAs	TGGTGTT**GGTACC**ATAATGCACTCATTAATAATTTTAAACCGTT
	prom3.1r1Nsi	
3.1-1354-LH	prom3.1f2Asp	TGGTGTT**GGTACC**GATGATATTTAATAATATTGATATCCGTA
	prom3.1r1Nsi	
3.1-2198-LH	prom3.1f1Asp	TGGTGTT**GGTACC**AGGAAAAAATAATGCAATCTAAGTAATG
	prom3.1r1Nsi	
3.1-3265-LH	prom3.1f0.5Asp#	TGGTGTT**GGTACC**TTCTACATTTTCTCCATCTGAAT
	3.1seq3rev#	TTTTCTTTTTTACAGTTTATTCTC
3.2-518-LH	prom3.2F0.5kbAsp	TGGTGTT**GGTACC**ATCAACGTTCACGGTTACGTTTCTGTT
	prom3.2rNsi	GGAG**ATGCAT**TTTTTACTAATACTTTATTCACGTGTAC
3.2-1021-LH	prom3.2F1kbAsp	TGGTGTT**GGTACC**ATAAATTATACCAAATATGTTAATAATA
	prom3.2rNsi	
3.2-1371-LH	prom3.2fAsp	TGGTGTT**GGTACC**TAGAATATATTTATTTATAGTCAAAATTTGC
	prom3.2rNsi	
140-803-LH	prom140F0.8kbAsp	TGGTGTT**GGTACC**AAATAAGAGATAAGAAAAAGGCTCAT
	Pr140rNsi	TGGTGTT**ATGCAT**TATATATAATTTATTGGTCTAATT
140-1080-LH	Pr140f2.5Asp	TGGTGTT**GGTACC**ATTATTGAGGATCAATTAGTG
	Pr140rNsi	
140-1332-LH	pr140f2Asp	TGGTGTT**GGTACC**AATACACACATTTTTGATTAGG
	Pr140rNsi	
140-1925-LH	pr140f0.5Asp	TCGTCT**GGTACC**AAGCTTTTTGTTGTTATTACAC
	Pr140rNsi	
140-2377-LH	Pr140f6Asp$	TGGTGTT**GGTACC**CATCAATAAGAGAAAATATATGTC
	Pr140r1.2$	CTCTTATATAAAAATAGCTAAAT
140-3279-LH	Pr140f3Asp$	TGGTGTT**GGTACC**TGTATGTATGTATGTATATTCCTT
	Pr140r1.2$	
ama1-LH and ama1-RH	promAma1fAsp	TGGTGTT**GGTACC**ACGATTTTTTAATCAGATCCC
	promAma1rNsi	TGGAG**ATGCAT**TTTTGTACAATTTATAACAAGTAC
linker-LH	LinkAspNsiFw	GTACCAGTTATTTTAAAAAATGCA
	LinkAspNsiRev	TTTTTTAAAATAACTG

Primer Name	Sequence 5' – 3'
Primers used to generate probes for Southern blot analysis	
Probe140F	GACCAATAAATTATATATAATGAAAG
Probe140R	TGAAGGTGGAATTATATCATATGA
Cl3.1fATG	ATGGTTTCATTTTTTAAAACTCCA
3.1r0.5	GTTCAATATTGTCATTATCATTTACATT
Kan3.2f (Iriko *et al.*, 2008)	AATAGTTGAGTACGCACTAATATGTC

3.2r0.5	TTCAATATTGTCATACTCATTTAGAAG	
Pfg377f1	TTTGTTAGATGCATATACAAATTTG	
Pfg377r1	CCATCCTTCTTATTATCATATAG	
qChIP Primers		**Final Primer Concentration (nm)**
clag3.1 – 1F	GATTTTATAATGCACTCATTAATAA	500
clag3.1 – 1R	AATAAATGTATACGTAATTAGAACA	1000
clag3.1 – 2F*	TTTCCATTGCACCGATATTAAAA	200
clag3.1 – 2R*	GAAGTGAGATAAAAATACATATTC	1000
clag3.1 – 3F	CATAAACATAAGAGGGTGATTTA	200
clag3.1 – 3R	CCACATTTTGTTTTAAAGTGTAC	200
clag3.1 – 4F*	CATATATTTGTGTACTCCTTTTT	1000
clag3.1 – 4R*	TTGCATGTACACTTTGATATATT	500
clag3.1 – 5F	TCAAAATGAAAATGATACCATTAGT	200
clag3.1 clag3.2 – 5R	CATGGATTTTAATTGTTCAATATTG	200
clag3.1 clag3.2 – 6F†	TAGTAATGAGAATTAGTTGGACA	200
clag3.1 –6R†	ATAAATATTTGGATGCTTCAGCA	200
clag3.2 – 1F	CATTACATATAAATAAATACGTGC	200
clag3.2 – 1R	AGTGTTTTAAGGCAATAATTATATT	200
clag3.2 – 2F*	TTCACGGTTACGTTTCTGTTTT	500
clag3.2 – 2R*	GTTGATTTTCCATTTGTTTAGTC	1000
clag3.2 – 3F	ATCGGTATTGTACAAAAATGTGA	200
clag3.2 – 3R	GTATTCAAAAAGGAGTACACAAT	200
clag3.2 – 4F*	ATTCTGGGTTATTTTCTTCTGTA	500
clag3.2 – 4R*	ACTAATACTTTATTCACGTGTAC	500
clag3.2 – 5F	TAGGCGAAAATAAAAACGAAAATG	200
clag3.2 – 6R†	ACAAATATGTTTCTGAACTAGGA	200
eba-140 – 1F	GAAATAGTTAGTGAATAATTAAAATC	1000
eba-140 – 1R	TTTAATGCTTAACATAATGCGACT	250
eba-140 – 2F*	GCTATTTTTATATAAGAGGGAAAT	1000
eba-140 – 2R*	ATAATGAGCCTTTTTCTTATCTC	1000
eba-140 – 3F*	TGAGAAATAAAAAGAGAGAGTAAGTTAAGAA	500
eba-140 – 3R*	ACAACCCAAAAATTGAATTAAAAAT	500
eba-140 – 4F	GAATCAATTATAGGTAGAACACT	1000
eba-140 – 4R	TTACATCTTCGTAATCTGAATCA	1000
eba-140 – 5F	TATGCAGATTTGGCAGATATTAT	500
eba-140 – 5R	CTTCAAAGATTCTTCATTACGTT	500
eba-140 – 6F	TGAATAGAACAACGAATGCACAAGA	500
eba-140 – 6R	CCGATTGTCAAAGAACCTGAATT	500
eba-140 – 7F†	TTGTTCTAGAGAAGAAAGAAAGA	1000
eba-140 – 7R†	ACCCTCGCTTTTAAAACATTTAT	500
MAL13P1.59 5' – 1F	TCTTTATTCTGTTTGAAAGACCA	500
MAL13P1.59 5' – 1R	GTTATAGACACATTTAAACCTCA	500
MAL13P1.59 5' – 2F	GCATTTGTAGAACGCAATTTTAT	500
MAL13P1.59 5' – 2R	CCATAAGTTATTTTACACATGAAA	500
MAL13P1.59 5' – 3F	TAGGAAACACTGAATAAATTATAAA	3000
MAL13P1.59 5' – 3R	ACCTACACTTAATCCATAATAAG	500
MAL13P1.59 5' – 4F	GAATTTTAGGATGAATGAAATACA	500
MAL13P1.59 5' – 4R	AAAAATATCTCCCTCCTTAATTG	500
(A1) ama1P1F	TTCACTTTTGTTAGAGTCTCTTA	200
(A1) ama1P1R	CCATTTTTCTTCTTTTTTCCCTT	200
(A2) ama1P2R	CGCCTTTGAGTTTACATATATGA	200
(A2) ama1P2R	TGGATATTCGTATTCTTTTGGAT	200
(R1) rhopH2P1R	ACTTTAAAATGACATGTTGGTAAA	200

(R1) rhopH2P1R	AAACCACATACATATAACACACA	200
(R2) rhopH2P2R	TGTTGCTGTCCATATTTAGTTTT	200
(R2) rhopH2P2R	AATATATCGCTACATAACTTCGT	200
(VA) MAL7P1.55 F (Salanti *et al.*, 2003)	ACGTGGTGGAGACGTAAACA	500
(VA) MAL7P1.55 R (Salanti *et al.*, 2003)	CCTTTGTTGTTGCCACTTTG	500
(VB) PFL1950w R (Jiang *et al.*, 2010)	CTATGTTGTATTATTCGATATTTTC	500
(VB) PFL1950w F (Jiang *et al.*, 2010)	AGAATAGGAAAATACAAATTATAGC	500
Primers for transcriptional analysis (in addition to primers above marked with †)		
MAL13P1.59 F	AATTCCGTCAGATGAAGATCTTG	
MAL13P1.59 R	TTGCATTTCTTCATATTCATCCA	
PF07_0073 F (seryl tRNA synthetase) (Ref. 2)	AAGTAGCAGGTCATCGTGGTT	
PF07_0073 R (seryl tRNA synthetase) (Ref. 2)	TTCGGCACATTCTTCCATAA	

Restriction sites are indicated in bold and are underlined
#The PCR product obtained with these primers was not cloned directly into the HLH plasmid. Instead, it was digested with *Asp718*I and *Bcl*I and cloned into the 3.1-2198-LH plasmid digested with the same enzymes.
$ The PCR product obtained with these primers was not cloned directly into the HLH plasmid. Instead, it was digested with *Asp718*I and *Bgl*II and cloned into the 140-1925-LH plasmid digested with the same enzymes.
*Primers used to determine the number of transcripts initiated upstream of the mapped TSSs
†qChIP primers that were also used for transcriptional analysis

REFERENCES

Campbell, T.L., De Silva, E.K., Olszewski, K.L., Elemento, O., and Llinas, M. (2010) Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog* **6**: e1001165.

Elemento, O., Slonim, N., and Tavazoie, S. (2007) A universal framework for regulatory element discovery across all genomes and data types. *Mol Cell* **28**: 337-350.

Iriko, H., Kaneko, O., Otsuki, H., Tsuboi, T., Su, X.Z., Tanabe, K., and Torii, M. (2008) Diversity and evolution of the *rhoph1/clag* multigene family of *Plasmodium falciparum*. *Mol Biochem Parasitol* **158**: 11-21.

Jiang, L., Lopez-Barragan, M.J., Jiang, H., Mu, J., Gaur, D., Zhao, K., *et al.* (2010) Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. *Proc Natl Acad Sci USA* **107**: 2224-2229.

Polson, H.E., and Blackman, M.J. (2005) A role for poly(dA)poly(dT) tracts in directing activity of the *Plasmodium falciparum* calmodulin gene promoter. *Mol Biochem Parasitol* **141**: 179-189.

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

Young, J.A., Johnson, J.R., Benner, C., Yan, S.F., Chen, K., Le Roch, K.G*., et al.* (2008) In silico discovery of transcription regulatory elements in *Plasmodium falciparum*. *BMC Genomics* **9**: 70.

4.3 Additional results

4.3.1 Verification that commercial antibodies for qChIP cross-react with *P. falciparum* histones

Since the commercial antibodies purchased for use in qChIP analyses were not raised against *Plasmodium* histones, the antibodies were checked for cross-reactivity with *P. falciparum* samples via fixed immunolocalisation assays (Figure 4.1). Segmented schizonts were used in these analyses because of their distinctive morphology, and DAPI was used to fluorescently stain DNA. The signal obtained with the commercial antibodies colocalised with the DAPI stain as seen in the column where the two signals are merged. The punctate staining pattern reflects individual merozoites that develop within the schizont, each of which contains a nucleus. This indicated that the commercial antibodies raised against a core histone and to modified histones from other species cross-react with *P. falciparum* histones and that the signal is localised to the appropriate cellular compartment, suggesting that they are suitable for use in qChIP experiments. The samples in Figure 4.1 were fixed with formaldehyde, however similar

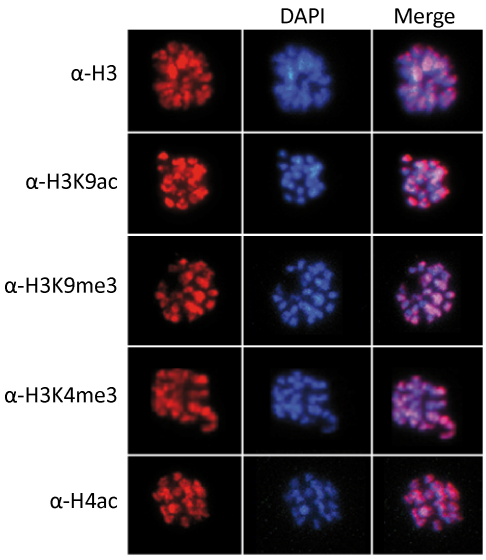


Figure 4.1 Immunolocalisation assays with antibodies used in qChIP analyses in *P. falciparum* segmented schizonts. These samples were fixed with formaldehyde, using a 1:100 dilution of primary antibody. The reference numbers for the antibodies used are listed in Table 3.4

results were obtained with acetone-methanol fixation (data not shown). Information regarding the antibodies used, including reference numbers and details about the immunogens used are listed in Table 3.4.

4.3.2 qChIP data using antibodies against core histone H3 were highly variable

Often qChIP analyses against histone modifications are normalized by nucleosome density. This refers to normalization of the %input obtained with the antibody against a specific histone modification by the %input obtained with an antibody against a core histone. As histone modifications can only be detected at a particular DNA region if the DNA region is wrapped into nucleosomes, this normalisation method would correct for differences in the density of nucleosomes, rather than changes in histone modification levels. qChIP was performed with an antibody raised against unmodified histone H3 (Figure 4.2). However, the %input levels that were recovered were often low and highly variable between different sample preparations and thus we did not normalise our qChIP results in Article 1, Figure 5 by these data. Similar results were obtained with an antibody against core histone H4 (Millipore, 04-858, data not shown). However, we observe antagonist enrichments with antibodies against different modifications of the same histone residue, i.e. trimethylation and acetylation of H3K9, ruling out the possibility that the observed differences between the active and repressed states of these genes is due to differences in nucleosome densities.

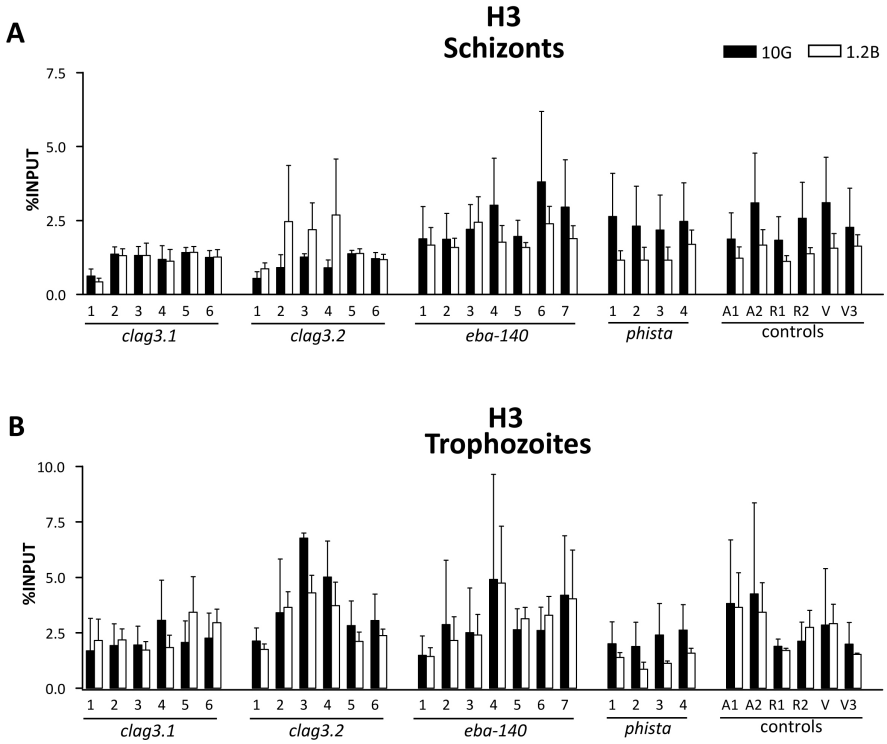


Figure 4.2 Raw qChIP data for the analysis of core histone H3 in A. schizonts and B. trophozoites. ChIP data is expressed as %input, defined as the percentage of immunoprecipitated chromatin relative to the input sample. Values are the average of three independent experiments, with standard error. The position of the primers for *clag3.1*, *clag3.2*, *eba-140* and *phista* (MAL13P1.59) can be found in Article 1, Figure 5A. A description of the control genes can be found in Article 1, Figure S6.

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

5. Article 2

5.1 Introduction to Article 2

Our study of the epigenetic regulation of CVGE in genes other than the *var* genes began with the study of *clag3.1*, *clag3.2* and *eba-140*. We demonstrated that akin to the *var* gene family, CVGE of the *clag* and *eba* families was controlled by the formation of H3K9me3-faculatative heterochromatin. During the course of our analyses, we observed that a member of the *phista* family also showed CVGE that was controlled in the same fashion. If the epigenetic regulation of CVGE is common to four unrelated gene families that are expressed at different stages along the IEC and participate (when the function is known) in unrelated processes: how common is CVGE? are all clonally variant genes regulated by H3K9me3-facultative heterochromatin? what is the biological function of CVGE? In order to begin to address these questions, we need to know which genes show CVGE. Towards this aim, a microarray-based study was undertaken to identify the genes that show CVGE in *P. falciparum* (Rovira-Graells et al submitted, Article 2).

Characterisation of the level of transcriptional variation across the *P. falciparum* genome has revealed that this mode of regulation is widespread. The gene families that show CVGE are implicated in many host-parasite interactions apart from antigenic variation and therefore, CVGE appears to be central to *P. falciparum* biology. Despite the critical role of CVGE for survival and virulence of *P. falciparum*, the molecular mechanisms that control it are largely unexplored. The striking correlation between CVGE and heterochromatic marks is suggestive of an

H3K9me3-faculatative heterochromatin based mechanism that contributes to phenotypic variation.

As CVGE appears to be a fundamental process in parasite biology, the molecular mechanisms involved in establishing this phenomenon need to be understood. The study of CVGE would be greatly facilitated if it could be studied in an episomal system. For example, sophisticated episomal systems could be established where genes could be tested for their ability to be epigenetically silenced, to test sequences for potential barrier insulator activity, to calculate the switch rates of clonally variant genes and to test candidate proteins for their involvement in CVGE. Also, some aspects of CVGE, such as its spontaneous nature or whether it is a stochastic process, cannot be studied using the endogenous genes.

It has previously been demonstrated that the CVGE expression of *var* genes can be reproduced in an episomal system (Voss et al 2006), however, it is not known whether the CVGE for genes other than *var* can be mimicked using episomes. The *var* genes are a peculiar gene family, where most members are repressed. Thus, it is not known if smaller gene families where most members are expressed will demonstrate episomal CVGE. Therefore, we began a series of proof-of-principle experiments, to determine if the CVGE for genes other than *var* can be reproduced in an episomal system, using the clonally variant gene *clag3.2* as a model in these experiments.

We hypothesize that CVGE is a spontaneous process that is mediated by the promoter of the variant gene and is controlled by the formation of H3K9me3-facultative heterochromatin, see figure legend for further

details (Figure 5.1). If the CVGE of *clag3.2* can be mimicked in this episomal system, the system could be expanded to analyse multiple different promoters and be used as a tool to study CVGE in the future.

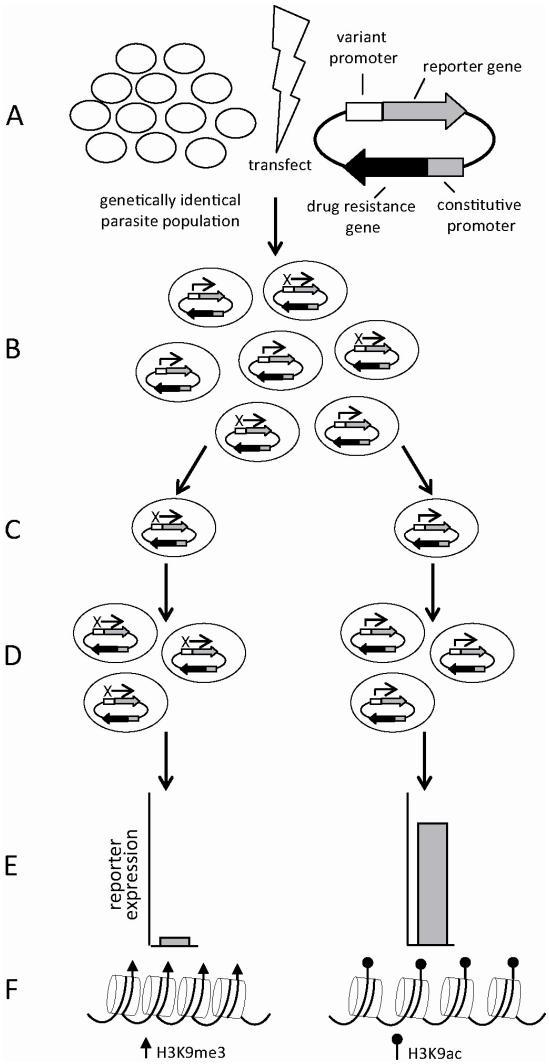


Figure 5.1. Simplified model to recreate episomal *clag3.2* CVGE. A. A population of genetically identical parasites is transfected with a plasmid that contains a reporter gene under the control of a variant promoter (*clag3.2*). This plasmid is stably maintained in the transfected population by the expression of a gene that confers drug resistance. **B.** If CVGE is a spontaneous process, under homogenous growth conditions, we expect the transfected population to consist of individual parasites that express or repress the reporter gene. **C**. We will isolate individual parasites by subcloning, grow them for a limited number of cycles **(D)** and assay for reporter activity **(E)**. **F.** We predict that the variant episomal promoter that represses the reporter gene will be enriched in H3K9me3, whereas the variant episomal promoter that expresses the reporter gene will be enriched in H3K9ac.

5.2 Article 2

The variantome of the malaria parasite *Plasmodium falciparum*

Núria Rovira-Graells, Archna P. Gupta, Evarist Planet, **Valerie M. Crowley**, Sachel Mok, Lluís Ribas de Pouplana, Peter R. Preiser, Zbynek Bozdech and Alfred Cortés

Submitted manuscript currently under review in Genome Research

The variantome of the malaria parasite *Plasmodium falciparum*

Núria Rovira-Graells[1,4], Archna P. Gupta[2], Evarist Planet[1], Valerie M. Crowley[1], Sachel Mok[2], Lluís Ribas de Pouplana[1,3], Peter R. Preiser[2], Zbynek Bozdech[2,5] and Alfred Cortés[1,3,4,5]

[1]Institute for Research in Biomedicine (IRB), 08028 Barcelona, Catalonia, Spain

[2]School of Biological Sciences, Nanyang Technological University, 637551, Singapore

[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain

[4]Present address: Barcelona Centre for International Health Research (CRESIB, Hospital Clínic-Universitat de Barcelona), 08036 Barcelona, Catalonia, Spain

[5]To whom correspondence should be addressed.

ZB: zbozdech@ntu.edu.sg, tel. +65 6316 2925

AC: alfred.cortes@cresib.cat, tel. +34 93 2275400 ext. 4519

Running title: epigenetic variation in malaria parasites

Keywords: malaria; *Plasmodium falciparum*; epigenetics; transcriptomics; heterochromatin; adaptation; bet-hedging strategies

ABSTRACT

Malaria genetic variation has been extensively characterized, but the level of epigenetic plasticity remains largely unexplored. Here we provide a comprehensive characterization of transcriptional variation in the most lethal malaria parasite, *Plasmodium falciparum*, based on highly accurate transcriptional analysis of isogenic parasite lines grown under homogeneous conditions. This analysis revealed extensive transcriptional heterogeneity within genetically-homogeneous clonal parasite populations. We show that clonally variant expression is an intrinsic property of specific genes and gene families, the majority of which participate in host-parasite interactions. Intrinsic transcriptional variability, controlled at the epigenetic level, is not restricted to genes involved in immune evasion, but also affects genes linked to lipid metabolism, protein folding or transcriptional regulation, among many others. The degree of genetic diversity and the predicted functions of variantly-expressed gene families indicate that epigenetic variation results in both antigenic and functional variation. We observed a general association between heterochromatin marks and clonally variant expression, extending previous observations for specific genes to essentially all variantly-expressed gene families. These results suggest that phenotypic variation of functionally unrelated *P. falciparum* gene families is mediated by a common mechanism based on reversible formation of H3K9me3-based heterochromatin. In changing environments, diversity confers fitness to a population. Our results support the idea that *P. falciparum* uses a bet-hedging strategy, as an alternative to directed transcriptional responses, to adapt to the most common fluctuations in its environment.

Supplemental material is available. The microarray data presented in this study have been submitted to Array Express under accession number E-MTAB-673 with the following accession details: Username: Reviewer_E-MTAB-673, Password: A7R009ai.

INTRODUCTION

Adaptation to environmental changes is essential for survival. In the case of *P. falciparum* asexual blood stages, which are responsible for all clinical malaria symptoms and complications, the intraerythrocytic niche provides a relatively stable environment compared to the environments where free-living unicellular organisms reside, but there are still challenges that the parasite must confront (Mackinnon and Marsh 2010). Among the conditions that can change during the course of a blood infection or between different human hosts are immune pressure, drug pressure, host metabolic and nutritional conditions, presence and magnitude of febrile episodes, host genetics, and presence of competing parasites.

Long-term, slow adaptation typically occurs at the genetic level. In *P. falciparum*, the genetic basis of adaptation have been extensively studied (Mu et al. 2010; Ochola et al. 2010; Van Tyne et al. 2011). However, adaptation to fast-fluctuating environmental pressures requires a much faster time-scale and reversibility, and operates at the transcriptional level. In the majority of organisms, including bacteria, yeast and higher eukaryotes, rapid adaptation typically occurs through directed transcriptional responses: an external condition is sensed, which triggers a signaling cascade that results in transcription factor-driven activation or repression of specific genes that mediate adaptation to the new condition. Some evidence suggests that *P. falciparum* may have a limited capacity to mount directed transcriptional responses: i) there is a marked paucity of specific transcription factors in the *P. falciparum* genome (Coulson et al. 2004), in spite of the recent identification of an expanded family of AP2 domain-containing transcription factors (ApiAP2), which probably play a role in controlling life cycle progression (Balaji et al. 2005; Campbell et al. 2010). The number of identified specific

transcription factors seems insufficient to mount a wide range of transcriptional responses; ii) analysis of transcriptional responses upon environmental perturbations failed to detect large alterations in the levels of transcripts related with the challenge, in contrast to similar studies in other eukaryotes (Gunasekera et al. 2007; Ganesan et al. 2008; Le Roch et al. 2008; Young et al. 2008). This led some authors to conclude that *P. falciparum* has an unresponsive, hard-wired transcriptome (Ganesan et al. 2008; Le Roch et al. 2008). This is controversial because other studies detected changes in transcription upon various challenges (Functional Genomics Workshop Group: Deitsch et al. 2007; Oakley et al. 2007; Natalang et al. 2008; Tamez et al. 2008; Hu et al. 2010), but the changes observed where of low amplitude, and could be explained by cell cycle arrest or sexual differentiation, or may represent pathways leading to parasite death, rather than being authentic protective transcriptional responses (Functional Genomics Workshop Group: Deitsch et al. 2007; Natalang et al. 2008).

Regardless of the ability of *P. falciparum* to mount directed transcriptional responses, one source of transcriptional variability in malaria parasites is clonally variant gene expression. Variantly-expressed genes can be found in different transcriptional states in genetically identical parasites at the same stage of life cycle progression and grown under homogeneous conditions. The transcriptional state is clonally transmitted by epigenetic mechanisms (Scherf et al. 2008). Clonally variant gene expression has a clear adaptative potential, because stochastic changes in expression lead to transcriptionally heterogeneous populations. It is well-established in the population biology field that heterogeneous populations are fitter than homogeneous populations in changing environments, as diversity provides the grounds for natural selection of individuals with phenotypes that confer maximum fitness upon an environmental change (Kussell and

Leibler 2005; Veening et al. 2008). Clonally variant gene expression in *P. falciparum* was initially described for *var* genes (Scherf et al. 2008), a family of about 60 genes encoding PfEMP-1, which plays a major role in immune evasion. Other *P. falciparum* gene families also show clonally variant expression, including gene families encoding proteins exported to the erythrocyte surface, and families linked to erythrocyte invasion. Variant expression of these gene families may also play a role in immune evasion (Cortés et al. 2007; Lavazec et al. 2007; Petter et al. 2007; Cortés 2008; Scherf et al. 2008; Gomez-Escobar et al. 2010).

Adaptation through clonally variant gene expression would imply that parasite populations spontaneously show transcriptional diversity for genes relevant for adaptation to common environmental pressures. While genetic variation in *P. falciparum* has been extensively characterized (Jeffares et al. 2007; Mu et al. 2007; Volkman et al. 2007; Van Tyne et al. 2011), the genome-wide extent of clonally variant gene expression remains unknown. Some studies have compared global blood stages transcription between different *P. falciparum* lines, but their sample size and experimental design resulted in the identification of only a small number of variantly-expressed genes: one study compared two isogenic lines selected to display different antigenic and adhesive properties (Mok et al. 2007), whereas another study compared the transcriptome of three genetically-different parasite lines, focusing on the comparison of patterns of expression along the life cycle rather than comparing transcript levels (Llinas et al. 2006). Genome-wide transcriptional differences have also been analyzed in field isolates (Daily et al. 2007; Lemieux et al. 2009; Mackinnon et al. 2009), but these studies suffer intrinsic limitations such as the inability to distinguish between spontaneous variation and selection, and the common occurrence of mixed

infections. Furthermore, some of the transcript level differences observed may be attributable to genetic polymorphism. A recent study established that transcriptional diversity in *P. falciparum* has a strong genetic component (Gonzales et al. 2008).

Here we comprehensively characterized the asexual blood stages transcriptome of 21 parasite lines, and compared transcript levels within sets of parasite lines sharing a common clonal origin. With this approach, transcriptional variation was studied independently of the genetic component. We identified the majority of inherently-variant gene families, and observed extensive transcriptional heterogeneity among essentially isogenic parasites. Variantly-expressed gene families participate in multiple aspects of host-parasite interactions, including but not limited to immune evasion. We also show that functionally diverse variantly-expressed gene families are regulated by a common epigenetic mechanism involving heterochromatin formation.

RESULTS

Identification of genes under clonally variant expression.

Transcriptional variability was first assessed for two stocks of the clonal parasite line 3D7 maintained in different labs for several years, 3D7-A and 3D7-B, and three subclones of 3D7-A (Fig. 1A, 3D7). 3D7-B differs from 3D7-A and subclones in its ability to use different invasion pathways and to invade red cells with the malaria-protective trait SAO (Cortés et al. 2004; Cortés 2005; Cortés et al. 2007). Subclones were cultured for the minimal number of cycles necessary to obtain sufficient material for transcriptional analysis. For genes with stable clonally-transmitted expression patterns, this is akin to characterizing transcript levels in single parasites within the clonal population from which the subclones were derived (Cortés et al. 2007). The five

parasite lines were tightly synchronized to a 0-5 h window and RNA was obtained at seven time points along the asexual blood cycle. Relative transcript levels were determined by hybridizing samples against a common reference pool in a long-oligonucleotides microarray (Hu et al. 2007). Principal components analysis (PCA) revealed clustering of samples collected at the same time point, as expected for well-synchronized samples. Samples collected at the latest time points approached samples collected at earlier time points, reflecting the cyclic nature of asexual blood growth (Fig. 1B).

The majority of *P. falciparum* genes show stage-specific expression, and even small differences in the stage of cycle progression can lead to large differences in observed transcript levels (Bozdech et al. 2003). This is a major confounder for microarray-based studies comparing expression between different parasite lines. To account for this, we estimated the stage of the cycle for each microarray sample using a statistical likelihood-based method (Lemieux et al. 2009) (Supplemental Fig. S1) and plotted expression levels against estimated times post-invasion (Fig. 1C-D). To quantify, for each gene, the level of transcriptional variability among the parasite lines compared, we developed a score termed aMAFC (adjusted maximum average fold change across half the time interval compared, see Methods), which is a conservative estimate of the expression fold-change between the parasite line that expresses the gene at the highest level and the parasite line that expresses it at the lowest level. Only about 5% of the genes showed substantial transcriptional variation (Fig. 1E). The median aMAFC corresponds to the first gene in Fig. 1C.

We generated a high-confidence list of genes showing different transcript levels among

3D7-derived parasite lines, based on criteria dependent on the aMAFC and permutation tests (see Methods). Comparative genome hybridization (CGH) analysis revealed that in 6 of the 144 genes selected, different transcript levels could be explained by genetic differences (copy number variations, CNVs). Five of these 6 genes fell within a duplication of approximately 56 kb in 3D7-B, which resulted in approximately two-fold transcript levels relative to parasite lines without the duplication (Fig. 1F). After excluding these genes, we generated a list of 138 genes that we called the 3D7 variantome (Supplemental Table S1), defined as the set of genes with clonally variant expression in 3D7. Some of these genes (76 of 138) belong to the large hypervariable families *var*, *rif*, *stevor* and *pfmc-2tm*, for which variant expression was previously described (Scherf et al. 2008) (Supplemental Text and Supplemental Fig. S2-3).

Variant expression in parasites of different genetic backgrounds.

We followed the same approach to identify genes under clonally variant expression in other sets of parasite lines sharing a clonal origin. Each comparison included the parental parasite line, of clonal origin, and several subclones (Fig. 1A). We also compared transcript levels among the genetically different parental (original) lines 7G8, D10, and HB3A (Fig. 1A, parasite lines in bold). 3D7 parental line could not be included in this comparison because 3D7-derived samples were hybridized against a different reference pool (see Methods). In total, we identified genes under variant expression based on five comparisons, four between parasite lines sharing a clonal origin (3D7, 7G8, D10 and HB3 comparisons), and one between genetically different parasites (parentals comparison) (Fig. 1A).

As in the case of the 3D7 comparison, we excluded from the lists of variant genes all

cases where deletions, duplications or other genetic alterations (as determined by CGH) may account for the differences in transcript levels observed. The number of genes excluded was 8, 1, 22 and 15 for the 7G8, D10, HB3 and parental comparisons, respectively (Supplemental Table S1). Furthermore, all genes from the highly polymorphic gene families *var*, *rif*, *stevor* and *pfmc-2tm* were excluded from the analysis of non-3D7 parasite lines, because CGH and blast analysis revealed that this microarray (designed against the 3D7 reference genome) is unsuitable to study expression of these families in non-3D7 parasite lines (Methods and Supplemental Fig. S4). However, it is already well-established from our data with 3D7 parasites and by others (Scherf et al. 2008) that these gene families show extensive clonally variant expression. Even after excluding these large variant gene families from the analysis and further filtering the lists of variant genes (Supplemental Text and Supplemental Fig. S5-7), we identified a substantial number of variant genes in all comparisons (Fig. 2 and Supplemental Table S1). No two single parasite lines analyzed were transcriptionally identical. Transcriptional differences between isogenic subclones indicate that all the parental parasite lines from which they were derived, of clonal origin but grown for many generations after cloning, consist of a mixture of single parasites with different transcriptional patterns, in spite of growing under homogeneous conditions. These results demonstrate that genetically homogeneous clonal populations show high levels of spontaneous epigenetic variation. Importantly, qPCR analysis on cDNA from independent parasite preparations revealed that transcriptional patterns were stable for the majority of genes tested (dots in Fig. 2 and Supplemental Fig. S8), implying stable epigenetic heritability, with expression switches occurring at low frequency.

Overlap in the variant genes identified in the different comparisons.

There was extensive overlap in the genes that showed variant expression in the different comparisons (Fig. 3A), indicating that variant expression is an intrinsic property of certain genes. Eight genes showed variant expression in four or five of the comparisons. More than 40% of the genes that showed variant expression in the 3D7 comparison also showed variant expression in at least one other comparison, and the same was true for the variant genes identified in each comparison (Fig. 3B). Remarkably, transcriptional variability between genetically different parasites seems to affect similar genes as transcriptional variability within clonal parasites populations: 70% of the genes differentially expressed between the genetically different 7G8, D10 and HB3 parental parasite lines (parentals comparison) were also variantly expressed in at least one of the comparisons involving isogenic parasite lines. The cumulative number of variant genes identified increased with the number of comparisons (Fig. 3C). In total, 193 different variant genes were identified (excluding gene families that were only analyzed in 3D7-derived parasite lines; 269 variant genes were identified including these families, Supplemental Table S2).

The overlap between the gene families with variant expression in the different comparisons was even more extensive (Fig. 3D), and the cumulative number of variant gene families identified clearly approached a plateau (Fig. 3E). This result indicates that we may have captured the majority of *P. falciparum* gene families under clonally variant expression.

Gene families under clonally variant expression.

The majority of the variantly-expressed genes (180 out of 269) are members of previously described gene families known or predicted to participate in host-parasite

interactions. In total, we identified 30 gene families with at least two members listed as variantly expressed, or one member listed as variant in two comparisons (Table 1). Even after excluding the dominant variant gene families *var*, *rif*, *stevor* and *pfmc2tm* from the analysis, 104 of 193 variant genes were members of gene families, and 40% had paralogs according to PlasmoDB/OrthoMCL, compared to only 9% in the full genome (two-tailed Fischer's exact test, P=0.000). Functional redundancy in some gene families probably results in permissivity for variant expression.

Control gene sets encoding single-copy, essential genes showed minimal levels of transcriptional heterogeneity in all sets of parasite lines compared (Fig. 4). In contrast, we identified several gene families that showed extensive transcriptional variability in all five comparisons, and gene families such as *apiAP2* or lysophospholipases that showed stable expression levels for the majority of genes, with substantial variant expression restricted to specific members (Fig. 4 and Table 1). These results indicate again that variant expression is an intrinsic property of some genes and gene families.

Many of the variant gene families are subtelomeric and encode proteins exported to the erythrocyte (Table 1), including exported HSP40-type co-chaperones (*dnaj*) (Sargeant et al. 2006), predicted exported protein kinases (*fikk*) (Nunes et al. 2007), and exported proteins of unknown function (such as *phist*, *hyp*, and *gbp* families) (Sargeant et al. 2006). Additionally, variant expression was observed in gene families that encode proteins that participate in lipid metabolism, such as acyl-CoA synthetases (*acs*), acyl-CoA binding proteins (*acbp*) (Bethke et al. 2006) and lysophospholipases; proteins involved in nutrients import and possibly in erythrocyte invasion (*clag*) (Cortés 2008; Nguitragool et al. 2011); transcription factors (*apiAP2*) (Campbell et al. 2010); proteins

involved in transmission (*6-cys*) (Templeton and Kaslow 1999); and proteins located at the parasite-host cell interface (*etramp*) (Spielmann et al. 2003) (Table 1). Surprisingly, many tRNAs were also expressed at different levels between different parasite lines. Other genes were not part of any of the variant gene families included in Table 1, but also showed clonally variant gene expression. Among those, there were 16 genes encoding exported proteins (Sargeant et al. 2006); the gene *eba-140*, and the pseudogene *PfRh3* from families linked to erythrocyte invasion (Cortés 2008); *surfin4.1* (Winter et al. 2005); *hrpIII* from a gene family encoding secreted proteins with functions in hemozoin formation and with procoagulant activity (Ndonwi et al. 2011); three genes encoding putative Zn-finger proteins; genes related with gametocyte formation, for which expression levels in asexual parasites may determine the propensity to enter sexual differentiation (Supplemental Text); two more enzymes linked to lipid metabolism (beta-hydroxyacyl-ACP dehydratase and dephospho-CoA kinase); two tubulin-tyrosine ligases; and 46 genes encoding proteins annotated as "unknown function" (Supplemental Table S2).

Gene set enrichment analysis (GSEA) (Subramanian et al. 2005) did not reveal any additional gene family, metabolic pathway, biochemical process or protein domain associated with high levels of variant expression (not shown). However, genes that have paralogs, lack identified orthologs in other species, are located in subtelomeric regions, are pseudogenes, or encode proteins with transmembrane domains, showed a significantly skewed distribution towards high levels of transcriptional variability (Supplemental Fig. S9).

Epigenetic mechanisms regulating clonally variant gene expression.

Many of the genes that showed variant expression were found to be positive for the heterochromatin marks H3K9me3 and HP1 in previous chip on ChIP-based mapping studies (Flueck et al. 2009; Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009). Only 4 of the 30 variant gene families included in Table 1 were HP1 and H3K9me3 negative. Likewise, of 35 gene families reported to be marked by HP1 or H3K9me3, only 5 small families did not have any member showing clonally variant expression (Fig. 5A). The striking correlation between heterochromatin marks and variant expression at the gene family level was also observed at the individual gene level within some families such as lysophospholipase, *apiAP2*, *acbp* (Fig. 4), or *clag*. These results strongly suggest a general mechanistic link between H3K9me3-based heterochromatin and phenotypic variation. Previous studies showed that variant expression of *var* genes and genes encoding merozoite proteins is mediated by reversible formation of heterochromatin-like structures, with H3K9me3 associated to the repressed state (Chookajorn et al. 2007; Lopez-Rubio et al. 2007; Jiang et al. 2010; Comeaux et al. 2011; Crowley et al. 2011). Our data suggest that this is a general mechanism also controlling variant expression of other functionally unrelated gene families, including members of non-subtelomeric gene families such as *apiAP2*. However, the majority of the clonally variant genes that are not members of gene families and do not encode exported proteins were not marked by H3K9me3 or HP1, and may not be regulated by heterochromatin formation (Supplemental Table S2). Specific transcription factors, such as an ApiAP2 (Campbell et al. 2010) that showed variant expression (PFL1085w) and is HP1/H3K9me3 positive (Flueck et al. 2009; Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009) (Fig. 5B), may be master regulators controlling the variant expression of other genes.

The stage of expression during the cycle is not coregulated along the chromosomes

(Bozdech et al. 2003). However, our results indicate that, at least for gene families involved in host-parasite interactions, and in contrast with developmental regulation, clonally variant expression is regulated by heterochromatin formation. H3K9me3-based heterochromatin has the capacity to spread into adjacent regions (Grewal and Jia 2007), which may result in coordinated repression of neighboring genes. Many of the genes under variant expression cluster together in subtelomeric regions, but we observed little evidence for coordinated expression patterns. Different patterns of expression were often observed for genes within the same variant expression cluster (Fig. 5B), and stand alone variant genes were observed in pairwise parasite lines comparisons (Fig. 5C-D). Possible coordinated repression along a large chromosome region was only observed in a subtelomeric region of chromosome 14 (Fig. 5E), which includes several genes involved in gametocyte formation (Eksi et al. 2005), but further along the chromosome neighboring genes showed opposite expression patterns in some parasite lines (Fig. 5E). Our results are consistent with the fact that activation of a *var* gene does not de-repress adjacent genes (Scherf et al. 2008), and with recent data indicating that *var* genes are repressed at the level of an individual gene, rather than by chromatin spreading (Swamy et al. 2011). The general lack of coordinated expression in clusters of neighboring genes that can be repressed by formation of heterochromatin indicates that subtelomeric regions are organized into small chromatin domains, possibly delimitated by barrier insulators (Gaszner and Felsenfeld 2006), as previously suggested by us (Crowley et al. 2011). This would allow independent variant expression of neighboring genes, maximizing the number of possible combinations of expressed and repressed genes and consequently the adaptative potential of clonally variant gene expression.

DISCUSSION

Here we present, to our knowledge, the most comprehensive characterization of transcriptional variation in the malaria parasite *Plasmodium falciparum*. Our experimental design and analysis resulted in unprecedented accuracy in the transcriptional comparison of different parasite lines, which allowed us to distinguish between authentic clonally transmitted transcriptional differences, and variation attributable to genetic diversity or to different relative abundance of parasites at different stages. This approach led to the identification of the majority of gene families under variant expression, and demonstrated that genetically homogeneous parasite populations derived from a single parasite (clonal) are epigenetically highly heterogeneous. The extensive overlap in the genes and gene families that show variant expression in comparisons involving different sets of parasites indicate that spontaneous clonally variant expression is an intrinsic property of some genes and gene families. Data showing expression variability for each *P. falciparum* gene, and the subclones with defined transcriptional patterns, are important resources for both basic malariologists and vaccine developers that will be available through PlasmoDB (www.plasmodb.org) and MR4 (www.mr4.org).

The critical aspects of our study design and analysis that allowed identification of the majority of inherently-variant gene families were: i) transcriptional comparisons involving parasite lines of the same clonal origin, including recently subcloned parasite lines; ii) transcriptional comparison along the full asexual blood cycle for a large number of parasite lines (over 150 microarray hybridizations), belonging to four different genetic backgrounds; iii) tight synchronization of parasites to a well-defined narrow age window of 0-5 h post-invasion, completely avoiding contaminating parasites

from the previous cycle; iv) identification of transcriptional differences based on expression plots with statistically-estimated times, and calculation of adjusted expression fold-changes from the area underneath the curve of the plots (aMAFC score); v) exclusion of different probes for the analysis of parasites of different genetic backgrounds, based on CGH and blast analysis against partially-sequenced genomes; and vi) CGH analysis of all parasite lines, to avoid the potential confounding effect of genetic alterations. This workflow provides the necessary level of accuracy to identify with high-confidence and sensitivity epigenetic differences between independent parasite lines.

The vast majority of gene families that are positive for heterochromatin marks also show variant expression, demonstrating a general association between H3K9me3-based heterochromatin gene repression and phenotypic variation. This observation reveals a scenario where specific regions of the largely euchromatic *P. falciparum* genome are organized into bistable chromatin domains. These regions can be found in two alternative clonally transmitted states, euchromatin or heterochromatin, which are associated with transcriptional activity or repression, respectively. Bistable chromatin is prone to infrequent spontaneous switches between its two possible states, as a consequence of molecular noise (Dodd et al. 2007; Crowley et al. 2011). The striking correlation between variant expression and heterochromatin marks allows us to predict that the majority of the genes that were positive for heterochromatin marks in previous studies may be variantly expressed, implying that about 10% of the genome may be under clonally variant expression. Considering that we observed only very few cases of co-regulated variant expression, this represents an unlimited number of possible combinations of expressed and repressed genes, with the exception of families such as

var or *clag* where specific rules of mutually-exclusive expression apply (Cortés et al. 2007; Scherf et al. 2008; Comeaux et al. 2011; Crowley et al. 2011).

Our results provide a snapshot of transcriptional diversity within clonal parasite populations, revealing that variable expression mainly affects gene families involved in host-parasite interactions. The degree of genetic diversity between isolates in these gene families provides insights into the physiological role of transcriptional variability. Gene families specifically amplified in the *P. falciparum* lineage can be roughly divided into those that show hypervariability between parasite isolates, such as *var*, *rif*, *stevor* and *pfmc-2tm*, and those that are conserved between isolates, including the majority of other amplified families (eg *phist*, *dnaj*, *fikk*, *etramp*, *acs*, lysophospholipases, and many others) (Jeffares et al. 2007; Mu et al. 2007; Volkman et al. 2007; Templeton 2009). Amplification of these two types of gene families was probably driven by immune evasion and functional diversification, respectively (Templeton 2009). We clearly observed clonally variant expression for both types of gene families, suggesting that transcriptional variability results not only in antigenic variation but also in functional variation.

This idea is further supported by the known or predicted function of the gene families for which transcriptional variability is an intrinsic property, and is suggestive for a role of variant expression in *P. falciparum* adaptation. Variant expression of *var* genes mediates immune evasion, and the same may be true for other hypervariable gene families such as *rif*, *stevor* and *pfmc-2tm* (Scherf et al. 2008). Other gene families for which we observed variant expression participate in metabolic processes such as lipid metabolism or nutrients import (*acs*, *acbp*, lysophospholipases and *clag*), protein

folding/stability (predicted co-chaperones of families *dnajIII* and *phistb/dnaj*), or erythrocyte invasion (*eba* and possibly *clag*). Expression of alternative members of these families, possibly with different functional abilities, has the potential to mediate adaptation to changes in host metabolic/nutritional conditions, changes in temperature associated to febrile episodes, or host genetic polymorphisms in erythrocyte receptors, respectively, which are probably the major fluctuating environmental pressures that *P. falciparum* faces in its intraerytrhocytic niche beyond immune pressure (Merrick and Duraisingh 2010). Importantly, two of the transcriptionally different parasite lines characterized in this study differ in their ability to use alternative invasion pathways or to invade erythrocytes with the malaria protective mutation SAO (Cortés et al. 2004). Of special interest is variant expression of the expanded gene family *acs*. This gene family encodes enzymes that participate in host-parasite interactions by mediating activation of fatty acids scavenged from the host, with different specificities predicted for different members of the family (Tellez et al. 2003; Bethke et al. 2006). Also supporting the idea that variant expression can modulate fitness and hence permit adaptation upon challenge, switches in the expression of *clag3* genes were recently shown to result in altered sensitivity to chemical inhibitors (Nguitragool et al. 2011). Several other conserved (between isolates) gene families, predominantly encoding proteins at the host-parasite interface or exported to the erythrocyte, show clear evidence of variant expression. The poor functional knowledge of these gene families precludes prediction of the functional role and adaptative potential of their variant expression. In any case, the extensive number and the characteristics of the gene families under clonally variant gene expression indicate that they have the potential to mediate adaptation to the major fluctuating conditions encountered by *P. falciparum* asexual blood stages, clearly beyond immune evasion.

Adaptation through clonally variant gene expression is fundamentally different from adaptation through directed transcriptional responses (Supplemental Fig. S10). We propose that spontaneous, stochastic switches between the active and repressed states for a large number of genes generates transcriptional diversity within clonal parasite populations before any challenge is applied, thus representing a bet-hedging (risk-spreading) strategy (Veening et al. 2008). Upon a change in the environment, transcriptional heterogeneity provides the grounds for natural selection of pre-existing parasites with transcriptional patterns that confer maximum fitness under the new conditions. Diversity confers fitness to a population in changing environments, and mathematical models predict that diversity generated by stochastic variation is favored over sensing followed by transcriptional responses when the environment changes infrequently (Kussell and Leibler 2005), as in the case of the environment where *P. falciparum* blood stages reside. Adaptation through a stochastic phenotype-switching mechanism, controlled at the epigenetic level, occurs at a much faster time-scale than genetic adaptation, and is both heritable and reversible. Our results showing extensive transcriptional variation for genes involved in the interaction of the parasite with its environment provide strong support to the idea that bet-hedging strategies play a predominant role in *P. falciparum* adaptation. While adaptation via spontaneous clonally variant expression/selection and adaptation via directed transcriptional responses are not mutually exclusive, the former strategy may compensate for the reported limited ability of *P. falciparum* to mount immediate directed responses. This is a testable hypothesis that has important implications not only to understand how this devastating parasite adapts to natural fluctuating environmental conditions, but also to predict how it may evolve in front of renewed efforts to control or eradicate the disease

(Mackinnon and Marsh 2010). Carefully designed future studies will need to identify the variantly-expressed genes that may confer a growth advantage under specific physiologically-relevant conditions. Any studies aimed at identifying genes that mediate adaptation to specific environmental conditions should be designed and interpreted considering that in *P. falciparum* adaptation may require a time-scale consistent with selection of preexisting epigenetic variants.

METHODS

Parasites.

Clonal parasite lines 3D7, 7G8, D10 and HB3 have been previously described (Anders et al. 1983; Burkot et al. 1984; Walliker et al. 1987). 3D7 stocks 3D7-A and 3D7-B, and 3D7-A subclones have also been described before (Cortés et al. 2004; Cortés 2005; Cortés et al. 2007). HB3A and HB3B are stocks of the same HB3 clonal parasite line obtained before (HB3A) or after (HB3B) mosquito and chimpanzee passage (Walliker et al. 1987), kindly provided by O. Kaneko (Nagasaki University, Japan). 7G8 parental parasite line and 7G8 subclones identified as 7G8 self-fertilization events in a 7G8xGB4 genetic cross (Hayton et al. 2008) were obtained from MR4 (www.mr4.org). Other subclones were generated in this study by limiting dilution. We cultured subclones for as few cycles as possible between subcloning and transcriptional analysis.

Parasites were cultured under standard conditions in media containing Albumax II (Invitrogen) and no human serum, in a 5% CO_2, 3% O_2, 92% N_2 atmosphere. Tight synchronization was achieved by Percoll (3D7 parasite lines) or magnet (other parasite lines) purification of schizonts followed by sorbitol lysis five hours later, to obtain a population of a defined age window of 0-5 h post invasion. Parasites were split into 6 or

7 dishes that were cultured undisturbed for different periods of time (10, 20, 30, 34, 37, 40 or 43h) before collecting in Trizol and freezing at -80ºC.

Microarray hybridizations.

RNA was purified using the Trizol method, and cDNA synthesized by reverse transcription, purified, labeled and hybridized as previously described (Bozdech et al. 2003). Microarray images were analyzed using GenePix Pro 6.0 (Axon Instruments), with careful visual inspection to refine automatic spot definitions and to eliminate poor quality spots. All test samples (labeled with Cy5) were hybridized against a reference pool (labeled with Cy3), which consisted of a mixture of cDNA from 3D7-A and 3D7-B rings, trophozoites and schizonts in equal amounts for samples from 3D7-derived parasite lines, or from D10, 7G8 and HB3 for all other samples. For CGH analysis, sonicated gDNA of test parasite lines was labeled with Cy5 and hybridized against the same amount of 3D7-A gDNA labeled with Cy3.

Data analysis.

All data was processed using Bioconductor in an R environment. Spots with Cy3 and Cy5 intensities lower than 1.5 fold times the microarray background were eliminated. When a spot had intensity higher than 1.5 times the background for only Cy3 or Cy5, the other channel was assigned a value of 1.5 times its microarray background, to avoid extreme ratio values. After background subtraction, the $\log_2$ of the Cy5/Cy3 intensity ratio was normalized using LOESS. Gene level $\log_2$(Cy5/Cy3) expression ratios were computed from probe level values using median polish. We calculated the most likely time post-invasion along the asexual cycle for each sample using a statistical likelihood method (Lemieux et al. 2009), with the HB3 transcriptome as reference (Bozdech et al.

2003). Expression values for each gene [$\log_2$(Cy5/Cy3)] were plotted against estimated times for each parasite line. For genes with values for at least five time points, missing value(s) where estimated using linear models predictors. Estimated values were less than 1% of the total.

To compare expression levels among a given set of parasite lines and to identify genes under variant expression, the area under the curve of the time-course plots was computed for the time interval between the latest first time point and the earliest last time point among the parasite lines compared (time compared, TC). For each gene, areas were only calculated for parasite lines that had values for all time points (either experimental or estimated). For developmentally-regulated genes that are only expressed during part of the asexual cycle (Bozdech et al. 2003), it is expected that potential differences in expression levels would be observed only at stages when the genes are expressed. To prevent a bias against such genes, areas were also computed for the following overlapping half-TC intervals: 0 to 0.5x TC, 0.5x to1x TC, 0.25x to 0.75x TC, and the complementary to the latter, 0 to 0.25x TC + 0.75x to 1x TC. The area difference (AD) for a given half-TC interval was the difference between the largest and smallest areas among the parasite lines compared.

Different sets of parasite lines compared showed different levels of background AD attributable to technical variability, and background AD was also different between the four half-TC intervals. To account for this, we calculated for each gene an adjusted maximum average fold-change across half-TC (aMAFC). Because expression values were in $\log_2$ scale, aMAFC was calculated as: aMAFC=2^[max(AD-B)/0.5xTC], where max refers to the maximum of AD-B among the four half-TC intervals, and B is the

background level of area differences. B was estimated as the 80th percentile of AD for each comparison and each half-TC interval (calculated excluding *var*, *rif*, *stevor* and *pfmc2tm* genes, which were eliminated for the analysis of non-3D7 parasites). The 80th percentile was selected based on the assumption that in all comparisons less than 20% of the genes have authentic variant expression. Thus, the 80th percentile reflects the high-end of area differences attributable to experimental variability. Comparison of the distribution of aMAFC with non-adjusted maximum average fold-changes (MAFC, calculated as aMAFC but without the B parameter) demonstrates that the adjustment corrects for the different levels of technical variability in the different comparisons (Supplemental Fig. S11).

We set up empirical criteria to produce high-confidence lists of genes showing variant transcript levels. These criteria were based on aMAFC and permutation tests with 10,000 permutations of parasite line labels within each time point, not allowing two consecutive labels from the same parasite line. The *P*-values, which correspond to the probability of a random permutation generating a higher value than the observed AD, were corrected for multiple testing using Benjamini and Hochberg method. For the comparison of sets of parasite lines with the same clonal origin, all genes with aMAFC higher than 2.5 were included in the lists of variantly-expressed genes. In all cases, these differences were larger than the average plus 3 standard deviations of aMAFC in non-variant genes. Genes with aMAFC between 1.75 and 2.5 were only included in the lists if the adjusted *P* value of the permutation test was $P<0.05$. Genes with aMAFC lower than 1.75 were not included in the lists, even if the permutation test gave a significant result, because such small changes in expression are unlikely to have biological relevance. For the comparison of non-isogenic parasite lines (parentals

comparison), which show higher levels of technical variability, as expected, and a different distribution of aMAFC (Supplemental Fig. S11), only genes with aMAFC > 3.5 were retained in the list of variant genes. Finally, plots for variant genes were visually inspected and genes with differences in expression based only on data points with low absolute Cy5 signal intensity (less than 3x background) were removed from the high-confidence lists of variantly-expressed genes, as near background values provide lower confidence.

We also normalized Cy5 intensity values (no ratio) using quantile normalization among samples collected at the same time point. These values only give a semi-quantitative estimation of the level of expression of a gene, because different oligonucleotides have intrinsically different hybridization efficiencies, and in these arrays each gene is represented by a small number of probes (average 1.9 probes/gene). Gene expression values were calculated as the median of values for the different probes.

Heat maps and hierarchical clustering were performed using TMEV 4.4 (Saeed et al. 2006). Chromosome plots were constructed using IGV (Robinson et al. 2011). GSEA analysis was conducted using GSEA Preranked (Subramanian et al. 2005).

Microarray reannotation.

The microarray used in this study is based on long oligonucleotides (70 nt) designed against the PlasmoDB 4.4 release of the 3D7 reference *P. falciparum* genome (Hu et al. 2007). We reannotated the microarray against the 3D7 PlasmoDB 7.0 genome release. The oligonucleotides in the array were blasted against PlasmoDB 7.0 transcripts (excluding non-coding RNAs) using blast 2.2 default parameters. Based on the

comparison of blast scores with CGH values (Supplemental Fig. S4A), we selected stringent cut-off blast scores to exclude probes with low first blast scores (<100, 70 oligonucleotides) or high second blast scores (>70, 258 oligonucleotides), because they may not hybridize efficiently or they may cross-react with other genes, respectively. Furthermore, 424 of the remaining probes were reassigned to a new gene ID (Supplemental Table S5).

Additional probes were excluded for the analysis of parasites genetically different from 3D7, as some probes fall within highly polymorphic regions of the genome. All oligonucleotides in the microarray were blasted against supercontigs of the incomplete genomes of HB3, D10 and 7G8 (http://www.broadinstitute.org/annotation/genome/plasmodium_falciparum_spp) (Volkman et al. 2007). Probes for the gene families *var*, *rif*, *stevor* and *pfmc-2tm* were excluded, because both blast and CGH analysis revealed that this microarray is not suitable for the analysis of these hypervariable gene families in non-3D7 parasite lines (Supplemental Fig. S4B-C). Furthermore, for each non-3D7 genetic background we excluded probes with low blast score and low CGH value (exclusion criteria are detailed in Supplemental Fig. S12). We performed a stringent test for the criteria used to exclude probes, and found that they resulted on correct exclusion of probes even for genes that are polymorphic between parasite isolates, but have very similar paralogs (Supplemental Table S5 and Supplemental Fig. S12).

ACKNOLEDGMENTS

This research was funded by Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I), Instituto de Salud Carlos III – Subdirección General de Evaluación y Fomento de la Investigación grant PI07/0891 and Spanish Ministry of

Science and Education grant SAF2010-20111 to AC, Singaporean National Medical Research Council grant #IRG07Nov030 to ZB, and European Union FP7 grant HEALTH-F3-2009-223024 MEPHITIS to LRdP. We thank D. Rossell for useful discussion on data analysis, J.E. Lemieux for assistance with the statistical estimation of parasite age, O. Kaneko for providing parasite lines HB3A and HB3B, and MR4 for providing 7G8-derived parasite lines. We are also grateful to A. Casali, A. Vaquero, S. Pagans and M. Llinas for useful comments on the manuscript.

AUTHOR CONTRIBUTIONS

AC and ZB conceived and designed the experiments. NR, AC, VMC and APG performed the experiments. AC and EP analyzed the data. SM, PRP and LRdP contributed materials/analysis tools/discussion of the data. AC wrote the manuscript.

FIGURE LEGENDS

Figure 1. Identification of genes under clonally variant expression. **(A)** Parasite lines used in this study. Parental parasite lines 3D7, 7G8, D10 and HB3A are of clonal origin. 3D7-A and 3D7-B are stocks of 3D7 maintained in different labs for some years (Cortés et al. 2004). 7G8 subclones were identified as self-fertilization events after passage through mosquito and splectomized chimpanzee (Hayton et al. 2008). HB3B was derived from self-crossing HB3A (Walliker et al. 1987). Parentals comparison included the three parasite lines shown in bold (7G8, D10 and HB3A parental parasite lines). **(B)** Principal components analysis (PCA) of the time-course transcriptional analysis of 3D7 parasite lines. Samples from the same parasite line are represented by the same symbol, whereas samples collected at the same time point are represented by the same color. **(C)** Time-course expression plots of representative genes not showing clonally variant

expression. Expression levels (relative to the reference pool), in $\log_2$ scale, are plotted against statistically-estimated culture age (in hours post-invasion). The first gene has the median aMAFC, whereas the other three genes are single-copy genes commonly used as controls (*rhopH2*, *seryl tRNA synthetase* and *ptex150*, in this order). **(D)** Time-course expression plots, as in C, but for genes showing clonally variant expression. The genes belong to the *phista*, *clag*, *acs* and *hrp* families, in this order. **(E)** Distribution of aMAFC. Genes are ranked by their aMAFC in descending order. **(F)** Duplication of a region of chromosome 10 in 3D7-B. Transcript (blue) and gDNA (red) levels in 3D7-B relative to 3D7-A ($\log_2$ ratio) are shown for the second half of chromosome 10. The names of the first and last genes duplicated in 3D7-B are shown. Time-course expression plots for two of the genes within the duplication are shown.

Figure 2. Overview of genes showing variant expression. Expression patterns are shown for the variant genes identified in the 3D7 **(A)**, 7G8 **(B)**, D10 **(C)**, HB3 **(D)**, and parental lines **(E)** comparisons. The numbers in parenthesis indicate the number of variant genes identified in each comparison. Genes from the families *var*, *rif*, *stevor* and *pfmc-2tm* are excluded from the 3D7 heatmap, to make it comparable with the other parasite lines. Values are expression fold-change (in $\log_2$ scale) relative to the average of the parasites lines compared within each set. Black dots indicate genes showing the same transcriptional patterns in qPCR analysis of independent biological preparations, whereas red dots are genes that showed different patterns (Supplemental Fig. S8 and Supplemental Table S4).

Figure 3. Overlap in the variantly-expressed genes identified in the different comparisons. **(A)** Genes (in rows) that showed variant expression in the 3D7, 7G8, D10,

HB3 or parentals comparisons are shown in yellow in the corresponding lane. Genes from the families *var*, *rif*, *stevor* and *pfmc2tm* were excluded from this analysis, because they were only analyzed in 3D7-derived parasite lines. **(B)** Percentage of variant genes identified in each comparison that showed variant expression in at least one other comparison. **(C)** Cumulative number of variantly-expressed genes identified with increasing number of comparisons. Values are the average, with 95% ci, of all possible combinations of comparisons. **(D** and **E)** Same as panels A and C, but at the gene family level. Gene families are defined in Supplemental Table S3.

Figure 4. Distribution of variant expression in *P. falciparum* gene families. The aMAFC for each comparison is shown for the genes of selected families. Aminoacyl-tRNA-synthetases (*ars*) and genes encoding the basal transcription machinery (Transcr.) (Bischoff and Vaquero 2010) provide appropriate controls as gene sets encoding essential genes, for which variant expression is not expected. For *phist*, *dnaj* and *hyp* families, the specific family identity is indicated at the right of the heatmaps. Gene families are detailed in Supplemental Table S3. The tracks at the right indicate whether a gene was included in the list of variant genes, whether it was positive for heterochromatin markers HP1 or H3K9me3 in any of the published reports (Flueck et al. 2009; Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009), and whether it is subtelomerically located (less than 150 kb from a chromosome end). Genes that show high aMAFC in some of the tracks but are not in the list of variant genes were excluded because changes in expression were attributable to CNVs, or for other reasons detailed in the text.

Figure 5. Heterochromatin-based regulation of variant expression. **(A)** Correlation

between variant expression and heterochromatin marks at the gene family level. Previously described *P. falciparum* gene families (Supplemental Table S3) are considered variant if at least one member showed variant expression, and HP1/H3K9me3 positive if at least one member had one of these heterochromatin marks in any of the published reports (Flueck et al. 2009; Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009). Dark yellow indicates gene families with only one gene showing variant expression in only one of the comparisons (gene families excluded from Table 1). **(B-E)** Chromosomal distribution of variantly-expressed genes. The top lane (green) is the aMAFC in the 3D7 comparison. The next lanes represent the relative transcript (blue) and gDNA (red) levels in pairwise parasite lines comparisons, expressed as the $\log_2$ ratio of the fold-change. The bottom lane (black) indicates HP1 or H3K9me3 positive genes in previous studies (Flueck et al. 2009; Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009).

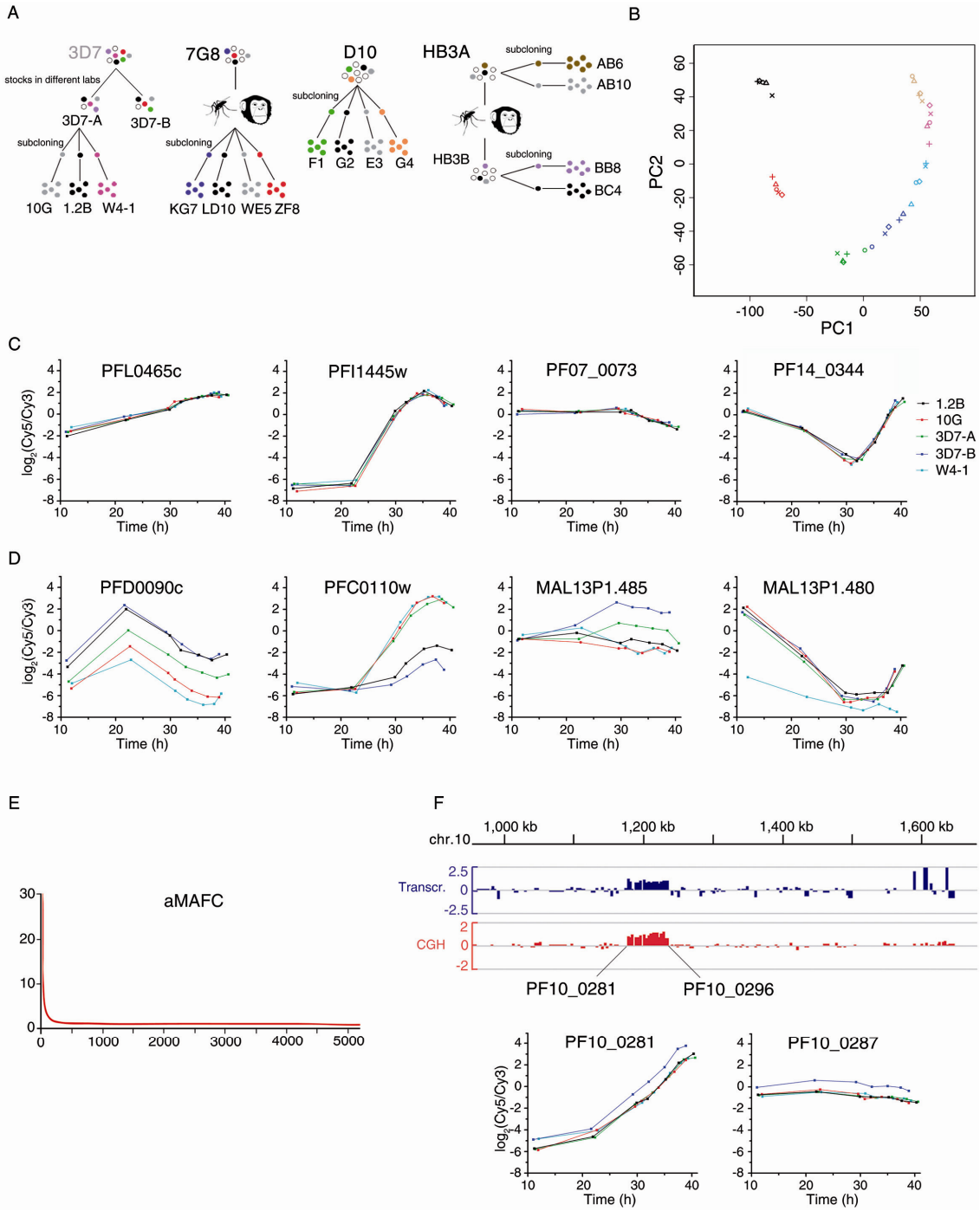

A
3D7
stocks in different labs
3D7-A
3D7-B
subcloning
10G 1.2B W4-1
7G8
subcloning
KG7 LD10 WE5 ZF8
D10
subcloning
F1 G2 E3 G4
HB3A
subcloning
AB6
AB10
HB3B
subcloning
BB8
BC4
B
PC2
PC1
C
PFL0465c
PFI1445w
PF07_0073
PF14_0344
log2(Cy5/Cy3)
Time (h)
1.2B
10G
3D7-A
3D7-B
W4-1
D
PFD0090c
PFC0110w
MAL13P1.485
MAL13P1.480
E
aMAFC
F
chr.10
1,000 kb
1,200 kb
1,400 kb
1,600 kb
Transcr.
CGH
PF10_0281
PF10_0296
PF10_0281
PF10_0287

FIGURE 1

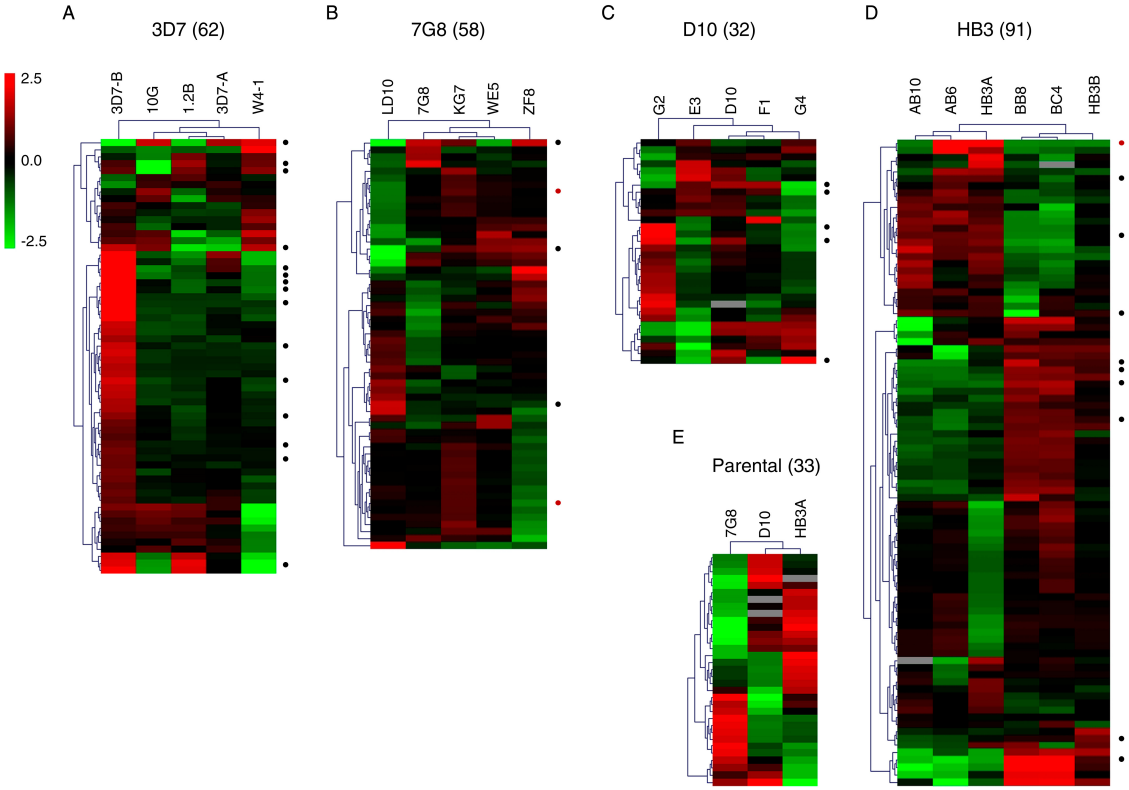

A
3D7 (62)
2.5
0.0
-2.5
3D7-B
10G
1.2B
3D7-A
W4-1
B
7G8 (58)
LD10
7G8
KG7
WE5
ZF8
C
D10 (32)
G2
E3
D10
F1
G4
D
HB3 (91)
AB10
AB6
HB3A
BB8
BC4
HB3B
E
Parental (33)
7G8
D10
HB3A

FIGURE 2

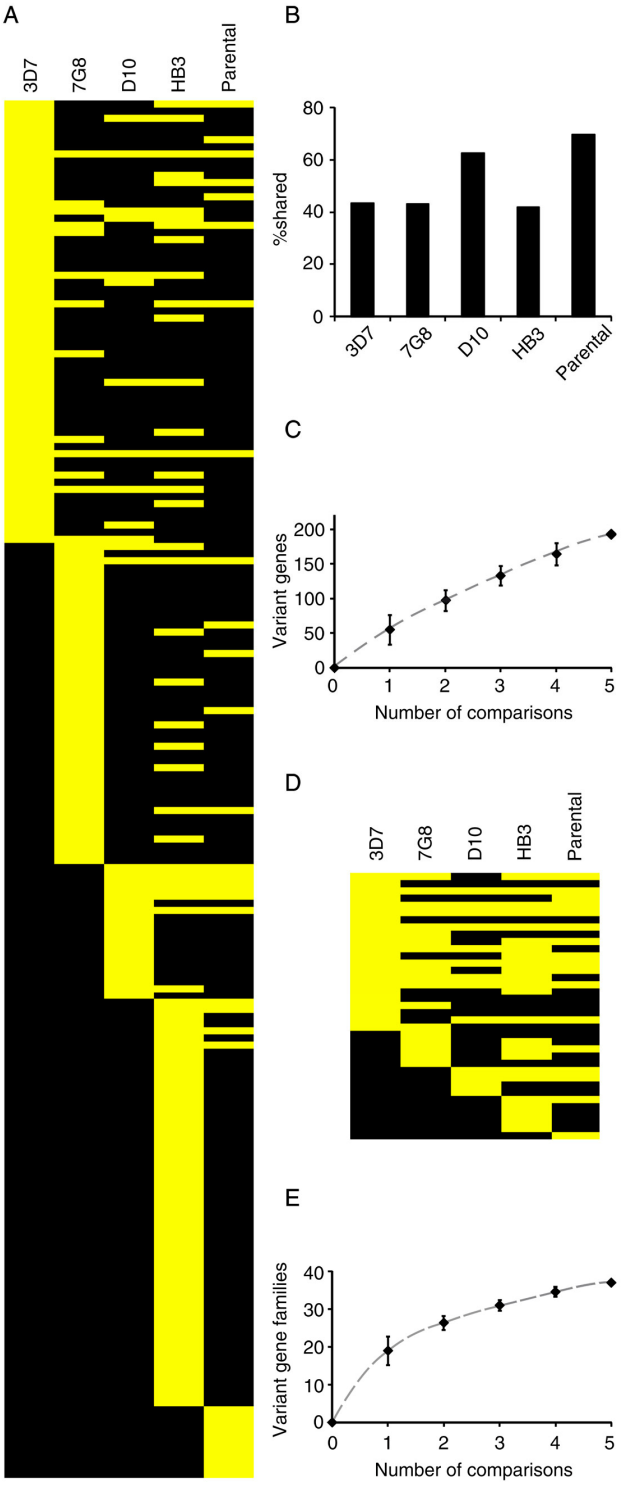

A
3D7
7G8
D10
HB3
Parental
B
%shared
80
60
40
20
0
3D7
7G8
D10
HB3
Parental
C
Variant genes
200
150
100
50
0
0
1
2
3
4
5
Number of comparisons
D
3D7
7G8
D10
HB3
Parental
E
Variant gene families
40
30
20
10
0
0
1
2
3
4
5
Number of comparisons

FIGURE 3

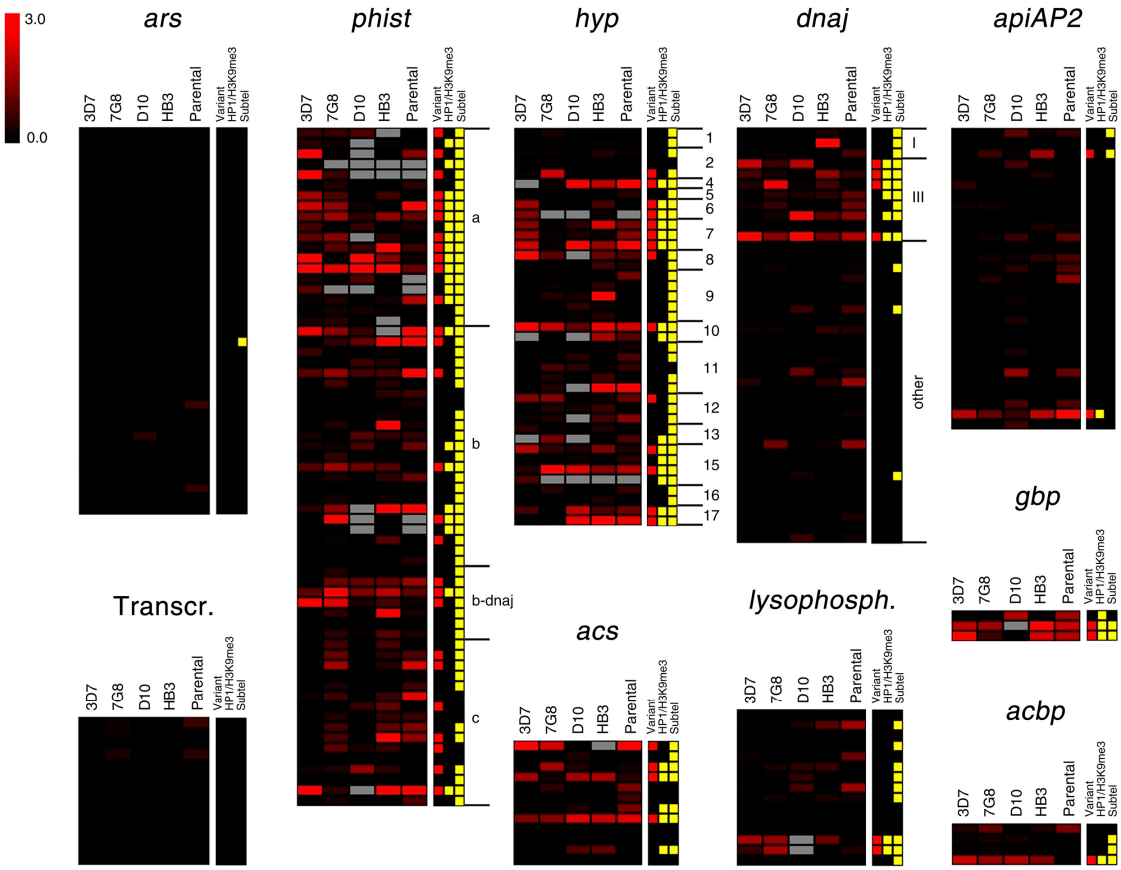

3.0
0.0
ars
phist
hyp
dnaj
apiAP2
Transcr.
acs
lysophosph.
gbp
acbp
3D7
7G8
D10
HB3
Parental
Variant
HP1/H3K9me3
Subtel
a
b
b-dnaj
c
1
2
4
5
6
7
8
9
10
11
12
13
15
16
17
I
III
other

FIGURE 4

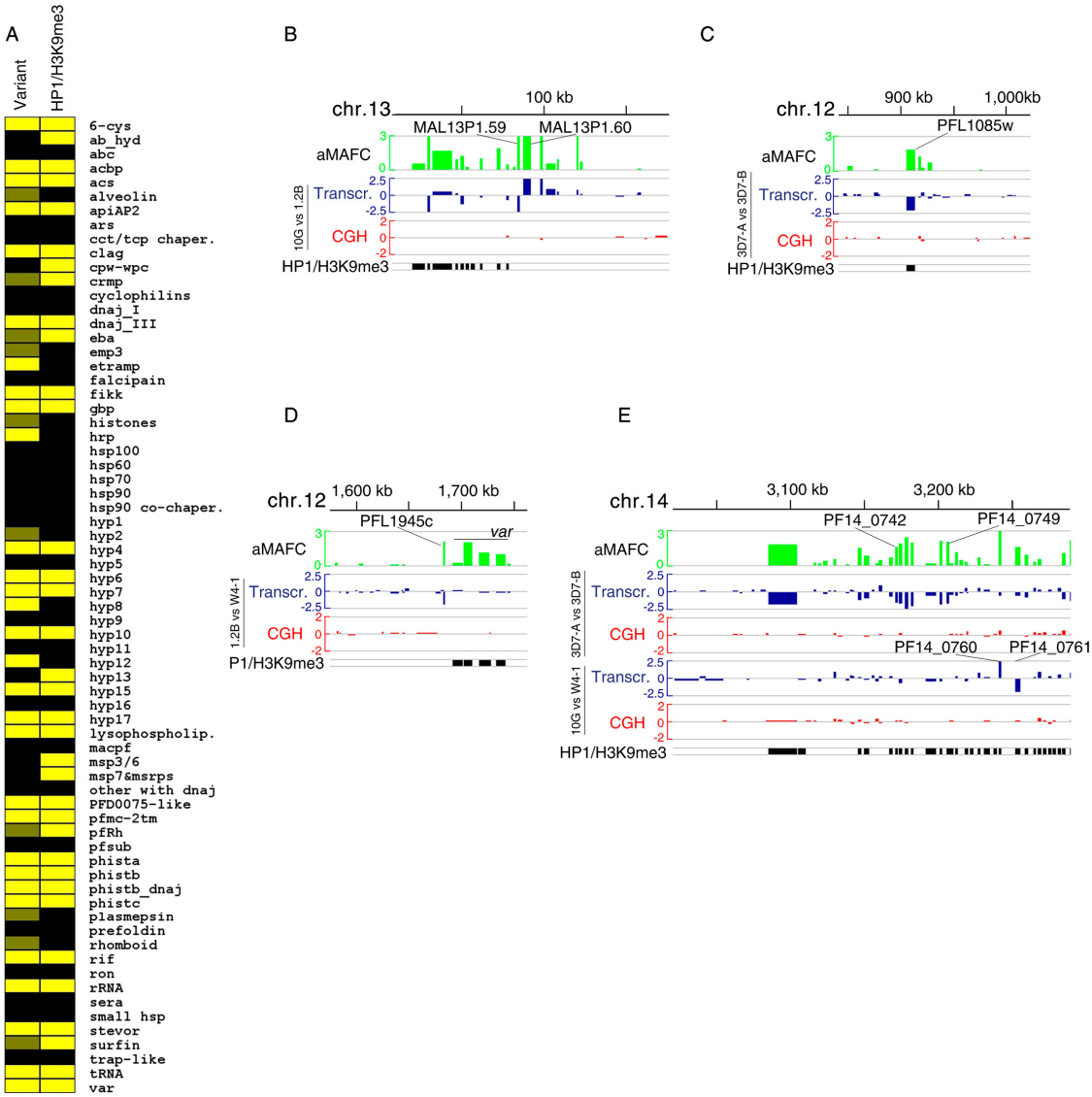

A
Variant
HP1/H3K9me3
6-cys
ab_hyd
abc
acbp
acs
alveolin
apiAP2
ars
cct/tcp chaper.
clag
cpw-wpc
crmp
cyclophilins
dnaj_I
dnaj_III
eba
emp3
etramp
falcipain
fikk
gbp
histones
hrp
hsp100
hsp60
hsp70
hsp90
hsp90 co-chaper.
hyp1
hyp2
hyp4
hyp5
hyp6
hyp7
hyp8
hyp9
hyp10
hyp11
hyp12
hyp13
hyp15
hyp16
hyp17
lysophospholip.
macpf
msp3/6
msp7&msrps
other with dnaj
PFD0075-like
pfmc-2tm
pfRh
pfsub
phista
phistb
phistb_dnaj
phistc
plasmepsin
prefoldin
rhomboid
rif
ron
rRNA
sera
small hsp
stevor
surfin
trap-like
tRNA
var
B
chr.13
100 kb
MAL13P1.59
MAL13P1.60
aMAFC
Transcr.
CGH
10G vs 1.2B
HP1/H3K9me3
C
chr.12
900 kb
1,000kb
PFL1085w
aMAFC
Transcr.
CGH
3D7-A vs 3D7-B
HP1/H3K9me3
D
chr.12
1,600 kb
1,700 kb
PFL1945c
var
aMAFC
Transcr.
CGH
1.2B vs W4-1
P1/H3K9me3
E
chr.14
3,100 kb
3,200 kb
PF14_0742
PF14_0749
aMAFC
Transcr.
CGH
3D7-A vs 3D7-B
PF14_0760
PF14_0761
Transcr.
CGH
10G vs W4-1
HP1/H3K9me3

FIGURE 5

Table 1. *P. falciparum* gene families showing clonally variant expression.

Gene families	N*	Variant expression^	HP1/ H3K9me3[†]	Subtelom.[§]	Exported[#]
6-cys	10	4	+/-	-	-
ApiAP2-TF	29	2	+/-	-	-
clag	5	3	+	+	-
etramp	14	4	-	+	-
PFD0075-like	4	3	+	+	-
rRNA	9	2	+/-	-	-
tRNA	53	20	+/-	-	-
acbp	4	1	+/-	+	+/-
acs	12	4	+	+	+/-
Lysophospholipase	15	2	+/-	+	+/-
FIKK	21	2	+/-	+	+
gbp	3	2	+	+	+
hrp	1	1	-	+	+
hyp4	1	1	+	+	+
hyp6	2	2	+	+	+
hyp7	3	3	+	+	+
hyp8	2	1	-	+	+
hyp10	2	1	+	+	+
hyp12	3	1	-	+	+
hyp15	4	2	+	+	+
hyp17	2	2	+	+	+
PHISTa	20	11	+	+	+
PHISTb	23	6	+/-	+	+
PHISTb_dnaj	7	3	+/-	+	+
PHISTc	16	7	+/-	+	+
dnaj_III	8	4	+	+	+
pfmc-2tm[$]	9	7	+	+	+
rif[$]	157	26	+	+	+
stevor[$]	36	9	+	+	+
var[$]	75	34	+	+	+

Only gene families with more than one member showing variant expression, or a single member showing variant expression in more than one comparison, are included. Gene families are described in Supplemental Table S3.

*Number of genes of the family that were analyzed.

^Number of genes showing variant expression (included in the list of variant genes, Supplemental Table S2).

[†]Gene families with more than 30% (+) or some, but less than 30% (+/-) of the members marked by HP1 or H3K9me3 (Flueck et al. 2009; Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009).

[§]Gene families with more than 50% of the members located at subtelomeric regions (less than 150 kb from the chromosome end) are indicated as positive.

[#]Gene families encoding proteins predicted to be exported to the erythrocyte (Sargeant et al. 2006) are indicated as positive, whereas +/- indicates gene families without a PEXEL motif that may be exported (Templeton 2009).

[$]Analyzed only in 3D7 parasite lines.

REFERENCES

Anders RF, Brown GV, Edwards A. 1983. Characterization of an S antigen synthesized by several isolates of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **80:** 6652-6656.

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

Bethke LL, Zilversmit M, Nielsen K, Daily J, Volkman SK, Ndiaye D, Lozovsky ER, Hartl DL, Wirth DF. 2006. Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of *Plasmodium falciparum*. *Mol Biochem Parasitol* **150:** 10-24.

Bischoff E , Vaquero C. 2010. *In silico* and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics* **11:** 34.

Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*. *PLoS Biol* **1:** E5.

Burkot TR, Williams JL, Schneider I. 1984. Infectivity to mosquitoes of *Plasmodium falciparum* clones grown in vitro from the same isolate. *Trans R Soc Trop Med Hyg* **78:** 339-341.

Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinas M. 2010. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog* **6:** e1001165.

Chookajorn T, Dzikowski R, Frank M, Li F, Jiwani AZ, Hartl DL, Deitsch KW. 2007. Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci USA* **104:** 899-902.

Comeaux CA, Coleman BI, Bei AK, Whitehurst N, Duraisingh MT. 2011. Functional analysis of epigenetic regulation of tandem *RhopH1/clag* genes reveals a role in *Plasmodium falciparum* growth. *Mol Microbiol* **80:** 378-390.

Cortés A. 2005. A chimeric *Plasmodium falciparum Pfnbp2b*/*Pfnbp2a* gene originated during asexual growth. *Int J Parasitol* **35:** 125-130.

Cortés A. 2008. Switching *Plasmodium falciparum* genes on and off for erythrocyte invasion. *Trends Parasitol* **24:** 517-524.

Cortés A, Benet A, Cooke BM, Barnwell JW, Reeder JC. 2004. Ability of *Plasmodium falciparum* to invade Southeast Asian ovalocytes varies between parasite lines. *Blood* **104:** 2961-2966.

Cortés A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA. 2007. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog* **3:** e107.

Coulson RM, Hall N, Ouzounis CA. 2004. Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res* **14:** 1548-1554.

Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. 2011. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. *Mol Microbiol* **80:** 391-406.

Daily JP, Scanfeld D, Pochet N, Le Roch K, Plouffe D, Kamal M, Sarr O, Mboup S, Ndir O, Wypij D et al. 2007. Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. *Nature* **450:** 1091-1095.

Dodd IB, Micheelsen MA, Sneppen K, Thon G. 2007. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* **129:** 813-822.

Eksi S, Haile Y, Furuya T, Ma L, Su X, Williamson KC. 2005. Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum*. *Mol Biochem Parasitol* **143:** 90-99.

Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, Ehlgen F, Ralph SA, Cowman AF, Bozdech Z et al. 2009. *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* **5:** e1000569.

Functional Genomics Workshop Group: Deitsch K, Duraisingh M, Dzikowski R, Gunasekera A, Khan S, Le Roch K, Llinas M, Mair G, McGovern V, Roos D et al. 2007. Mechanisms of gene regulation in *Plasmodium*. *Am J Trop Med Hyg* **77:** 201-208.

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

Gaszner M , Felsenfeld G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7:** 703-713.

Gomez-Escobar N, Amambua-Ngwa A, Walther M, Okebe J, Ebonyi A, Conway DJ. 2010. Erythrocyte invasion and merozoite ligand gene expression in severe and mild *Plasmodium falciparum* malaria. *J Infect Dis* **201:** 444-452.

Gonzales JM, Patel JJ, Ponmee N, Jiang L, Tan A, Maher SP, Wuchty S, Rathod PK, Ferdig MT. 2008. Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Biol* **6:** e238.

Grewal SI , Jia S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8:** 35-46.

Gunasekera AM, Myrick A, Le Roch K, Winzeler E, Wirth DF. 2007. *Plasmodium falciparum*: genome wide perturbations in transcript profiles among mixed stage cultures after chloroquine treatment. *Exp Parasitol* **117:** 87-92.

Hayton K, Gaur D, Liu A, Takahashi J, Henschen B, Singh S, Lambert L, Furuya T, Bouttenot R, Doll M et al. 2008. Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host Microbe* **4:** 40-51.

Hu G, Cabrera A, Kono M, Mok S, Chaal BK, Haase S, Engelberg K, Cheemadan S, Spielmann T, Preiser PR et al. 2010. Transcriptional profiling of growth perturbations of the human malaria parasite *Plasmodium falciparum*. *Nat Biotechnol* **28:** 91-98.

Hu G, Llinas M, Li J, Preiser PR, Bozdech Z. 2007. Selection of long oligonucleotides for gene expression microarrays using weighted rank-sum strategy. *BMC Bioinformatics* **8:** 350.

Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, Ingle CE, Thomas A, Quail MA, Siebenthall K, Uhlemann AC et al. 2007. Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nat Genet* **39:** 120-125.

Jiang L, Lopez-Barragan MJ, Jiang H, Mu J, Gaur D, Zhao K, Felsenfeld G, Miller LH. 2010. Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. *Proc Natl Acad Sci USA* **107:** 2224-2229.

Kussell E , Leibler S. 2005. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309:** 2075-2078.

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

Le Roch KG, Johnson JR, Ahiboh H, Chung DW, Prudhomme J, Plouffe D, Henson K, Zhou Y, Witola W, Yates JR et al. 2008. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, *Plasmodium falciparum*. *BMC Genomics* **9:** 513.

Lemieux JE, Gomez-Escobar N, Feller A, Carret C, Amambua-Ngwa A, Pinches R, Day F, Kyes SA, Conway DJ, Holmes CC et al. 2009. Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in ex vivo culture. *Proc Natl Acad Sci USA* **106:** 7559-7564.

Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL. 2006. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* **34:** 1166-1173.

Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, Scherf A. 2007. 5' flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol* **66:** 1296-1305.

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

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

Mackinnon MJ , Marsh K. 2010. The selection landscape of malaria parasites. *Science* **328:** 866-871.

Merrick CJ , Duraisingh MT. 2010. Epigenetics in *Plasmodium*: what do we really know? *Eukaryot Cell* **9:** 1150-1158.

Mok BW, Ribacke U, Winter G, Yip BH, Tan CS, Fernandez V, Chen Q, Nilsson P, Wahlgren M. 2007. Comparative transcriptomal analysis of isogenic *Plasmodium falciparum* clones of distinct antigenic and adhesive phenotypes. *Mol Biochem Parasitol* **151:** 184-192.

Mu J, Awadalla P, Duan J, McGee KM, Keebler J, Seydel K, McVean GA, Su XZ. 2007. Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat Genet* **39:** 126-130.

Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, Chotivanich K, Wilairatana P, Krudsood S, White NJ et al. 2010. *Plasmodium falciparum* genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet* **42:** 268-271.

Natalang O, Bischoff E, Deplaine G, Proux C, Dillies MA, Sismeiro O, Guigon G, Bonnefoy S, Patarapotikul J, Mercereau-Puijalon O et al. 2008. Dynamic RNA profiling in *Plasmodium falciparum* synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics* **9:** 388.

Ndonwi M, Burlingame OO, Miller AS, Tollefsen DM, Broze GJ, Jr., Goldberg DE. 2011. Inhibition of antithrombin by *Plasmodium falciparum* histidine-rich protein II. *Blood* **117:** 6347-6354.

Nguitragool W, Bokhari AA, Pillai AD, Rayavara K, Sharma P, Turpin B, Aravind L, Desai SA. 2011. Malaria Parasite *clag3* Genes Determine Channel-Mediated Nutrient Uptake by Infected Red Blood Cells. *Cell* **145:** 665-677.

Nunes MC, Goldring JP, Doerig C, Scherf A. 2007. A novel protein kinase family in *Plasmodium falciparum* is differentially transcribed and secreted to various cellular compartments of the host cell. *Mol Microbiol* **63:** 391-403.

Oakley MS, Kumar S, Anantharaman V, Zheng H, Mahajan B, Haynes JD, Moch JK, Fairhurst R, McCutchan TF, Aravind L. 2007. Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic *Plasmodium falciparum* parasites. *Infect Immun* **75:** 2012-2025.

Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, Conway DJ. 2010. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. *Mol Biol Evol* **27:** 2344-2351.

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

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29:** 24-26.

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 microarray software suite. *Methods Enzymol* **411:** 134-193.

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. 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **106:** 9655-9660.

Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, Speed TP, Cowman AF. 2006. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* **7:** R12.

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

Spielmann T, Fergusen DJ, Beck HP. 2003. *etramps*, a new *Plasmodium falciparum* gene family coding for developmentally regulated and highly charged membrane proteins located at the parasite-host cell interface. *Mol Biol Cell* **14:** 1529-1544.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545-15550.

Swamy L, Amulic B, Deitsch KW. 2011. *Plasmodium falciparum var* gene silencing is determined by *cis* DNA elements that form stable and heritable interactions. *Eukaryot Cell* **10:** 530-539.

Tamez PA, Bhattacharjee S, van Ooij C, Hiller NL, Llinas M, Balu B, Adams JH, Haldar K. 2008. An erythrocyte vesicle protein exported by the malaria parasite promotes tubovesicular lipid import from the host cell surface. *PLoS Pathog* **4:** e1000118.

Tellez M, Matesanz F, Alcina A. 2003. The C-terminal domain of the *Plasmodium falciparum* acyl-CoA synthetases PfACS1 and PfACS3 functions as ligand for ankyrin. *Mol Biochem Parasitol* **129:** 191-198.

Templeton TJ. 2009. The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, *Plasmodium falciparum*. *Mol Biochem Parasitol* **166:** 109-116.

Templeton TJ , Kaslow DC. 1999. Identification of additional members define a *Plasmodium falciparum* gene superfamily which includes Pfs48/45 and Pfs230. *Mol Biochem Parasitol* **101:** 223-227.

Van Tyne D, Park DJ, Schaffner SF, Neafsey DE, Angelino E, Cortese JF, Barnes KG, Rosen DM, Lukens AK, Daniels RF et al. 2011. Identification and Functional Validation of the Novel Antimalarial Resistance Locus PF10_0355 in *Plasmodium falciparum*. *PLoS Genet* **7:** e1001383.

Veening JW, Smits WK, Kuipers OP. 2008. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **62:** 193-210.

Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, Milner DA, Jr., Daily JP, Sarr O, Ndiaye D, Ndir O et al. 2007. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat Genet* **39:** 113-119.

Walliker D, Quakyi IA, Wellems TE, McCutchan TF, Szarfman A, London WT, Corcoran LM, Burkot TR, Carter R. 1987. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236:** 1661-1666.

Winter G, Kawai S, Haeggstrom M, Kaneko O, von Euler A, Kawazu S, Palm D, Fernandez V, Wahlgren M. 2005. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *J Exp Med* **201:** 1853-1863.

Young JA, Johnson JR, Benner C, Yan SF, Chen K, Le Roch KG, Zhou Y, Winzeler EA. 2008. In silico discovery of transcription regulatory elements in *Plasmodium falciparum*. *BMC Genomics* **9:** 70.

5.3 Additional results

To clarify some of our assumptions regarding CVGE and to establish whether episomal CVGE can be reproduced for a gene family other than the *var* family of genes, a number of proof-of-principle experiments were performed, using the clonally variant gene *clag3.2.*

5.3.1 Generation of a parasite line that stably maintains an episome with the *clag3.2* promoter

To begin these analyses, a parasite line that stably maintains a plasmid containing the *clag3.2* 5'UTR was required. 10G parasites were transfected with the plasmid HLHprom3.2bsdR to generate the parental parasite line 10Gluc (Figure 5.2A). HLHprom3.2bsdR contains the entire 5' UTR of *clag3.2* driving the expression of the firefly luciferase gene. The plasmid is stably maintained in this parasite line by continuous drug selection with blasticidin through the expression of the blasticidin deaminase gene driven by the constitutive *hsp86* promoter. This plasmid also contains a Rep20 sequence which is known to aid in plasmid segregation (O'Donnell et al 2002).

5.3.2 10Gluc is a population of individual parasites that differentially express luciferase

To determine if the 10Gluc population is a mixture of individual parasites that differentially express luciferase under the control of the *clag3.2* 5' UTR, the 10Gluc parental population was subcloned by limiting dilution yielding 16 subclones. Gene expression analysis, when performed on recent subclones, is akin to studying the gene expression pattern of a single parasite. These 16 subclones and the 10Gluc parental population

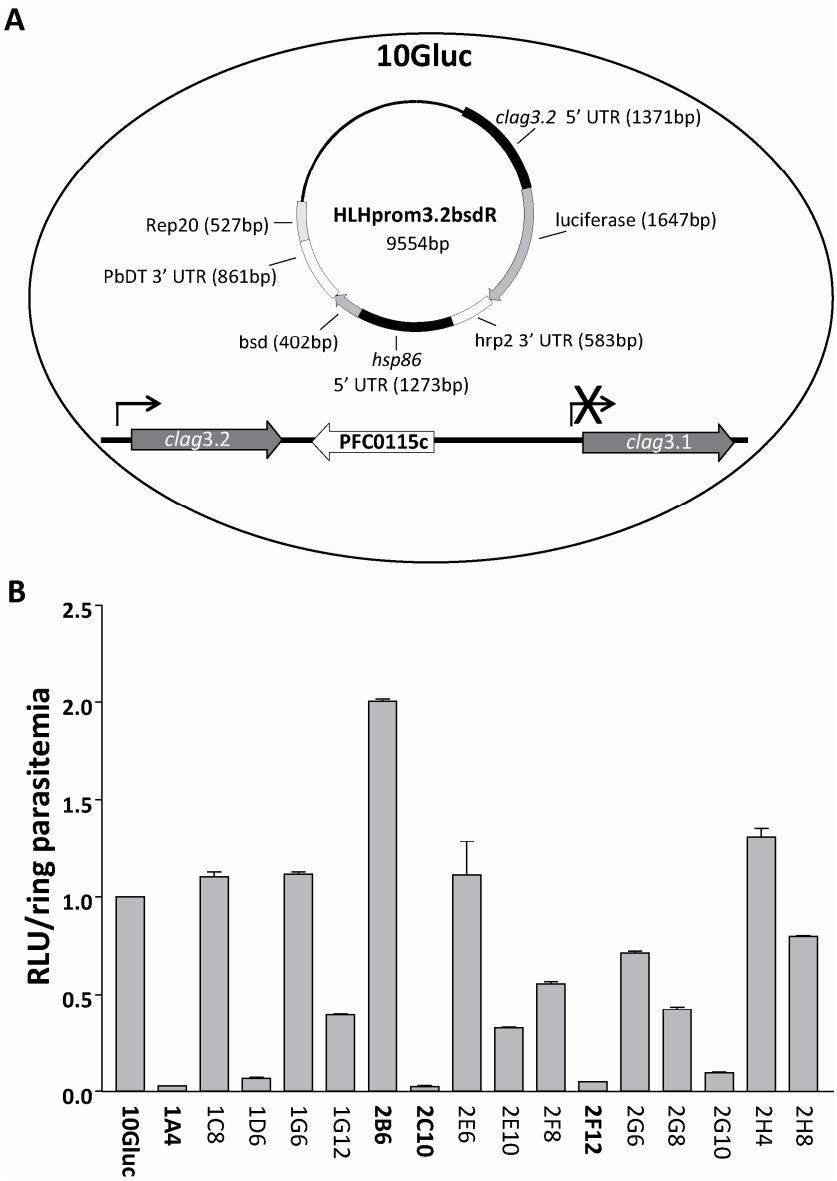


Figure 5.2 Preliminary luciferase assay screen on 10Gluc and 10Gluc-derived subclones. A. 10Gluc parasite line. The plasmid HLHprom3.2bsdR was transfected into the 10G parasite line. 10G parasites express endogenous *clag*3.2 and repress endogenous *clag*3.1. **B.** Luciferase activity in schizont stages in 10Gluc and subclones derived from 10Gluc. The subclones indicated in bold were selected for further analysis. Values correspond to luciferase activity (in relative light units, RLU) assayed in duplicate and normalized to 1% parasitemia as determined by flow cytometry. Values are relative to 10Gluc which has been set to 1.

were grown in parallel and under homogeneous conditions.

To determine whether there was heterogeneity in levels of luciferase expression amongst the subclones, an initial screen was performed whereby the 16 subclones and the parental 10Gluc population were synchronised twice with sorbitol (6 h apart) and assayed for luciferase activity at the schizont stage (Figure 5.2B). In this initial screen, we observed a broad range of luciferase expression levels in the 10Gluc-derived subclones when compared to the parental 10Gluc population. One subclone, 2B6, showed higher levels of luciferase expression, whereas 1A4, 1D6, 2C10, 2F12, 2G10, showed markedly reduced levels of luciferase expression when compared to the parental 10Gluc parasite line. The remaining subclones either showed similar levels or modest reductions in luciferase expression when compared to 10Gluc.

This result suggests that the 10Gluc parental population is composed of individual parasites that maintain the plasmid in transcriptionally different states and that this is controlled by the 5' UTR of *clag3.2*. Based on this preliminary screen, four subclones, 1A4, 2B6, 2C10 and 2F12 were selected for further analysis as the results suggested that 2B6 parasite line may maintain the plasmid in a transcriptionally active state, whereas the 1A4, 2C10 and 2F12 parasite lines may maintain the plasmid in a transcriptionally repressed state.

To complement the subcloning experiment, an immunolocalisation-based approach was also undertaken to demonstrate that the 10Gluc parental line is a population composed of individual parasites that maintain the episome in transcriptionally different states.

Immunolocalisation assays were performed with commercial antibodies against luciferase (Abcam, ab64564) and blasticidin deaminase (EMC Biosciences, BP1231). The rationale behind this approach was that all parasites within the 10Gluc parental population would give a signal for blasticidine deaminase (as the parasite line is grown with continuous blasticidin pressure), but that the population would contain a mixture of luciferase expressing and repressing individuals and this heterogeneity would be visualised by immunofluorescence. However, using both formaldehyde and acetone-methanol fixation, various secondary antibodies and different antibody dilutions, the luciferase antibody did not produce a signal (results not shown).

5.3.3 Tightly synchronised 10Gluc-derived subclones differentially express luciferase as determined by gene reporter assays and confirmed by qPCR

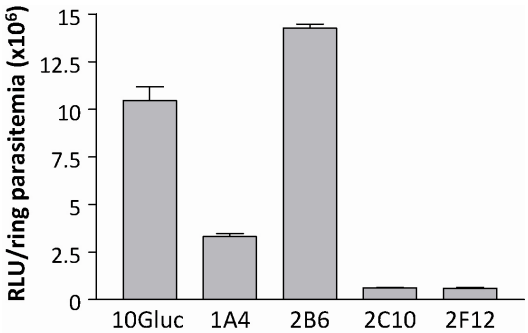


Figure 5.3 Luciferase assay on tightly synchronised schizonts (37-44 hpi) in 10Gluc and 10Gluc-derived subclones. Values correspond to the average luciferase activity (in relative light units, RLU), measured in duplicate, normalized to 1% parasitemia as determined by flow cytometry. Bars represent standard deviation.

Comparative gene expression analysis is only reliable when parasites are well synchronised; small differences in life stage between parasite lines may lead to artefactual differences in gene expression levels that are solely attributable to parasite synchronicity between different cultures.

To determine whether the observed differences in luciferase expression after double-sorbitol synchronisation was due to differences in parasite stage, 1A4, 2B6, 2C10, 2F12 and 10Gluc were tightly synchronised to a 7 h window and schizonts were assayed for luciferase activity (Figure 5.3).

The result of this luciferase assay confirmed that these parasite lines express luciferase at different levels and that these differences are not attributable to differences in parasite stage. The parasite line 2B6 expresses luciferase at a higher level than the 10Gluc parental line, whereas the 1A4, 2C10 and 2F12 parasite lines express lower levels of luciferase than the 10Gluc parental line.

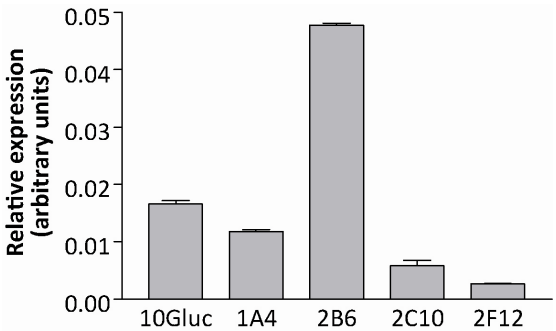


Figure 5.4 qPCR analysis of luciferase transcript levels on tightly synchronised schizonts (37-44 hpi) in 10Gluc and 10Gluc-derived subclones. Results are expressed in arbitrary units relative to a standard curve using genomic DNA. Values were normalized by rhopH2. Bars represent standard deviation from qPCRs conducted in triplicate.

To confirm whether the observed differences in luciferase activity were also reflected in luciferase transcript levels, RNA was extracted from the same tightly synchronised samples, reverse transcribed with a mixture of oligo $(dT)_{15}$ and random primers and relative luciferase transcript levels were determined via qPCR (Figure 5.4).

These results indicate that the expression of luciferase transcripts is higher in the 2B6 parasite line compared to the 10Gluc parental line, and luciferase transcript levels in the 1A4, 2C10 and 2F12 parasite lines is lower compared to the 10Gluc parental line, confirming the results obtained by the luciferase assay in Figure 5.3.

5.3.4 Differences in luciferase expression observed between 10Gluc-derived subclones does not increase with the relative of amount of plasmid

To determine if the observed difference in luciferase expression is due to different amounts of episomal DNA in the different parasite lines, genomic DNA was extracted and genomic (seryl tRNA synthetase) and episomal specific sequences (luciferase) were amplified by qPCR and the relative amount of episomal to genomic content was calculated (Figure 5.6).

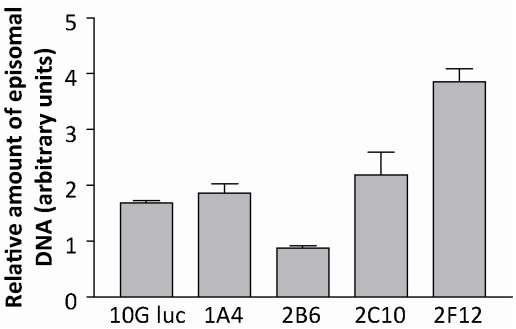


Figure 5.6 Relative amount of episomal DNA to genomic DNA in 10Gluc and 10Gluc-derived subclones. Results are expressed as the ratio of episomal to genomic DNA. The amount of episomal and genomic DNA was calculated relative to a standard curve using genomic DNA. Bars represent standard deviation from qPCRs conducted in triplicate.

Stable episomes introduced in *Plasmodium* are known to be maintained as large concatamers, ranging in size from 9-15 copies of the monomer, arranged in a head-to-tail array (Williamson et al 2002). The different parasite lines show different relative amounts of episomal to genomic DNA content

(Figure 5.6). However, a negative correlation is observed between relative episomal DNA amount and luciferase expression, as the parasite line with the highest luciferase expression, 2B6, has the lowest amount of episomal DNA, whereas the 2F12 parasite line displays the lowest amount of luciferase expression and contains the highest amount of episomal DNA. Therefore, differences in luciferase levels between 10Gluc and its derived subclones cannot be explained by higher episome content.

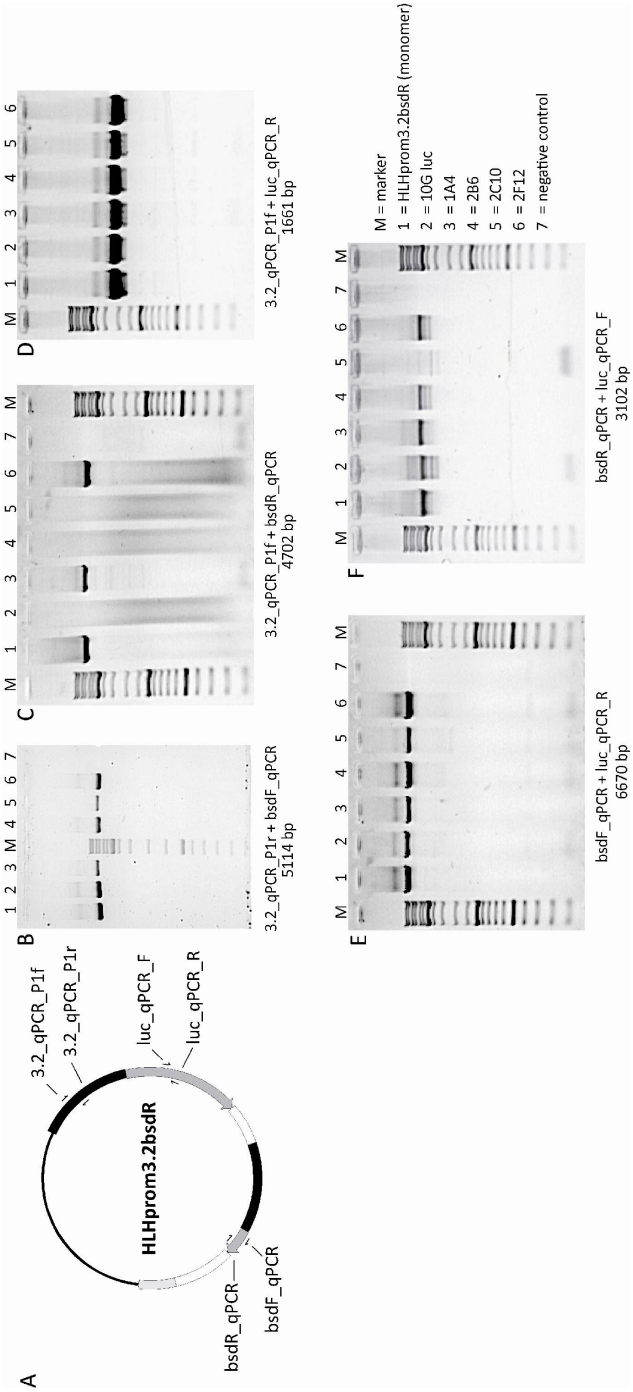


Figure 5.7 Preliminary PCR screen for episomal recombination events in 10Gluc and 10G-luc derived subclones. A. Plasmid HLHprom3.2bsdR and position of the primers used in PCR analysis. **B-F.** PCR screen for episomal recombination events. The primers used and the size of the expected amplification product are listed in each panel. The prominent band in each lane corresponds to the expected size.

5.3.5 Differences in luciferase expression between 10Gluc-derived subclones are not due to an episomal recombination

HLHprom3.2bsdR contains two promoter activities, that of *clag3.2* and *hsp86*. Episomal *var* genes become active when they are unpaired from a second promoter activity (Frank & Deitsch 2006). Consequently, one possibility for the luciferase expression observed in the 2B6 parasite line is that the episome has undergone some recombination event that has uncoupled the *clag3.2* promoter activity from that of *hsp86*. A recombination event would likely be detected by PCR by the amplification of additional bands of different sizes.

To determine if episomal recombination events in 10Gluc and 10Gluc-derived subclones is responsible for the observed differences in luciferase expression, preliminary PCR analysis was performed on genomic DNA extracted from these parasite lines (Figure 5.7). Monomeric HLHprom3.2bsdR served as a positive control.

The positions of the primers used in these analyses are depicted in Figure 5.7A. No differences were observed between the monomeric plasmid and any of the parasite lines using primers 3.2_qPCR_P1r and bsdF_qPCR (Figure 5.7B). The other half of the plasmid was screened with the primers 3.2_qPCR_P1f and bsdR_qPCR (Figure 5.7C). A band of the appropriate size was amplified from the positive control, and in parasite lines 1A4 and 2F12 (Figure 5.7C, lanes 1,3 and 6 respectively). This band was not amplified from 10Gluc, 2B6 or 2C10 samples (Figure 5.7C, lanes 2,4 and 5 respectively).

This half of the plasmid was analysed further by PCR using three other primer pair combinations, however no differences were observed between the monomeric plasmid and any of the parasite lines (Figure 5.7D,E,F). This suggested that perhaps the bands that did not amplify using 3.2_qPCR_P1f and bsdR_qPCR (Figure 5.7C) from the 10Gluc, 2B6 and 2C10 parasite lines was the result of a PCR amplification problem, as extensive smearing can be observed in these lanes. Cumulatively, these preliminary analyses did not detect a recombination event in 2B6, nor any of the other parasite lines, but we cannot exclude the possibility of other recombination events that would not be detected by these particular experiments. These results do suggest however that the observed differences in luciferase activity between 10Gluc and its derived subclones is attributable to factors other than episomal rearrangement.

5.3.6 The episome that is expressed is enriched in H3K9ac, whereas the episome that is repressed is enriched in H3K9me3

To determine whether the repressed or active state of the episome was correlated with the enrichment of H3K9me3 and H3K9ac, respectively, qChIP was performed in schizonts on two parasite lines that differentially express luciferase, 2B6 and 2F12 (Figure 5.8). In previous experiments using isogenic subclones that differentially express *clag*3.2, we showed that the 5' UTR of *clag*3.2 is enriched in H3K9me3 when repressed and enriched with H3K9ac when expressed (Article 1, Figure 5C,E). We are

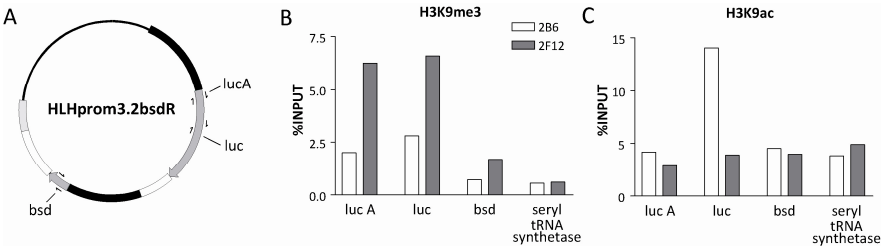


Figure 5.8 Preliminary qChIP analysis in schizonts of the 10Gluc-derived subclones, 2B6 and 2F12. A. Position of the fragments amplified by qPCR. **B.** qChIP using antibodies against H3K9me3 **C.** qChIP using antibodies against H3K9ac. ChIP data is expressed as % input, defined as the percentage of immunoprecipitated chromatin relative to the input sample. serARS refers to the chromosomally located seryl tRNA synthetase.

not able to analyse the 5' UTR of *clag*3.2 directly, as the results would be confounded by the endogenous *clag*3.2 5' UTR. However, we observe an enrichment of H3K9me3 on the luciferase gene in 2F12 parasites, the line that has the reporter gene silenced, whereas the luciferase gene in 2B6, where the reporter gene is expressed, is enriched in H3K9ac (Figure 5.8B,C). We do not observe a difference in either parasite line for the chromosomally located seryl tRNA synthetase. We do observe an enrichment of H3K9me3 over the blasticidine deaminase gene in 2F12 (Figure 5.8B). These are the results of a single experiment and further experiments are required for confirmation. However, these results suggest that the 5' UTR of *clag*3.2 is sufficient to recruit the enzyme activities necessary for the formation of H3K9me3-facultative heterochromatin. It is also suggests that in the absence of a boundary element, H3K9me3-facultative heterochromatin spreads from the *clag*3.2 5' UTR nucleation site, into the luciferase gene and even into the blasticidine deaminase gene. However we know that the blasticidine deaminase gene is not entirely repressed, as the parasites are grown in the presence of blasticidine. These preliminary results also suggest that

silencing via the formation of H3K9me-heterochromatin happens in a spontaneous fashion, as these subclones are grown in homogeneous conditions, yet contain the plasmid in different transcriptional states. Furthermore, this suggests that CVGE for *clag3.2* may be mimicked using an episomal-based system.

5.3.7 In transient transfection assays, HLHprom3.2bsdR appears to be repressed

According to the level of luciferase transcripts observed in Figure 5.4, the level of luciferase transcripts for the parental 10Gluc population is more similar to the parasite lines 1A4, 2C10 and 2F12, where the episome appears to be repressed. The opposite trend was observed with 10Gluc in the gene reporter assay, where the level of luciferase expression for the 10Gluc parental population is most similar to 2B6, the parasite line that expressed the highest amount of luciferase (Figure 5.3). The gene reporter assay was performed on tightly synchronised parasites and normalised by parasitemia. However, qPCR can quantitatively normalise for both the number of parasites and for parasite stage. Consequently, the qPCR data could suggest that the majority of the parasites within the parental 10Gluc population preferentially maintain the plasmid in its repressed state and that the luciferase expression in the 10Gluc population is contributed by a small proportion of parasites that maintain the plasmid in its active state.

To determine the relative expression status of the HLHprom3.2bsdR plasmid compared with other *clag3.2* 5'UTR containing plasmids, transient transfections were performed. The composition of each plasmid is shown in detail in Figure 5.5A-C. HLHprom3.2bsdR contains a

drug resistance cassette that is driven by the *hsp86* promoter. 3.2-LH-1371-R20 and 3.2-LH-1371 do not contain drug resistance cassette and thus the only promoter activity contained within these plasmids is that of *clag3.2*.

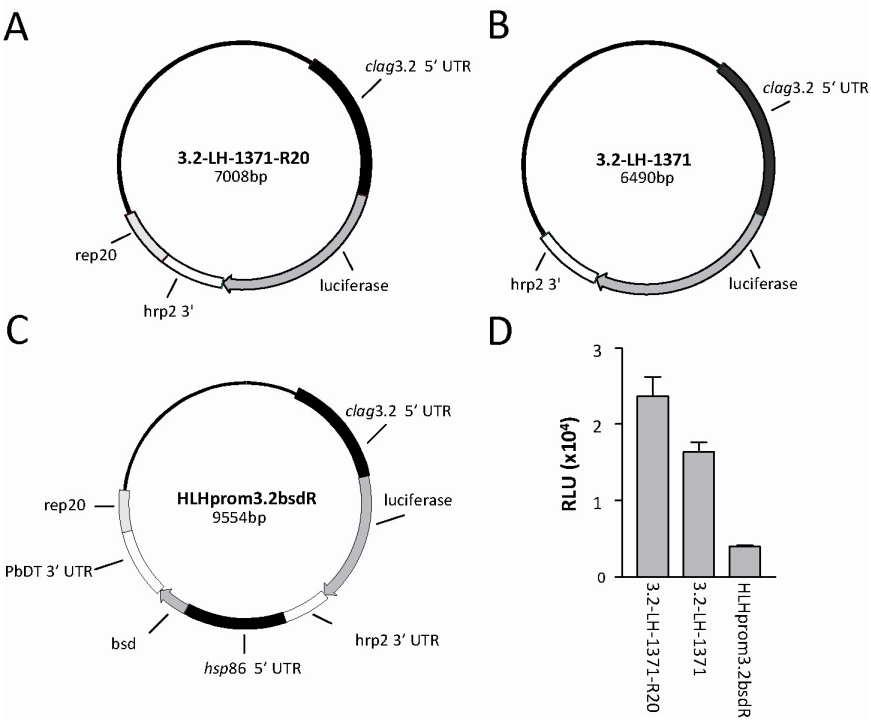


Figure 5.5 Transient transfection using plasmids containing the *clag*3.2 5' UTR in combination with other elements in 3D7-A parasites. A-C. Schematic of the plasmids used in the transient transfection assay. **D.** Luciferase activity was assayed in schizonts. Each construct was tested in duplicate and values correspond to average luciferase activity (in relative light units, RLU), with standard deviation.

In the transient transfection assay, the plasmid HLHprom3.2bsdR shows a markedly lower level of luciferase expression when compared to 3.2-LH-1371-R20 and 3.2-LH-1371 (Figure 5.5D). This supports the observation with the stable transfectants that the majority of parasites within the

10Gluc parental population maintain the episome in a repressed state (Figure 5.4).

This plasmid contains two promoter activities, that of *clag3.2* and that of *hsp86*. Other elements apart from the second promoter activity could also contribute to the decreased luciferase expression with this plasmid. However, one hypothesis for the observed silencing of this plasmid is that it is repressed via the promoter-pairing mechanism that has been previously described for *var* genes. Whether this is a mechanism for all variant promoters, mutually-exclusive promoters or is an artefact of the episomal system is yet to be determined and requires further investigation.

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

6. Mutually exclusive gene expression in *P. falciparum*

6.1 Hypotheses regarding *clag3* mutual exclusion

10Gluc and 10Gluc-derived subclones provide a unique opportunity to assess the mutually exclusive expression of *clag3* genes in the presence of multiple stable episomal *clag3.2* promoters. We do not know whether the episomal promoters will influence the mutually exclusive expression program of the endogenous *clag3* genes. However in the *var* gene family, the expression of an episomal *var* promoter, coupled with a second episomal promoter, leads to the repression of all endogenous *var* genes (Voss et al 2006). Whether the mutually exclusive expression of the *clag3* genes functions in a similar way is currently unknown but will be addressed by these experiments.

ncRNAs play a predominant role in the establishment and/or maintenance of monoallelic and mutually exclusive gene expression. Apart from the lncRNA-TARES transcribed from the telomeres (Broadbent et al 2011), only two long ncRNAs are known to be transcribed from chromosome 3: PF03TR001 and PF03TR002 (Otto et al 2010). Both of these ncRNAs are located between *clag3.1* and *clag3.2*. PF03TR001 is a ncRNA that corresponds to a *var* pseudogene (PFC0115c). Given the involvement of ncRNAs in the regulation of monoallelic and mutually exclusive gene expression, we hypothesize that the ncRNA PF03TR002 may be involved in the mutually exclusive expression of *clag3* genes and consequently, we expect the expression of PF03TR002 to

correlate with the expression/repression status of the *clag3* genes in parasite lines that differentially express these genes.

Our study of the mutual exclusive expression of *clag3.1* and *clag3.2* has been based on parasite lines derived from the 3D7 parasite line. Just as CVGE is an inherent property of many *P. falciparum* genes, we hypothesize that the mutually exclusive expression of the *clag3* genes is an inherent property of *clag3.1* and *clag3.2* and not a phenomenon associated with a particular genetic background. We will assess the pattern of *clag3.1* and *clag3.2* in parasite lines other than 3D7 and predict that the *clag3* genes will show mutually exclusive expression in these parasite lines.

We also hypothesize that the mutually exclusive expression of *clag3.1* and *clag3.2* is controlled by competition of their promoters for the interaction with a single, limiting enhancer or repressor element, located somewhere in the genome. We have come to this limiting enhancer or repressor hypothesis for the mutually exclusive expression of *clag3* genes for numerous reasons.

We know that these genes are mutually exclusive at the level of transcription and are encoded by monocistronic transcripts; thus excluding post-transcriptional regulatory mechanisms such as RNAi (a pathway that *P. falciparum* lacks) and regulation by other ORFs within a polycistronic transcript. There is no appreciable DNA methylation in *P. falciparum*, excluding it as a possible regulator of gene expression.

We have shown that the mutually exclusive expression of *clag3.1* and *clag3.2* in 10G and 1.2B parasites is not due to the presence or absence of *trans*-acting factors present in one parasite line or the other, as there is no difference in reporter gene activity in transient transfections with plasmids containing either *clag3.1* or *clag3.2* 5' UTRs in 10G and 1.2B parasite lines (Article 1, Figure 2D).

We also have data in the lab that the mutually exclusive expression of the *clag3* genes is not due to limiting space in a transcriptionally permissive subnuclear compartment that can only spatially accommodate one *clag3* promoter at a time. A colleague has generated a transgenic parasite that has extra copies of the *clag3.2* promoter integrated in its genome and when forced by drug pressure, the transgenic parasite line can express both *clag3.1* and *clag3.2* (N. Rovira-Graells, unpublished data).

We know that PTMs are involved in the maintenance of *clag3* mutual exclusion as the expression of a particular *clag3* is maintained for many generations by epigenetic memory (Article 1, Figure5). However, these PTMs cannot explain how mutually exclusive gene expression is established.

The active and repressed states of the *clag3* genes in 10G and 1.2B are not due to major genetic rearrangements, ruling out recombination as a mechanism in their mutually exclusive expression (Cortés et al 2007). The coding region of *clag3.1* and *clag3.2* in 10G and 1.2B parasites are identical (A. Cortés, unpublished data). We cannot currently exclude the possibility that small differences in nucleotide sequence between the 5'

flanking regions of *clag3.1* and *clag3.2* contribute to their mutually exclusive expression in 10G and 1.2B parasites. However, these two parasite lines are subclones from the same parasite line and should be genetically identical. However, to exclude this possibility, we will sequence the 5' flanking regions of *clag3.1* and *clag3.2* in these two parasite lines.

We also cannot exclude the possibility that the mutually exclusive expression of these two genes is established by other molecular entities, such as RNA or proteins. However, proteins and RNA are not typically single copy elements, whereas a single DNA enhancer or repressor element could exist in the genome. Mutually exclusive expression implies that there is a molecular counting mechanism that permits for the expression of only a single gene while simultaneously repressing the expression of all others. A single DNA enhancer or repressor, because it would be limiting, could function as this counting mechanism in the case of the *clag3* genes.

Hence, by the exclusion of several factors, we predict that the mutually exclusive expression of the *clag3* genes is mediated by competition for a limiting enhancer or repressor. We predict that the mutually exclusive expression of the *var* gene family is also mediated by competition for a limiting enhancer element (albeit a different one than the *clag3* genes), as has been previously proposed by the Scherf group (Lopez-Rubio et al 2007b). It would be more parsimonious for their expression to be controlled by a single enhancer element rather than ≈59 repressors. Other factors that contribute to the mutually exclusive expression for the *var* gene family were previously discussed in sections 1.5.2 and 1.5.3.

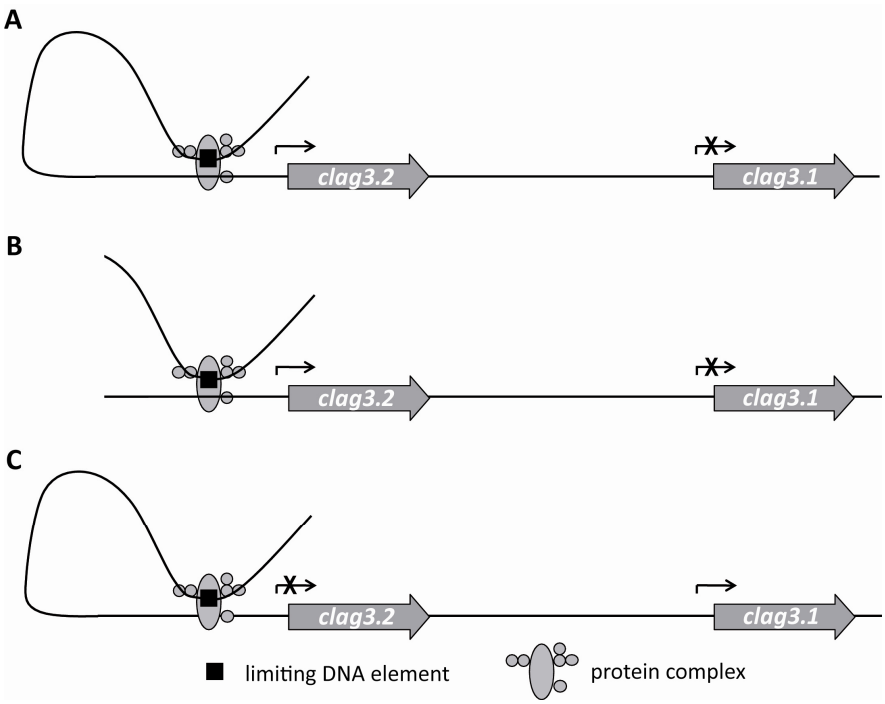


Figure 6.1 Schematic for the limiting enhancer or repressor hypothesis for the mutually exclusive expression of *clag3.1* and *clag3.2*. A. A limiting enhancer in *cis* interacts with the *clag3.2* promoter, permitting for the transcription of *clag3.2* but preventing the transcription of *clag3.1*. This could occur via intrachromosomal contacts and the looping out of the intervening chromatin fiber. **B.** A limiting enhancer in *trans* would involve interchromosomal contacts. **C.** The converse situation could involve a limiting repressor that interacts with one *clag3* promoter, permitting for the expression of the other. A limiting repressor in *cis* is depicted above but the repressor-promoter interaction could also occur in *trans* (not depicted).

The mutual exclusion of two *clag* genes, as opposed to ≈60 *var* genes, provides an amenable system in which to study mutually exclusive gene expression and isogenic parasite lines that differential express the *clag3* genes will be useful tools in these analyses.

We predict that the interaction between a mutually exclusive gene and an enhancer or repressor would be mediated by protein complexes. The enhancer or repressor and mutually exclusive gene could be physically located on different chromosomes and thus would require the formation of interchromosomal contacts. It is also possible that the two DNA elements are located on the same chromosome and could be brought into close physical proximity by the looping out of the intervening chromatin fibre (Figure 6.1).

The limiting enhancer or repressor hypothesis is testable by the techniques of chromosome conformation capture (3C) and circular chromosome confirmation capture (4C, Figure 6.2). These are two techniques that are able to identify physical interactions between sequence elements that are located on the same or on different chromosomes. 3C can be used to analyse chromosomal regions where interactions are predicted to exist, whereas 4C is employed to detect unknown interactions that could occur between any two regions in the genome. To the best of our knowledge, neither of these techniques have been published in *Plasmodium*. As these are technically challenging experiments, our focus has been to optimise the initial steps including restriction enzyme (RE) digestion of cross-linked nuclei , (indicated with a star, Figure 6.2) a major bottleneck in 3C and 4C experiments and initially in the 3D7-A parasite line (Hagege et al 2007) .

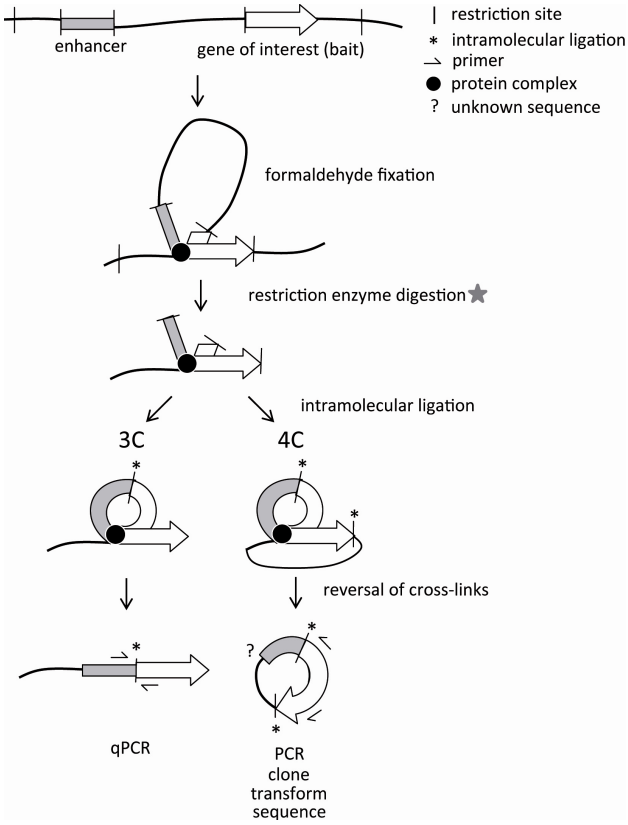


Figure 6.2 Simplified overview of chromosome conformation capture (3C) and circular chromosome conformation capture (4C). In both techniques, cells are treated with formaldehyde, creating reversible cross-links between proteins and between proteins and nearby DNA. Cross-linked chromatin is digested with appropriate restriction enzymes before undergoing intramolecular ligation. In the case of 3C, after ligation, the cross-links are reversed and ligation products are detected by qPCR. The cross-linking frequency of two specific restriction fragments is measured by the amount of the corresponding ligation product and is proportional to the frequency in which these genomic sites interact. In the case of 4C, intramolecular ligations are performed under conditions to favour two intermolecular ligation events to form a circular template. The cross-links are reversed and the unknown sequence is amplified by inverse primers designed against the bait sequence. These PCR products are cloned, transformed into bacteria and the unknown sequence is identified via sequencing. This figure had been adapted from Simonis et al (2007) and Zhao et al (2006).

Incomplete RE digestion leads to a reduced number of restriction ends available for ligation, lowering the efficiency of 3C and 4C experiments, and potentially introducing biases. To obtain meaningful data, it is recommended that the RE digestion efficiency be at least 60-70%, but preferentially higher (Hagege et al 2007). RE digestions for 3C and 4C analyses need to be performed in the context of cross-linked, intact nuclei. In order for the RE to have access to the target chromatin, the nuclei are permeabilised with the detergent sodium dodecyl suphate (SDS). SDS also serves to inactivate any proteins that were not cross-linked during the formaldehyde treatment and to relax chromatin structure to facilitate enzyme digestion. However, after permeablisation with SDS, the nuclei become distorted and form aggregates that may hamper the accessibility of the RE. And in addition, SDS will inactivate RE, so prior to RE digestion, SDS is sequestered by the addition of Triton X-100. It is in these less than optimal conditions that the RE are required to function, thus RE digestion efficiencies need to be carefully checked and optimised.

We have decided to use *clag3.2* in 3C and 4C analyses and the *var* gene PF07_0048 in 4C analysis. PF07_0048 is the *var* gene that is expressed in the 1.2B parasite line.

6.2 Results

6.2.1 Mutually exclusive expression of *clag3.1* and *clag3.2*

6.2.1.1 10Gluc and 10Gluc-derived subclones have switched *clag3* expression

To determine if the presence of multiple episomal *clag3.2* 5' UTRs altered the mutually exclusive expression of the endogenous *clag3.1* and *clag3.2* genes, qPCR analysis of the chromosomally located *clag3.1* and *clag3.2* genes was performed (Figure 6.3). 10Gluc and 10Gluc-derived subclones that stably maintain the episome present an opportunity to assess the mutually exclusive expression status of the endogenous genes as each individual parasite stably expresses the plasmid. This analysis could not be performed using the transient transfections discussed in Article 1, Figure 2D and F, because the majority of these parasites do not contain the plasmid as the transformation efficiency of *P. falciparum* is on the order of ≈10^{-6} (O'Donnell et al 2002).

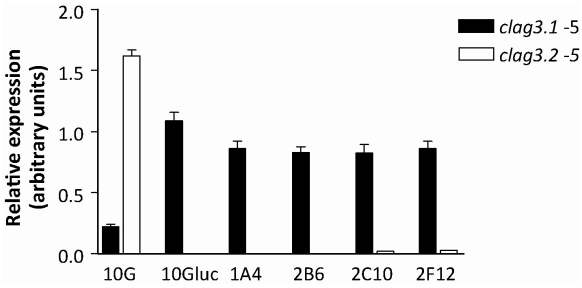


Figure 6.3 qPCR analysis of endogenous *clag*3.1 and *clag*3.2 transcript levels in schizonts (37-44 h post invasion) in 10G, 10Gluc and 10Gluc-derived subclones. The region amplified by qPCR corresponds to the *clag3.1* and *clag3.2* coding region. Results are expressed in arbitrary units relative to a standard curve using genomic DNA. Values were normalized by rhopH2. Bars represent standard deviation from qPCRs conducted in triplicate. The position of the primers used are indicated in Article 1, Figure 5A.

We know that the parasite line into which the plasmid was transfected, 10G, expresses *clag3.2* and represses *clag3.1* (Article 1, Figure 2B), however, 10Gluc and these four 10Gluc-derived subclones show the opposite pattern of expression where *clag3.1* is expressed and *clag3.2* is repressed. It is for the moment unclear whether the switch in 10Gluc and its derived subclones is due to the presence of the plasmid, or whether, over time, the 10G parasite line would have switched to express *clag3.1* in the absence of the plasmid.

The activation of an episomal *var* promoter paired with a second promoter activity, results in the repression of all endogenous *var* genes (Voss et al 2006). Previous experiments have shown that in the parasite line 2B6, the episomal *clag3.2* promoter is expressed in a construct that also contains a second promoter (Figures 5.3, 5.4, 5,6 and 5.8). However, the continuous expression of an endogenous *clag3* gene (*clag3.1*) in 2B6 indicates that the mutual exclusive expression of *clag3.1* and *clag3.2* is not the same as observed in the *var* gene family.

6.2.1.2 The ncRNA PF03TR002 is correlated with the expression of *clag3.1* and the repression of *clag3.2*

To characterise the expression of PF03TR001 and PF03TR002 in four different parasite lines with different *clag3.1* and *clag3.2* expression profiles, qPCR was performed at different stages along the IEC (Figure 6.4). Monoallelic and mutually exclusive gene expression often involves the participation of ncRNAs and we wanted to determine if there was a correlation of either of these two transcripts with *clag3* expression status. The position of the fragments amplified by qPCR can be found in Figure 6.4A. Four different parasite lines, 3D7-A, 3D7-B, 10G and 1.2B

(Article 1, Figure 2A) were analysed and the expression of *clag3.1* and *clag3.2* in these parasite lines can be found in Figure 6.4B. 3D7-A, although of clonal origin, is a transcriptionally heterogeneous population that expresses both *clag3* genes. 3D7-B expresses *clag3.1* and represses *clag3.2*, the repression of which is almost complete. 10G expresses *clag3.2* while 1.2B expresses *clag3.1* and both of these parasite lines are subclones of 3D7-A. In 10G and 1.2B, the repression of the repressed *clag3* is either not complete or some parasites within the population have switched *clag3* expression, as low levels of the repressed *clag3* can be detected (Article 1, Figure 2B).

The *var* pseudogene, PF03TR001, is very high expressed during the ring stage in 3D7-B parasites (Figure 6.4G), whereas it is expressed at very low levels in rings of 3D7-A, 10G and 1.2B (Figure 6.4C, J and M). PF03TR001 is somewhat expressed in schizont stages in 3D7-B (6.4H and I), however the levels are dramatically reduced from its expression level in rings (Figure 6.4G). This is consistent with the expected timing of expression for *var* genes. However, the expression of PF03TR001 does not correlate with the expression status of the *clag3* genes.

PF03TR002 is expressed at moderate levels in the schizont stage in both 3D7-A and 1.2B (Figure 6.4E, F and O), whereas it is not expressed in 10G schizonts (Figure 6.4L). PF03TR002 is not expressed outside of the schizont stage in 3D7-A, 10G, nor 1.2B (Figure 6.4C, D, J, K, M and N). However, PF03TR002 shows a different expression profile in the 3D7-B parasite line, where it is highly expressed in the schizont stage (Figure 6.4H and I), and shows moderate expression during the ring stage (Figure 6.4G). The PF03TR002 transcript also appears to be of a different length

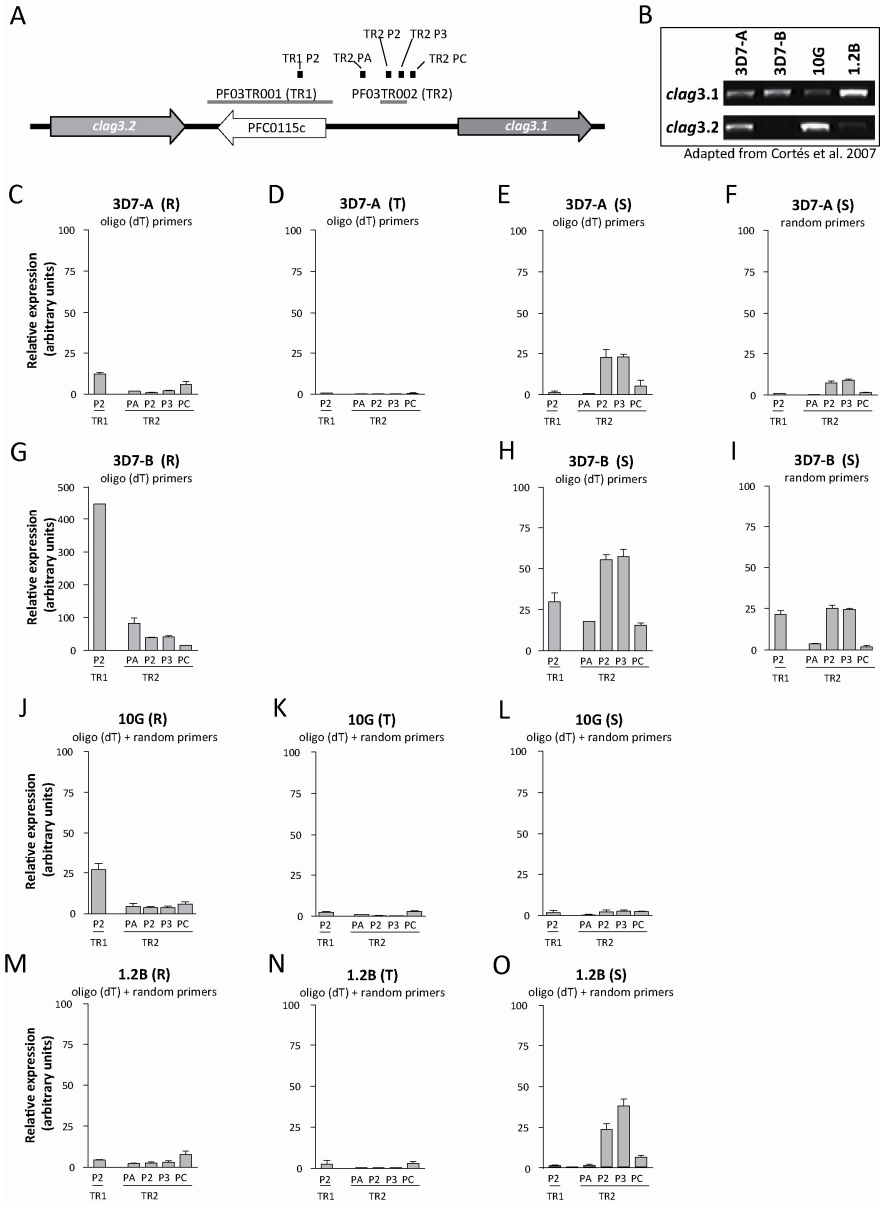


Figure legend on next page

in 3D7-B, or multiple transcripts of different lengths are transcribed from this region, as there is amplification with the TR2 PA and TR2 PC primers that is not observed in the other three parasite lines. However, in contrast to the expression of PF03TR001, PF03TR002 expression consistently correlates with the expression of *clag3.1* and the repression of *clag3.2*.

To determine whether a switch in *clag3* expression is correlated with PF03TR002 expression, qPCR was performed on 10Gluc and 10Gluc-derived subclones (Figure 6.5). These parasite lines switched *clag3* expression from *clag3.2* to *clag3.1*. (Figure 6.3). 10G parasites do not transcribe PF03TR002 to any appreciable level (Figure 6.4 J ,K and L).

Figure 6.4 qPCR analysis of PF03TR001 and PF03TR002 in various stages along the IEC in 3D7-A, 3D7-B, 10G and 1.2B parasite lines. A. Genomic organization and position of the fragments amplified by qPCR. PF03TR001 corresponds to the transcript derived from the *var* pseudogene PFC0115c. PF03TR002 is a ncRNA situated between *clag3.1* and *clag3.2*. It is currently unknown which DNA strand these transcripts are transcribed from. **B.** Expression of *clag3.1* and *clag3.2* in schizont stages in 3D7-A, 3D7-B, 10G and 1.2B parasite lines. This figure has been adapted from (Cortés et al 2007). **C-O.** qPCR analysis of PF03TR001 and PF03TR002 in 3D7-A, 3D7-B, 10G and 1.2B parasite lines. R, T and S refers to rings (10-15 hpi), trophozoites (20-25 hpi) and schizonts (40-45 hpi) respectively. Often ncRNAs are not polyadenylated, therefore, for cDNA synthesis, RNA was primed using either oligo (dT) primers, random primers or a mixture of both, and the priming method used is indicated above each graph. Results are expressed in arbitrary units relative to a standard curve using genomic DNA. Please note that the scale is different in panel G and expression of these two transcripts was not assessed in 3D7-B trophozoites. Values were normalized against seryl tRNA synthetase which is expressed throughout the IEC.

However, 10Gluc and four of its derived subclones that express *clag3.1* all express high levels of PF03TR002 (Figure 6.5). This is consistent with the previous correlation of PF03TR002 expression with the expression of *clag3.1* and the repression of *clag3.2*, suggesting that this ncRNA could play a role in maintaining the mutually exclusive expression of these two genes.

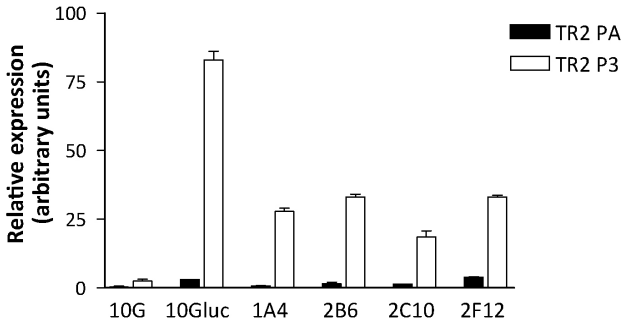


Figure 6.5 qPCR analysis of the ncRNA PF03TR002 transcript levels in schizonts (37-44 h post invasion) in 10G, 10Gluc and 10Gluc-derived subclones. Results are expressed in arbitrary units relative to a standard curve using genomic DNA. Values were normalized by seryl tRNA synthetase. Bars represent standard deviation from qPCRs conducted in triplicate. The position of the TR2 PA and TR2 PB primers can be found in Figure 6.4A.

6.2.1.3 The differential expression of the *clag3* genes in 10G and 1.2B parasite lines is not the result of nucleotide differences in their 5' intergenic regions

To determine if nucleotide differences in the 5' intergenic regions of *clag3.1* and *clag3.2* in the parasite lines 10G and 1.2B could account for the mutually exclusive expression of these two genes in these two parasite lines, we amplified the 5' intergenic regions of both of these genes from the two different parasite lines by PCR and sequenced the PCR products by single primer extension (Figure 6.6). We know that the

active and repressed states of *clag3* genes in 10G and 1.2B are not due to major DNA rearrangements (Cortés et al 2007). However, the 5' intergenic regions of both *clag3* genes contain heteropolymeric dA:dT tracts , the lengths of which were correlated with transcriptional efficiency in the *P. falciparum* calmodulin gene (Polson & Blackman 2005). However, we confirmed that the entire 5' intergenic region of *clag3.2* and ≈ 2.3 kb of the *clag3.1* 5' intergenic are identical between 10G and 1.2B and identical to that of the reference strain 3D7, indicating that the differential expression of these two genes in these two parasite lines is not due to small difference in DNA sequence that could have arisen from spontaneous mutations or from slipped-strand mispairing.

Figure 6.6 Sequencing results of *clag3.1* and *clag3*.2 5' intergenic regions from 10G and 1.2B parasites. *clag3.1* and *clag3.2* were sequenced by single primer extension from 10G and 1.2B parasites and were found to be identical to each other and to the reference strain 3D7. 2384 bp of the *clag3.1* 5' intergenic region and the entire *clag3.2* 5' intergenic region (1371 bp) were sequenced. The position of the primers used for sequencing are highlighted in grey, forward primers are in bold, whereas reverse primers are underlined. The figure is on the next page.

A

clag3.1 (2384 bp)

AGGAAAAAATAATGCAATCTAAGTAATGATCATATAAAAATTGAATTTTAATTAAAGAAGATGAT
TAAAAAGAAACATGATATATATTTAAAAAAAAAAAAAAATAATGAAAACAAAAAAATATAATATA
ATTATAATGAATTATAGGAACAAACATATAAAATTTTTTAATAAATTACTTCTGAAAAAATATAA
AATAATTATATATTTATACAAAAAATAAATATATAAATGAATAATAAAAAAATATTATGTTAAAA
ATAATATTTTTTTTTTTTTTTTTTTAATAAATAAATAAATAAAATATATTAAAAACCAAATGAAT
TATCTATCCAATAAAATATAAAAAAATAAAATAATTTCAAATTAGAAAAATTTATATATTAATCA
AAGTGCAATAATAAAATAAAAGTGCATTAATTATAATTCTTTTTGAATAAATAAAAAGAAAAGAA
ATAGAATCAACAAAGATTCCTATGAAAAAATATATATTTATATATATA**GAGAATAAACTGTAAAA**
AAGAAAAAAAAACGAAAAGACAATATATCAAAATATTTATATATAATTTTATTTTACAAACATTA
GAAAAATTAAAAGAATCCTTGTGCAATACAATATATTAAATATAACAAAAAAACAAACAACGATA
AAATATCAAATAAAAAATAAAAAAAAAAAAAGAAAAACAATTAATAAAAATTAACATAAAAAATAA
GTCAACAAGAATTAAGTGCATAATTTTTTATTAAATAAACAAATTTTGTGCATTAAAAAAAAAAA
AAAAAAATTCTTATTAAAATATATAGAAAGAAAAAAATAAAAAAAAAAAAAAAAAG**AAGAAAAAAG**
ATGATATTTAATAATATTGATATCCGTAATGAAATTAATTATTTCAAATATTTATTTATTGATTT
ATTTTATTATTATTATTTTTTTTTTTTTTAATGAAAAAAAATATATAAAAATATTTATAAACATAT
TTATTAAAATAATAAATTAAAAATTAAAATATCCATTAATTATTTTTTTTTTTTTAATGTCGAAA
AAAAATATATACGTATCGCAATATAATTTAATATAATATTTAAGTATTATGGATTTTTATATATT
TTTGAAAATTAAGAATTCTAAATAATTAGAAATGTATTCATAACTAATTTATTTTAAATATTATA
TAGATTTTA**TAATGCACTCATTAATAATTTTAAACCGTT**TTATTTATTATATTTATTTATTATTA
TTTATTTTTTCTTTTTTTTCTCTTATTTATTTTTTTTTTTTTTGTTCTAATTACGTATACATTT
ATTTAATAATTGAATAACCAATAAAAAAAATAGCTCACATAAATACATATATATAATATTATATT
TTTATTAATATAAATTTATTCAATAACGTATTAGGAAATATTACATAACTTATAATAATATATTT
TAACAATACATAAAGGAAGAAATGGAAAGAAAGAAACGTTTTACGTAACTAACATATATAATTAA
AACGCTACGTAAATATTAAAATAATTCATTTTGTATTTTTTTCTCCCTTTAACTTTACTTTGAAT
TCTTACATATTATAATACTCTAAATTCAATAAATGAAAATTTCCATAATTTATTTTCTAATAATT
ATCTTCTTTTCTTTCCATTGCACCGATATTAAAAACAAAAACAAAAAAAAAAAAAACCAAAAAACA
CCATTCATATTTATATTGAATATGTATTTTTATCTCACTTCTTTTTATTTACCTAAGTAAATATA
TATATAATCATATAACAAATAACATGTTATGCAAAAATATGCATAAACATAAGAGGGTGATTTAA
AATAAAAATAAAAATATACTTTACAATATGTTCAAAATATAATTCGTATATTCATATATTCTCGA
ATAGTTTTATTTCCTTTTATATTACATAATGTACACTTTAAAACAAAATGTGGACAAATATGATT
TTGTACTAAAAGAAAAAAAAAAAAAAAGTACAACATTAAAAAAATTCTTTTTGATATATATATATA
TGTTTAAACATATATTTGTGTACTCCTTTTTGTACATGGTTTTTGATTTTGTTATATTTATGTAC
TATTCAATTCTGTACTTTTTTTTAATAATAACATAATAATAATTACTTATATATTTGTGCAATAT
ATCAAAGTGTACATGCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGGTTTCATTT
TTTAAAACTCCAATCTTTATTTTAATTATCTTTTTATACTTAAATGAAAAGGTAATATGTTCAAT
AAATGAAAATCAAAATGAAAATGATACCATTAGTCAAAATGTCAACCAACATGAAAATATTAATC
AAAATGTAAATGATAATGACAATATTGAACAATTAAAATCCATG

B

clag3.2 (1371 bp)

TAGAATATATTTATTTATAGTCAAAATTTGCAAAAAAAATTTTAACTACAAGTAAATATATACAG
TTTATCATATATATTAATTTTTACAATTAATAAATAAAACAGCATGCACATAATATTTTATTATA
TTATTAAACCTAACGGGTACATTTTTTTACTTATTAAATATATCACATTATTTTTAATTAAAAGA
TTTATTGAATATCCATAACCTTTTTATATTAACATAAAATTGTGTGTTTCATTAAAATCATATTT
TTAAATGTTACAAAAATATAAATAGGGTCGATTAATAATATACATATATAAATATGTATATGTAA
TATTTCGTTTTACGTAATAATGTTTATAAATTATACCAAATATGTTAATAATATATATAATTAAG
TTTATATATATATATATTTCATTACATATAAATAAATACGTGCAATTATAATTTTATTTTATTTT
TTTTTAATTGTAAAATATAAATTTATTTTAATACAATATAATTATTGCCTTAAAACACTTAATAT
ATATTATAAAATAACTAAATAATATTATTTCAATTTATTTTATATATTTAAAATATATAATATAT
ATATATTTTGAAAATTATGTATAAAAAATTTTGAAATATAAATGAGAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAATTAAATATAAAATTAAAATTAAAATTAAATTATATATTTTAATTATTAATTTAA
TTTTTTTGTAAATTGATAAAATTCAATTTCCCAAAAAAATTTATTGTCTTATATTGTTTTATTTT
ATTTTTTTATTTTTTGTTTTTATGGAGTTGCACCGACATTAAAAAGAAACGAAACGATATATACT
TTTTTATAATCAACG**TTCACGGTTACGTTTCTGTTTTT**CTTTCTTTTTTTTTAATTATTTATTTA
TGTGTATTAACTTAAATATTTGGAAATATGCAAAATTATGCATAATACTGAAAATAAAGACTAAA
CAAATGGAAAATCAACATATATTTTATGTCTATTATATATGTCATTTCGATATTGTTATTGTTTA
TTAACAAATGGGTAAAATAATAAAAATGGTCATCATCGGTATTGTACAAAAATGTGAAAATTAAA
AAATATGTACAAAAGTAAAAATATGTTATTCAATTATATATATATATATATATATATAATATTTC
ACATACACAAACATATATATAACATATAATTGTGTACTCCTTTTTGAATACGGTTTTTATTTTTG
TTATTGTTTTTTCTTTTTATTCTGGGTTATTTTCTTCTGTACTTTTTTAAATATATATCGTAATA
ATAGTTGAGTACGCACTAATATGTCAATTTGAAAAAAAATATTGTACACGTGAATAAAGTATTAG
TAAAAAATGGTTTCATTTTTTAAGACTCCGATCATTATTTTTTTTTTCCTCTTATGTTTAAATGA
AAAGGTATTATGTTCAATAAATGAAAATGAAAATTTAGGCGAAAATAAAAACGAAAATGCAAATG
TAAACACACCTGAAAATTTAAATAAACTTCTAAATGAGTATGACAATATTGAACAATTAAAATCC
ATG

Figure legend on preceding page

6.2.1.4 *clag3.1* and *clag3.2* are mutually exclusive in parasite lines other than 3D7

Our analyses of the mutually exclusive expression of *clag3.1* and *clag3.2* thus far have focused on parasite lines that are all derived from 3D7. To determine whether *clag3.1* and *clag3.2* display mutually exclusive expression in genetic backgrounds other than 3D7, we looked at the relative expression of *clag3.1* and *clag3.2* in the other parasite lines 7G8, D10 and HB3 using the microarray data from Article 2 (Table 6.1). Quantitative data are not provided as the microarrays were designed against the 3D7 reference genome, and many of the probes were cross reactive. However, the expression of many samples was confirmed by qPCR and these samples are denoted with an asterisk. The parasite line D10 contains a single *clag3* hybrid and PFC0115c, which is situated between the two *clag3* genes, has been deleted. This is reflected in all D10-derived subclones. This was confirmed in our comparative genome hybridisation analysis (CGH) and is consistent with reports by (Iriko et al 2008). A single *clag3* and a deletion of PFC0115c are also observed in LD10, a subclone of 7G8.

In the remaining parasite lines where both *clag3.1* and *clag3.2* are present, mutually exclusive expression of the *clag3* genes is consistently observed with the exception of two parasites lines, KG7 and WE5, subclones of 7G8. *clag3.1* and *clag3.2* are expressed at similar levels in KG7, whereas *clag3.1* expression predominates in WE5, but *clag3.2* expression is also detected. One possible explanation for these two observations is that after subcloning, some of the parasites have switched to express *clag3.2*. As previously observed in Figure 6.4, the transcript associated with the pseudogene, PFC0115c, does not correlate

with the expression status of the *clag3* genes. The microarray does not contain probes that recognise PF03TR002.

This suggests that the mutually exclusive expression of *clag3.1* and *clag3.2* is not solely a phenomenon associated with the 3D7 genetic background, but rather that it is an inherent property of these two genes.

Table 6.1 Qualitative summary of *clag3.1, clag3.2* and a *var* pseudogene (PFC0115c) expression in different parasite lines and their subclones. The data were compiled from the microarray hybridisations described in Article 2. These parasite lines are described in the methods section of Article 2. Please refer to the legend for further details.

		clag3.2	PFC0115c[a]	*clag3.1*
3D7	3D7-A	+*	+*	+*
	10G	+*	+*	X*
	1.2B	X*	X*	+*
	W41	+	X	X
	3D7-B	X	+*	+
7G8	7G8	X*	+	+*
	KG7[!]	+*	+	+*
	LD10	#	-#	#
	WE5[!]	+*	+	+*
	ZF8	X*	+	+*
D10	D10	#&	-&	#&
	E3	#	-	#
	F1	#	-	#
	G2	#	-	#
	G4	#	-	#
HB3	HB3A	+*	+	X*
	AB6	+*	+	X*
	AB10	+*	+	X*
	HB3B	X*	+	+*
	BB8	X*	+	+*
	BC4	X*	+	+*

LEGEND

+	expressed
+	low level of expression
X	not expressed
-	deletion
	one *clag3* hybrid

* - confirmed by qPCR
\# - confirmed by CGH
& - consistent with Irikio et al. 2008
a - microarray data based solely on one probe
! - discrepancy in *clag3.1* and *clag3.2* mutual exclusion

6.2.2 Chromosome conformation capture (3C) and circular chromosome conformation capture (4C)

6.2.2.1 3C strategy for *clag3.2* and 4C strategy for *clag3.2* and the *var* gene PF07_0048

Since we have no *a priori* knowledge to guide our 3C analysis, we have directed our 3C analysis to ≈65 kb upstream and ≈17 kb downstream of *clag3.2* (Figure 6.7). Based on RE analysis, *Bcl*I and *Bgl*II appear to be suitable restriction enzymes for the 3C analysis of *clag3.2* (Figure 6.7A). Using these two enzymes, the entire *clag3.2* 5' intergenic region, ORF and 3' intergenic region can be used as the bait (Figure 6.7B). The digestion with *Bcl*I and *Bgl*II results in 13 suitable restriction fragments (Figure 6.7A, named 1-10 and A-C).

Primers specific to each restriction fragment and to the bait are designed in order to determine the interaction frequency between each fragment with the bait. When the interaction frequencies of each fragment with the bait are plotted against genomic distance, an interaction is identified by a local peak in interaction frequency (Dekker 2006).

The interaction frequency is determined by quantifying the amount of the corresponding ligation product in the 3C sample by qPCR. To determine the interaction frequency for each potential interacting fragment with the bait, each primer pair (i.e. one primer specific to the bait plus one primer specific to the fragment of interest) is divided by the amount of PCR product obtained in the 3C template by the amount of PCR product obtained from the control template.

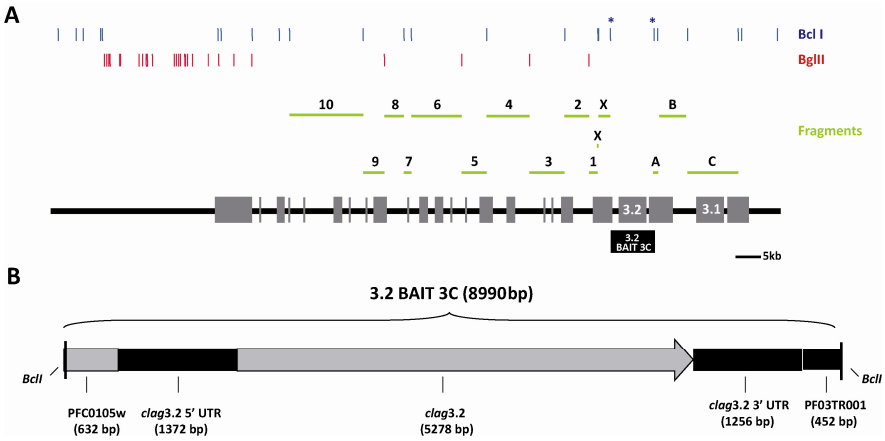


Figure 6.7 Schematic of the 3C strategy, using *clag3.2* as the bait and *Bcl*I and *Bgl*II as the restriction enzymes. A. The position of the *Bcl*I and *Bgl*II restriction sites are indicated in blue and red respectively. The two *Bcl*I restrictions sites that delineate the *clag3.2* bait are indicated with asterisks. The green fragments numbered 1-10 and A-C are included in the analysis, whereas the green fragments marked with an X are excluded either because the fragment is too small or specific primers could not be designed in these regions. **B.** A more detailed description of what constitutes the 3.2 BAIT 3C fragment.

The control template refers to a sample that contains, in equimolar amounts, all possible ligation products that could occur between the fragments of interest and the bait. This template also serves as a control for primer amplification efficiencies (Splinter et al 2004). This control template for *clag3.2*, containing all fragment-bait interactions, was generated by PCR amplification of DNA fragments that span each of the restriction sites of interest (Figure 6.7A). These fragments were gel purified, digested with the appropriate restriction enzyme (either *Bcl*I or *Bgl*II) and ligated to the digested bait fragment (data not shown).

Based on restriction enzyme analysis, *Dpn*II appears to be a suitable enzyme for the 4C analysis of *clag.3.2* and the *var* gene PF07_0048, as the baits encompass their entire 5' UTRs (Figure 6.8A-D). Upon the formation of circular ligation products, 4C is able to detect interactions between the bait of interest and unknown sequences that could reside anywhere in the genome. Consequently, this technique relies on the amplification of the unknown sequence by inverse primers that are specific to the bait. The position of these primers for the *clag3.2* and *var* baits are indicated as arrows in Figures 6.8B and D.

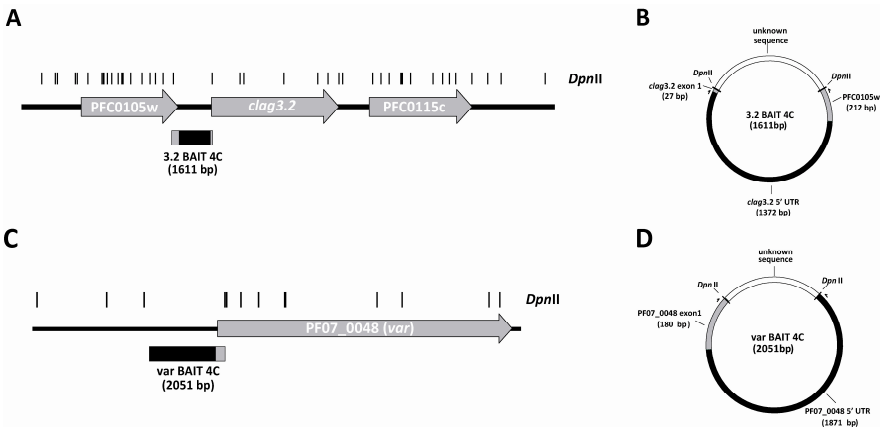


Figure 6.8 Schematic of the 4C strategy for *clag3.2* and the *var* gene PF07_0048, using *Dpn*I as the restriction enzyme. **A.** *Dpn*II restriction sites around the *clag3.2* locus. *Dpn*II restriction sites are indicated with black lines **B.** A more detailed description of what constitutes the 3.2 BAIT 4C fragment (ligated to a hypothetical interacting fragment). **C.** *Dpn*II restriction sites around the PF07_0048 locus. **D.** A more detailed description of what constitutes the var BAIT 4C fragment (ligated to a hypothetical interacting fragment). Arrows in panels B and D represent primers.

6.2.2.2 Design of Southern blot experiments for the analysis of restriction enzyme digestion efficiency in cross-linked nuclei

To determine the efficiency of the restriction enzyme digestions in cross-linked nuclei for 3C and 4C experiments, Southern blot analysis was performed. A schematic and description of each Southern blot experiment can be found in Figures 6.9 – 6.13.

Figure 6.9 Schematic for Southern blot analysis for *Bcl*I digestion of the 3.2 BAIT 3C. *Bcl*I restriction sites are indicated with blue lines and the particular *Bcl*I restriction site under analysis is indicated with an asterisk. The restriction enzymes used to digest the sample for Southern blot analysis were *Nco*I and *Nde*I, their restriction sites indicated with green and orange lines respectively. The position of the radiolabelled probe is indicated along with the size of the expected bands.

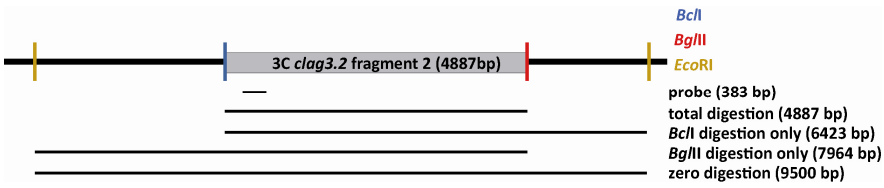


Figure 6.10 Schematic for Southern blot analysis for *BclI* and *Bgl*II digestion of 3C *clag3.2* fragment 2. *Bcl*I and *Bgl*II restriction sites are indicated with blue and red lines respectively. The restriction enzyme used to digest the sample for Southern blot analysis was EcoRI, its restriction sites are indicated with yellow lines. The position of the radiolabelled probe is indicated along with the size of the expected bands. The location of fragment 2 can be found in Figure 6.7.

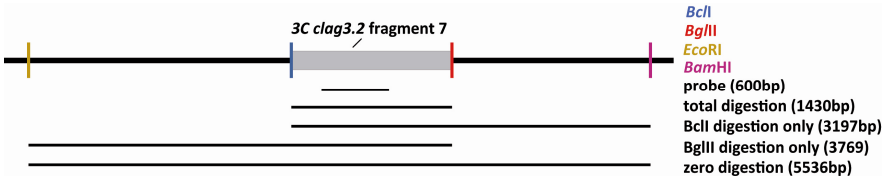


Figure 6.11 Schematic for Southern blot analysis for *Bcl*I and *Bgl*II digestion of 3C clag3.2 fragment 7. *Bcl*I and *Bgl*II restriction sites are indicated with blue and red lines respectively. The restriction enzymes used to digest the sample for Southern blot analysis were *EcoR*I and *BamH*I, their restriction sites are indicated with yellow and purple lines respectively. The position of the radiolabelled probe is indicated along with the size of the expected bands. The location of fragment 7 can be found in Figure 6.7.

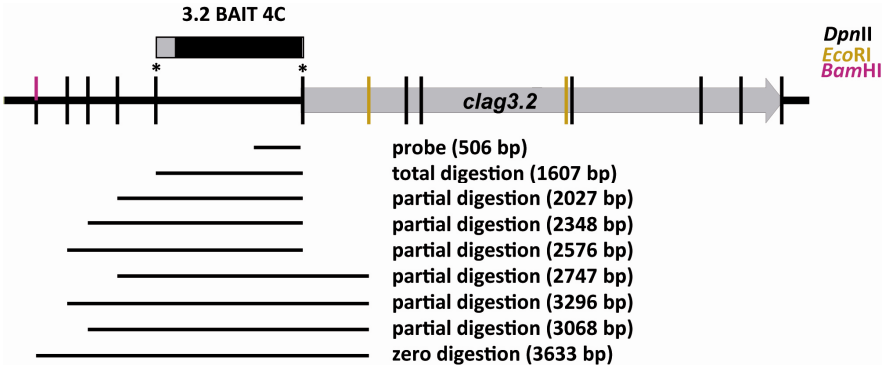


Figure 6.12 Schematic for Southern blot analysis for *Dpn*II digestion of 3.2 BAIT 4C. *Dpn*II restriction sites are indicated with black lines and the two *Dpn*II sites that create the bait fragment for *clag3.2* are indicated with asterisks. The restriction enzymes used to digest the sample for Southern blot analysis were *EcoR*I and *BamH*I, their restriction sites are indicated with yellow and purple lines respectively. The position of the radiolabelled probe is indicated along with the size of the expected bands.

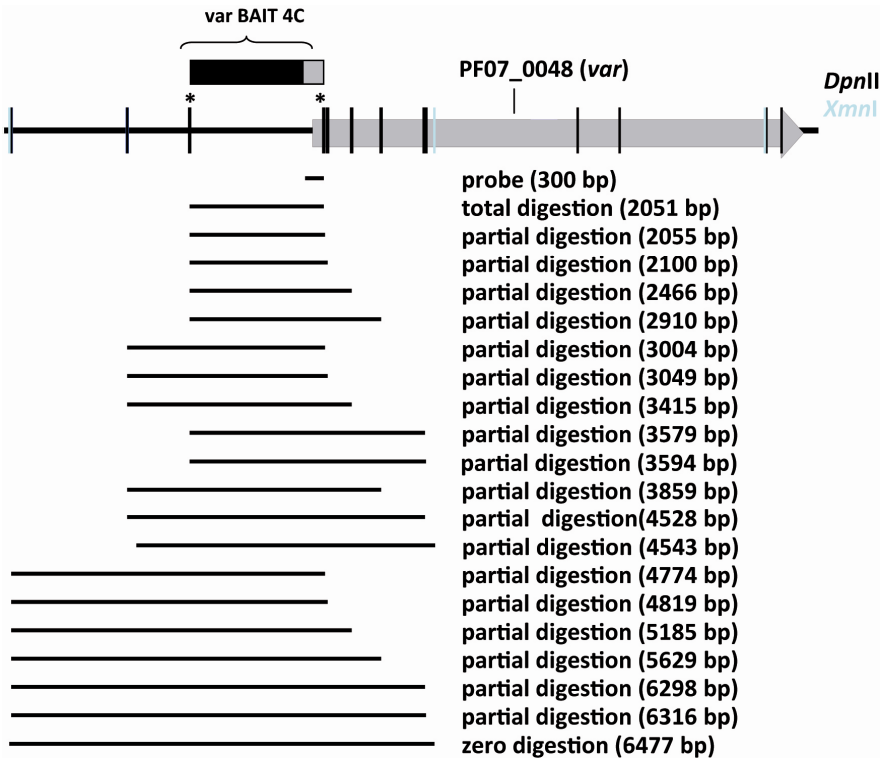


Figure 6.13 Schematic for Southern blot analysis for *Dpn*II digestion of var BAIT 4C. *Dpn*II restriction sites are indicated with black lines and the two *Dpn*II sites that create the bait fragment for PF07_0048 are indicated with asterisks. The restriction enzyme used to digest the sample for Southern blot analysis was *XmnI*, its restriction sites are indicated with light blue lines. The position of the radiolabelled probe is indicated along with the size of the expected bands.

6.2.2.3 With optimization, cross-linked nuclei can be efficiently digested by *Bcl*I, *Bgl*II and *Dpn*II

To determine the digestion efficiency for the 3C and 4C RE, Southern blot analysis was performed (Figure 6.14). Lanes denoted with a U represent undigested controls and those denoted with a grey circle denote the original digestion conditions. Southern blot analysis revealed that under the original digestion conditions, *Bcl*I digested the DNA in formaldehyde cross-linked nuclei

efficiently. This is observed in Figure 6.14A for 3.2 BAIT 3C in the lane denoted with a circle. This band corresponds to the expected size for total digestion by *Bcl*I (see Figure 6.9 for details). However, Southern blot analysis revealed that *Bgl*II and *Dpn*II did not efficiently digest the DNA in formaldehyde cross-linked nuclei. For *Bgl*II, this is observed in Figure 6.14B and C in the lanes denoted with a circle, where the prominent band corresponds to digestion solely by *Bcl*I and not *Bgl*II (see Figure 6.10 and 6.11 for details). For *Dpn*II, this is observed in Figure 6.14D and E in the lanes denoted with grey circles, as the bands in these lanes correspond to partial *Dpn*II digestion products (see Figure 6.12 and 6.13 for more details).

Changes were made to the restriction enzyme digestion protocol to optimise the digestion efficiency for *Bgl*II and *Dpn*II, and these lanes are denoted with a star. The lowermost bands in the starred lanes in Figure 6.14B, C, and D correspond to the expected sizes for total digestion by both *Bcl*I and *Bgl*II (B, C) or *Dpn*II (D). The starred lane in Figure 6.14E contains two bands, the lowermost band corresponds to the expected size for total digestion by *Dpn*II. It is possible that the upper band is one of the numerous partial *Dpn*II digestion products; however, comparing the intensities of the two bands, the majority of the sample appears to be fully digested by *Dpn*II.

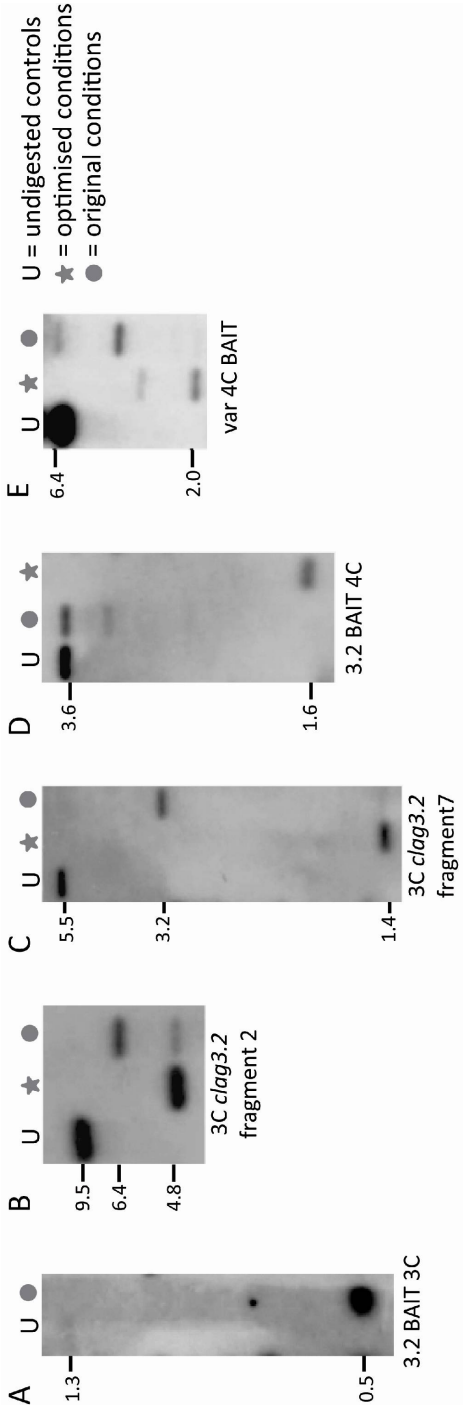


Figure 6.14 Optimization of 3C and 4C restriction enzyme digestions. Southern blot analysis comparing the restriction enzyme digestion efficiencies between the original digestion conditions and the optimized digestion conditions. Schematics showing the restriction sites used for Southern blot analysis and position of the probes are located in Figures 6.9-6.13. The sizes correspond to kb.

Now that sufficient digestion of cross-linked nuclei can be achieved, we can begin to optimise ligation conditions in order to test our limiting enhancer or repressor hypothesis for the mutually exclusive expression of *clag3.1* and *clag3.2* by 3C and 4C.

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

7. General Discussion

Collectively, the work described in this thesis has contributed to our understanding of the epigenetic regulation of clonally variant gene expression in *P. falciparum*. We have demonstrated that *clag3.1*, *clag3.2* and *eba-140* are good models to study transcriptional silencing associated with life cycle progression, clonally variant and mutually exclusive gene expression and that isogenic parasite lines that differentially express these genes are useful tools in the study of variant gene expression. Importantly, we provide insights that implicate epigenetics in the regulation of genes apart from the *var* family of genes, and provide evidence that this mode of regulation is an inherent biological property of the parasite that regulates processes apart from antigenic variation.

Initial promoter characterisation revealed that *clag3.1*, *clag3.2* and *eba-140* contain multiple TSSs and gene reporter assays revealed that multiple proximal *cis* elements (enhancers) contribute to maximal promoter activity during the schizont stage. This data indicates that these genes conform to the typical bipartite structure of eukaryotic promoters. When assayed in trophozoites, their 5' flanking regions did not drive the expression of the reporter gene, indicating that these regions contain the necessary information to dictate correct temporal expression, and that they contain binding sites for stage-specific transcription factors. qPCR analysis confirmed the differential expression of *clag3.1*, *clag3.2* and *eba-140* in the isogenic parasite lines 10G and 1.2B, and additional luciferase assays in these two parasite lines confirmed that the differential expression of these genes in 10G and 1.2B is not due to the

presence or absence of STFs that are present in one parasite line or the other. When these genes are expressed, changes in nucleosome position were observed in comparisons between 10G and 1.2B, indicating that chromatin remodelling complexes are involved in granting the transcriptional machinery access to the underlying DNA. When *clag3.1*, *clag3.2* and *eba-140* are repressed during schizont stages and bookmarked for repression at the trophozoite stage, we observe reduced restriction enzyme accessibility, suggesting that these regions are in a heterochromatic conformation and that that this conformation persists along the IEC.

Histone modifications are associated with the CVGE of clag3.1, clag3.2 and eba-140 in schizonts and with the epigenetic memory of these genes in trophozoites

qChIP analysis was performed to characterise the PTMs associated with the clonally variant expression status of *clag3.1*, *clag3.2* and *eba-140* in 10G and 1.2B parasites at the schizont stage. This analysis demonstrated that the clonally repressed status of these three genes is associated with an enrichment of H3K9me3, whereas the clonally active status of these three genes is associated with enrichment of H3K9ac. As these genes can adopt both open and compact conformations, the chromatin architecture of these regions can be described as being organised into facultative heterochromatin. Modest enrichments of H3K4me3 and H4ac are also associated with the clonally active state of these three genes.

At the molecular level, facultative heterochromatin in *P. falciparum*, which is hypoacetylated and enriched in H3K9me3, is more similar to the constitutive heterochromatin (hypoacetylated and enriched in H3K9me3, H4K20me) than the facultative heterochromatin (hypoacetylated and

enriched in H3K27me3, H3K27me2, H3K9me2) found in other species. Perhaps constitutive heterochromatin has been evolutionarily co-opted by *P. falciparum* into a form that is functionally facultative without the use of H3K27 methylation, PTMs that the parasite lacks.

Concurrent with the publication of Article 1, the Duraisingh lab published a study where they also characterised the PTMs H3K9me3, H3K9ac and H3K4me3 associated with the active and silent states of *clag3.1* and *clag3.2* in schizonts (Comeaux et al 2011). The two data sets are in agreement as they also report enrichments of H3K9ac and H3K9me3 with the active and repressed states, respectively, of the *clag3* genes. Their study reports a larger enrichment of H3K4me3 near the TSSs of the active *clag3*, where we observe a more modest enrichment of this modification. In addition to these PTMs, they also analysed the distribution of H3K4me2, a PTM that we did not include in our analyses. This PTM is enriched in active *var* genes and *var* genes that are bookmarked for activation (Lopez-Rubio et al 2007a), however, they did not observe a correlation of H3K4me2 with *clag3* expression status, suggesting that perhaps different clonally variant gene families use different PTMs to bookmark for activation.

We performed qChIP analysis on 10G and 1.2B parasites at the trophozoite stage to characterise the pattern of PTMs associated with *clag3.1*, *clag3.2* and *eba-140* when bookmarked either for expression or repression. These analyses revealed that these genes, when bookmarked for expression, are also enriched in H3K9ac and when bookmarked for repression are enriched in H3K9me. The persistence of H3K9me3 and H3K9ac along the IEC also suggests that they are capable of clonally transmitting epigenetic memory, discussed further below.

Collectively, our study and that by Comeaux and colleagues indicate that akin to the *var* genes, the repressed and active states of other clonally variant gene families, such as *clag* and *eba,* are associated with enrichments in H3K9me3 and H3K9ac respectively. Similarly, the *clag*, *eba* and *var* gene families are enriched in H3K9me3 when they are bookmarked for repression.

H3K9me3 and H3K9ac may function as epigenetic signals for the transmission of clag3.1, clag3.2 and eba-140 expression status

Not all factors that contribute to the maintenance of an epigenetic phenotype function as epigenetic signals capable of transmitting the requisite information for gene expression to the next generation, and it is often difficult to discern what molecular signals carry epigenetic information and which are effectors of certain processes (Campos & Reinberg 2009).

For example, the histone variant H3.3 and the PTMs H3K4me3 and H3K9ac are all characteristic of euchromatin; H3K9ac has been shown to directly recruit TFIID, H3K4me3 is often found at the TSS of actively transcribed genes and H3.3 is often incorporated at the site of active transcription (Agalioti et al 2002; Santos-Rosa et al 2002; Talbert & Henikoff 2010). These recurrent euchromatic signals may create characteristic chromatin landscapes that facilitate a certain process, such as transcription, and might even be considered hallmarks of such processes, but whether they transmit cellular memory is another matter.

With the observation that both H3K9me3 and H3K9ac persist along the 48 h IEC, we propose that these PTMs fulfil the criteria necessary to be considered true epigenetic signals that transmit cellular memory for the

expression status of *clag3.1*, *clag3.2* and *eba-140*. To be classified as an epigenetic signal, the signal must i) self perpetuate ii) be heritable iii) affect gene expression iv) be reversible. Although we do not know the mechanism by which these two PTMs self perpetuate, we do know that they are maintained after DNA replication and that these PTMs are inherited within clonal populations (criteria i and ii). We also know that H3K9ac and H3K9me3 have positive and negative effects on gene expression, and that enzyme activities exist to remove (and deposit) these modifications (criteria iii and iv).

Additionally, both states (activation and repression) are stable under homogeneous conditions and thus their transmission is independent from a triggering event. Consequently, H3K9me3 and H3K9ac fit well into the model of heritable bistability for the transmission of PTM-mediated epigenetic memory (Dodd et al., Figure 1.8). The control of gene expression by the acetylation/methylation balance at a single residue, H3K9, also fits well into this bistable model.

This suggests that H3K9ac can function both as an epigenetic signal capable of transmitting cellular memory and as an effector of transcription. Unlike the highly localised distribution of H3K9me3 in *P. falciparum*, ChIP-on-chip analysis of H3K9ac show that it has a broad distribution (Lopez-Rubio et al 2009; Salcedo-Amaya et al 2009), and its role as an epigenetic signal of activation for some clonally variant gene families could not have been implied from these studies and could only be determined by comparing subclones (or selected parasite lines) that stably express and repress genes of interest. Alternatively, this could also suggest that H3K9ac might be the default PTM that marks many genes. *P. falciparum* is often described as having a largely euchromatic genome as

it lacks many of the PTMs associated with silencing, such as methylation of H3K27me3, H3K36 and H3K79 (Salcedo-Amaya et al 2009; Trelle et al 2009). Perhaps H3K9ac is only absent when methylation has been deposited on the same residue in clonally variant genes.

CVGE is controlled by PTMs, whereas progression along the IEC is controlled by stage-specific transcription factors

The persistence of H3K9ac along the IEC, indicated that when bookmarked for activation, the chromatin is in a transcriptionally competent conformation; however, *clag3.1*, *clag3.2* and *eba-140* are not transcribed because of the absence of the appropriate stage-specific transcription factor. This also suggests that transcription associated with progression along the IEC is independent of the PTMs that are associated with CVGE.

The conclusion that correct temporal expression of genes along the IEC is dictated by stage-specific transcription factors is consistent with other studies where plasmids containing *P. falciparum* promoters transcribed reporter genes in an appropriate temporal manner, such as in the falcipain genes and PfPCNA, just to name a few (Sunil et al 2008; Wong et al 2011). However, as we see temporal expression of the *clag* and *eba* families that is not accompanied by changes in chromatin architecture, our observations are not consistent with a recent study that suggests that expression along the IEC is regulated by global changes in chromatin conformation (Ponts et al 2010).

Collectively, data from the gene reporter, nuclease accessibility and qChIP assays lead us develop a model whereby CVGE is controlled by the formation of H3K9me3-facultative heterochromatin, and that correct

temporal expression of clonally variant genes is dependent on the presence or absence of the appropriate stage-specific transcription factor (Article 1, Figure 7).

Our model is supported by experiments conducted in the Horrocks laboratory where the transcription of PfPCNA, a gene expressed in trophozoites, was studied upon treatment of parasite cultures with sublethal doses (IC_2-IC_7) of SAHA (suberoylanilide hydroxamic acid) and SBHA (suberohydroxamic acid), type I and type II HDAC inhibitors (Wong et al 2011). With these sublethal doses of inhibitor, they observed a change in the absolute expression of a reporter gene driven by the PfPCNA 5' flanking region, but did not observe expression in rings or schizonts, indicating that global hyperacetylation did not result in changes to the temporal expression pattern of PfPCNA.

Our model could be tested if we knew the identity of the transcription factor for a clonally variant gene, for example *eba-140*, constitutively expressed this transcription factor in parasite lines that differentially express *eba-140* (i.e. in 10G and 1.2B) and assay for *eba-140* expression along the IEC. According to our model, we would expect *eba-140* expression throughout the IEC in a 10G parasites where the chromatin is maintained in a transcriptionally permissive conformation along the 48 h IEC. In contrast, we would not expect *eba-140* expression in 1.2B parasites at any point along the IEC.

Barrier insulators could separate neighbouring clonally variant genes into distinct chromatin domains

The complex pattern of PTMs observed for *eba-140* suggests that a barrier insulator is located within its first exon. The large increase in marks associated with activation adjacent to a heterochromatic region is consistent with barrier insulator activity. In yeast, large increases in both histone acetylation and H3K4me3 have been documented adjacent to barrier insulators, due to the active recruitment of the barrier for HATs and H3K4me3 HMTs (Oki et al 2004; West et al 2004; Zhou & Berger 2004). The observation that the inflection point where the transition from repressive to active PTMs occur is found almost at the same position in both 10G and 1.2B parasite lines is also suggestive of barrier activity. In the absence of a barrier, as in position-effect variegation (PEV), the balance of heterochromatic and euchromatic activities along a chromosome is what dictates the point of transition (Gaszner & Felsenfeld 2006), and consequently, we would expect to see different points of transition between different parasite clones . This suggests that a barrier is present to shield against the silencing of two neighbouring genes, *eba-140* and MAL13P1.59 from heterochromatin spreading. The differential expression of *clag3.1* and *clag3.2*, two other neighbouring genes could perhaps be maintained by a barrier element.

The position of the potential barrier insulator activity within an ORF as observed in *eba-140* is atypical when compared to what is known in other systems. For example, genome wide mapping of *Su(Hw)* (a component of the *D. melanogaster* gypsy insulator) and of CTCF in the human genome, revealed that their respective binding sites were predominately in intergenic regions or within large introns of large genes,

with only 3-4% of their binding sites corresponding to ORFs (Ramos et al 2006; Cuddapah et al 2009). However, now that the site of a potential barrier insulator has been identified within the coding region of *eba-140*, studies can begin to try to characterise the nature of this barrier and attempt to find others in the *P. falciparum* genome.

Genome wide analysis of CVGE reveals that the P. falciparum genome is epigentically plastic

Characterisation of the level of transcriptional variation in genetically identical parasites revealed that *P. falciparum* is epigenetically plastic, even in the absence of selection, suggesting that CVGE is an inherent property of *P. falciparum* biology.

Two studies published during the course of this work mapped the global distribution of H3K9me3 and HP1 in *P. falciparum* and it was observed that these heterochromatic marks were restricted to gene families that mediate host-parasite interactions and it was hypothesized that these genes could show CVGE (Lopez-Rubio et al 2009; Flueck et al 2009). Our work complements these studies, confirming a striking correlation between genes that show CVGE and heterochromatic marks. Not only was this phenotypic variation observed in antigenically important gene families but also in gene families that are never exposed to host immunity.

Since CVGE is not solely associated with evasion of host immunity, and observed in both conserved and hypervariable gene families, we propose the CVGE is a bet-hedging strategy employed by the parasite to mediate fast, short time adaptation (Figure 7.1). We propose that CVGE caused

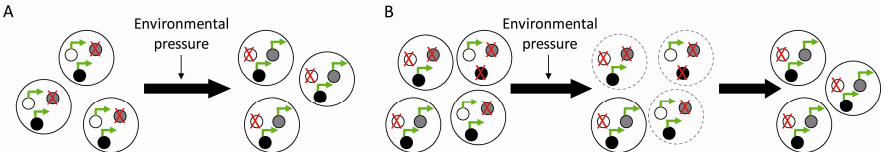


Figure 7.1. Model for adaptation through directed transcriptional responses (A) versus adaptation through clonally variant gene expression (CVGE) (B). Small circles represent individual genes that can be either expressed (green arrows) or repressed (red). In A, a change in the environment results in an immediate protective transcriptional response, such that the genes are activated or repressed to mediate survival under the new conditions. In contrast, in (B), the population is transcriptionally heterogeneous to being with as a consequence of spontaneous CVGE. Upon a change in the environment, pre-existing parasite with the combination of expressed and repressed genes that confer fitness under the conditions are selected and survive, whereas those that do not, die (discontinuous line). The end result is similar, but mechanism A results in immediate adaptation where the individuals survive, whereas in B, some of the parasites are eliminated.

by spontaneous switches that result from the molecular noise in bistable systems (previously described in section 1.4.3, Figure 1.8), is exploited by the parasite to generate transcriptional heterogeneity in a clonal population. This could confer an adaptive advantage to a parasite population, if the individuals that posses the appropriate combination of expressed and repressed genes survive a change in the environment (Figure 7.1, see legend for details). This is suggestive of an H3K9me3-facultative heterochromatin based mechanism that contributes to phenotypic variation and consequently parasite fitness. To test this hypothesis the adaptive potential of transcriptionally homogeneous and transcriptionally heterogeneous parasite lines would have to be compared upon different challenges to determine if transcriptional heterogeneity confers an adaptive advantage.

Not only is the knowledge regarding epigenetic plasticity important in understanding the basic biology of *P. falciparum*, but knowing which

genes and families are capable of CVGE has implications regarding control and eradication strategies, such as vaccine design. We suggest that CVGE is an inherent property of certain genes and gene families. For example, the five member EBL family that mediates erythrocyte invasion has the capacity for CVGE as demonstrated by *eba-140*. Although other members have not been captured in a repressed state in our microarray studies, *eba-165* has been shown to be marked by HP1, strongly suggesting that it also has the capacity to be repressed by facultative heterochromatin (Flueck et al. 2009). Another member of the EBL family that we did not detect CVGE for *in vitro* is *eba-175*. A knockout of *eba-175* by the Beeson lab demonstrated that its disruption resulted in a slightly reduced growth rate (Persson et al 2008), and consequently, *eba-175* repressing parasites would be outcompeted by parasites that express it and thus, the repressed state of *eba-175* would not be captured in our microarrays. Antibodies acquired by malaria-exposed children in Kenya against EBA-175 have been shown to inhibit invasion of *P. falciparum in vitro* (Persson et al 2008) and EBA-175 is a target for two blood-stage vaccines currently in clinical development, the EBA-175-RII NG and JAIVAC-1 vaccines that either contain the EBA-175 antigen alone or in combination with other antigens (Crompton et al 2010; El Sahly et al 2010). Although EBA-175 it is a dominant invasion ligand in field isolates sampled from The Gambia, a cluster of the samples taken and cultured *ex vivo* did not express *eba-175*. Additional evidence for the repression of *eba-175* comes from some field isolates from Kenya where it was expressed at very low levels and the authors could not attribute this to differences in copy number between samples, (Nery et al 2006), strongly suggestive that *eba-175* could be repressed in these samples. Therefore, even if vaccines targeting EBA-175 prove to be strongly immunogenic,

doubts regarding their widespread applicability could arise in light of CVGE.

Proof-of-principle experiments indicate that the CVGE of clag3.2 can be reproduced in an episomal system

Since CVGE is emerging as a fundamental process in parasite biology that could play an important role in parasite adaptation and affect control and eradication strategies, the molecular mechanism involved in establishing this phenomenon and how switches are mediated need to be understood. The CVGE of *var* genes has been mimicked in an episomal system (Voss et al 2006), and the study of CVGE for other gene families would be greatly facilitated if the CVGE of other gene families could also be reproduced in an episomal system. An episomal system would permit us to quantify switch rates, test possible insulator elements and allow for the tagging and identification of proteins that are involved in these molecular processes.

However, before a sophisticated episomal system can be developed as a tool to study CVGE, proof-of-principle experiments were required to determine whether CVGE could be mimicked in an episomal for other gene families. The large *var* family of genes represent a peculiar family where most members are repressed and it could not be assumed that CVGE could be reproduced for other smaller gene families where most members are expressed.

We began these proof of principle experiments by transfecting the 10G parasite line with HLHprom3.2bsdR, a plasmid containing the entire *clag3.2* 5' flanking region driving the expression of luciferase, and a drug

resistance cassette driven by the constitutive promoter *hsp86*. Subcloning of this transfected population by limiting dilution resulted in distinct subpopulations that maintained the plasmid in different transcriptional states as determined by gene reporter assays. These subcloned parasite lines were grown in parallel and in homogeneous conditions.

This preliminary experiment suggested that the episomal *clag3.2* is capable of switching between the ON and OFF state and demonstrates that the *clag3.2* 5´region is stably transmitted in both the active and repressed states. However, we cannot exclude the possibility that alternative chromatin states were induced or established at the time of transfection and are thus not the result of switching events. To demonstrate that the episomal *clag3.2* is capable of switching between the ON and OFF state would require assaying gene reporter activity in second generational clones (i.e. subclones of 10Gluc-derived subclones).

Subclones that differentially expressed the episome were selected and tightly synchronised to ensure that the observed differences in luciferase expression was not due to difference in parasite stage between cultures, and we confirmed by gene reporter assays that indeed, tightly synchronised subclones differentially expressed the episome. These results were confirmed by qPCR, indicating that these differences occur at the level of transcription.

Furthermore, the differences could not be attributed to different relative amounts of episome maintained by each parasite line, as the parasite

line that showed the highest luciferase expression contains that lowest relative amount of episome compared to the other parasite lines.

Although the origin of the parasite lines and subclones used in the microarray analysis indicated that genetically identical parasite populations are transcriptionally heterogeneous, we have been able to provide experimental evidence of this with our reporter system. These experiments also show that CVGE is mediated solely by the promoter of the clonally variant gene, as the only component of a variant gene contained within this plasmid is the 5' flanking region of *clag3.2*.

Based on the qPCR results on the tightly synchronised parasite lines, it appeared that the parental population predominately maintains the HLHprom3.2bsdR episome in a repressed state. To confirm this, HLHprom3.2bsdR was used in transient transfections in 3D7-A parasites in parallel with two other plasmids that contain the *clag3.2* 5' flanking regions but do not contain drug resistance cassettes. In comparison with 3.2-LH-1371-R20 and 3.2-LH-1371, HLHprom3.2bsdR shows reduced luciferase expression, which is suggestive that the activity from the second transcriptional unit represses the activity of the episomal *clag3.2* promoter. A similar result was observed with *var* genes. When removed from their chromosomal context and placed on an episome, *var* genes are active, even after a DNA replication phase when naked episomes are formed into chromatin (Deitsch et al 1999). However, when placed in an episome in the presence of any second promoter activity and after an S-phase, *var* promoters are repressed (Deitsch et al 2001; Dzikowski et al 2007). Many more experiments are required to determine if promoter-pair silencing is a phenomenon for all variant promoters, of mutually

exclusive promoters or an artefact of these transfection assays. This would require testing another variant promoter, such as *eba-140*, and a non-variant promoter, such as a*ma1*, coupled to a second promoter. Apart from the genes in this study and *var* genes, no other functional promoter characterisation has been conducted with plasmids containing two promoter activities to allow for comparisons (Horrocks et al 2009 and sources therein).

To determine if episomal recombination events between the subclones are what contribute to the observed differences in luciferase expression, we conducted a preliminary PCR screen on the episomes. Of particular interest was the parasite line 2B6, the subclone that expresses the episome, since unpaired *var* promoters become active, perhaps 2B6 was active due to a recombination event that uncoupled the promoter activities of *clag3.2* and *hsp86*. However, we did not detect any recombination events with these particular analyses, suggesting that the episomes may not be rearranged and that the observed differences in episomal expression are attributable to other factors, which could include the spontaneous establishment of different chromatin states.

To determine if differential PTMs are associated with the active and repressed state of the episome, qChIP was performed on 2B6 and 2F12, parasite lines that respectively express and repress the episome. In accordance with our hypothesis, we observed an enrichment of H3K9me3 in 2F12, and an enrichment of H3K9ac in 2B6. This suggests that the 5' flanking region of *clag3.2* is sufficient to nucleate the formation of H3K9me3-facultative heterochromatin. Cumulatively, our data suggests that the CVGE of the *clag* family can be mimicked in an

episomal system. Our data indicates that episomal CVGE for *clag3.2* happens in a spontaneous fashion. This episomal CVGE is mediated by the 5' flanking region of *clag3.2*, the expression status of the episome is clonally inherited, and the repressed and active episomes are correlated with enrichments in H3K9me3 and H3K9ac respectively. Therefore, we provide support for the development of episomal-based systems to be used as tools to probe the biology of CVGE.

Mutually exclusive expression of clag3.1 and clag3.2

The mechanism of mutually exclusive gene expression observed in different systems remains elusive. The *clag3* genes are an amenable system in which to study the molecular mechanism of mutually exclusive gene expression in *P. falciparum,* and isogenic parasite lines that differentially express these two genes will serve as valuable tools towards the elucidation of this phenomenon. The *clag3* genes are the only other genes currently known that are subject to mutual exclusion apart from the *var* family, which consists of ≈60 members.

The molecular mechanism involved in maintaining mutually exclusive *var* expression has been the subject of intense interest, mainly because this phenomenon, coupled with a low switching frequency, is believed to contribute to waves of parasitemia and persistent infections. Therefore, not only would elucidating the molecular mechanism of mutual exclusion for *clag3* contribute to the understanding of this phenomenon in general, but it could be reasonable to hypothesize that similar mechanisms could apply to the antigenically important *var* gene family, where one could envision that interventions that disrupt their mutual exclusion could reduce chronic infections.

The 10Gluc and derived subclones used to model episomal CVGE also present an opportunity to assess what happens to the mutually exclusive expression status of endogenous *clag3.1* and *clag3.*2 in the presence of multiple episomal *clag3.2* promoters. The HLHprom3.2bsdR plasmid was transfected into the 10G parasite line, which is known to express *clag3.2*, however, upon qPCR analysis, 10Gluc and its derived subclones all expressed *clag3.1*.

The transfection and maintenance of the HLHprom3.2bsd plasmid could have induced a switch in 10Gluc and 10Gluc-derived subclones. It is also possible that over time, these parasite lines would have switched to *clag3.1* expression in the absence of the plasmid. To exclude this second possibility, nontransfected 10G would need to be grown in parallel and for the same number of cycles and then check the expression status of the *clag3* genes. We know from previous experiments that the expression pattern of *clag3.1, clag3.2* and *eba-140* is stable for at least 25-30 cycles (Cortés et al 2007). However, these parasites were grown longer, for approximately 54 asexual cycles, and the expression of these genes has not been checked in clonal untransfected parasite lines cultivated *in vitro* for this duration of time. Random switching to express *clag3.1* is unlikely as it occurred in all five parasite lines. It has been observed that *clag3.2* expressing subclones (derived from 3D7) maintained in the Duraisingh lab, quickly, even after approximately 15 asexual cycles, begin to switch to *clag3.1* expression (Comeaux et al 2011).

Whether the expression of *clag3.1* confers a growth advantage over *clag3.2* in a 3D7 genetic background is unknown, but if that were the case, we would expect solely *clag3.1* expressing clones *in vitro*, and subclones of 3D7-A produce *clag3.1* and *clag3.2* clones in a roughly equal proportion that are stable for many asexual cycles (Cortés et al 2007). The fact that the Duraisingh lab also have stable 3D7 clones that express *clag3.2* indicates that this is not a phenomenon associated with 3D7-A. The Duraisingh lab reports that they could readily disrupt the *clag3.2* locus but not *clag3.1* (Comeaux et al 2011). Possible advantages of expressing *clag3.1* remain puzzling and experiments designed to address this question are required.

Interestingly, the Duraisingh lab generated a transgenic parasite with a disrupted *clag3.2* that also has *clag3.1* repressed, effectively resulting in a double *clag3* knockout line (Comeaux et al. 2011). This parasite line was generated by a single homologous recombination event that introduced a drug resistance gene driven by a calmodulin promoter into the *clag3.2* ORF, keeping the endogenous *clag3.2* promoter intact. With the addition and removal of drug pressure, they observe that the repression of *clag3.2* was reversible. They also performed a parasite competition assay whereby they grew this *clag3* null line together with another parasite line and observed that the null line was out competed. In comparison with the other parasite line, the *clag3* null line had a slight reduction in growth rate of ≈2.5% and the authors concluded that the RhopH complex is required for efficient parasite growth. However, they previously observed that the drug resistance gene is spontaneously repressed in the absence of drug selection, and the reduced growth efficiency could be attributed to the formation of H3K9me3-facultative

heterochromatin at the *clag3.2* promoter that spreads and silences the drug resistance gene. Therefore, the reduced growth rate could be due to a loss of drug resistance as opposed to loss of *clag3* function in this parasite line. Undoubtedly, this *clag3* null parasite line will serve as a valuable tool in further CLAG3, RhopH complex and PSAC analyses, however, *clag3* expression levels must be carefully monitored by qPCR, considering the possibility of spontaneous *clag3.1* activation and the aforementioned possibility of the repression of drug resistance at the disrupted *clag3.2* locus.

The expression of the endogenous *clag3.1* promoter and the concurrent expression of the episomal *clag3.2* promoter in the 2B6 parasite line, a 10Gluc-derived subclone, indicates that both *clag3* promoters can be active simultaneously and are thus not strictly mutually exclusive. This supports other data obtained in our lab with a transgenic parasite that can express both *clag3.1* and *clag3.2* when induced by drug pressure (Rovira-Graells, unpublished data). This simultaneous express of the episomal *clag3.2* and the endogenous *clag3.1* is unlike the situation observed with the *var* gene family, where the episomal expression of a *var* promoter silences all endogenous family members (Voss et al 2006; Dzikowski et al 2006).

var genes are found on different chromosomes and in both subtelomeric and chromosome internal regions, hence, the mechanism that permits for the expression of only one member is required to work in *trans*. Therefore, perhaps the introduction of another *var* promoter in *trans* (on an episome) is able to be recognised by the counting mechanism because the artificial situation is sufficiently similar to what occurs endogenously.

Since the *clag3* genes are in a tandem arrangement on the same chromosome and the episome did not result in the silencing of the endogenously expressed *clag3*, could suggest that the *clag3* counting mechanism solely functions in *cis*, and is not capable of silencing a *clag3* in *trans*, as it is a situation that does not occur naturally. This could suggest that in addition to or as part of the counting mechanism, local chromosome conformation also plays a role in maintaining the mutually exclusive expression of the *clag3* genes as observed for the *var* family of genes.

We can also conclude that the active transcription of *clag3.1,* does not prevent transcription of the episomal *clag3.2* promoter, nor does the CLAG3.1 protein prevent *clag3.2* promoter activity, conclusions that we could infer from the transient transfections of plasmids containing 5' flanking regions in 10G and 1.2B parasites, but could not confirm.

Correlation is not causation, but the consistent association of the ncRNA PF03TR002 with *clag3.1* expression and *clag3.2* repression, even in parasites that have switched which *clag*3 is expressed, suggests a potential role for this ncRNA in their mutually exclusive expression. It has been previously shown that in 3D7-B, the repression of *clag3.2* is nearly complete (Cortés et al 2007). It is also the parasite line that has the highest levels of PF03TR002 in schizonts compared to the other parasite lines tested. This ncRNA is also present, although to a much lower degree, outside of the schizont stage in 3D7-B. qPCR analysis also suggests that the PF03TR002 transcript of 3D7-B is of a different composition when compared to this transcript in 1.2B or 3D7-A parasites, either longer or perhaps transcripts of different lengths are produced

from this region, possibilities that would have to be investigated further by Northern blot analyses. And although we did not check the PTMs associated with the PF03TR002 locus, the Duraisingh group looked at one primer pair located within this region in parasite lines that differentially express *clag3*, and there were no differences in enrichment for H3K9me3, H3K9ac, H3K4me3, nor H3K4me2 between the two different *clag3* expression clones. PF03TR002 could function as other ncRNAs do in processes, such as genetic imprinting, by physically interacting with proteins that may either contribute to the repression of *clag3.2* or allow for the expression of *clag3.1*.

The correlation of PF03TR002 with the expression of *clag3.1* could also be the result of their physical proximity, as they could share the same chromatin conformation. Consequently, when *clag3.1* is in a transcriptionally competent confirmation, the chromatin of the PF03TR002 locus could likewise be in a conformation permissive for transcription, resulting in the transcription of both in the presence of their appropriate stage-specific TFs.

Genetic disruption of this ncRNA locus could demonstrate if it is required for the mutual exclusion of the *clag3* genes, but considerations would have to be made in order not to alter the local chromatin conformation of *clag3.1* and *clag3.2* that may lead to erroneous conclusions. It would also be interesting to determine the subnuclear localization of PF03TR002 by RNA-FISH to determine if it remains at the site of transcription, and thus more proximal to *clag3.1*, or if it is also found upstream of *clag3.2*.

We have data from transgenic parasites (with stably maintained and integrated episomes) that indicates that both *clag3.1* and *clag3.2* can be simultaneously expressed when forced by drug pressure to form the same open chromatin architecture. Thus, if the chromatin conformation of both *clag3.1* and *clag3.2* are permissive for transcription, they can both be transcribed and are thus not strictly mutually exclusive. However, in their endogenous setting, the most favourable situation is to express only one of the *clag3* genes. Therefore, a mechanism is in place that restricts the simultaneous transcription of both *clag3* genes, one that does not depend on large or small scale genetic differences between parasite lines. Thus the limiting enhancer or repressor hypothesis is still feasible and awaits testing. It is tempting to speculate that the location from which PF03TR002 is transcribed contains the limiting enhancer or repressor element that works in conjunction with the ncRNA to regulate the mutually exclusive expression of *clag3.1* and *clag3.2*. For the moment this remains purely speculative, nevertheless, now that some of the critical steps have been optimised, it is an interesting hypothesis that can be explored by 3C and 4C experiments in the future.

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

8. Conclusions

The key conclusions of this work can be summarised as follows:

-the 5' UTRs of *clag3.1*, *clag3.2* and *eba-140* contain sufficient information to direct their stage-specific expression

-the clonally variant expression pattern of *clag*3.1, *clag3.2* and *eba-140* is not controlled by the presence or absence of *trans*-acting factors

-*clag3.1*, clag*3.2* and *eba-140* are enriched in H3K9ac when transcriptionally active and when bookmarked for activation, suggesting that H3K9ac is an epigenetic signal for activation

-*clag3.1*, clag*3.2* and *eba-140* are enriched in H3K9me3 when transcriptionally repressed and when bookmarked for repression, suggesting that H3K9me3 is an epigenetic signal for repression

-*eba-140* shows the histone modifications associated with two independent chromatin domains, suggesting that a barrier insulator may be located within its first exon

-the clonally variant expression status of *clag*3.1, *clag3.2* and *eba-140* is controlled by the formation of H3K9me-facultative heterochromatin

-CVGE is a widespread phenomenon in *P. falciparum* that regulates many gene families that participate in numerous host-parasite interactions apart from immune evasion

8. Conclusions

-correlation of clonally variant genes with heterochromatic marks suggests that phenotypic variation is mediated by the formation of H3K9me3-facultative heterochromatin

-the expression of the ncRNA PF03TR002 consistently correlates with the expression of *clag3.1* and the repression of *clag3.2*

1. Introduction
2. Objectives
3. Materials and Methods
4. Article 1
5. Article 2
6. Mutually exclusive gene expression in *P. falciparum*
7. General Discussion
8. Conclusions
9. Bibliography

9. Bibliography

Agalioti T, Chen G, Thanos D. 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 111:381-92.

Amulic B, Salanti A, Lavstsen T, Nielsen MA, Deitsch KW. 2009. An upstream open reading frame controls translation of var2csa, a gene implicated in placental malaria. *PLoS Pathog* 5:e1000256.

Annunziato AT. 2005. Split decision: what happens to nucleosomes during DNA replication? *J Biol Chem* 280:12065-8.

Aravind L, Iyer LM, Wellems TE, Miller LH. 2003. Plasmodium biology: genomic gleanings. *Cell* 115:771-85.

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

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* 21:381-95.

Bannister L, Mitchell G. 2003. The ins, outs and roundabouts of malaria. *Trends Parasitol* 19:209-13.

Bartfai R, Hoeijmakers WA, Salcedo-Amaya AM, Smits AH, Janssen-Megens E, Kaan A, Treeck M, et al. 2010. H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3. *PLoS Pathog* 6:e1001223.

Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, Feldman M, Taraschi TF, Howard RJ. 1995. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* 82:77-87.

Baum J, Maier AG, Good RT, Simpson KM, Cowman AF. 2005. Invasion by P. falciparum merozoites suggests a hierarchy of molecular interactions. *PLoS Pathog* 1:e37.

Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, Waters AP, Cowman AF, Crabb BS, de Koning-Ward TF. 2009. Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res* 37:3788-98.

Becker PB, Horz W. 2002. ATP-dependent nucleosome remodeling. *Annu Rev Biochem* 71:247-73.

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405:482-5.

Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98:387-96.

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. 2009. An operational definition of epigenetics. *Genes Dev* 23:781-3.

Bird AP, Wolffe AP. 1999. Methylation-induced repression--belts, braces, and chromatin. *Cell* 99:451-4.

Bischoff E, Vaquero C. 2010. In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum. *BMC Genomics* 11:34.

Bonasio R, Tu S, Reinberg D. 2010. Molecular signals of epigenetic states. *Science* 330:612-6.

Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol* 1:E5.

Briquet S, Boschet C, Gissot M, Tissandie E, Sevilla E, Franetich JF, Thiery I, et al. 2006. High-mobility-group box nuclear factors of Plasmodium falciparum. *Eukaryot Cell* 5:672-82.

Broadbent K, Park D, Wolf AR, Van Tyne D, Sims JS, Ribacke U, Volkman S, et al. 2011. A global transcriptional analysis of Plasmodium falciparum malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol* 12:R56.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71:515-26.

Bull PC, Lowe BS, Kortok M, Molyneux CS, Newbold CI, Marsh K. 1998. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med* 4:358-60.

Calderwood MS, Gannoun-Zaki L, Wellems TE, Deitsch KW. 2003. Plasmodium falciparum var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J Biol Chem* 278:34125-32.

Callaway E. 2007. Malaria research should go 'back to basics'. *Nature* 449:266.

Callebaut I, Prat K, Meurice E, Mornon JP, Tomavo S. 2005. Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes. *BMC Genomics* 6:100.

Cam HP, Sugiyama T, Chen ES, Chen X, FitzGerald PC, Grewal SI. 2005. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat Genet* 37:809-19.

Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinas M. 2010. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog* 6:e1001165.

Campos EI, Reinberg D. 2009. Histones: annotating chromatin. *Annu Rev Genet* 43:559-99.

Carroll J, Rossi EA, Porter J, Neitz J, Roorda A, Williams DR, Neitz M. 2010. Deletion of the X-linked opsin gene array locus control region (LCR) results in disruption of the cone mosaic. *Vision Res* 50:1989-99.

Chaal BK, Gupta AP, Wastuwidyaningtyas BD, Luah YH, Bozdech Z. 2010. Histone deacetylases play a major role in the transcriptional regulation of the Plasmodium falciparum life cycle. *PLoS Pathog* 6:e1000737.

Chaiyaroj SC, Coppel RL, Magowan C, Brown GV. 1994. A Plasmodium falciparum isolate with a chromosome 9 deletion expresses a trypsin-resistant cytoadherence molecule. *Mol Biochem Parasitol* 67:21-30.

Choi SW, Keyes MK, Horrocks P. 2006. LC/ESI-MS demonstrates the absence of 5-methyl-2'-deoxycytosine in Plasmodium falciparum genomic DNA. *Mol Biochem Parasitol* 150:350-2.

Chookajorn T, Dzikowski R, Frank M, Li F, Jiwani AZ, Hartl DL, Deitsch KW. 2007. Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci U S A* 104:899-902.

Comeaux CA, Coleman BI, Bei AK, Whitehurst N, Duraisingh MT. 2011. Functional analysis of epigenetic regulation of tandem RhopH1/clag genes reveals a role in Plasmodium falciparum growth. *Mol Microbiol* 80:378-90.

Cooper JA, Ingram LT, Bushell GR, Fardoulys CA, Stenzel D, Schofield L, Saul AJ. 1988. The 140/130/105 kilodalton protein complex in the rhoptries of Plasmodium falciparum consists of discrete polypeptides. *Mol Biochem Parasitol* 29:251-60.

Cortés A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA. 2007. Epigenetic silencing of Plasmodium falciparum genes linked to erythrocyte invasion. *PLoS Pathog* 3:e107.

Coulson RM, Hall N, Ouzounis CA. 2004. Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. *Genome Res* 14:1548-54.

Crompton PD, Pierce SK, Miller LH. 2010. Advances and challenges in malaria vaccine development. *J Clin Invest* 120:4168-78.

Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 19:24-32.

Cui L, Fan Q, Miao J. 2008a. Histone lysine methyltransferases and demethylases in Plasmodium falciparum. *Int J Parasitol* 38:1083-97.

Cui L, Miao J, Furuya T, Fan Q, Li X, Rathod PK, Su XZ. 2008b. Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro Plasmodium falciparum development. *Eukaryot Cell* 7:1200-10.

Czermin B, Schotta G, Hulsmann BB, Brehm A, Becker PB, Reuter G, Imhof A. 2001. Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila. *EMBO Rep* 2:915-9.

Darkin-Rattray SJ, Gurnett AM, Myers RW, Dulski PM, Crumley TM, Allocco JJ, Cannova C, et al. 1996. Apicidin: a novel antiprotozoal agent that inhibits parasite histone deacetylase. *Proc Natl Acad Sci U S A* 93:13143-7.

Day KP, Karamalis F, Thompson J, Barnes DA, Peterson C, Brown H, Brown GV, Kemp DJ. 1993. Genes necessary for expression of a virulence determinant and for transmission of Plasmodium falciparum are located on a 0.3-megabase region of chromosome 9. *Proc Natl Acad Sci U S A* 90:8292-6.

De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, Bulyk ML, Llinas M. 2008. Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci U S A* 105:8393-8.

Deeb SS. 2006. Genetics of variation in human color vision and the retinal cone mosaic. *Curr Opin Genet Dev* 16:301-7.

Deeb SS, Bisset D, Fu L. 2010. Epigenetic control of expression of the human L- and M- pigment genes. *Ophthalmic Physiol Opt* 30:446-53.

Deitsch KW. 2009. A mark of silence in malaria parasites. *Cell Host Microbe* 5:112-3.

Deitsch KW, Calderwood MS, Wellems TE. 2001. Malaria. Cooperative silencing elements in var genes. *Nature* 412:875-6.

Deitsch KW, del Pinal A, Wellems TE. 1999. Intra-cluster recombination and var transcription switches in the antigenic variation of Plasmodium falciparum. *Mol Biochem Parasitol* 101:107-16.

Dekker J. 2006. The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods* 3:17-21.

Dodd IB, Micheelsen MA, Sneppen K, Thon G. 2007. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* 129:813-22.

Donze D, Kamakaka RT. 2001. RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. *EMBO J* 20:520-31.

Duraisingh MT, Triglia T, Ralph SA, Rayner JC, Barnwell JW, McFadden GI, Cowman AF. 2003. Phenotypic variation of Plasmodium falciparum merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO J* 22:1047-57.

Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, et al. 2005. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. *Cell* 121:13-24.

Dzikowski R, Deitsch K. 2006. Antigenic variation by protozoan parasites: insights from Babesia bovis. *Mol Microbiol* 59:364-6.

Dzikowski R, Deitsch KW. 2008. Active transcription is required for maintenance of epigenetic memory in the malaria parasite Plasmodium falciparum. *J Mol Biol* 382:288-97.

Dzikowski R, Frank M, Deitsch K. 2006. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathog* 2:e22.

Dzikowski R, Li F, Amulic B, Eisberg A, Frank M, Patel S, Wellems TE, Deitsch KW. 2007. Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO Rep* 8:959-65.

El Sahly HM, Patel SM, Atmar RL, Lanford TA, Dube T, Thompson D, Sim BK, Long C, Keitel WA. 2010. Safety and immunogenicity of a recombinant nonglycosylated erythrocyte binding antigen 175 Region II malaria vaccine in healthy adults living in an area where malaria is not endemic. *Clin Vaccine Immunol* 17:1552-9.

Epp C, Li F, Howitt CA, Chookajorn T, Deitsch KW. 2009. Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite Plasmodium falciparum. *RNA* 15:116-27.

Fan Q, An L, Cui L. 2004. PfADA2, a Plasmodium falciparum homologue of the transcriptional coactivator ADA2 and its in vivo association with the histone acetyltransferase PfGCN5. *Gene* 336:251-61.

Figueiredo LM, Freitas-Junior LH, Bottius E, Olivo-Marin JC, Scherf A. 2002. A central role for Plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J* 21:815-24.

Figueiredo LM, Janzen CJ, Cross GA. 2008. A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS Biol* 6:e161.

Fletcher TM, Hansen JC. 1996. The nucleosomal array: structure/function relationships. *Crit Rev Eukaryot Gene Expr* 6:149-88.

Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BT, Moes S, Bozdech Z, et al. 2010. A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog* 6:e1000784.

Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, Ehlgen F, et al. 2009. Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 5:e1000569.

Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, Bozdech Z. 2011. Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics*

Frank M, Deitsch K. 2006. Activation, silencing and mutually exclusive expression within the var gene family of Plasmodium falciparum. *Int J Parasitol* 36:975-85.

Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellems TE, Scherf A. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature* 407:1018-22.

Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP, Mancio-Silva L, et al. 2005. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 121:25-36.

Fried M, Duffy PE. 1996. Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta. *Science* 272:1502-4.

Fuss SH, Omura M, Mombaerts P. 2007. Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell* 130:373-84.

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

Gannoun-Zaki L, Jost A, Mu J, Deitsch KW, Wellems TE. 2005. A silenced Plasmodium falciparum var promoter can be activated in vivo through spontaneous deletion of a silencing element in the intron. *Eukaryot Cell* 4:490-2.

Gardiner DL, Spielmann T, Dixon MW, Hawthorne PL, Ortega MR, Anderson KL, Skinner-Adams TS, Kemp DJ, Trenholme KR. 2004. CLAG 9 is located in the rhoptries of Plasmodium falciparum. *Parasitol Res* 93:64-7.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, et al. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 419:498-511.

Gaszner M, Felsenfeld G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 7:703-13.

Gaur D, Mayer DC, Miller LH. 2004. Parasite ligand-host receptor interactions during invasion of erythrocytes by Plasmodium merozoites. *Int J Parasitol* 34:1413-29.

Gerald N, Mahajan B, Kumar S. 2011. Mitosis in the human malaria parasite Plasmodium falciparum. *Eukaryot Cell* 10:474-82.

Giacani L, Lukehart S, Centurion-Lara A. 2007. Length of guanosine homopolymeric repeats modulates promoter activity of subfamily II tpr genes of Treponema pallidum ssp. pallidum. *FEMS Immunol Med Microbiol* 51:289-301.

Gissot M, Briquet S, Refour P, Boschet C, Vaquero C. 2005. PfMyb1, a Plasmodium falciparum transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *J Mol Biol* 346:29-42.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74:481-514.

Gomez-Escobar N, Amambua-Ngwa A, Walther M, Okebe J, Ebonyi A, Conway DJ. 2010. Erythrocyte invasion and merozoite ligand gene expression in severe and mild Plasmodium falciparum malaria. *J Infect Dis* 201:444-52.

Gondor A, Ohlsson R. 2008. Chromatin insulators and cohesins. *EMBO Rep* 9:327-9.

Grewal SI, Elgin SC. 2002. Heterochromatin: new possibilities for the inheritance of structure. *Curr Opin Genet Dev* 12:178-87.

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

Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2:1722-33.

Hampsey M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* 62:465-503.

Hattman S. 2005. DNA-[adenine] methylation in lower eukaryotes. *Biochemistry (Mosc)* 70:550-8.

Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, Kabaria CW, et al. 2009. A world malaria map: Plasmodium falciparum endemicity in 2007. *PLoS Med* 6:e1000048.

Helm S, Lehmann C, Nagel A, Stanway RR, Horstmann S, Llinas M, Heussler VT. 2010. Identification and characterization of a liver stage-specific promoter region of the malaria parasite Plasmodium. *PLoS One* 5:e13653.

Hervouet E, Vallette FM, Cartron PF. 2009. Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics* 4:487-99.

Hill DA, Desai SA. 2010. Malaria parasite mutants with altered erythrocyte permeability: a new drug resistance mechanism and important molecular tool. *Future Microbiol* 5:81-97.

Hiller NL, Akompong T, Morrow JS, Holder AA, Haldar K. 2003. Identification of a stomatin orthologue in vacuoles induced in human erythrocytes by malaria parasites. A role for microbial raft proteins in apicomplexan vacuole biogenesis. *J Biol Chem* 278:48413-21.

Holder AA, Freeman RR, Uni S, Aikawa M. 1985. Isolation of a Plasmodium falciparum rhoptry protein. *Mol Biochem Parasitol* 14:293-303.

Holt DC, Gardiner DL, Thomas EA, Mayo M, Bourke PF, Sutherland CJ, Carter R, et al. 1999. The cytoadherence linked asexual gene family of Plasmodium falciparum: are there roles other than cytoadherence? *Int J Parasitol* 29:939-44.

Horn D, McCulloch R. 2010. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol* 13:700-5.

Horrocks P, Dechering K, Lanzer M. 1998. Control of gene expression in Plasmodium falciparum. *Mol Biochem Parasitol* 95:171-81.

Horrocks P, Wong E, Russell K, Emes RD. 2009. Control of gene expression in Plasmodium falciparum - ten years on. *Mol Biochem Parasitol* 164:9-25.

Hu G, Cabrera A, Kono M, Mok S, Chaal BK, Haase S, Engelberg K, et al. 2010. Transcriptional profiling of growth perturbations of the human malaria parasite Plasmodium falciparum. *Nat Biotechnol* 28:91-8.

Huber SM, Duranton C, Lang F. 2005. Patch-clamp analysis of the "new permeability pathways" in malaria-infected erythrocytes. *Int Rev Cytol* 246:59-134.

Hughes K, Wand M, Foulston L, Young R, Harley K, Terry S, Ersfeld K, Rudenko G. 2007. A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. *EMBO J* 26:2400-10.

Iriko H, Kaneko O, Otsuki H, Tsuboi T, Su XZ, Tanabe K, Torii M. 2008. Diversity and evolution of the rhoph1/clag multigene family of Plasmodium falciparum. *Mol Biochem Parasitol* 158:11-21.

Ishii K, Arib G, Lin C, Van Houwe G, Laemmli UK. 2002. Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell* 109:551-62.

Issar N, Ralph SA, Mancio-Silva L, Keeling C, Scherf A. 2009. Differential sub-nuclear localisation of repressive and activating histone methyl modifications in P. falciparum. *Microbes Infect* 11:403-7.

Iyer LM, Anantharaman V, Wolf MY, Aravind L. 2008. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 38:1-31.

Jackson V, Chalkley R. 1985. Histone segregation on replicating chromatin. *Biochemistry* 24:6930-8.

Jacobson RH, Ladurner AG, King DS, Tjian R. 2000. Structure and function of a human TAFII250 double bromodomain module. *Science* 288:1422-5.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* 293:1074-80.

Jiang L, Lopez-Barragan MJ, Jiang H, Mu J, Gaur D, Zhao K, Felsenfeld G, Miller LH. 2010. Epigenetic control of the variable expression of a Plasmodium falciparum receptor protein for erythrocyte invasion. *Proc Natl Acad Sci U S A* 107:2224-9.

Joergensen L, Bengtsson DC, Bengtsson A, Ronander E, Berger SS, Turner L, Dalgaard MB, et al. 2010. Surface co-expression of two different PfEMP1 antigens on single plasmodium falciparum-infected erythrocytes facilitates binding to ICAM1 and PECAM1. *PLoS Pathog* 6:e1001083.

Joshi MB, Lin DT, Chiang PH, Goldman ND, Fujioka H, Aikawa M, Syin C. 1999. Molecular cloning and nuclear localization of a histone deacetylase homologue in Plasmodium falciparum. *Mol Biochem Parasitol* 99:11-9.

Juillerat A, Lewit-Bentley A, Guillotte M, Gangnard S, Hessel A, Baron B, Vigan-Womas I, et al. 2011. Structure of a Plasmodium falciparum PfEMP1 rosetting domain reveals a role for the N-terminal segment in heparin-mediated rosette inhibition. *Proc Natl Acad Sci U S A* 108:5243-8.

Kaneko O. 2007. Erythrocyte invasion: vocabulary and grammar of the Plasmodium rhoptry. *Parasitol Int* 56:255-62.

Kaneko O, Tsuboi T, Ling IT, Howell S, Shirano M, Tachibana M, Cao YM, Holder AA, Torii M. 2001. The high molecular mass rhoptry protein, RhopH1, is encoded by members of the clag multigene family in Plasmodium falciparum and Plasmodium yoelii. *Mol Biochem Parasitol* 118:223-31.

Kaneko O, Yim Lim BY, Iriko H, Ling IT, Otsuki H, Grainger M, Tsuboi T, et al. 2005. Apical expression of three RhopH1/Clag proteins as components of the Plasmodium falciparum RhopH complex. *Mol Biochem Parasitol* 143:20-8.

Klose RJ, Yamane K, Bae Y, Zhang D, Erdjument-Bromage H, Tempst P, Wong J, Zhang Y. 2006. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* 442:312-6.

Koerner MV, Pauler FM, Huang R, Barlow DP. 2009. The function of non-coding RNAs in genomic imprinting. *Development* 136:1771-83.

Kornberg RD. 1977. Structure of chromatin. *Annu Rev Biochem* 46:931-54.

Kornberg RD, Lorch Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98:285-94.

Kouzarides T. 2002. Histone methylation in transcriptional control. *Curr Opin Genet Dev* 12:198-209.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* 128:693-705.

Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, Smith LM, et al. 2007. Patterns of gene recombination shape var gene repertoires in Plasmodium falciparum: comparisons of geographically diverse isolates. *BMC Genomics* 8:45.

Kraemer SM, Smith JD. 2006. A family affair: var genes, PfEMP1 binding, and malaria disease. *Curr Opin Microbiol* 9:374-80.

Kulakova L, Singer SM, Conrad J, Nash TE. 2006. Epigenetic mechanisms are involved in the control of Giardia lamblia antigenic variation. *Mol Microbiol* 61:1533-42.

Kumar K, Singal A, Rizvi MM, Chauhan VS. 2008. High mobility group box (HMGB) proteins of Plasmodium falciparum: DNA binding proteins with pro-inflammatory activity. *Parasitol Int* 57:150-7.

Kyes S, Christodoulou Z, Pinches R, Kriek N, Horrocks P, Newbold C. 2007. Plasmodium falciparum var gene expression is developmentally controlled at the level of RNA polymerase II-mediated transcription initiation. *Mol Microbiol* 63:1237-47.

Kyes SA, Christodoulou Z, Raza A, Horrocks P, Pinches R, Rowe JA, Newbold CI. 2003. A well-conserved Plasmodium falciparum var gene shows an unusual stage-specific transcript pattern. *Mol Microbiol* 48:1339-48.

Kyes SA, Rowe JA, Kriek N, Newbold CI. 1999. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. *Proc Natl Acad Sci U S A* 96:9333-8.

LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, et al. 2005. A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature* 438:103-7.

Lanzer M, de Bruin D, Ravetch JV. 1992. Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates. *EMBO J* 11:1949-55.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-8.

Lavazec C, Sanyal S, Templeton TJ. 2007. Expression switching in the stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. *Mol Microbiol* 64:1621-34.

Lavrov SA, Kibanov MV. 2007. Noncoding RNAs and chromatin structure. *Biochemistry (Mosc)* 72:1422-38.

Le Roch KG, Johnson JR, Ahiboh H, Chung DW, Prudhomme J, Plouffe D, Henson K, et al. 2008. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum. *BMC Genomics* 9:513.

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, et al. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301:1503-8.

Lee JT. 2009. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* 23:1831-42.

Lewcock JW, Reed RR. 2004. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc Natl Acad Sci U S A* 101:1069-74.

Li F, Sonbuchner L, Kyes SA, Epp C, Deitsch KW. 2008. Nuclear non-coding RNAs are transcribed from the centromeres of Plasmodium falciparum and are associated with centromeric chromatin. *J Biol Chem* 283:5692-8.

Li G, Reinberg D. 2011. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* 21:175-86.

Li WB, Bzik DJ, Gu HM, Tanaka M, Fox BA, Inselburg J. 1989. An enlarged largest subunit of Plasmodium falciparum RNA polymerase II defines conserved and variable RNA polymerase domains. *Nucleic Acids Res* 17:9621-36.

Ling IT, Florens L, Dluzewski AR, Kaneko O, Grainger M, Yim Lim BY, Tsuboi T, et al. 2004. The Plasmodium falciparum clag9 gene encodes a rhoptry protein that is transferred to the host erythrocyte upon invasion. *Mol Microbiol* 52:107-18.

Ling IT, Kaneko O, Narum DL, Tsuboi T, Howell S, Taylor HM, Scott-Finnigan TJ, Torii M, Holder AA. 2003. Characterisation of the rhoph2 gene of Plasmodium falciparum and Plasmodium yoelii. *Mol Biochem Parasitol* 127:47-57.

Lobo CA, Rodriguez M, Reid M, Lustigman S. 2003. Glycophorin C is the receptor for the Plasmodium falciparum erythrocyte binding ligand PfEBP-2 (baebl). *Blood* 101:4628-31.

Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. 2006. Interchromosomal interactions and olfactory receptor choice. *Cell* 126:403-13.

Longhurst HJ, Holder AA. 1997. The histones of Plasmodium falciparum: identification, purification and a possible role in the pathology of malaria. *Parasitology* 114 (Pt 5):413-9.

Lopez-Estrano C, Gopalakrishnan AM, Semblat JP, Fergus MR, Mazier D, Haldar K. 2007. An enhancer-like region regulates hrp3 promoter stage-specific gene expression in the human malaria parasite Plasmodium falciparum. *Biochim Biophys Acta* 1769:506-13.

Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, Scherf A. 2007a. 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol* 66:1296-305.

Lopez-Rubio JJ, Mancio-Silva L, Scherf A. 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* 5:179-90.

Lopez-Rubio JJ, Riviere L, Scherf A. 2007b. Shared epigenetic mechanisms control virulence factors in protozoan parasites. *Curr Opin Microbiol* 10:560-8.

Lucchini R, Wellinger RE, Sogo JM. 2001. Nucleosome positioning at the replication fork. *EMBO J* 20:7294-302.

Lustigman S, Anders RF, Brown GV, Coppel RL. 1988. A component of an antigenic rhoptry complex of Plasmodium falciparum is modified after merozoite invasion. *Mol Biochem Parasitol* 30:217-24.

Mackinnon MJ, Marsh K. 2010. The selection landscape of malaria parasites. *Science* 328:866-71.

Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans ZA, et al. 2011. An Epigenetic Signature for Monoallelic Olfactory Receptor Expression. *Cell*

Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM. 2006. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev* 20:1268-82.

Mancio-Silva L, Rojas-Meza AP, Vargas M, Scherf A, Hernandez-Rivas R. 2008. Differential association of Orc1 and Sir2 proteins to telomeric domains in Plasmodium falciparum. *J Cell Sci* 121:2046-53.

Marcello L, Barry JD. 2007. Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res* 17:1344-52.

Margueron R, Reinberg D. 2010. Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* 11:285-96.

Maston GA, Evans SK, Green MR. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 7:29-59.

Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJ. 2007. Targets of RNA-directed DNA methylation. *Curr Opin Plant Biol* 10:512-9.

McAndrew MB, Read M, Sims PF, Hyde JE. 1993. Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T-rich human malaria parasite Plasmodium falciparum. *Gene* 124:165-71.

McMahon C, Carroll J, Awua S, Neitz J, Neitz M. 2008. The L:M cone ratio in males of African descent with normal color vision. *J Vis* 8:5 1-9.

McMahon C, Neitz J, Neitz M. 2004. Evaluating the human X-chromosome pigment gene promoter sequences as predictors of L:M cone ratio variation. *J Vis* 4:203-8.

Merrick CJ, Duraisingh MT. 2007. Plasmodium falciparum Sir2: an unusual sirtuin with dual histone deacetylase and ADP-ribosyltransferase activity. *Eukaryot Cell* 6:2081-91.

Miao J, Fan Q, Cui L, Li J. 2006. The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation. *Gene* 369:53-65.

Militello KT, Dodge M, Bethke L, Wirth DF. 2004. Identification of regulatory elements in the Plasmodium falciparum genome. *Mol Biochem Parasitol* 134:75-88.

Miller LH, Baruch DI, Marsh K, Doumbo OK. 2002. The pathogenic basis of malaria. *Nature* 415:673-9.

Miller LH, Good MF, Milon G. 1994. Malaria pathogenesis. *Science* 264:1878-83.

Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C. 2010. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* 137:2493-9.

Molineaux L, Diebner HH, Eichner M, Collins WE, Jeffery GM, Dietz K. 2001. Plasmodium falciparum parasitaemia described by a new mathematical model. *Parasitology* 122:379-91.

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322:1717-20.

Natalang O, Bischoff E, Deplaine G, Proux C, Dillies MA, Sismeiro O, Guigon G, et al. 2008. Dynamic RNA profiling in Plasmodium falciparum synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics* 9:388.

Nelson HC, Finch JT, Luisi BF, Klug A. 1987. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* 330:221-6.

Nery S, Deans AM, Mosobo M, Marsh K, Rowe JA, Conway DJ. 2006. Expression of Plasmodium falciparum genes involved in erythrocyte invasion varies among isolates cultured directly from patients. *Mol Biochem Parasitol* 149:208-15.

Nguitragool W, Bokhari AA, Pillai AD, Rayavara K, Sharma P, Turpin B, Aravind L, Desai SA. 2011. Malaria Parasite clag3 Genes Determine Channel-Mediated Nutrient Uptake by Infected Red Blood Cells. *Cell* 145:665-77.

Nguyen MQ, Zhou Z, Marks CA, Ryba NJ, Belluscio L. 2007. Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell* 131:1009-17.

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

Nishizumi H, Kumasaka K, Inoue N, Nakashima A, Sakano H. 2007. Deletion of the core-H region in mice abolishes the expression of three proximal odorant receptor genes in cis. *Proc Natl Acad Sci U S A* 104:20067-72.

O'Donnell RA, Freitas-Junior LH, Preiser PR, Williamson DH, Duraisingh M, McElwain TF, Scherf A, Cowman AF, Crabb BS. 2002. A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of Plasmodium falciparum chromosomes. *EMBO J* 21:1231-9.

Oakley MS, Kumar S, Anantharaman V, Zheng H, Mahajan B, Haynes JD, Moch JK, et al. 2007. Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic Plasmodium falciparum parasites. *Infect Immun* 75:2012-25.

Oki M, Valenzuela L, Chiba T, Ito T, Kamakaka RT. 2004. Barrier proteins remodel and modify chromatin to restrict silenced domains. *Mol Cell Biol* 24:1956-67.

Olivieri A, Silvestrini F, Sanchez M, Alano P. 2008. A 140-bp AT-rich sequence mediates positive and negative transcriptional control of a Plasmodium falciparum developmentally regulated promoter. *Int J Parasitol* 38:299-312.

Osta M, Gannoun-Zaki L, Bonnefoy S, Roy C, Vial HJ. 2002. A 24 bp cis-acting element essential for the transcriptional activity of Plasmodium falciparum CDP-diacylglycerol synthase gene promoter. *Mol Biochem Parasitol* 121:87-98.

Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, Lemieux J, et al. 2010. New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol Microbiol* 76:12-24.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32:232-46.

Perez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernandez-Cuevas NA, Delgadillo DM, Vargas M, Martinez-Calvillo S, Scherf A, Hernandez-Rivas R. 2009. Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res* 37:2596-606.

Persson KE, McCallum FJ, Reiling L, Lister NA, Stubbs J, Cowman AF, Marsh K, Beeson JG. 2008. Variation in use of erythrocyte invasion pathways by Plasmodium falciparum mediates evasion of human inhibitory antibodies. *J Clin Invest* 118:342-51.

Petter M, Lee CC, Byrne TJ, Boysen KE, Volz J, Ralph SA, Cowman AF, Brown GV, Duffy MF. 2011. Expression of P. falciparum var genes involves exchange of the histone variant H2A.Z at the promoter. *PLoS Pathog* 7:e1001292.

Pollack Y, Kogan N, Golenser J. 1991. Plasmodium falciparum: evidence for a DNA methylation pattern. *Exp Parasitol* 72:339-44.

Polson HE, Blackman MJ. 2005. A role for poly(dA)poly(dT) tracts in directing activity of the Plasmodium falciparum calmodulin gene promoter. *Mol Biochem Parasitol* 141:179-89.

Ponger L, Li WH. 2005. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* 22:1119-28.

Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG. 2010. Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res* 20:228-38.

Pospelov V, Russev G, Vassilev L, Tsanev R. 1982. Nucleosome segregation in chromatin replicated in the presence of cycloheximide. *J Mol Biol* 156:79-91.

Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, Saura A, Carranza PG, Lujan HD. 2008. Antigenic variation in Giardia lamblia is regulated by RNA interference. *Nature* 456:750-4.

Ralph SA, Bischoff E, Mattei D, Sismeiro O, Dillies MA, Guigon G, Coppee JY, David PH, Scherf A. 2005. Transcriptome analysis of antigenic variation in Plasmodium falciparum--var silencing is not dependent on antisense RNA. *Genome Biol* 6:R93.

Ramos E, Ghosh D, Baxter E, Corces VG. 2006. Genomic organization of gypsy chromatin insulators in Drosophila melanogaster. *Genetics* 172:2337-49.

Recker M, Buckee CO, Serazin A, Kyes S, Pinches R, Christodoulou Z, Springer AL, Gupta S, Newbold CI. 2011. Antigenic variation in Plasmodium falciparum malaria involves a highly structured switching pattern. *PLoS Pathog* 7:e1001306.

Richards JS, Stanisic DI, Fowkes FJ, Tavul L, Dabod E, Thompson JK, Kumar S, et al. 2010. Association between naturally acquired antibodies to erythrocyte-binding antigens of Plasmodium falciparum and protection from malaria and high-density parasitemia. *Clin Infect Dis* 51:e50-60.

Rosario V. 1981. Cloning of naturally occurring mixed infections of malaria parasites. *Science* 212:1037-8.

Rothman A, Feinstein P, Hirota J, Mombaerts P. 2005. The promoter of the mouse odorant receptor gene M71. *Mol Cell Neurosci* 28:535-46.

Rowe A, Obeiro J, Newbold CI, Marsh K. 1995. Plasmodium falciparum rosetting is associated with malaria severity in Kenya. *Infect Immun* 63:2323-6.

Rowe JA, Moulds JM, Newbold CI, Miller LH. 1997. P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* 388:292-5.

Ruvalcaba-Salazar OK, del Carmen Ramirez-Estudillo M, Montiel-Condado D, Recillas-Targa F, Vargas M, Hernandez-Rivas R. 2005. Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions. *Mol Biochem Parasitol* 140:183-96.

Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AM, Cohen AM, Janssen-Megens EM, et al. 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum. *Proc Natl Acad Sci U S A* 106:9655-60.

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J, Kouzarides T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* 419:407-11.

Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, Gysin J, Lanzer M. 1998. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum. *EMBO J* 17:5418-26.

Scherf A, Lopez-Rubio JJ, Riviere L. 2008. Antigenic variation in Plasmodium falciparum. *Annu Rev Microbiol* 62:445-70.

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G. 2002. Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J* 21:1121-31.

Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, van Lohuizen M, Tanay A, Cavalli G. 2009. Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol* 7:e13.

Serizawa S, Miyamichi K, Nakatani H, Suzuki M, Saito M, Yoshihara Y, Sakano H. 2003. Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* 302:2088-94.

Serizawa S, Miyamichi K, Sakano H. 2004. One neuron-one receptor rule in the mouse olfactory system. *Trends Genet* 20:648-53.

Shykind BM, Rohani SC, O'Donnell S, Nemes A, Mendelsohn M, Sun Y, Axel R, Barnea G. 2004. Gene switching and the stability of odorant receptor gene choice. *Cell* 117:801-15.

Simon JA, Kingston RE. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* 10:697-708.

Simonis M, Kooren J, de Laat W. 2007. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4:895-901.

Smallwood PM, Wang Y, Nathans J. 2002. Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *Proc Natl Acad Sci U S A* 99:1008-11.

Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. 2000. Classification of adhesive domains in the Plasmodium falciparum erythrocyte membrane protein 1 family. *Mol Biochem Parasitol* 110:293-310.

Splinter E, Grosveld F, de Laat W. 2004. 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol* 375:493-507.

Sproul D, Gilbert N, Bickmore WA. 2005. The role of chromatin structure in regulating the expression of clustered genes. *Nat Rev Genet* 6:775-81.

Struhl K. 1996. Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell* 84:179-82.

Stubbs J, Simpson KM, Triglia T, Plouffe D, Tonkin CJ, Duraisingh MT, Maier AG, Winzeler EA, Cowman AF. 2005. Molecular mechanism for switching of P. falciparum invasion pathways into human erythrocytes. *Science* 309:1384-7.

Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, Ravetch JA, Wellems TE. 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. *Cell* 82:89-100.

Sullivan WJ, Jr., Naguleswaran A, Angel SO. 2006. Histones and histone modifications in protozoan parasites. *Cell Microbiol* 8:1850-61.

Sunil S, Chauhan VS, Malhotra P. 2008. Distinct and stage specific nuclear factors regulate the expression of falcipains, Plasmodium falciparum cysteine proteases. *BMC Mol Biol* 9:47.

Swamy L, Amulic B, Deitsch KW. 2011. Plasmodium falciparum var gene silencing is determined by cis DNA elements that form stable and heritable interactions. *Eukaryot Cell* 10:530-9.

Talbert PB, Henikoff S. 2010. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* 11:264-75.

Tamez PA, Bhattacharjee S, van Ooij C, Hiller NL, Llinas M, Balu B, Adams JH, Haldar K. 2008. An erythrocyte vesicle protein exported by the malaria parasite promotes tubovesicular lipid import from the host cell surface. *PLoS Pathog* 4:e1000118.

Taylor HM, Triglia T, Thompson J, Sajid M, Fowler R, Wickham ME, Cowman AF, Holder AA. 2001. Plasmodium falciparum homologue of the genes for Plasmodium vivax and Plasmodium yoelii adhesive proteins, which is transcribed but not translated. *Infect Immun* 69:3635-45.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 10:1453-65.

Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M, Duffy MF, et al. 2009. Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum. *PLoS Biol* 7:e84.

Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG, Jensen ON. 2009. Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum. *J Proteome Res* 8:3439-50.

Trenholme KR, Gardiner DL, Holt DC, Thomas EA, Cowman AF, Kemp DJ. 2000. clag9: A cytoadherence gene in Plasmodium falciparum essential for binding of parasitized erythrocytes to CD36. *Proc Natl Acad Sci U S A* 97:4029-33.

Trojer P, Reinberg D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* 28:1-13.

Vaquero A. 2009. The conserved role of sirtuins in chromatin regulation. *Int J Dev Biol* 53:303-22.

Vignali M, Armour CD, Chen J, Morrison R, Castle JC, Biery MC, Bouzek H, et al. 2011. NSR-seq transcriptional profiling enables identification of a gene signature of Plasmodium falciparum parasites infecting children. *J Clin Invest*

Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, Tonkin CJ, Langer C, Crabb BS, Cowman AF. 2010. Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in Plasmodium falciparum. *Int J Parasitol* 40:109-21.

Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, Beeson JG, Reeder JC, Crabb BS, Cowman AF. 2006. A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. *Nature* 439:1004-8.

Voss TS, Kaestli M, Vogel D, Bopp S, Beck HP. 2003. Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters. *Mol Microbiol* 48:1593-607.

Watanabe J, Sasaki M, Suzuki Y, Sugano S. 2002. Analysis of transcriptomes of human malaria parasite Plasmodium falciparum using full-length enriched library: identification of novel genes and diverse transcription start sites of messenger RNAs. *Gene* 291:105-13.

West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G. 2004. Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol Cell* 16:453-63.

WHO. 2010. World Malaria Report 2010. *World Health Organization*: 238p.

Wickstead B, Ersfeld K, Gull K. 2003. Repetitive elements in genomes of parasitic protozoa. *Microbiol Mol Biol Rev* 67:360-75.

Wigler M, Levy D, Perucho M. 1981. The somatic replication of DNA methylation. *Cell* 24:33-40.

Williamson DH, Janse CJ, Moore PW, Waters AP, Preiser PR. 2002. Topology and replication of a nuclear episomal plasmid in the rodent malaria Plasmodium berghei. *Nucleic Acids Res* 30:726-31.

Winter G, Chen Q, Flick K, Kremsner P, Fernandez V, Wahlgren M. 2003. The 3D7var5.2 (var COMMON) type var gene family is commonly expressed in non-placental Plasmodium falciparum malaria. *Mol Biochem Parasitol* 127:179-91.

Wolffe AP, Pruss D. 1996. Targeting chromatin disruption: Transcription regulators that acetylate histones. *Cell* 84:817-9.

Wong EH, Hasenkamp S, Horrocks P. 2011. Analysis of the molecular mechanisms governing the stage-specific expression of a prototypical housekeeping gene during intraerythrocytic development of P. falciparum. *J Mol Biol* 408:205-21.

Yamamoto K, Sonoda M. 2003. Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. *Biochem Biophys Res Commun* 301:287-92.

Young JA, Johnson JR, Benner C, Yan SF, Chen K, Le Roch KG, Zhou Y, Winzeler EA. 2008. In silico discovery of transcription regulatory elements in Plasmodium falciparum. *BMC Genomics* 9:70.

Yuda M, Iwanaga S, Shigenobu S, Kato T, Kaneko I. 2010. Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol Microbiol* 75:854-63.

Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, Waters AP, Kato T, Kaneko I. 2009. Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol* 71:1402-14.

Zhang Y, Reinberg D. 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* 15:2343-60.

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322:750-6.

Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, et al. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 38:1341-7.

Zhou J, Berger SL. 2004. Good fences make good neighbors: barrier elements and genomic regulation. *Mol Cell* 16:500-2.

Zhu B, Reinberg D. 2011. Epigenetic inheritance: uncontested? *Cell Res* 21:435-41.

Zuccala ES, Baum J. 2011. Cytoskeletal and membrane remodelling during malaria parasite invasion of the human erythrocyte. *Br J Haematol* 10.1111/j.1365-2141.2011.08766.x.

Acknowledgments

Blood, sweat and tears.

First and foremost I would like to extend my deepest gratitude to Alfred for his resolute mentorship over the last four years and Lluís for being such a great model of scientific leadership. Gentlemen, it has been a privilege.

I would like to thank the members of my thesis advisory committee, Hernando del Portillo, Alejandro Vaquero, Ferran Azorín and Francesc Posas for their support and advice. I would also like to thank the IRB for funding me over the past four years.

Núrietttttta, what would I do without you? Thank you for your (24 h) help and support.

Colleagues, past (Ana, Francesc, Laia, Manuel, Renaud, Teresa, Tom, Yaiza) and present (Anna, Daria, Eva, Noelia, Paco, Patricia, Tanit), thank you for your tireless help and support. I was fortunate to have such generous, patient, conscientious (and fun!) lab mates as yourselves.

Thank you IRR, BM, JBA, PB, TBS, HG, MPS, IGA, TO, RP and SR for sharing so many good times with me in (vibrant!) Barcelona. Thank you for the love and support from those who have been far away and sorely missed (SF, JJ, KK, AH and GL).

Above all, thank you to my family for all the love.