

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/191135>

Please be advised that this information was generated on 2019-06-01 and may be subject to change.

Chromatin regulation in malaria parasites

Sabine Anne-Kristin Fraschka

ISBN

978-94-6295-927-9

Cover & Layout

S.A. Fraschka

Printed by

Digiforce || ProefschriftMaken

The research presented in this thesis was performed at the Department of Molecular Biology of the Science Faculty, Radboud Institute for Molecular Life Science (RIMLS), Radboud University Nijmegen, The Netherlands, and was supported by a Ph.D. fellowship from the European Community's Seventh Framework Program [grant no. FP7/2007–2013] under grant agreement no. 242095 and no. ParaMet 290080 and The Netherlands Organization for Scientific Research [NWO-Vidi 864.11.007].

Copyright © 2018 S.A. Fraschka

All rights reserved. No part of this book may be reproduced, stored in a retrieval database or published in any form or by any means, without prior permission of the author or the publisher holding the copyright of the published articles.

Chromatin regulation in malaria parasites

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op

donderdag 5 juli 2018

om 14.30 uur precies

door

Sabine Anne-Kristin Fraschka

geboren op 19 april 1986

te Mosbach, Duitsland

Promotor

Prof. dr. ir. H.G. Stunnenberg

Copromotor

Dr. R. Bártfai

Manuscriptcommissie

Prof. dr. ir. J.H.L.M. van Bokhoven

Prof. dr. A.P. Waters (University of Glasgow, Verenigd Koninkrijk)

Dr. B.M.D. Franke-Fayard (LUMC)

Chromatin regulation in malaria parasites

Doctoral thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of Deans
to be defended in public on

Thursday, July 5, 2018

at 14.30 hours

by

Sabine Anne-Kristin Fraschka

born on April 19, 1986

in Mosbach, Germany

Supervisor

Prof. dr. ir. H.G. Stunnenberg

Co-supervisor

Dr. R. Bártfai

Doctoral Thesis Committee

Prof. dr. ir. J.H.L.M. van Bokhoven

Prof. dr. A.P. Waters (University of Glasgow, United Kingdom)

Dr. B.M.D. Franke-Fayard (LUMC)

Contents

| | |
|---|-----|
| CHAPTER 1 | 9 |
| General introduction | |
| CHAPTER 2 | 29 |
| Epigenetic gene regulation: key to development and survival of malaria parasites | |
| CHAPTER 3 | 51 |
| H3.3 demarcates GC-rich coding and subtelomeric regions and serves as potential memory mark for virulence gene expression in <i>Plasmodium falciparum</i> | |
| CHAPTER 4 | 85 |
| Comparative heterochromatin profiling reveals conserved and unique epigenome signatures linked to adaptation and development of malaria parasites | |
| CHAPTER 5 | 129 |
| GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing | |
| CHAPTER 6 | 167 |
| Chromatin accessibility-based characterization of the gene regulatory network underlying <i>Plasmodium falciparum</i> blood stage development | |
| CHAPTER 7 | 207 |
| General discussion | |
| SUMMARY..... | 223 |
| SAMENVATTING..... | 225 |
| ACKNOWLEDGEMENTS..... | 227 |
| CURRICULUM VITAE..... | 230 |
| PUBLICATIONS..... | 231 |

CHAPTER 1

General introduction

The global impact of malaria

Transmitted by *Anopheles* mosquitoes, unicellular parasites of the genus *Plasmodium* yearly infect millions of people who consequently fall sick with a disease that has a potential fatal outcome: Malaria.

Widespread in **tropical and subtropical countries**, malaria jeopardizes nearly **half of the world's population**. In 2016 alone, 216 million people suffered and an estimated 445,000 people died from malaria, despite the painstakingly achieved decline in global incidences (21%) and mortality rates (29%) since 2010 [1]. **Children** under five years of age accounted for more than 70% of all malaria mortalities, and, together with **pregnant women, HIV patients, and travellers** from non-endemic or low-transmission countries are most susceptible due to insufficient immunity against the parasite [1]. Malaria also has a **huge economic burden** on these countries. Costs due to healthcare can reach up to 40% of all public health expenses in high-burden countries and together with the costs due to lost productivity sum up to millions of dollars. In (Sub-Saharan) Africa alone, where nine out of ten disease cases occur, yearly malaria related costs have been estimated to be at least US\$ 12 billion [1]. Current strategies to fight malaria encompass diagnosis and treatment of patients as well as preventive treatment of pregnant women and infants, usage of insecticide-treated (bed) nets, indoor residual spraying and occasional larval control or other vector control interventions [1, 2]. To date, a vaccine against malaria is not commercially available, and the most advanced vaccine candidate RTS, S/AS01 provides, unfortunately, only partial and transient protection amongst young children and infants [3]. Furthermore, **resistance of parasites** to nearly all currently available antimalarial drugs emerges. Even to artemisinin, a compound of the herb *Artemisia annua*, which has been used to treat malaria for more than 1,000 years [4]. Additionally, **insecticide resistance** in *Anopheles* mosquitoes spreads rapidly, and has been reported in 60 countries since 2010 [1, 5].

Thus, the alarming development of drug-resistant malaria parasites and insecticide-resistant mosquitoes, together with the lack of an effective and long-term protective malaria vaccine, poses a tremendous threat to current malaria control interventions and demands better understanding of the biology of this malaria causing parasite and novel intervention strategies.

The genus *Plasmodium*

Malaria parasites are **obligate intracellular** parasites, which depend on female blood-sucking **mosquitoes** for **sexual reproduction** and transmission, and **vertebrate hosts** for **asexual proliferation**. They have been traditionally distinguished by basic features such as morphology, host specificity and geographic location and belong to the **genus *Plasmodium***, the family Plasmodiidae, the order haemosporida, the class haematozoa and the **phylum**

Apicomplexa [6, 7]. The existence of *Plasmodium* parasites can be traced back to the Cretaceous period (~66-145 million years ago) [6] and until now, approximately 200 different *Plasmodium* species have been discovered. Most of them infect reptiles and birds, but also a variety of mammals such as ungulates, bats, rodents and primates [7, 8]. Reptile- and bird-infecting *Plasmodium* parasites are transmitted by a wide range of culicine and anopheline mosquitoes from the family Culicidae and the order Diptera [9]. Mammal-infecting *Plasmodium* parasites, by contrast, seem to be solely transmitted by anopheline mosquitoes [10], suggesting a link between vector specialization and parasite radiation in mammals [11].

Most *Plasmodium* species are rather host specific [12]. But some of them, such as certain avian malaria species, infect hosts from different families and orders [13, 14] and under experimental conditions even adapted to erythrocytes of phylogenetically distant hosts, for example, of the mammalian class [15, 16]. Interestingly, a phylogenetically early host switch from birds to bats has previously been identified as the evolutionary key event that lead to the radiation of *Plasmodium* parasites together with their respective mammalian hosts [17, 18], also giving rise to the current human-infecting *Plasmodium* species [18-20].

Table 1: Human-infecting *Plasmodium* species and their typical features (reviewed in [21-25]).

| Species | Pre-erythrocytic infection | Hypnozoite formation | Asexual cycle*/RBC or RC# | Severity of the disease | Endemic areas |
|--|----------------------------|----------------------|---------------------------|--|--|
| <i>P. falciparum</i> | 5-10 days | no | 2 days / both | Most severe, cerebral malaria, highest mortality | Predominantly in Africa |
| <i>P. vivax</i> | 6-12 days | yes | 2 days / only RC | Moderate to severe | Predominantly in South-East Asia |
| <i>P. malariae</i> | 14-16 days | no | 3 days / prefers RBC | Mild to moderate | Found worldwide in endemic countries |
| <i>P. ovale curtisi</i> , <i>P. ovale wallikeri</i> | 9 days | yes | 2 days / only RC | Mild | Found in Africa, western Pacific islands |
| <i>P. knowlesi</i> (Macaque monkeys and humans) | 6-8 days | no | 1 day / RBC | Severe | Frequent zoonosis in Southeast Asia |

*approximate duration, # reticulocyte

Human-infecting *Plasmodium* species and their pathogenesis

Five *Plasmodium* species parasitize humans as their natural hosts (*P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale curtisi* and *P. ovale wallikeri* [26]) and the Macaque monkey-infecting species *P. knowlesi* infects humans frequently [27] (Table 1). Zoonosis (the disease transmission from animals to humans) with other primate-infecting *Plasmodium* species has been documented, but just diagnosed for few individual cases [23]. The vast majority of human malaria cases is caused by *P. falciparum* and *P. vivax*: in **Sub-Saharan Africa** primarily by *P. falciparum*, in **South America** primarily by *P. vivax* and in South-East Asia and western Pacific by both [24, 28]. *P. o. curtisi*, *P. o. wallikeri* and *P. malariae* infections are widespread, but much less abundant, and mostly occur as co-infection with *P. falciparum* [20, 24]. *P. knowlesi* infections are also less common and only occur in certain forested areas of **South-East Asia**. However, they might be underrated due to misdiagnoses by microscopy [25, 27].

After being infected, **malaria symptoms** appear when the parasites asexually proliferate within the human blood. The course of the disease varies depending on the infecting *Plasmodium* species. At the onset of disease, all species cause rather unspecific symptoms such as headache, slight fever, muscle pain and nausea. With increasing parasite load, these are followed by the **characteristic periodic fever** attacks with chills (cold stage), feeling of intense heat, burning skin and severe headache (hot stage) and profuse sweating and exhaustion (sweating stage). At the same time, additional symptoms can occur such as enlargement of the spleen and liver, slight jaundice and hemolytic anemia. The periodicity of the fever attacks occurs every **24 hours** for *P. knowlesi*, every **48 hours** for *P. vivax*, *P. o. curtisi* and *P. o. wallikeri*, and every **72 hours** for *P. malariae*. *P. falciparum* causes rather **irregular attacks** with a 36-48-hour periodicity. Even daily attacks or continuous fever are possible [21, 25, 29, 30] (Table 1).

P. falciparum causes the most life-threatening form of malaria and evolved an exceptional strategy to evade the host's immune system. Specific proteins anchored at the surface of infected erythrocytes mediate **adherence to endothelial cells** of blood vessels and aggregation with other infected and uninfected erythrocytes. In this way, the parasites can escape the clearance by the spleen. However, it also leads to reduced blood flow causing **dysfunction of vital organs** such as kidneys, lung and liver, and the brain (**cerebral malaria**) resulting in severe malaria with acute anemia, respiratory distress and coma. Untreated severe *P. falciparum* infections are often lethal and account for the vast **majority of malaria mortalities** worldwide [21, 29, 30] (Table 1).

The life cycle of *P. falciparum* parasites

A single bite of an infected mosquito can be sufficient to cause malaria. While the mosquito probes for blood, saliva and parasites residing in the mosquito's salivary glands, so-called **sporozoites**, are injected into the skin of the human host (Figure 1). Within the skin the sporozoites randomly migrate until they reach a blood vessel and after invading it rapidly home to the liver. In the liver, just a couple of hours after the mosquito bite, they infect hepatocytes by invagination and grow into mature **liver schizonts** within the next ten days. Each mature liver schizont gives rise to tens of thousands infectious **merozoites** which are shed as vesicles, the merozoites, into the bloodstream. This pre-erythrocytic phase of the infection remains clinically asymptomatic [21]. In the pulmonary capillaries, the merozoites are released from the merozoites and within seconds recognize, attach and enter erythrocytes by invagination through specific ligand-receptor interactions. Once intracellular again, the parasite immediately turns into a **ring** stage and remodels the host cell for about 24 hours, before it enters the metabolically very active **trophozoite** stages. In the trophozoite stage, the parasite grows tremendously digesting hemoglobin as a primary source of nutrition. After roughly 12 hours the parasite enters the schizont stage and undergoes asynchronous rounds of endoreplication within the next 12 hours, while forming a multinucleated **schizont**. The mature schizont ruptures and releases up to 32 invasive **merozoites** as well as fever-causing endotoxins into the bloodstream. Within seconds fresh erythrocytes are invaded and a new cycle of **asexual replication** starts. Importantly, all malaria-related symptoms occur due to asexual proliferation and as mentioned before *P. falciparum* trophozoite and schizont stages additionally adhere to the vascular endothelium reducing the blood circulation in vital organs. Merozoites descending from a low number of sexually committed schizonts, undergo **sexual differentiation** and within two weeks they develop into mature **male** and **female gametocytes**, which might become part of a mosquito's blood meal [21]. In the mosquito midgut, male gametocytes produce eight flagellated **microgametes** and female gametocytes mature into **macrogametes**. Male and female gametes fuse and form a diploid **zygote** which differentiates into a motile **ookinete**. The ookinete traverses the midgut wall and forms an oocyst, which undergoes several cycles of replication producing thousands of **sporozoites**. After approximately 17 days these are released into the hemocoel and migrate to the salivary glands where they become infectious and reside until the female mosquito takes its next blood meal, thereby closing the cycle. [21, 31] (Figure 1).

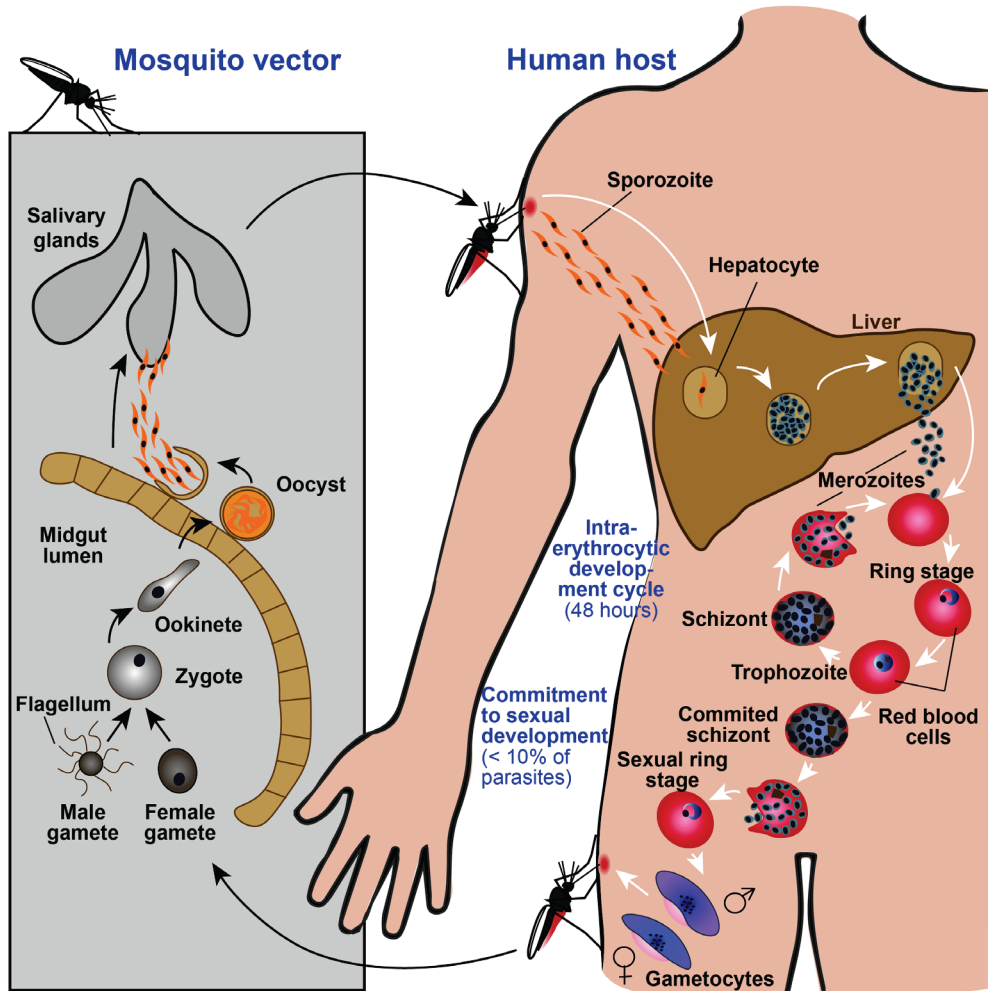


Figure1: Life cycle of *P. falciparum* parasites.

Chromatin and gene expression regulation in eukaryotic organisms

All information needed for an organism to develop, survive and to reproduce is encoded in its DNA. In eukaryotes – to which malaria parasites also belong – the DNA is packaged into a highly organized structure, the chromatin, and stored in a special compartment, the nucleus. To convert the information encoded by the DNA into various functional products, sections of the DNA, the genes, are transcribed into RNA, which are subsequently translated into proteins. By expressing different sets of genes, cells carrying identical DNA can give rise to very different phenotypes with strikingly distinct morphologies and functions. Accordingly, the expression of genes needs to be tightly regulated. This is accomplished through mechanisms which regulate the accessibility of the DNA (epigenetic mechanisms), the transcription of the DNA (transcriptional mechanisms), the stability and translation of the RNA (post-transcriptional mechanisms) as well as further processing of the proteins (post-translational mechanisms). In this thesis, I mainly discuss transcriptional and epigenetic mechanisms.

Transcriptional gene regulation

In eukaryotes, most protein coding genes are transcribed by the DNA-dependent RNA polymerase II (Pol II). The activity of this multi-protein complex is regulated both at the level of transcriptional initiation and elongation [32-34].

The transcriptional initiation occurs at the **promoter**, which is located upstream of the gene body and contains the so-called **core (or minimal) promoter**. The core promoter consists of the **transcription start site (TSS)** and diverse regulatory elements (e.g. the TFIIB recognition element (BRE) and/or the TATA box), which are essential for the binding of general transcription factors (GTFs) and their assembly into the so-called **pre-initiation complex (PIC)** [35]. The PIC is crucial for the positioning of Pol II at the TSS and its advance to the pause site from where, after additional signaling events and recruitment of further factors, Pol II is released ('promoter clearance') and progresses into a phase of productive elongation [32, 33]. The promoter sequence upstream of the core promoter contains further DNA elements, the so-called proximal *cis*-regulatory elements (*cis* = acting on the same DNA molecule) [36]. These contain distinct DNA sequences, so-called **motifs**, which can be unique for individual genes or subsets of genes and are recognized and bound by the DNA-binding domain of specific **transcription factors (TFs)** [37]. Through their *trans*-activating domain (*trans* = acting on a different molecule) specific TFs interact with co-factors and/or directly with the PIC, which in turn alter the local chromatin structure and/or influence the assembly and/or the activity of the PIC. Depending on the respective activity, specific TFs can promote (**activators**) or inhibit (**repressors**) the transcription of their **target genes** [38]. In addition, TF binding sites can be located outside of the gene promoter and are able to influence gene transcriptions over large distances, for instance, by local DNA

looping or broader DNA reorganizations within the nucleus [39]. In mammals, long-range interactions between distant regulatory elements and promoters even occur over several thousands of base pairs. Distant regulatory sequences, which activate their target genes independent of their orientation in *cis* or in *trans* are often referred to as **enhancers** whereas distant regulatory elements which repress gene transcription as **silencers**. Silencers which restrict gene transcription to specific chromatin domains are often described as **insulators** [39]. The complex interplay between the above regulatory DNA elements and GTFs as well as specific TFs orchestrates the regulation of gene transcription, which is further fine-tuned by epigenetic mechanisms.

Epigenetics

In 1942, Conrad Waddington defined the term epigenetics as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being” [40]. Since then, many refined definitions have been suggested [41, 42], including the definition of an epigenetic trait as “a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” [43]. Nowadays, however, the term epigenetics often refers to any chromatin-based regulatory processes, ranging from the methylation of the DNA to the localization of chromatin domains within the nucleus, whereas the term epigenetic memory often refers to the stable propagation of an altered gene expression pattern, potentially induced by cell internal or external stimuli [44].

Epigenetic regulatory mechanisms

Nucleosomes and their modifications

In eukaryotes, the basic unit of the **chromatin** is the **nucleosome** which consists of 145- 147 base pairs (bp) of DNA wrapped around a core particle composed of eight **canonical histones** (2x H2A, 2x H2B, 2x H3 and 2x H4) (Figure 2). Additionally, in many eukaryotes, a linker histone (H1) binds the histone octamer at the entry and exit sites of the DNA and stabilizes the chromatin structure [45]. Together, the nucleosomes form a continuous array across the chromatin fiber. Their density (**nucleosome occupancy**) and their location in relation to DNA elements or other nucleosomes (**nucleosome positioning**) influence and/or regulate the accessibility of the underlying DNA. For instance, by masking or exposing TF binding sites. ATP-dependent **chromatin remodelers** contribute to these mechanisms and actively remove nucleosomes (**nucleosome eviction**) or slide the histone octamers along the DNA (**nucleosome repositioning**). Furthermore, the continuous exchange of nucleosome core particles (**nucleosome turnover**) leads to a transient exposure of the underlying DNA to DNA-binding proteins. [45-48] (Figure 2).

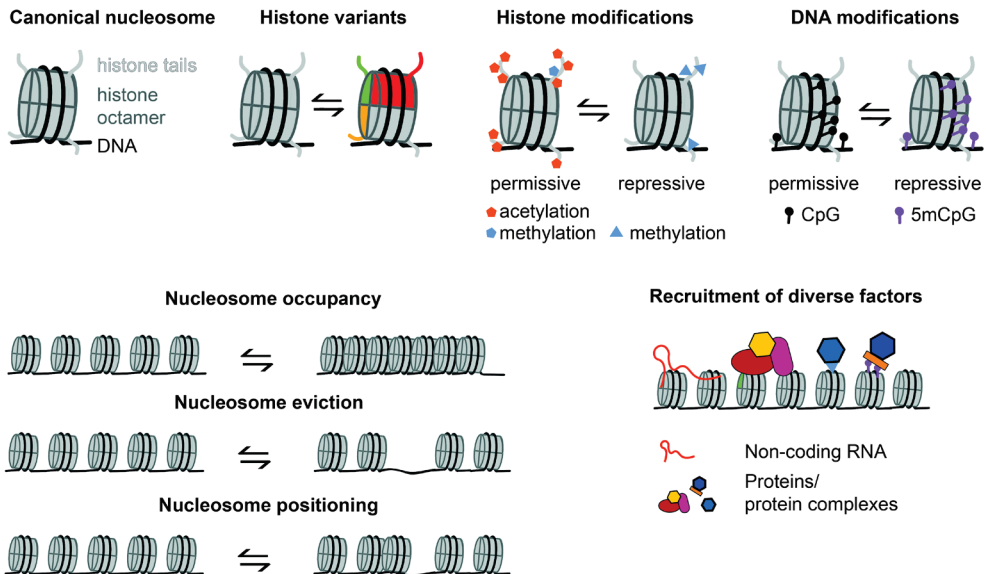


Figure 2: Examples of epigenetic mechanisms.

Importantly, epigenetic mechanisms act together to orchestrate chromatin regulating processes and should not be seen as separate entities. For instance, histone variant incorporation could enable distinct post-translational histone modifications (PTMs), which recruit effector proteins regulating nucleosome repositioning. This in turn could lead to TF binding, which results in the recruitment of other effector proteins, which in turn write or erase PTMs.

The synthesis of the canonical histones (and of several H1 variants) is commonly coupled to DNA replication, and their assembly into nucleosomes typically occurs behind the replication fork. Special **histone chaperones** participate in nucleosome assembly and disassembly and ensure proper histone folding as well as escort and deposition of histones into the chromatin [45]. The histone chaperone Caf-1, for instance, binds to the DNA replication complex ('replisome') where it is thought to assemble and deposit (H3–H4)_x2 tetramers onto freshly replicated DNA [45].

Canonical histones can be replaced by a wide array of **histone variants**, mainly on a replication independent manner (Figure 2 and Table 2). Their incorporation can alter the biophysical properties of the nucleosome and/or provide distinct sites for post-translational modifications, thereby giving rise to nucleosomes with unique structural and functional characteristics [45, 49, 50]. The incorporation of the H3-variant CenH3 (also known as CENPA or Cse4), for instance, is essential for the formation and functionality of the centromeres [51, 52] (Table 2). Furthermore, nucleosomes contain distinct sites for **post-translational histone modifications** (PTMs), which can alter the structure and function of the nucleosomes and the surrounding chromatin alike (Figure 2, Table 3). Histone PTMs can directly influence histone-DNA and histone-histone interactions and thus, for instance, alter the stability of the nucleosome and affect the pace of histone exchange (turnover rate), or

Table 2: Canonical histones and their variants as well as their function/properties and chaperones (reviewed in [45, 46, 53-56])

| Histone | Function and/or properties | Chaperone |
|--|---|---|
| H2A | Canonical histone | FACT, NAP1, Nucleolin (H2A-H2B) |
| H2A.Z/H2AV | Transcription initiation, DNA repair | CHZ1 (yeast), ANP32E (mammals) (H2A.Z-H2B) |
| H2A.X | DNA repair, genome stability, meiotic remodeling of sex chromosomes | FACT (H2A.X-H2B) |
| macroH2A | Gene silencing, X chromosome inactivation | APLF (core histones/macroH2A-H2B) |
| H2A.B/H2A.Bbd | Transcription initiation, testis and brain | NAP1 |
| H2B | Canonical histone | FACT (H2A-H2B) |
| TSH2B.1/H2B1A | Testis-specific | Not known |
| H2B.W/H2BFWT | Testis-specific | SWI-SNF(remodeler) |
| H3 | Canonical histone | Caf-1, NAP1, FACT, Asf1 (H3-H4) |
| CenH3 (i. a. plants) /CENP-A (vertebrates) /Cse4 (yeast) | Kinetochores attachment, chromosome segregation | Scm3 (yeast), HJURP (mammals) CenH3-H4 |
| H3.3 | Transcription, Transcriptional memory, genome integrity, establishment of heterochromatin | HIRA (deposition at active genes, DNA damage sites), ATRX-DAXX (deposition at heterochromatic regions), additional chaperones |
| H3.Z/H3.Y.2 | testis, brain and tumors (euchromatin) | Not known |
| H3.Y/H3.Y.1 | testis, brain and tumors (not known) | Not known |
| H4 | Canonical histone | Caf-1, FACT, Asf1 (H3-H4) |
| H4V | Trypanosomes (protozoan parasite) [56] | ? |

can recruit non-histone proteins to chromatin [46]. PTMs occur at histone tails and at the lateral surface of nucleosomes and primarily affect lysine, serine, arginine and tyrosine residues [58]. Common histone PTMs are modifications with chemical groups such as acetyl, methyl- and phosphate or small proteins such as ubiquitin and SUMO. They are placed, removed and recognized by special classes of proteins, commonly referred to as **'writers'**, **'erasers'** and **'readers'**. Acetyl groups, for instance, are placed by histone acetyltransferases (HATs), removed by histone deacetylases (HDACs) and recognized by proteins with unique structural motifs such as the bromodomain that binds acetylated lysine. Methyl groups, on the other hand, are placed by histone methyltransferases (HMTs), removed by histone

demethylases (HDMs) and recognized by, for instance, chromodomain (H3K9me2/me3, H3K27me2/me3), Tudor domain (H3K36me3, H3Rme2, H4Rme2), or PHD finger containing proteins (H3K4me2/me3, H3K9me3) [59, 60].

Besides histone PTMs, chemical modifications of the DNA, such as **methylation** and **hydroxymethylation** play important roles for gene regulation and chromatin organization in many eukaryotic organisms (Figure 2). In mammals, the most abundant modification is the methylation of the carbon-5 of cytosine (5mC) at sites where a cytosine (C) is followed

Table 3: The most prominent histone posttranslational modifications, their function, location and associated readers (reviewed in [58-62])

| Histone PTM | Function and/or location | Reader protein/domain |
|---------------------------------------|---|---|
| H2AK119ub (mammals) | Transcriptional inhibition at (facultative) heterochromatin | ? |
| H2BK123ub (yeast)/ H2B120ub (mammals) | Transcriptional activation, gene bodies | Cps35 (H3K4 methylation) |
| H3K4me3 | Active euchromatin (promoters), transcriptional activation; bivalent genes/gene poising | CHD1 (remodeler) via chromodomain, and others |
| H3K9me2/3 | (Constitutive) heterochromatin formation, transcriptional silencing; imprinting | HP1 via chromodomain. EED (PRC mediated repression) via WD40 domain |
| H3K27me3 | (Facultative) heterochromatin formation, transcriptional silencing; X-inactivation; bivalent genes/gene poising | PCR1 via chromodomain. EED (PRC mediated repression) via WD40 domain |
| H3K36me3 | Transcriptional elongation (coding region) | Tudor domain |
| H3K9ac | Histone deposition; transcriptional activation | Bromodomain |
| H3K14ac | Transcriptional activation; DNA repair | Rsc4 (remodeler) via bromodomain, double PHD finger (DPF) containing proteins |
| H3K27ac | Gene activation, active enhancers | by NURD complex |
| H3K56ac | Cell cycle progression, heterochromatin function | Bromodomain-, double pleckstrin homology (DPH) domain-containing proteins |
| H3K79me2/3 | Active chromatin, telomeric silencing | Tudor domain |
| H3Rme2 | Mitosis | Tudor domain |
| H3S10p | Mitosis, meiosis, transcriptional activation | GCN5 (histone acetylation), Bmh1/2 (yeast) via 14-3-3 domain |
| H4K20me3 | Transcription activation | TDRD3 via Tudor domain |
| H4K16ac | Transcriptional activation; DNA repair | Bromodomain |
| H4Rme2 | Transcription activation | TDRD3 via Tudor domain |

by a guanine (G) forming a CpG dinucleotide [63]. 5mC is involved in processes such as genomic imprinting and X-chromosome inactivation, transposon ‘control’ and regulation of gene expression by recruiting proteins such as members of the methyl-CpG-binding domain (MBD) protein family or influencing the binding of transcriptional regulators. MBDs, in turn, recruit chromatin modifier complexes, which act as transcriptional repressors [63, 64]. Methylation of the carbon-5 of cytosine is ‘written’ by DNA methyltransferases (DNMTs) and ‘erased’ by methylcytosine dioxygenases (Tet proteins) or passively by through DNA replication (dilution) [63, 64]. Besides PTMs and DNA modifications **non-coding RNAs** (ncRNAs) are also involved in regulating gene expression. In particular, long ncRNA (lncRNAs, >200 nucleotides) acting in *cis* such as *XIST* (involved in X-chromosome inactivation) or in *trans* such as *HOTAIR* (involved in *HOX* gene silencing) have been shown to promote gene inactivation by recruiting repressive chromatin modifiers [65]. Together, all these components of the epigenetic machineries orchestrate the composition, the structure and the dynamics of the chromatin.

Chromatin domains and higher order organization

‘Writer’, ‘eraser’ and ‘reader’ proteins as well as chromatin remodeling enzymes often reside in multi-protein complexes and collectively give rise to chromatin regions, often referred to as **chromatin domains**, with distinct physicochemical properties. The two most distinctive examples of chromatin organization are the highly compact and transcriptionally silent **heterochromatin** and the transcriptionally permissive **euchromatin** (Figure 3). Remarkably, during the interphase, euchromatic and heterochromatic domains are differently distributed within the nucleus: actively transcribed euchromatic domains locate more towards the center whereas transcriptional silent heterochromatic domains locate to the periphery of the nucleus where they often associate with the nuclear lamina [66, 67].

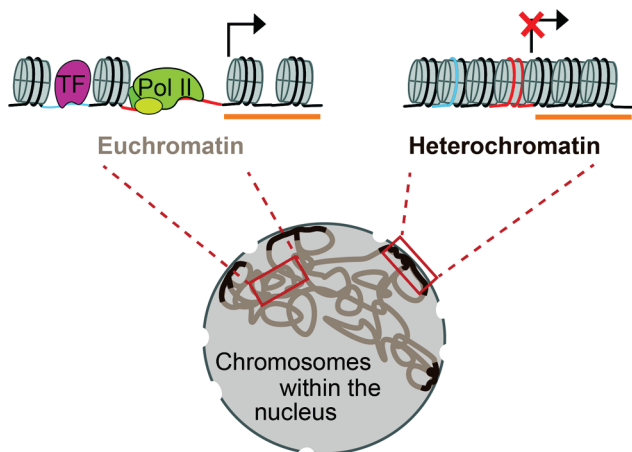


Figure 3: Heterochromatin and euchromatin are distinctly distributed within the nucleus. Densely packaged and transcriptional silent heterochromatin locates towards the periphery of the nucleus. More loosely packaged and transcriptional active euchromatin locates closer to the center of the nucleus. Orange bar represents the coding sequence of a gene. TF: Transcription factor, Pol II: Polymerase II.

In addition, intra- and inter-chromosomal interactions can further contribute to the **higher order organization** of the chromatin. For instance, in metazoans, chromatin is organized into sharply defined genomic regions (e.g. in vertebrates partly through the insulator CTCF and the cohesion complex), which are isolated from the neighboring regions and preferentially self-interact [68]. These so called **topologically associated domains** (TADs) range in size from hundreds of kilobases to several megabases and are either in a heterochromatic or euchromatic state, depending on the transcriptional activity of the underlying genes, and hence contribute to gene expression regulation [67-69].

Heterochromatin and euchromatin

Heterochromatin is formed by the combination of repressive epigenetic marks and effectors proteins that lead to chromatin compaction and typically spans multiple adjacent genes. A well investigated effector protein is the **heterochromatin protein 1** (HP1) that binds di- and tri-methylated lysine 9 of histone 3 (H3K9me2/me3) directly through its chromodomain (Table 3). HP1 molecules on nearby nucleosomes oligomerize and mediate chromatin condensation, which results in a compact chromatin structure in which the underlying genes are silenced. Remarkably, HP1 dimers/oligomers also serve as a scaffold to recruit the enzyme SU(VAR)39, a histone lysine methyltransferase (HKMT), that adds the H3K9me3 modification to adjacent nucleosomes and hence propagates chromatin condensation [70, 71]. In mammals, H3K9me3 typically marks **constitutive heterochromatin**, which remains condensed in all cells at the same regions such as telomeres and pericentromeric regions [72].

Next to constitutive heterochromatin, long term repression of genes can be established through Polycomb group (PcG) proteins, which are core components of two major protein complexes: Polycomb repressive complex 2 (PRC2), which 'writes' the heterochromatic mark H3K27me3, and PRC1, which is able to bind H3K27me3 and mediates mono-ubiquitination of H2A (e.g. H2AK119 in mammal), hence likely restricting Pol II elongation. The resulting **facultative heterochromatin** silences genes on a cell type specific manner and restricts the gene-expression program of differentiated cells [73-76]. In contrast to these repressive epigenetic marks, tri-methylation of H3K4 at promoters and H3K36 at coding regions mediated by Trithorax group (TrxG) proteins is closely linked to gene activation and often counteracts PcG activity. H3K4me3, for instance, likely provides a physical impediment for PRC2 but also recruits complexes involved in H3K27 acetylation. However, bivalent domains marked with H3K4me3 and H3K27me3 (same nucleosome, different H3 tail) allow an 'intermediate' chromatin state in which genes are primed for expression but kept silent [62, 77]. Together with abundant histone acetylation (e.g. H3ac mediated by GCN5), H3K4me3 and H3K36me3 are linked to the more 'open' (accessible) euchromatin which facilitates binding of transcriptional activators and transcription of the underlying genes.

Overview of the thesis

During their life cycle *Plasmodium* parasites develop into various morphologically and functionally distinct forms without altering the sequence of their DNA. As in other eukaryotic organisms, these phenotypic changes are based on the expression of distinct sets of genes. The expression of these genes is regulated by a wide range of mechanisms, which tightly interact with each other. In this thesis, I describe and discuss aspects of the epigenetic landscape of malaria parasites and the contribution of epigenetic and transcriptional mechanisms to gene expression regulation.

Chapter 2 reviews the chromatin landscape in *P. falciparum* and illustrates the complex interplay of epigenetic mechanisms and their contribution to gene expression regulation on the basis of two examples: the process of antigenic variation, which is critical for the evasion of the human immune system and the process of commitment to gametocytogenesis, which is central to parasite transmission.

In **Chapter 3** I explored the chromatin landscape of *P. falciparum* and characterize the genome-wide localization of the histone variant *PfH3.3*. We revealed that *PfH3.3* is incorporated in GC-rich coding and subtelomeric repeat regions on a base composition-dependent manner, and that incorporation of *PfH3.3* to the active *var* locus might contribute to the regulation of antigenic variation.

In **Chapter 4** we performed genome-wide profiling of the heterochromatic mark, HP1 and provide a detailed catalog of heritably silenced genes for different *Plasmodium* species, strains and stages. We unveiled conserved and specialized features of HP1-mediated gene silencing across the genus *Plasmodium*, explored variably silenced genes across different strains and unraveled a major role for HP1-mediated gene silencing in the process of sexual differentiation.

Chapter 5 provides initial mechanistic insights to the regulatory events that lead to derepression of the master transcription factor of gametocyte commitment in *P. falciparum*. Here, we demonstrate that binding of the nuclear protein *PfGDDV1* to the heterochromatic mark *PfHP1* at the *pfap2-g* locus counteracts HP1-mediated silencing and enables expression of *PfAP2-G*, which in turn triggers gametocyte conversion. Our study also revealed that the expression of *PfGDDV1* itself is regulated by its own antisense RNA.

In **Chapter 6** we explored the transcriptional regulatory network in *P. falciparum* during blood stage development. Using the Assay for Transposase Accessible Chromatin-sequencing (ATAC-seq) we identified *cis*-regulatory elements in a genome-wide scale and investigated their role in relation to transcription. Additionally, we identified 'stage-specific' *de novo* and previously identified (*Plasmodium*) motifs and uncovered yet unknown *cis-trans* regulatory interactions.

In **Chapter 7**, I discuss our findings, place them into the context of the state-of-the-art and discuss how they contribute to a better understanding of the biology of malaria parasites.

References

1. WHO, *World Malaria Report 2016*. 2017, World Health Organization.
2. Benelli, G. and J.C. Beier, *Current vector control challenges in the fight against malaria*. Acta Trop, 2017. **174**: p. 91-96.
3. Lyke, K.E., *Steady progress toward a malaria vaccine*. Curr Opin Infect Dis, 2017. **30**(5): p. 463-470.
4. Blasco, B., D. Leroy, and D.A. Fidock, *Antimalarial drug resistance: linking Plasmodium falciparum parasite biology to the clinic*. Nat Med, 2017. **23**(8): p. 917-928.
5. Alout, H., et al., *Malaria Vector Control Still Matters despite Insecticide Resistance*. Trends Parasitol, 2017. **33**(8): p. 610-618.
6. O'Donoghue, P., *Haemoprotozoa: Making biological sense of molecular phylogenies*. Int J Parasitol Parasites Wildl, 2017. **6**(3): p. 241-256.
7. Perkins, S.L., *Malaria's many mates: past, present, and future of the systematics of the order Haemosporida*. J Parasitol, 2014. **100**(1): p. 11-25.
8. Perkins, S.L. and J. Schaer, *A Modern Menagerie of Mammalian Malaria*. Trends Parasitol, 2016. **32**(10): p. 772-782.
9. Santiago-Alarcon, D., V. Palinauskas, and H.M. Schaefer, *Diptera vectors of avian Haemosporidian parasites: untangling parasite life cycles and their taxonomy*. Biol Rev Camb Philos Soc, 2012. **87**(4): p. 928-64.
10. Molina-Cruz, A., T. Lehmann, and J. Knöckel, *Could culicine mosquitoes transmit human malaria?* Trends Parasitol, 2013. **29**(11): p. 530-7.
11. Martinsen, E.S., J.L. Waite, and J.J. Schall, *Morphologically defined subgenera of Plasmodium from avian hosts: test of monophyly by phylogenetic analysis of two mitochondrial genes*. Parasitology, 2007. **134**(Pt 4): p. 483-90.
12. Garnham, P.C.C., *Malaria parasites and other haemosporidia*. 1966, Oxford: Blackwell.
13. Krizanauškiene, A., et al., *Variation in host specificity between species of avian hemosporean parasites: evidence from parasite morphology and cytochrome B gene sequences*. J Parasitol, 2006. **92**(6): p. 1319-24.
14. Santiago-Alarcon, D., et al., *Different meal, same flavor: cospeciation and host switching of haemosporidian parasites in some non-passerine birds*. Parasit Vectors, 2014. **7**: p. 286.
15. McGHEE, B., *The adaptation of the avian malaria parasite Plasmodium lophurae to a continuous existence in infant mice*. J Infect Dis, 1951. **88**(1): p. 86-97.
16. MCGHEE, R.B., *Comparative susceptibility of various erythrocytes to four species of avian Plasmodia*. J Infect Dis, 1957. **100**(1): p. 92-6.
17. Lutz, H.L., et al., *Diverse sampling of East African haemosporidians reveals chiropteran origin of malaria parasites in primates and rodents*. Mol Phylogenet Evol, 2016. **99**: p. 7-15.
18. Silva, J.C., et al., *A new method for estimating species age supports the coexistence of malaria parasites and their Mammalian hosts*. Mol Biol Evol, 2015. **32**(5): p. 1354-64.
19. Sundararaman, S.A., et al., *Genomes of cryptic chimpanzee Plasmodium species reveal key evolutionary events leading to human malaria*. Nat Commun, 2016. **7**: p. 11078.
20. Rutledge, G.G., et al., *Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution*. Nature, 2017. **542**(7639): p. 101-104.
21. Cowman, A.F., et al., *Malaria: Biology and Disease*. Cell, 2016. **167**(3): p. 610-624.
22. Antinori, S., et al., *Biology of Human Malaria Plasmodia Including Plasmodium Knowlesi*. Mediterranean Journal of Hematology and Infectious Diseases, 2012. **4**(1): p. e2012013.
23. Ramasamy, R., *Zoonotic malaria - global overview and research and policy needs*. Front Public Health, 2014. **2**: p. 123.
24. Autino, B., et al., *Epidemiology of malaria in endemic areas*. Mediterr J Hematol Infect Dis, 2012. **4**(1): p. e2012060.

25. Singh, B. and C. Daneshvar, *Human infections and detection of Plasmodium knowlesi*. Clin Microbiol Rev, 2013. **26**(2): p. 165-84.
26. Sutherland, C.J., et al., *Two nonrecombining sympatric forms of the human malaria parasite Plasmodium ovale occur globally*. J Infect Dis, 2010. **201**(10): p. 1544-50.
27. Barber, B.E., et al., *World Malaria Report: time to acknowledge Plasmodium knowlesi malaria*. Malar J, 2017. **16**(1): p. 135.
28. Rougemont, M., et al., *Detection of four Plasmodium species in blood from humans by 18S rRNA gene subunit-based and species-specific real-time PCR assays*. J Clin Microbiol, 2004. **42**(12): p. 5636-43.
29. Gazzinelli, R.T., et al., *Innate sensing of malaria parasites*. Nat Rev Immunol, 2014. **14**(11): p. 744-57.
30. Crutcher, J. and S. Hoffman, *Malaria.*, in *Medical Microbiology. 4th edition.*, S. Baron, Editor. 1996, University of Texas Medical Branch at Galveston: Galveston (TX).
31. Josling, G.A. and M. Llinás, *Sexual development in Plasmodium parasites: knowing when it's time to commit*. Nat Rev Microbiol, 2015. **13**(9): p. 573-87.
32. Luse, D.S., *Promoter clearance by RNA polymerase II*. Biochimica et biophysica acta, 2013. **1829**(1): p. 63-68.
33. Sainsbury, S., C. Bernecky, and P. Cramer, *Structural basis of transcription initiation by RNA polymerase II*. Nat Rev Mol Cell Biol, 2015. **16**(3): p. 129-143.
34. Jonkers, I. and J.T. Lis, *Getting up to speed with transcription elongation by RNA polymerase II*. Nature reviews. Molecular cell biology, 2015. **16**(3): p. 167-177.
35. Roy, A.L. and D.S. Singer, *Core Promoters in Transcription: Old Problem, New Insights*. Trends in biochemical sciences, 2015. **40**(3): p. 165-171.
36. Lenhard, B., A. Sandelin, and P. Carninci, *Metazoan promoters: emerging characteristics and insights into transcriptional regulation*. Nat Rev Genet, 2012. **13**(4): p. 233-45.
37. Inukai, S., K.H. Kock, and M.L. Bulyk, *Transcription factor–DNA binding: beyond binding site motifs*. Current Opinion in Genetics & Development, 2017. **43**(Supplement C): p. 110-119.
38. Cooper, G., *Regulation of Transcription in Eukaryotes*, in *The Cell: A Molecular Approach*. 2000, Sinauer Associates: Sunderland (MA).
39. Kolovos, P., et al., *Enhancers and silencers: an integrated and simple model for their function*. Epigenetics & Chromatin, 2012. **5**: p. 1-1.
40. Waddington, C.H., *Endeavour, 1 (1942) 1942*. p. 18–20.
41. Tronick, E. and R.G. Hunter, *Waddington, Dynamic Systems, and Epigenetics*. Front Behav Neurosci, 2016. **10**: p. 107.
42. Deans, C. and K.A. Maggert, *What do you mean, "epigenetic"?* Genetics, 2015. **199**(4): p. 887-96.
43. Berger, S.L., et al., *An operational definition of epigenetics*. Genes Dev, 2009. **23**(7): p. 781-3.
44. D'Urso, A. and J.H. Brickner, *Mechanisms of epigenetic memory*. Trends Genet, 2014. **30**(6): p. 230-6.
45. Talbert, P.B. and S. Henikoff, *Histone variants on the move: substrates for chromatin dynamics*. Nat Rev Mol Cell Biol, 2017. **18**(2): p. 115-126.
46. Venkatesh, S. and J.L. Workman, *Histone exchange, chromatin structure and the regulation of transcription*. Nat Rev Mol Cell Biol, 2015. **16**(3): p. 178-189.
47. Lai, W.K.M. and B.F. Pugh, *Understanding nucleosome dynamics and their links to gene expression and DNA replication*. Nat Rev Mol Cell Biol, 2017. **18**(9): p. 548-562.
48. Elsässer, S.J. and S. D'Arcy, *Towards A Mechanism for Histone Chaperones*. Biochimica et biophysica acta, 2013. **1819**(0): p. 211-221.
49. Weber, C.M. and S. Henikoff, *Histone variants: dynamic punctuation in transcription*. Genes Dev, 2014. **28**(7): p. 672-82.

-
50. Kalashnikova, A.A., R.A. Rogge, and J.C. Hansen, *Linker histone H1 and protein-protein interactions*. *Biochim Biophys Acta*, 2016. **1859**(3): p. 455-61.
 51. Müller, S. and G. Almouzni, *Chromatin dynamics during the cell cycle at centromeres*. *Nat Rev Genet*, 2017. **18**(3): p. 192-208.
 52. Steiner, F.A. and S. Henikoff, *Diversity in the organization of centromeric chromatin*. *Curr Opin Genet Dev*, 2015. **31**: p. 28-35.
 53. Ray-Gallet, D. and G. Almouzni, *Nucleosome dynamics and histone variants*. *Essays In Biochemistry*, 2010. **48**: p. 75.
 54. Volle, C. and Y. Dalal, *Histone variants: the tricksters of the chromatin world*. *Curr Opin Genet Dev*, 2014. **25**: p. 8-14,138.
 55. Hammond, C.M., et al., *Histone chaperone networks shaping chromatin function*. *Nat Rev Mol Cell Biol*, 2017. **18**(3): p. 141-158.
 56. Buschbeck, M. and S.B. Hake, *Variants of core histones and their roles in cell fate decisions, development and cancer*. *Nat Rev Mol Cell Biol*, 2017. **18**(5): p. 299-314.
 57. Siegel, T.N., et al., *Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei*. *Genes Dev*, 2009. **23**(9): p. 1063-76.
 58. Lawrence, M., S. Daujat, and R. Schneider, *Lateral Thinking: How Histone Modifications Regulate Gene Expression*. *Trends Genet*, 2016. **32**(1): p. 42-56.
 59. Yun, M., et al., *Readers of histone modifications*. *Cell Res*, 2011. **21**(4): p. 564-78.
 60. Andrews, F.H., B.D. Strahl, and T.G. Kutateladze, *Insights into newly discovered marks and readers of epigenetic information*. *Nat Chem Biol*, 2016. **12**(9): p. 662-8.
 61. Zhang, T., S. Cooper, and N. Brockdorff, *The interplay of histone modifications - writers that read*. *EMBO Rep*, 2015. **16**(11): p. 1467-81.
 62. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control*. *Nat Rev Genet*, 2016. **17**(8): p. 487-500.
 63. Spruijt, C.G. and M. Vermeulen, *DNA methylation: old dog, new tricks?* 2014. **21**: p. 949.
 64. Breiling, A. and F. Lyko, *Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond*. *Epigenetics Chromatin*, 2015. **8**: p. 24.
 65. Dykes, I.M. and C. Emanuelli, *Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA*. *Genomics, Proteomics & Bioinformatics*, 2017. **15**(3): p. 177-186.
 66. van Steensel, B. and A.S. Belmont, *Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression*. *Cell*, 2017. **169**(5): p. 780-791.
 67. Kuznetsova, T. and H.G. Stunnenberg, *Dynamic chromatin organization: Role in development and disease*. *Int J Biochem Cell Biol*, 2016. **76**: p. 119-22.
 68. Ramani, V., J. Shendure, and Z. Duan, *Understanding Spatial Genome Organization: Methods and Insights*. *Genomics, Proteomics & Bioinformatics*, 2016. **14**(1): p. 7-20.
 69. Ciabrelli, F. and G. Cavalli, *Chromatin-driven behavior of topologically associating domains*. *J Mol Biol*, 2015. **427**(3): p. 608-25.
 70. Al-Sady, B., H.D. Madhani, and G.J. Narlikar, *Division of labor between the chromodomains of HP1 and Suv39 methylase enables coordination of heterochromatin spread*. *Mol Cell*, 2013. **51**(1): p. 80-91.
 71. Canzio, D., et al., *Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly*. *Mol Cell*, 2011. **41**(1): p. 67-81.
 72. Saksouk, N., E. Simboeck, and J. Déjardin, *Constitutive heterochromatin formation and transcription in mammals*. *Epigenetics Chromatin*, 2015. **8**: p. 3.
 73. Aranda, S., G. Mas, and L. Di Croce, *Regulation of gene transcription by Polycomb proteins*. *Sci Adv*, 2015. **1**(11): p. e1500737.
 74. Golbabapour, S., et al., *Gene silencing and Polycomb group proteins: an overview of their structure, mechanisms and phylogenetics*. *OMICS*, 2013. **17**(6): p. 283-96.

75. Blackledge, N.P., N.R. Rose, and R.J. Klose, *Targeting Polycomb systems to regulate gene expression: modifications to a complex story*. Nat Rev Mol Cell Biol, 2015. **16**(11): p. 643-649.
76. van Kruijsbergen, I., S. Hontelez, and G.J. Veenstra, *Recruiting polycomb to chromatin*. Int J Biochem Cell Biol, 2015. **67**: p. 177-87.
77. Geisler, S.J. and R. Paro, *Trithorax and Polycomb group-dependent regulation: a tale of opposing activities*. Development, 2015. **142**(17): p. 2876-87.

CHAPTER 2

Epigenetic gene regulation: key to development and survival of malaria parasites

Sabine Anne-Kristin Fraschka and Richárd Bártfai*

Adapted from
Comprehensive Analysis of Parasite Biology,
From Metabolism to Drug Discovery, pp. 399-420 (2016)

* Corresponding author; r.bartfai@science.ru.nl

Abstract

Malaria parasites exhibit a very complex, multi-stage life cycle that requires extensive and accurate regulation of gene expression. Over the last decade, it has become increasingly clear that epigenetic mechanisms are a prerequisite for tightly controlled gene expression regulation by orchestrating chromatin composition, structure and dynamics; and hence play a key role in development and survival of the human malaria parasite *Plasmodium falciparum*. This chapter provides an overview of *P. falciparum*'s epigenetic machinery and its role in various biological processes. It summarizes the current knowledge on *P. falciparum*'s chromatin landscape and illustrates the complexity and importance of epigenetic regulation of gene expression in *P. falciparum* by means of two examples: the pathogenic process of antigenic variation and the transmission-ensuring process of gametocyte conversion. Although, up until now, the precise involvement of epigenetic mechanisms in these processes remains to be unraveled, it becomes evident that deciphering these mechanisms will contribute to understanding parasite pathogenicity and could help to find new strategies to fight this challenging parasite.

Introduction

Plasmodium is the unicellular parasite that causes malaria. It has a complex life cycle involving two hosts: the mosquito vector and a vertebrate host. Within both hosts, *Plasmodium* undergoes rapid transitions into several morphologically and functionally distinct forms. Unlike in multicellular eukaryotes, where the majority of cells differentiate into certain cell types with stable patterns of gene expression, *Plasmodium* constantly re-establishes its gene expression pattern for every single form within its life cycle. However, *Plasmodium* is also capable to maintain expression patterns of certain genes over several generations; for instance, the expression of a single antigenic variation gene is conserved during multiplication of asexual blood stage parasites. To achieve this, *Plasmodium* not only needs to control its gene expression in a perfectly timed manner, but also needs to “remember” patterns of gene expression over generations of parasites.

Over the last decades, scientists investigated the role of epigenetic mechanisms in controlling the regulation and memory of gene expression of this perplexing eukaryote. In this chapter, we provide an overview of the various epigenetic mechanisms and discuss why they are key to parasite survival. We mostly refer to findings obtained from the intra-erythrocytic cycle of the human malaria parasite *P. falciparum*, unless indicated otherwise.

Epigenetics – an overview

“Epigenetics” in its classical term refers to “causal interactions between genes and their products, which bring the phenotype into being” [1]. In other words, cells with an identical genome are able to transmit different phenotypic traits to their descendants. Nowadays, the term “epigenetics” is more commonly used to refer to any chromatin-based regulatory processes, for example, post-translational modifications of histones (PTMs) or DNA-methylation. Therefore, the classical definition is often paraphrased as “epigenetic memory”, meaning that genetically identical cells give rise to different phenotypes that are propagated over several generations.

The epigenetic machinery includes a wide-range of components, including DNA, non-coding RNA and numerous proteins. The well-known histones form the building block of chromatin, the nucleosome. A single nucleosome consists of a histone octamer composed of two of each canonical histone (H2A, H2B, H3 and H4) and ~147bp of DNA wrapped around it. Canonical histones can be exchanged by histone variants, which may alter the biophysical properties of nucleosomes and provide distinct sites for PTMs. All histones and their variants can be modified post-translationally with different chemical groups (e.g. acetylation, mono-, di-, or tri-methylation, phosphorylation), or they can be conjugated to small proteins (ubiquitination, sumoylation). Particularly, lysine, serine, arginine and tyrosine residues located in the N-terminal histone tail, protruding from the octamer core, are subject to these modifications. In addition, DNA itself can be modified with chemical groups (e.g. methylation or hydroxymethylation) [2]. Proteins that place (“writers”), remove (“erasers”), or bind to (“readers”) modifications, are also considered prime components of the epigenetic machinery and have key functions in dynamic modifications and interpretation of the epigenetic code [3]. “Writers”, for instance, attach acetyl groups to lysine residues (histone acetyltransferases – HATs) or methyl groups to lysine (histone lysine methyltransferases – HKMT) or arginine residues (histone arginine methyltransferase – HRMT). “Erasers” reverse these modifications by removing acetyl (histone deacetylases – HDACs) or methyl groups (histone demethylases). “Readers” contain domains that recognize and bind to specific modifications. These proteins often reside in multi-protein complexes or contain multiple distinct domains themselves, thereby exhibiting combinations of “reader”, “writer” or “eraser” activities. One well investigated “reader”, for example, is heterochromatin protein 1 (HP1) that binds tri-methylated lysine 9 of histone 3 (H3K9me3) directly through its chromodomain. Interestingly, HP1 also recruits the enzyme SU(VAR)39, a HKMT, that “writes” the H3K9me3 modification providing a prime example of how epigenetic processes are often interconnected [2].

All components of the epigenetic machinery together orchestrate the composition, the structure and the dynamics of chromatin. This in turn contributes to the regulation of chromatin structure dependent processes like chromosome segregation and transcriptional regulation of gene expression. For instance, “open” chromatin can facilitate the binding of

transcriptional activators and thus transcription of the underlying gene (loosely packed and active chromatin state = euchromatin) whereas denser “closed” chromatin has the opposite effect and silences gene expression (compact and inactive chromatin state = heterochromatin). Chromatin is also subjected to a “higher order”, 3D organization within the nucleus. Heterochromatic regions, for instance, often cluster and are localized close to the nuclear periphery (e.g., lamina) while euchromatin is more commonly associated with the nuclear lumen.

The unique genome and chromatin landscape of *Plasmodium falciparum*

P. falciparum has a haploid genome organized into 14 chromosomes consisting of approximately 23 million bp. More than half of the genome is covered by coding sequences that encode about 5300 proteins, and almost 80% of the genome is transcriptionally active during blood stage development [4, 5]. Remarkably, the *P. falciparum* genome is extremely AT-rich with an overall A+T content of 80.6%. Interestingly, the base composition differs in different chromosome regions. The centromeres are most A+T-rich (~97%), followed by intergenic regions (~86%) and coding sequences (~76%). Telomeric and subtelomeric regions, on the other hand, contain the most GC-rich sequences (~73% A+T) [5].

The “chromatome” of *P. falciparum* is composed of the four canonical histones (H2A, H2B, H3 and H4), four histone variants (H2A.Z, H2B.Z, H3.3 and CenH3), at least 70 different PTMs including methylation, acetylation, phosphorylation, sumoylation and ubiquitylation of certain amino acid residues [6-10], and numerous “reader”, “writer” and “eraser” proteins of which very little is known. Nonetheless, 4 HATs, 10 HKMTs, 3 HRMTs, 1 DNA (cytosine-5)-methyltransferase (DNMT), 5 HDACs, 3 lysine demethylases and at least 18 “readers” have been predicted based on domain and homology searches (Table 1) [11, 12]. Specific combinations of (variant) histones and their modifications “divide” the genome into functionally distinct chromatin domains (centromere, telomere, eu- and heterochromatin) [13-15]. The epigenetic makeup of these domains will be discussed in the in the following sections.

Centromeres

The smallest and most AT-rich chromatin domains, the centromeres, are essential for the proper segregation of the chromosomes during cell division. They are demarcated by the histone variant PfCenH3 throughout the entire intra-erythrocytic cycle [16] (Figure 1). On each chromosome, PfCenH3 covers a single region of 4-4.5 kb that encompasses a 2-2.5 kb repeat-rich core with an A+T content of ~97%. In addition to PfCenH3, centromeric nucleosomes may also contain the histone variants PfH2A.Z and PfH2B.Z [16]. Next to these histone variants, small non-coding RNAs transcribed from bidirectional promoters flanking the AT-rich centromere cores associate with centromeric DNA [17, 18] and were suggested

to play an important role in centromeric chromatin formation, for instance, as a tether to recruit PfCenH3 [13, 17, 18]. Interestingly, while centromeres are delineated by pericentromeric heterochromatin in many – including evolutionary-related – organisms [19], heterochromatic marks are apparently absent near *P. falciparum* centromeres [16].

Telomeres

P. falciparum's GC-richer sequences are the telomeres, consisting of degenerated G-rich repeats of which GGGTT(T/C)A is most frequent [20]. Their primary function is to protect the chromosome ends, and they are organized into two parts: an internal region associated with three to four nucleosomes and an outer nucleosome free region [20]. Telomeres of different chromosomes usually cluster together at the nuclear periphery forming four to seven heterologous chromosome end clusters [21]. Furthermore, several findings support the idea that the outer nucleosome free region in *P. falciparum* is similarly organized as in yeast. In yeast, the nucleosome free region of the telomere, the telosome, is occupied by multi-protein complexes, which protect the ends of linear chromosomes from degradation by exonucleases and play a role in anchoring telomeres to the nuclear periphery [22]. Orthologue proteins of these telosome complexes in *P. falciparum* are for instance PfSir2A, a member of the **NAD⁺-dependent**/class III HDACs family [23], "*P. falciparum* origin recognition complex 1 protein" (PFORC1) [24] and "*P. falciparum* telomerase reverse-transcriptase" (PfTERT) [25]. What the exact composition of this complex is, how it is recruited to the telomere and whether it plays a role in subtelomeric heterochromatin formation in *P. falciparum* remains to be investigated.

Table 1: The "writers", "erasers" and "readers" of *P. falciparum*'s epigenetic machinery [11].

| | Enzyme or Domain/PTM mark | Gene ID/Annotation/PTM mark* |
|-----------|---|---|
| 'writers' | Histone acetyltransferase (HAT) | PF3D7_0823300/PfGCN5 PF3D7_1118600/PfMYST PF3D7_1227800/PfElp3? PF3D7_0416400/PfHAT1 |
| | Histone lysine methyltransferase (HKMT) | PF3D7_0629700/PfSET1 PF3D7_1322100/PfSET2 PF3D7_0827800/PfSET3 PF3D7_0910000/PfSET4 PF3D7_1214200/PfSET5 PF3D7_1355300/PfSET6 PF3D7_1115200/PfSET7 PF3D7_0403900/PfSET8 PF3D7_0508100/PfSET9 PF3D7_1221000/PfSET10 |
| | Histone arginine methyltransferase (HRMT) | PF3D7_1426200/PfPRMT1 |

| | | |
|------------------|--|--|
| | DNA (cytosine-5) methyltransferase (DNMT) | PF3D7_1361000/PfPRMT5 PF3D7_0811500/PfCARM1 PF3D7_0727300/PfDNMT |
| 'erasers' | Histone deacetylase (HDAC) Lysine demethylase | PF3D7_0925700/PfHDAC1 PF3D7_1472200/PfHDA1 PF3D7_1008000/PfHDA2 PF3D7_1328800/PfSir2A PF3D7_1451400/PfSir2B PF3D7_0809900/PfJHDM1 PF3D7_0602800/PfJHDM2 PF3D7_1211600/PfLSD1 |
| 'readers' | Bromodomain/Kac Chromodomain/Kme Double chromodomain/Kme Tudor domain/Kme PHD fingers/Kme 14-3-3 proteins/Sph | PF3D7_0823300/PfGCN5 PF3D7_0629700/PfSET1 PF3D7_0110500 PF3D7_1212900 PF3D7_1234100 PF3D7_1033700 PF3D7_1475600 PF3D7_1220900/PfHP1/H3K9me 3 PF3D7_1118600/PfMYST PF3D7_1140700 PF3D7_1023900 PF3D7_1136300/PfTSN PF3D7_0323500 PF3D7_0629700/PfSET1 PF3D7_1322100/PfSET2 PF3D7_1221000/PfSET10/H3/ H3K4me PF3D7_1360700/SUMO ligase PF3D7_1211600/PfLSD1 PF3D7_0310200 PF3D7_1008100 PF3D7_1141800 PF3D7_1433400 PF3D7_0818200/H3S28ph/ H3S28pS32h PF3D7_1362100 PF3D7_1422900 |

* Only for "readers", if known.

Euchromatin

The largest part of *P. falciparum*'s epigenome is found in an open and transcriptional permissive euchromatic state. Throughout the intraerythrocytic cycle, AT-richer euchromatic intergenic regions are demarcated by PfH2A.Z/PfH2B.Z double-variant nucleosomes [26-28] and the classical "active" marks H3K4me3 and H3K9ac [26, 29-31] (Figure 1). H3K4me3 marks euchromatic intergenic regions in a stage-specific manner: early blood stages show only minor enrichment of H3K4me3 whereas at later stages intergenic regions of active genes are clearly marked [26]. H3K9ac, however, follows the expression pattern of most genes and seems to be placed in a transcription-coupled manner [26, 29]. Furthermore, the following marks were found to associate with euchromatic regions: H3K14ac, H3K56ac, H3K79me3, H4K5ac, H4K8ac, H4K12ac, H4K16ac, H4K20me and H4R3me2 [29, 30, 32, 33]. Intriguingly, "readers", "writers" or "erasers" of these modifications are still poorly investigated. However, two HATs, PfGCN5 and PfMYST and one HKMT, PfSET8, have been characterized to some extent [34-37] (Table 1). PfGCN5 preferentially acetylates H3 *in vitro*, particularly on H3K9 and H3K14, and interacts with the adaptor protein PfADA2; probably within the context of a large multi-protein complex similar to other eukaryotes [35, 38-40]. Its inhibition reduced H3K9ac and H3K14ac – but not H4ac levels – and negatively affected parasite growth, suggesting an important role in genome-wide gene activation [41-43]. PfMYST, in contrast, preferentially acetylates H4 *in vitro*, especially H4K5ac, H4K8ac, H4K12ac and H4K16ac and likely plays a role in transcriptional regulation of gene expression, cell cycle progression and DNA damage repair [34]. Finally, PfSET8, which is highly homologous to *Toxoplasma gondii*'s TgSET8, may mono-, di- and tri-methylate H4K20 in *P. falciparum* [36, 37].

Heterochromatin: Subtelomeric regions and intrachromosomal islands

Heterochromatin in *P. falciparum* can be found on all subtelomeric regions, some chromosome internal islands, as well as few singular genes, and encompasses approximately 400 genes, which account for approximately 8% of the parasite's coding genome [30, 31, 44-46]. The heterochromatic subtelomeres can be subdivided into two regions: (i) non-coding repeat-rich regions directly bordering the telomere, and (ii) protein-coding regions. The non-coding region is composed of a mosaic of six different blocks of repetitive sequences, the "telomere associated repetitive elements" (TAREs 1-6). These are positioned in the same order on all chromosomes and span around 20–40 kb [20, 47]. Interestingly, TAREs 1-6 give rise to long non-coding RNAs (lncRNAs) with yet unknown function, which were proposed to be involved in heterochromatin establishment [24, 48-50] (see below). Strikingly, both the protein-coding subtelomeres and the intrachromosomal heterochromatic islands encode primarily for *Plasmodium*-specific multigene families, e.g. *var*, *rifin*, *stevor* and *pfmc-2tm*, which are partly involved in antigenic variation and subject to mutually exclusive gene expression [51] (see 4.1).

Heterochromatic regions are defined by hypoacetylation of histones and reduced accessibility to MNase digestion, suggesting more compact nucleosomal packaging [30, 31, 45]. As a result, genes underlying heterochromatic regions are generally silenced. Other hallmarks of (constitutive) heterochromatin are H3K9me3 [30, 31] and heterochromatin protein PfHP1 [44, 45] (Figure 1). Orthologues of PfHP1 in *Schizosaccharomyces pombe*, *Drosophila melanogaster* and human contain a chromodomain that mediates H3K9me2 and H3K9me3 binding and a chromo shadow domain that enables oligomerization of HP1 proteins leading to nucleosome aggregation and hence the formation of densely packed heterochromatin [52, 53]. PfHP1 also contains both domains and localizes to H3K9me3-marked heterochromatin. Furthermore, a recombinant PfHP1 chromodomain binds both H3K9me2 and H3K9me3 *in vitro*, and PfHP1 depletion results in reduced H3K9me3 levels in blood-stage parasites. Hence, it is likely that PfHP1 and H3K9me3 may have a similar role in heterochromatin formation as shown for other organisms [44, 50, 54] (see the following text). A functional homolog of SU(VAR)39, the enzyme that places H3K9me3 in most organisms, could not be confirmed in *P. falciparum*. However, based on homology searches, PfSet3 (the only Plasmodium protein with a pre-SET domain), appears to be the most likely candidate to deposit this mark [30, 37, 55]. Recently, also unexpected enrichment of H3K36me3 was observed in heterochromatic subtelomeres and intrachromosomal islands [56-58]. Knockdown of PfSET2, the HKMT predicted to deposit H3K36me2/me3 in *P. falciparum*, results in decreased H3K36me3 levels and up-regulation of *var* gene expression [37, 56, 57]. However, H3K36me2/3 levels were only reduced in ring parasites, suggesting that alternative enzyme(s) might perform this methyltransferase activity in other stages [56]. Additionally, H4K20me3 was reported to be present in heterochromatic regions [29, 30, 36], but the repressive H3K27me3 mark, which hallmarks facultative heterochromatin in many multi-cellular organisms, has so far not been detected in *P. falciparum* [9].

Formation of the silent heterochromatic state is crucial for the regulation of several clonally variant gene families and single heterochromatized genes (see below). Thus, it is of great interest to understand the mechanisms of heterochromatin formation in *P. falciparum*. In model organisms, assembly of heterochromatic domains is thought to happen in three different steps: (i) establishment of heterochromatin at nucleation sites due to lncRNAs or DNA-binding proteins that recruit an initial HKMT to deposit H3K9me2/3, (ii) subsequent spreading of its silencing state in a self-perpetuating manner into neighboring regions via HP1 and HKMT interactions and (iii) its maintenance over several generations due to identical re-establishment of its state on newly replicated chromatin after each cell division, all as concerted actions of different histone modifying “writer”, “eraser” and “reader” proteins [59-64]. In *P. falciparum*, heterochromatin establishment may involve PfHDA2, a member of the class II HDAC family, and the two **NAD⁺-dependent**/class III HDACs family members, PfSir2A and PfSir2B. Together, they may allow spreading of H3K9me3/PfHP1 by

removing activating acetyl marks (especially H3K9ac), but also by recruiting additional proteins that are required for heterochromatin establishment [23, 65-68]. Additional proteins suggested to be involved in heterochromatin establishment are PfOrc1 and PfAlba. PfOrc1 binds to TAREs in a PfSir2A dependent-manner and PfAlba-containing molecular complexes associate with TARE6 and co-localize with PfSir2A [23, 24, 69, 70]. Subsequent heterochromatin spreading may require, as in other organisms, binding of PfHP1 to methylated H3K9 which in turn leads to the recruitment of a H3K9 HKMT [34, 54, 59, 61]. The HKMT may methylate H3K9 on the neighboring nucleosome, which in turn would result in PfHP1 binding and heterochromatin spreading from small nucleation sites into large chromosome domains, thereby silencing underlying genes [71, 72]. Despite identification of numerous components of *Plasmodium*'s epigenetic machinery involved in heterochromatin formation, it still remains unclear how these components work together to orchestrate heterochromatin establishment, spreading and maintenance in *P. falciparum*.

Epigenetic regulation of gene expression in *P. falciparum*

Over the last decade, the awareness that epigenetic mechanisms play an essential role for *P. falciparum*'s biology increased. They control pathogenic processes like antigenic variation [73, 74], sequestration of infected erythrocytes [75] and selection of erythrocyte invasion pathways [75-77]. Furthermore, they are also involved in antibiotic resistance [78-80], adaptation to changing environments, for instance increased blood lactate-levels or increased body temperatures of the host (fever) [81], and nutrition availability and uptake [78, 80, 82]. Lastly, epigenetic mechanisms were shown to be crucial for gametocyte conversion and thus for parasite transmission [54, 65, 83, 84]. In this section, we discuss two of these processes in detail to illustrate the complexity of epigenetic regulation of gene expression.

Epigenetic regulation of antigenic variation

Our current knowledge on epigenetic mechanisms in *P. falciparum* mainly derives from studying the process of antigenic variation. Antigenic variation is based on multigene families (e.g. *var*, *rifin*, *stevor* and *pfmc-2tm*) located in heterochromatic subtelomeres and intrachromosomal islands. Most of these families encode proteins that are exported and anchored to the erythrocyte membrane [85-92]. Interestingly, within these families only a small subset of genes is expressed while other family members remain transcriptionally silenced. Remarkably, the active state of one gene and the silent state of all other family members can be maintained over generations of blood stage parasites, being inherited from one parasite to its descendants by an epigenetic memory [46, 93]. However, a small minority of parasites begins to express a different member within the gene family giving the infected red blood cell a "fresh makeup" [94]. Through this phenomenon, some parasites

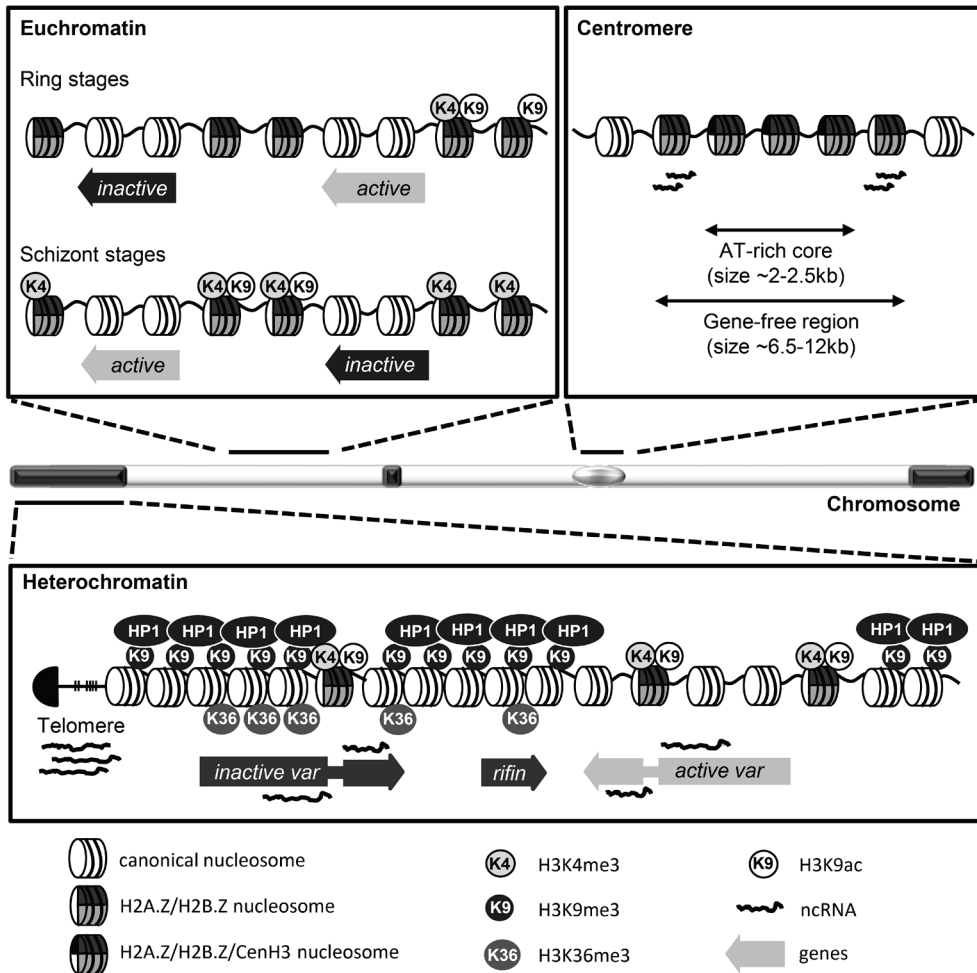


Figure 1: Schematic representation of *P. falciparum*'s chromatin landscape and stage-specific or transcription-coupled chromatin changes.

Euchromatin (white parts of the schematic chromosome): Euchromatic intergenic regions are demarcated by PfH2A.Z/PfH2B.Z double-variant nucleosomes, H3K4me3 and H3K9ac. H3K4me3 marks euchromatic intergenic regions in a stage-specific manner: low in early blood stages, high in late blood stages. H3K9ac is placed in a transcription-coupled manner: high in 5' upstream regions of active genes, low in 5' upstream regions of inactive genes. Centromere (silver ellipse): The centromere is a 4-4.5 kb region demarcated by PfH2A.Z/PfH2B.Z/PfCenH3 triple-variant nucleosomes; its 2-2.5 kb AT-rich core is flanked by bidirectional promoters that give rise to small ncRNAs. Heterochromatin (black blocks): Subtelomeres and intrachromosomal islands are marked by H3K9me3, PfHP1 and H3K36me3. lncRNAs are transcribed in sense and antisense orientation from bidirectional intronic *var* gene promoters. 5' upstream and bidirectional intronic *var* gene promoters of an active *var* gene are marked with H2A.Z/H2B.Z double-variant nucleosomes, H3K9ac and H3K4me3. In the subtelomeric non-coding region TAREs 1-6 express lncRNAs.

can escape the host's immune system, thus ensuring parasite survival within its host and sustained infection. The most extensively studied multigene family, the *var* gene family, encompasses approximately 60 members. Interestingly, *var* gene expression occurs in a mutually exclusive manner, meaning a parasite expresses only a single *var* gene at a time while all other family members remain transcriptionally silent [95]. Every *var* gene consists of two exons separated by a highly conserved intron. Exon 1 encodes an extremely diverse extracellular binding region, while exon 2 encodes the more conserved intracellular domain. Together, they assemble the variant adhesion surface molecule "*P. falciparum* erythrocyte membrane protein 1" (PfEMP1) whose extracellular binding region is exposed at the surface of the infected erythrocyte during trophozoite and schizont stages. Thereby, PfEMP1 exposes parasite antigens to the host immune system, which can be the target for an immune response. Nonetheless, PfEMP1 is an essential virulence factor that mediates adherence of trophozoite- and schizont-infected erythrocytes to the endothelium of capillaries and venules of the host, thus enabling the parasites to circumvent spleen-mediated clearance. Adverse effects associated with the infection are adhesion-based complications like cerebral malaria and pregnancy-associated malaria [96].

Transcription of *var* genes occurs in 8–20-h-old ring stages of *P. falciparum* [51, 95, 97-100]. During the trophozoite and schizont stage (25–48 h after invasion) the active *var* gene is in a "poised" state, meaning it is not actively transcribed but still retains an epigenetic make-up that prevents silencing and enables reactivation in the next generation [101]. With low frequency, a previously silenced *var* gene becomes active while the previously active gene becomes silent. It is currently unknown to which extent this infrequent switch is hard-wired into the parasite's genome or linked to external factors. However, there are few indications that external factors, for instance, starvation stress, high body temperature and lactate levels contribute to *var* gene switching; potentially through modulation of NAD⁺-dependent Sir2 activity [81, 102-104].

Over the last decade, tremendous progress has been made in unravelling the complex mechanisms of *var* gene regulation encompassing interaction of genetic elements, non-coding RNAs, chromatin modifications and modifiers and specific subnuclear localization [13, 74, 100, 105, 106]. Crucial genetic elements involved in *var* gene regulation are the 5' upstream *var* gene promoter that gives rise to PfEMP1-encoding transcripts as well as the *var* gene intron that contains bidirectional promoter activity [107-110]. Based on transgenic parasites expressing episomal or episomally integrated *var* promoters, the following observations were made: (i) 5' upstream *var* promoter and intron pairing is required for *var* gene silencing and (ii) loss of one-to-one *var* promoter and intron pairing as well as integration of an unpaired promoter leads to *var* gene activation [108, 109, 111, 112]. Important *cis*-acting elements within the *var* locus have also been unraveled, which is further discussed elsewhere [15, 100, 113].

Particular localization within the nucleus also influences *var* gene expression as telomeres, subtelomeric, and intrachromosomal *var* genes are tethered to the nuclear periphery and activation of a single *var* gene results in its “repositioning” into a special subnuclear expression site [15, 100]. Apart from genetic elements and subnuclear localization, lncRNAs were shown to play an important role in *var* gene regulation. In particular, lncRNAs transcribed in sense and antisense orientation from the bidirectional promoter within the *var* intron have gained a lot of attention [48, 74, 107, 114-116]. Sense lncRNAs are transcribed from the highly conserved exon2 and have a size of around 2–2.5 kb. They are expressed from all *var* genes at the time when the *var* upstream promoter is silent, meaning from the onset of DNA replication (~24 h post invasion) through schizogony and merozoite formation. They were shown to associate with chromatin [114-116] and were suggested to play a role in epigenetic memory of *var* expression [115]. Antisense lncRNA are transcribed from the variable exon 1 region of *var* genes at the same time as sense lncRNA and have a size of around 1.7 kb [114, 115]. Recently, it was shown that they associate specifically with the single active *var* gene during the ring stage and are incorporated into the chromatin. Moreover, expression of these lncRNAs *in trans* induces *var* gene activation in a sequence- and dose-dependent manner. Interference with antisense lncRNAs leads to silencing of the respective *var* gene, erases its epigenetic memory, and induces *var* gene switching [74]. Additionally, another family of lncRNAs named lncRNA-TARE-4L was detected [48]. These are transcribed from subtelomeric regions that contain a cluster of SIP2-binding motifs in the 5' upstream promoter of a subset of *var* genes and are thought to play an important role in silencing adjacent *var* genes, probably by contributing to the regulation of epigenetic *var* gene memory [48, 50]. Additionally, active, “poised” and silenced *var* genes are specially marked by histone modifications and variants implying an essential role for histone chaperones and chromatin “readers”, “writers” and “erasers” in regulating *var* gene expression (Figure 1). Both active 5' upstream and intronic *var* gene promoters are marked with H2A.Z/H2B.Z double-variant nucleosomes and the histone PTMs H3K9ac, H3K4me2/me3, as well as H4ac in case of the 5' upstream promoter [26, 28, 30, 34]. The “writers” of H4ac and H3K4me2/me3 are histone acetyltransferase PfMYST and methyltransferase PfSet10, respectively, and were shown to be enriched at the active 5' upstream *var* promoter [34, 101]. In the “poised” state, only H3K4me2 and PfSET10 association were observed, while 5' upstream promoter H2A.Z/H2B.Z incorporation was lost [28, 101, 117]. This implies a role for H3K4me2 in transmitting epigenetic memory, possibly by preventing H3K9me3 to spread into the *var* promoter region while the 5' upstream promoter is transcriptional inactive during trophozoite and schizont stages [101, 118]. In contrast, silenced *var* genes are constantly marked by PfHP1, H3K9me3 and H3K36me3, not only in their promoter regions but also over their coding sequences [30, 44, 45, 56]. Intriguingly, *var* introns are constitutively covered by PfH2A.Z/PfH2B.Z [26, 28]. Depletion of PfHP1 or PfSET2 results in simultaneous activation of all *var* genes, even

outside of the perinuclear *var* gene expression site [37, 54, 56, 58]. Similarly, depletion of deacetylating enzymes also leads to up-regulation of all (PfHDA2) or subsets of *var* genes (PfSir2A and PfSir2B) [65, 66, 68] (Table 1).

Although *var* genes represent the most investigated example of epigenetic regulation in the malaria parasite, we are far from understanding the complex interplay of various regulatory events that orchestrate mutually exclusive expression and switching. However, it seems evident that *var* gene expression depends on complex disassembly of heterochromatin at a single locus rather than on simple transcription factor activation.

Epigenetic mechanisms involved in gametocyte conversion

Asexual replication during the intra-erythrocytic cycle establishes and maintains infection of the host. To ensure transmission to other hosts, some parasites develop into gametocytes, through a process called *gametocytogenesis*. Unlike asexual parasites, gametocytes are unable to replicate within the mammalian host, but are infectious to the mosquito vector that subsequently transmits the parasites to new mammalian hosts [119]. To guarantee transmission and maintenance of infection at the same time, the switch from asexual replication to gametocyte conversion needs to be tightly controlled.

Recent studies uncovered AP2-G, a member of the ApiAP2 transcription factor family, as a key regulator of gametocyte conversion in both *P. falciparum* and *P. berghei* [83, 84, 120]. *P. berghei* strains that lost their ability to generate gametocytes showed mutations within their *pbap2-g* locus and targeted disruption of the *pbap2-g* locus caused the same phenotype [84]. Similarly, levels of gametocyte formation in *P. falciparum* strongly correlate with *pfap2-g* expression levels and disruption of PfAP2-G function resulted in loss of gametocyte formation [83]. Both in *P. falciparum* and *P. berghei*, potential AP2-G binding motifs were identified to be enriched in the upstream region of early gametocyte genes as well as *pfap2-g* itself [83, 84]. Additionally, a potential downstream target of PbAP2-G, also a member of the AP2 transcription factor family, was identified. This AP2 modulates gametocyte conversion rates, but disruption of its gene locus does not erase gametocyte formation completely [84]. Collectively, these observations strongly indicate a cascade-like regulation of gametocyte conversion with AP2-G as master regulator and, perhaps even with an auto-regulatory feedback loop.

Intriguingly, *pfap2-g* is the only AP2 family member marked by H3K9me3 and PfHP1, and its expression was recently revealed to be regulated on an epigenetic level [30, 31, 50, 54, 65]. Key regulators of *pfap2-g* expression are components of the epigenetic machinery involved in heterochromatin biology, PfHP1 and the histone deacetylase PfHDA2 [54, 65]. Conditional depletion of PfHP1 resulted in increased gametocyte conversion rates and transcriptional changes of genes associated with gametocyte development [54]. PfHP1 depletion also inhibited mitotic proliferation and caused up-regulated heterochromatic multigene families. Accordingly, Brancucci and co-workers suggest that epigenetic silencing

of *pfap2-g* allows continuous mitotic proliferation and counteracts gametocyte conversion. Local dissociation of PfHP1 from the *pfap2-g* locus, however, activates *pfap2-g* and induces gametocyte conversion. In addition, conditional depletion of PfHDA2 resulted in a similar phenotype with up-regulation of *pfap2-g* and genes associated with gametocyte development. In contrast, knockout of NAD⁺-dependent deacetylases Sir2A and Sir2B does not appear to influence gametocyte conversion rates. In conclusion, PfHP1 and PfHDA2 may cooperate to strictly control gametocyte conversion rates, potentially as part of a specific silencing complex as found in other eukaryotes [54, 121, 122].

Despite these current and exiting findings, much still remains to be uncovered on the precise regulation of *pfap2-g* expression and additional factors involved in this process. For instance, different *Plasmodium* isolates and clones exhibit highly variable gametocyte conversion rates and host signals such as anemia and cytokines influence gametocyte conversion rates [46, 54, 65, 83, 84, 123-125].

Concluding remarks and perspective

Plasmodium parasites exhibit an exquisitely complex life cycle alternating between intracellular, differentiating stages and extracellular, invasive stages, not only in different cell types but also within two different hosts. Prerequisite for such successful stage transition and environmental adaption is a perfectly and tightly controlled gene expression. In the previous sections, two processes with major impact on *Plasmodium*'s survival illustrated the importance of epigenetic mechanisms in controlling *P. falciparum*'s gene expression. However, the spectrum of epigenetically regulated processes during blood stage development is much broader, ranging from erythrocytes invasion [75-77] and nutrition uptake [78, 80, 82] to growth adaptation due to environmental changes [81]. Moreover, other life cycle stages are poorly investigated and may reveal even more epigenetically regulated processes, for instance, variable expression of sporozoite surface proteins to evade the host's immune system or variable expression of transport channels that might allow adjustments of nutrition uptake during liver stage development similar to *clag 3.1* and *clag 3.2* in blood stage development of the parasites [46]. Additionally, epigenetic mechanisms and components in other *Plasmodium* species are barely uncovered. Considering the key roles of epigenetic mechanisms in processes vital to parasite survival and transmission, identification of *Plasmodium* unique components of the epigenetic machinery provides great promise for drug development. Importantly, epigenetic drugs are successfully used in cancer treatment, for instance, **Vorinostat (Zolinza)** and **Romidepsin (Istodax) in cutaneous T-cell lymphoma treatment [126, 127]** and HDAC inhibitors were shown to inhibit malaria parasite growth efficiently [67, 128]. HAT inhibition by curcumin and anacardic acid resulted in more moderate *P. falciparum* blood-stage growth inhibition [37, 42]. However, when used in combination

therapy, they might still be advantageous in counteracting rapidly increasing drug resistance [129]. Recently, two potential HKMT inhibitors, BIX01294 and its structurally related analogue TM2-115, were shown to inhibit parasite growth at all stages of the intra-erythrocytic cycle and exhibit a rapid killing effect [130]. Despite these promising initial results, the major challenge will be to find drug targets and drugs unique for *Plasmodium* and to reduce off-target effects, which are known to occur for all current compounds as effectively as possible. Dissecting the intricate mechanisms of epigenetic regulation in detail will significantly contribute to this task.

References

1. Waddington, C.H., The Epigenotype *Endeavour*, 1 (1942) 1942. p. 18–20.
2. Felsenfeld, G. and M. Groudine, *Controlling the double helix*. Nature, 2003. **421**(6921): p. 448-453.
3. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. Nature, 2000. **403**(6765): p. 41-45.
4. Siegel, T.N., et al., *Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum*. BMC Genomics, 2014. **15**: p. 150.
5. Gardner, M.J., et al., *Genome sequence of the human malaria parasite Plasmodium falciparum*. Nature, 2002. **419**(6906): p. 498-511.
6. Lasonder, E., et al., *Insights into the Plasmodium falciparum schizont phospho-proteome*. Microbes Infect, 2012. **14**(10): p. 811-9.
7. Miao, J., et al., *The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation*. Gene, 2006. **369**: p. 53-65.
8. Treeck, M., et al., *The phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries*. Cell Host Microbe, 2011. **10**(4): p. 410-9.
9. Trelle, M.B., et al., *Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum*. J Proteome Res, 2009. **8**(7): p. 3439-50.
10. Ponts, N., et al., *Unraveling the ubiquitome of the human malaria parasite*. J Biol Chem, 2011. **286**(46): p. 40320-30.
11. Cui, L. and J. Miao, *Chromatin-mediated epigenetic regulation in the malaria parasite Plasmodium falciparum*. Eukaryot Cell, 2010. **9**(8): p. 1138-49.
12. Ponts, N., et al., *Genome-wide mapping of DNA methylation in the human malaria parasite Plasmodium falciparum*. Cell Host Microbe, 2013. **14**(6): p. 696-706.
13. Vembar, S.S., A. Scherf, and T.N. Siegel, *Noncoding RNAs as emerging regulators of Plasmodium falciparum virulence gene expression*. Curr Opin Microbiol, 2014. **20**: p. 153-61.
14. Duffy, M.F., et al., *Epigenetic regulation of the Plasmodium falciparum genome*. Brief Funct Genomics, 2014. **13**(3): p. 203-16.
15. Ay, F., et al., *Multiple dimensions of epigenetic gene regulation in the malaria parasite Plasmodium falciparum: gene regulation via histone modifications, nucleosome positioning and nuclear architecture in P. falciparum*. Bioessays, 2015. **37**(2): p. 182-94.
16. Hoeijmakers, W.A., et al., *Plasmodium falciparum centromeres display a unique epigenetic makeup and cluster prior to and during schizogony*. Cell Microbiol, 2012. **14**(9): p. 1391-401.

17. Li, F., et al., *Nuclear non-coding RNAs are transcribed from the centromeres of Plasmodium falciparum and are associated with centromeric chromatin*. J Biol Chem, 2008. **283**(9): p. 5692-8.
18. Hall, L.E., S.E. Mitchell, and R.J. O'Neill, *Pericentric and centromeric transcription: a perfect balance required*. Chromosome Res, 2012. **20**(5): p. 535-46.
19. Gissot, M., et al., *Toxoplasma gondii chromodomain protein 1 binds to heterochromatin and colocalises with centromeres and telomeres at the nuclear periphery*. PLoS One, 2012. **7**(3): p. e32671.
20. Figueiredo, L.M., et al., *Genomic organisation and chromatin structure of Plasmodium falciparum chromosome ends*. Mol Biochem Parasitol, 2000. **106**(1): p. 169-74.
21. Freitas-Junior, L.H., et al., *Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum*. Nature, 2000. **407**(6807): p. 1018-22.
22. Wright, J.H., D.E. Gottschling, and V.A. Zakian, *Saccharomyces telomeres assume a non-nucleosomal chromatin structure*. Genes Dev, 1992. **6**(2): p. 197-210.
23. Freitas-Junior, L.H., et al., *Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites*. Cell, 2005. **121**(1): p. 25-36.
24. Mancio-Silva, L., et al., *Differential association of Orc1 and Sir2 proteins to telomeric domains in Plasmodium falciparum*. J Cell Sci, 2008. **121**(Pt 12): p. 2046-53.
25. Figueiredo, L.M., et al., *The unusually large Plasmodium telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus*. Nucleic Acids Res, 2005. **33**(3): p. 1111-22.
26. Bártfai, R., et al., *H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3*. PLoS Pathog, 2010. **6**(12): p. e1001223.
27. Hoeijmakers, W.A., et al., *H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the Plasmodium falciparum genome*. Mol Microbiol, 2013. **87**(5): p. 1061-73.
28. Petter, M., et al., *H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite Plasmodium falciparum*. Mol Microbiol, 2013. **87**(6): p. 1167-82.
29. Gupta, A.P., et al., *Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite Plasmodium falciparum*. PLoS Pathog, 2013. **9**(2): p. e1003170.
30. Lopez-Rubio, J.J., L. Mancio-Silva, and A. Scherf, *Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites*. Cell Host Microbe, 2009. **5**(2): p. 179-90.
31. Salcedo-Amaya, A.M., et al., *Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum*. Proc Natl Acad Sci U S A, 2009. **106**(24): p. 9655-60.
32. Chen, H., S. Lonardi, and J. Zheng, *Deciphering histone code of transcriptional regulation in malaria parasites by large-scale data mining*. Comput Biol Chem, 2014. **50**: p. 3-10.
33. Issar, N., et al., *Differential sub-nuclear localisation of repressive and activating histone methyl modifications in P. falciparum*. Microbes Infect, 2009. **11**(3): p. 403-7.
34. Miao, J., et al., *The MYST family histone acetyltransferase regulates gene expression and cell cycle in malaria parasite Plasmodium falciparum*. Mol Microbiol, 2010. **78**(4): p. 883-902.
35. Fan, Q., L. An, and L. Cui, *Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling*. Eukaryot Cell, 2004. **3**(2): p. 264-76.
36. Sautel, C.F., et al., *SET8-mediated methylations of histone H4 lysine 20 mark silent heterochromatic domains in apicomplexan genomes*. Mol Cell Biol, 2007. **27**(16): p. 5711-24.

37. Cui, L., Q. Fan, and J. Miao, *Histone lysine methyltransferases and demethylases in Plasmodium falciparum*. *Int J Parasitol*, 2008. **38**(10): p. 1083-97.
38. Fan, Q., L. An, and L. Cui, *PfADA2, a Plasmodium falciparum homologue of the transcriptional coactivator ADA2 and its in vivo association with the histone acetyltransferase PfGCN5*. *Gene*, 2004. **336**(2): p. 251-61.
39. Horiuchi, J., et al., *ADA3, a putative transcriptional adaptor, consists of two separable domains and interacts with ADA2 and GCN5 in a trimeric complex*. *Mol Cell Biol*, 1995. **15**(3): p. 1203-9.
40. Ruiz-García, A.B., et al., *Gcn5p is involved in the acetylation of histone H3 in nucleosomes*. *FEBS Lett*, 1997. **403**(2): p. 186-90.
41. Cui, L., et al., *PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in Plasmodium falciparum*. *Eukaryot Cell*, 2007. **6**(7): p. 1219-27.
42. Cui, L. and J. Miao, *Cytotoxic effect of curcumin on malaria parasite Plasmodium falciparum: inhibition of histone acetylation and generation of reactive oxygen species*. *Antimicrob Agents Chemother*, 2007. **51**(2): p. 488-94.
43. Cui, L., et al., *Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro Plasmodium falciparum development*. *Eukaryot Cell*, 2008. **7**(7): p. 1200-10.
44. Pérez-Toledo, K., et al., *Plasmodium falciparum heterochromatin protein 1 binds to trimethylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes*. *Nucleic Acids Res*, 2009. **37**(8): p. 2596-606.
45. Flueck, C., et al., *Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors*. *PLoS Pathog*, 2009. **5**(9): p. e1000569.
46. Voss, T.S., Z. Bozdech, and R. Bártfai, *Epigenetic memory takes center stage in the survival strategy of malaria parasites*. *Curr Opin Microbiol*, 2014. **20**: p. 88-95.
47. Scherf, A., L.M. Figueiredo, and L.H. Freitas-Junior, *Plasmodium telomeres: a pathogen's perspective*. *Curr Opin Microbiol*, 2001. **4**(4): p. 409-14.
48. Broadbent, K.M., et al., *A global transcriptional analysis of Plasmodium falciparum malaria reveals a novel family of telomere-associated lncRNAs*. *Genome Biol*, 2011. **12**(6): p. R56.
49. Sierra-Miranda, M., et al., *Two long non-coding RNAs generated from subtelomeric regions accumulate in a novel perinuclear compartment in Plasmodium falciparum*. *Mol Biochem Parasitol*, 2012. **185**(1): p. 36-47.
50. Flueck, C., et al., *A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology*. *PLoS Pathog*, 2010. **6**(2): p. e1000784.
51. Kyes, S.A., S.M. Kraemer, and J.D. Smith, *Antigenic variation in Plasmodium falciparum: gene organization and regulation of the var multigene family*. *Eukaryot Cell*, 2007. **6**(9): p. 1511-20.
52. Lomberk, G., L. Wallrath, and R. Urrutia, *The Heterochromatin Protein 1 family*. *Genome Biol*, 2006. **7**(7): p. 228.
53. Al-Sady, B., H.D. Madhani, and G.J. Narlikar, *Division of labor between the chromodomains of HP1 and Suv39 methylase enables coordination of heterochromatin spread*. *Mol Cell*, 2013. **51**(1): p. 80-91.
54. Brancucci, N.M., et al., *Heterochromatin protein 1 secures survival and transmission of malaria parasites*. *Cell Host Microbe*, 2014. **16**(2): p. 165-76.
55. Volz, J., et al., *Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in Plasmodium falciparum*. *Int J Parasitol*, 2010. **40**(1): p. 109-21.
56. Jiang, L., et al., *PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum*. *Nature*, 2013. **499**(7457): p. 223-7.

57. Kishore, S.P., J.W. Stiller, and K.W. Deitsch, *Horizontal gene transfer of epigenetic machinery and evolution of parasitism in the malaria parasite Plasmodium falciparum and other apicomplexans*. BMC Evol Biol, 2013. **13**: p. 37.
58. Ukaegbu, U.E., et al., *Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in P. falciparum*. PLoS Pathog, 2014. **10**(1): p. e1003854.
59. Grewal, S.I. and D. Moazed, *Heterochromatin and epigenetic control of gene expression*. Science, 2003. **301**(5634): p. 798-802.
60. Rusche, L.N., A.L. Kirchmaier, and J. Rine, *The establishment, inheritance, and function of silenced chromatin in Saccharomyces cerevisiae*. Annu Rev Biochem, 2003. **72**: p. 481-516.
61. Grewal, S.I. and S. Jia, *Heterochromatin revisited*. Nat Rev Genet, 2007. **8**(1): p. 35-46.
62. Cohen, A.L. and S. Jia, *Noncoding RNAs and the borders of heterochromatin*. Wiley Interdiscip Rev RNA, 2014. **5**(6): p. 835-47.
63. Alabert, C., et al., *Two distinct modes for propagation of histone PTMs across the cell cycle*. Genes Dev, 2015. **29**(6): p. 585-90.
64. Raganathan, K., G. Jih, and D. Moazed, *Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment*. Science, 2015. **348**(6230): p. 1258699.
65. Coleman, B.I., et al., *A Plasmodium falciparum histone deacetylase regulates antigenic variation and gametocyte conversion*. Cell Host Microbe, 2014. **16**(2): p. 177-86.
66. Duraisingh, M.T., et al., *Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum*. Cell, 2005. **121**(1): p. 13-24.
67. Merrick, C.J. and M.T. Duraisingh, *Plasmodium falciparum Sir2: an unusual sirtuin with dual histone deacetylase and ADP-ribosyltransferase activity*. Eukaryot Cell, 2007. **6**(11): p. 2081-91.
68. Tonkin, C.J., et al., *Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum*. PLoS Biol, 2009. **7**(4): p. e84.
69. Goyal, M., et al., *Identification and molecular characterization of an Alba-family protein from human malaria parasite Plasmodium falciparum*. Nucleic Acids Res, 2012. **40**(3): p. 1174-90.
70. Chêne, A., et al., *PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites*. Nucleic Acids Res, 2012. **40**(7): p. 3066-77.
71. Talbert, P.B. and S. Henikoff, *Spreading of silent chromatin: inaction at a distance*. Nat Rev Genet, 2006. **7**(10): p. 793-803.
72. Wang, J., et al., *Chromosome boundary elements and regulation of heterochromatin spreading*. Cell Mol Life Sci, 2014. **71**(24): p. 4841-52.
73. Scherf, A., J.J. Lopez-Rubio, and L. Riviere, *Antigenic variation in Plasmodium falciparum*. Annu Rev Microbiol, 2008. **62**: p. 445-70.
74. Amit-Avraham, I., et al., *Antisense long noncoding RNAs regulate var gene activation in the malaria parasite Plasmodium falciparum*. Proc Natl Acad Sci U S A, 2015. **112**(9): p. E982-91.
75. Jiang, L., et al., *Epigenetic control of the variable expression of a Plasmodium falciparum receptor protein for erythrocyte invasion*. Proc Natl Acad Sci U S A, 2010. **107**(5): p. 2224-9.
76. Stubbs, J., et al., *Molecular mechanism for switching of P. falciparum invasion pathways into human erythrocytes*. Science, 2005. **309**(5739): p. 1384-7.
77. Crowley, V.M., et al., *Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant Plasmodium falciparum genes linked to erythrocyte invasion*. Mol Microbiol, 2011. **80**(2): p. 391-406.
78. Nguitragool, W., et al., *Malaria parasite clag3 genes determine channel-mediated nutrient uptake by infected red blood cells*. Cell, 2011. **145**(5): p. 665-77.
79. Mira-Martínez, S., et al., *Epigenetic switches in clag3 genes mediate blasticidin S resistance in malaria parasites*. Cell Microbiol, 2013. **15**(11): p. 1913-23.

80. Sharma, P., et al., *An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake*. J Biol Chem, 2013. **288**(27): p. 19429-40.
81. Merrick, C.J., et al., *Epigenetic dysregulation of virulence gene expression in severe Plasmodium falciparum malaria*. J Infect Dis, 2012. **205**(10): p. 1593-600.
82. Pillai, A.D., et al., *Solute restriction reveals an essential role for clag3-associated channels in malaria parasite nutrient acquisition*. Mol Pharmacol, 2012. **82**(6): p. 1104-14.
83. Kafsack, B.F., et al., *A transcriptional switch underlies parasite commitment to sexual development in malaria parasites*. Nature, 2014. **507**(7491): p. 248-52.
84. Sinha, A., et al., *A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium*. Nature, 2014. **507**(7491): p. 253-7.
85. Kyes, S.A., et al., *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, 1999. **96**(16): p. 9333-8.
86. Joannin, N., et al., *Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family*. BMC Genomics, 2008. **9**: p. 19.
87. Petter, M., et al., *Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns*. Mol Biochem Parasitol, 2007. **156**(1): p. 51-61.
88. Petter, M., I. Bonow, and M.Q. Klinkert, *Diverse expression patterns of subgroups of the rif multigene family during Plasmodium falciparum gametocytogenesis*. PLoS One, 2008. **3**(11): p. e3779.
89. Khattab, A., et al., *Plasmodium falciparum variant STEVOR antigens are expressed in merozoites and possibly associated with erythrocyte invasion*. Malar J, 2008. **7**: p. 137.
90. Blythe, J.E., et al., *Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites*. Infect Immun, 2008. **76**(7): p. 3329-36.
91. Niang, M., X. Yan Yam, and P.R. Preiser, *The Plasmodium falciparum STEVOR multigene family mediates antigenic variation of the infected erythrocyte*. PLoS Pathog, 2009. **5**(2): p. e1000307.
92. Sam-Yellowe, T.Y., et al., *A Plasmodium gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling*. Genome Res, 2004. **14**(6): p. 1052-9.
93. Chookajorn, T., et al., *Epigenetic memory at malaria virulence genes*. Proc Natl Acad Sci U S A, 2007. **104**(3): p. 899-902.
94. Roberts, D.J., et al., *Rapid switching to multiple antigenic and adhesive phenotypes in malaria*. Nature, 1992. **357**(6380): p. 689-92.
95. Scherf, A., et al., *Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum*. EMBO J, 1998. **17**(18): p. 5418-26.
96. Miller, L.H., et al., *The pathogenic basis of malaria*. Nature, 2002. **415**(6872): p. 673-9.
97. Dzikowski, R., T.J. Templeton, and K. Deitsch, *Variant antigen gene expression in malaria*. Cell Microbiol, 2006. **8**(9): p. 1371-81.
98. Schieck, E., et al., *Nuclear run-on analysis of var gene expression in Plasmodium falciparum*. Mol Biochem Parasitol, 2007. **153**(2): p. 207-12.
99. Kyes, S.A., et al., *A well-conserved Plasmodium falciparum var gene shows an unusual stage-specific transcript pattern*. Mol Microbiol, 2003. **48**(5): p. 1339-48.
100. Guizetti, J. and A. Scherf, *Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum*. Cell Microbiol, 2013. **15**(5): p. 718-26.
101. Volz, J.C., et al., *PfSET10, a Plasmodium falciparum methyltransferase, maintains the active var gene in a poised state during parasite division*. Cell Host Microbe, 2012. **11**(1): p. 7-18.

102. Ghosh, S., et al., *NAD: a master regulator of transcription*. *Biochim Biophys Acta*, 2010. **1799**(10-12): p. 681-93.
103. Naimi, M., C. Arous, and E. Van Obberghen, *Energetic cell sensors: a key to metabolic homeostasis*. *Trends Endocrinol Metab*, 2010. **21**(2): p. 75-82.
104. Rosenberg, E., et al., *Differential, positional-dependent transcriptional response of antigenic variation (var) genes to biological stress in Plasmodium falciparum*. *PLoS One*, 2009. **4**(9): p. e6991.
105. Dzikowski, R. and K.W. Deitsch, *Genetics of antigenic variation in Plasmodium falciparum*. *Curr Genet*, 2009. **55**(2): p. 103-10.
106. Kirkman, L.A. and K.W. Deitsch, *Antigenic variation and the generation of diversity in malaria parasites*. *Curr Opin Microbiol*, 2012. **15**(4): p. 456-62.
107. Calderwood, M.S., et al., *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*, 2003. **278**(36): p. 34125-32.
108. Dzikowski, R., M. Frank, and K. Deitsch, *Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production*. *PLoS Pathog*, 2006. **2**(3): p. e22.
109. Frank, M., et al., *Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite Plasmodium falciparum*. *J Biol Chem*, 2006. **281**(15): p. 9942-52.
110. Voss, T.S., et al., *A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria*. *Nature*, 2006. **439**(7079): p. 1004-8.
111. Swamy, L., B. Amulic, and K.W. Deitsch, *Plasmodium falciparum var gene silencing is determined by cis DNA elements that form stable and heritable interactions*. *Eukaryot Cell*, 2011. **10**(4): p. 530-9.
112. Avraham, I., J. Schreier, and R. Dzikowski, *Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite Plasmodium falciparum*. *Proc Natl Acad Sci U S A*, 2012. **109**(52): p. E3678-86.
113. Brancucci, N.M., et al., *A var gene upstream element controls protein synthesis at the level of translation initiation in Plasmodium falciparum*. *PLoS One*, 2014. **9**(6): p. e100183.
114. Su, X.Z., et al., *The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes*. *Cell*, 1995. **82**(1): p. 89-100.
115. Epp, C., et al., *Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite Plasmodium falciparum*. *RNA*, 2009. **15**(1): p. 116-27.
116. Kyes, S., et al., *Stage-specific merozoite surface protein 2 antisense transcripts in Plasmodium falciparum*. *Mol Biochem Parasitol*, 2002. **123**(1): p. 79-83.
117. Petter, M., et al., *Expression of P. falciparum var genes involves exchange of the histone variant H2A.Z at the promoter*. *PLoS Pathog*, 2011. **7**(2): p. e1001292.
118. Lopez-Rubio, J.J., et al., *5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites*. *Mol Microbiol*, 2007. **66**(6): p. 1296-305.
119. Baton, L.A. and L.C. Ranford-Cartwright, *Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito*. *Trends Parasitol*, 2005. **21**(12): p. 573-80.
120. Balaji, S., et al., *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*, 2005. **33**(13): p. 3994-4006.
121. Yamada, T., et al., *The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast*. *Mol Cell*, 2005. **20**(2): p. 173-85.

122. Zhang, C.L., T.A. McKinsey, and E.N. Olson, *Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation*. *Mol Cell Biol*, 2002. **22**(20): p. 7302-12.
123. Liu, Z., J. Miao, and L. Cui, *Gametocytogenesis in malaria parasite: commitment, development and regulation*. *Future Microbiol*, 2011. **6**(11): p. 1351-69.
124. Kooij, T.W. and K. Matuschewski, *Triggers and tricks of Plasmodium sexual development*. *Curr Opin Microbiol*, 2007. **10**(6): p. 547-53.
125. Baker, D.A., *Malaria gametocytogenesis*. *Mol Biochem Parasitol*, 2010. **172**(2): p. 57-65.
126. Mann, B.S., et al., *FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma*. *Oncologist*, 2007. **12**(10): p. 1247-52.
127. Campas-Moya, C., *Romidepsin for the treatment of cutaneous T-cell lymphoma*. *Drugs Today (Barc)*, 2009. **45**(11): p. 787-95.
128. Prusty, D., et al., *Nicotinamide inhibits Plasmodium falciparum Sir2 activity in vitro and parasite growth*. *FEMS Microbiol Lett*, 2008. **282**(2): p. 266-72.
129. Vathsala, P.G., et al., *Curcumin-arteether combination therapy of Plasmodium berghei-infected mice prevents recrudescence through immunomodulation*. *PLoS One*, 2012. **7**(1): p. e29442.
130. Sundriyal, S., et al., *Development of diaminoquinazoline histone lysine methyltransferase inhibitors as potent blood-stage antimalarial compounds*. *ChemMedChem*, 2014. **9**(10): p. 2360-73.

CHAPTER 3

H3.3 demarcates GC-rich coding and subtelomeric regions and serves as potential memory mark for virulence gene expression in *Plasmodium falciparum*

Sabine Anne-Kristin Fraschka¹, Rob Wilhelmus Maria Henderson^{1,#}, and Richárd Bártfai^{1,*}

Adapted from
Sci Rep 6, 31965 (2016)

¹ Department of Molecular Biology, Radboud University, Nijmegen, the Netherlands

* Corresponding author; r.bartfai@science.ru.nl

Current address: TropIQ Health Sciences, Nijmegen, the Netherlands

Contribution:

Designed and performed experiments, analysed data, prepared illustrations and wrote the manuscript.

Abstract

Histones, by packaging and organizing the DNA into chromatin, serve as essential building blocks for eukaryotic life. The basic structure of the chromatin is established by four canonical histones (H2A, H2B, H3 and H4), while histone variants are more commonly utilized to alter the properties of specific chromatin domains. H3.3, a variant of histone H3, was found to have diverse localization patterns and functions across species but has been rather poorly studied in protists. Here we present the first genome-wide analysis of H3.3 in the malaria-causing, apicomplexan parasite, *P. falciparum*, which revealed a complex occupancy profile consisting of conserved and parasite-specific features. In contrast to other histone variants, *Pf*H3.3 primarily demarcates euchromatic coding and subtelomeric repetitive sequences. Stable occupancy of *Pf*H3.3 in these regions is largely uncoupled from the transcriptional activity and appears to be primarily dependent on the GC-content of the underlying DNA. Importantly, *Pf*H3.3 specifically marks the promoter region of an active and poised, but not inactive antigenic variation (*var*) gene, thereby potentially contributing to immune evasion. Collectively, our data suggest that *Pf*H3.3, together with other histone variants, indexes the *P. falciparum* genome to functionally distinct domains and contribute to a key survival strategy of this deadly pathogen.

Introduction

In all eukaryotes, chromatin enables high-degree compaction of the genetic material while allowing access to the DNA for essential cellular processes such as transcription, replication, DNA damage response and repair. Canonical histones, namely H2A, H2B, H3 and H4, are mainly synthesized and assembled into nucleosomes in a replication-coupled manner to form the basic layout of the newly synthesized chromatin [1-3]. In contrast, histone variants (non-allelic histone isoforms [1,3]), and most posttranslational histone modifications (e.g. acetylation, methylation [4]) are locally incorporated into chromatin in a replication-independent fashion to alter its properties and function.

Given their essential function, histones are among the most conserved proteins and their evolutionary origins can be traced back to archaea, where one or two histone-like proteins are present [5]. In Eukarya, the histone family has expanded into canonical histones, the linker histone H1 and a great variety of histone variants for H2A, H2B and H3 [6]. Strikingly, the amino acid sequences of the four eukaryotic canonical histones and of many histone variants are extremely similar among distantly related species, although histone variants have emerged by convergent evolution [7-9]. The requirement of convergent evolution of histone variants indicates a universal theme in chromatin regulation; and indeed, some histone variants display universal functions. Histone H3 variant CenH3, for instance, demarcates the centromere and functions in chromosome

segregation [10-13]. The function of H3.3, another member of the H3 histone family is however more diverse among the species.

Phylogenetically earlier organisms, such as *Saccharomyces cerevisiae* and the algal protist, *Cyanidioschyzon merolae*, possess only one form of non-centromeric H3 that is needed for both replication-coupled and replication-independent assembly [11,14]. This single form corresponds to variant H3.3 of higher eukaryotes, which suggests that canonical forms of H3 evolved recurrently in eukaryotic evolution, most likely by divergence from H3.3-like forms [15,16]. In most eukaryotes H3.3 differs from H3 only by few amino acid substitutions; for example, by four in most mammals and plants and by 16 in the ciliate *Tetrahymena thermophila* (Superphylum Alveolata) [11,17]. Yet, these few amino acid substitutions are sufficient to determine H3.3 and H3 chaperone selectivity and their respective nucleosome deposition pathways [18-21].

Genome-wide studies performed mainly in animals, but also in *Arabidopsis thaliana* revealed that H3.3 is mainly deposited into the coding sequence of transcribed genes, promoters of active and inactive genes, transcription start and end sites as well as further gene regulatory elements in euchromatic regions [8,16,22,23]. H3.3 is incorporated into transcription start sites and coding regions of genes in a transcription-coupled manner [24,25]. However, enrichment of H3.3 has also been observed in regions of the genome that are presumed to be transcriptionally silent [26-30]. For example, in embryonic stem cells H3.3 contributes to the maintenance of the condensed chromatin state of telomeres [26,28,31,32], while it appears to be absent from the telomeres of *A. thaliana* [33,34]. Surprisingly, deletions of H3.3 genes are not always lethal, probably due to compensation by other H3 genes. For instance, increased H3 expression compensates for H3.3 lacking in *D. melanogaster* [35] and overexpression of H3.3 makes canonical H3 dispensable in *T. thermophila* [17]. In contrast, H3.3 disruption in *D. melanogaster* and in *T. thermophila* results in meiotic defects [17,36,37]. Furthermore, in mammals, H3.3 is not only required for reproduction but also for early development [32,38] and mutation in H3.3 can lead to various pediatric cancer types [39,40]. Hence, the function of H3 variants has been extensively studied in animals, fungi and plants - where it appears to have several conserved, but also specialized function.

In this article, we focus on the localization and role of H3.3 in the apicomplexan parasite, *Plasmodium falciparum*. Besides the canonical histones four histone variants are encoded in the *Plasmodium* genome, i.e. PfH2A.Z, PfH2B.Z, PfCenH3 and PfH3.3 [41, 42]. Interestingly, *P. falciparum* has an extremely AT-rich genome (on average ~80% adenine and thymine bases) where different genomic regions have distinctly different base composition. Earlier we observed an intriguing correlation between the base composition of these genomic regions and the incorporation of various histone variants [41]. Centromeres are formed via incorporation of PfCenH3 into a ~2kb long extremely AT-rich (~97% AT) sequence present once on every chromosome [13]. On the other end

of the spectrum are H3K9me3 and HP1-marked heterochromatic regions with distinctly higher GC content (~73% AT). These regions are present at the chromosome ends and some chromosome internal islands, and so far, have found to contain mainly canonical nucleosomes [43-46]. Within the euchromatic domain intergenic regions having distinctly higher AT-content (~87% AT) compared to coding sequences (~78% AT), are demarcated by double-variant nucleosomes containing *PfH2A.Z* and the *Plasmodium*-specific variant *PfH2B.Z* [46-49]. Collectively these observations support a model in which histone variants help indexing the *Plasmodium* epigenome into functionally distinct domains and the rough blueprint of the epigenome is sketched by the AT-content of the underlying sequence. Here, in further support of this model, we show that *PfH3.3* demarcates euchromatic coding sequences and subtelomeric repeat regions in a GC-content coupled-manner. Furthermore, we observe that *PfH3.3* specifically associates with the promoter region of the single expressed or poised - but not inactive - antigenic variation gene (*var* gene). This suggests that *PfH3.3* might contribute to the epigenetic memory of *var* gene expression and hence could underlie a major defense mechanism of this deadly human pathogen.

Results

***PfH3.3* primarily localizes to GC-rich euchromatic coding sequences and subtelomeric repetitive regions**

Plasmodium falciparum's histone variant *PfH3.3* (PF3D7_0617900) and its single copy canonical counterpart *PfH3* (PF3D7_0610400) both encode for a 136 amino-acid protein. They only differ in eight amino acids including an Apicomplexan specific feature for *PfH3.3*, namely the substitution of the conserved amino acids RY with KF at position 54 and 55 [42, 50]. With an identity of 94% they are hence difficult to investigate individually employing antibodies.

We circumvented this problem by generating a *P. falciparum* 3D7 parasite line that expresses an episomal Ty1-tagged copy of *PfH3.3* under the control of the chloroquine resistance transporter promoter (similar to pARL-1a-Ty1-H2A in ref. [47]). Additionally, to avoid any artefacts due to episomal (over)expression of the tagged protein, we engineered a construct (pHH1-*PfH3.3*-Ty1) that, after single crossover integration, resulted in a *P. falciparum* NF54-DCJ parasite line [51] expressing *PfH3.3*-Ty1 under control of the endogenous promoter (Fig. S1). Notably, these tagged lines show a very similar multiplication rate (data not shown) and expression pattern (Fig. S2) to wild type parasites, suggesting that the tag does not interfere with normal function of *PfH3.3*.

To determine the genome-wide localization of *PfH3.3*, we performed chromatin immunoprecipitation of mono-nucleosomes using an anti-Ty1 antibody followed by next generation sequencing of the immunoprecipitated DNA (ChIP-seq). First, we performed

a test ChIP-seq on native chromatin isolated from the episomal Ty1-*PfH3.3* line for two intraerythrocytic stages (30 and 40 hours post-invasion) in duplicates (Fig. S3). The respective ChIP input samples were sequenced as control. For detailed analysis we performed ChIP-seq on formaldehyde cross-linked chromatin from the endogenously *PfH3.3*-Ty1 line for four intraerythrocytic stages (10, 20, 30 and 40 hpi). Additionally, we performed ChIP-seq using an anti-H3 antibody that recognizes the C-terminal end of both *PfH3.3* and *PfH3* (referred to as H3core control) and RNA-seq on the same four parasite populations. The resulting *PfH3.3* ChIP-seq data were normalized either to input or H3core ChIP-seq data to correct for any biases due to differential cross-linking, PCR amplification efficiency and reduced mapability of the highly AT-rich sequences, typical to the *P. falciparum* genome [47]. The normalized *PfH3.3* ChIP-seq data from episomal and endogenously *PfH3.3*-tagged parasites displayed a highly similar localization pattern (Fig. S3A) and good genome-wide correlation (Fig. S3B; $r = 0.71$ – 0.86), despite numerous technical differences between the two experiments (N-versus

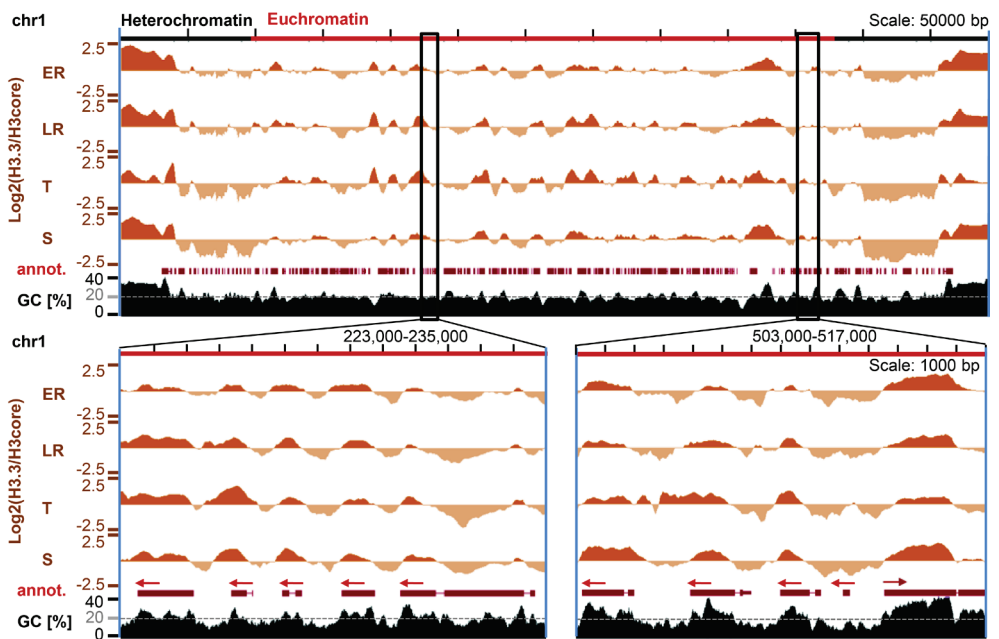


Figure 1: *PfH3.3*-Ty1 ChIP-seq profiles from different stages of intraerythrocytic development. ChIP-seq ratio tracks displaying relative enrichment of the endogenous *PfH3.3*-Ty1 compared to H3core (antibody capturing histone *PfH3* and *PfH3.3*) over entire chromosome 1 (upper panel) or two smaller sections of it (lower panels). Red arrows indicate the directionality of the respective gene (“annot.”). Stages: ER: early rings (10 hpi), LR: late rings (20 hpi), T: trophozoites (30 hpi), S: schizonts (40 hpi). Heterochromatic regions are defined based on HP1 occupancy in an earlier study [43].

C-terminal tagging, cross-linked versus native ChIP, different parasite lines, different library preparation methods) (Fig. S3). Hence, our ChIP-seq analysis provides a robust measure of *PfH3.3* localization, which we analyzed in detail using the data from the endogenously tagged line.

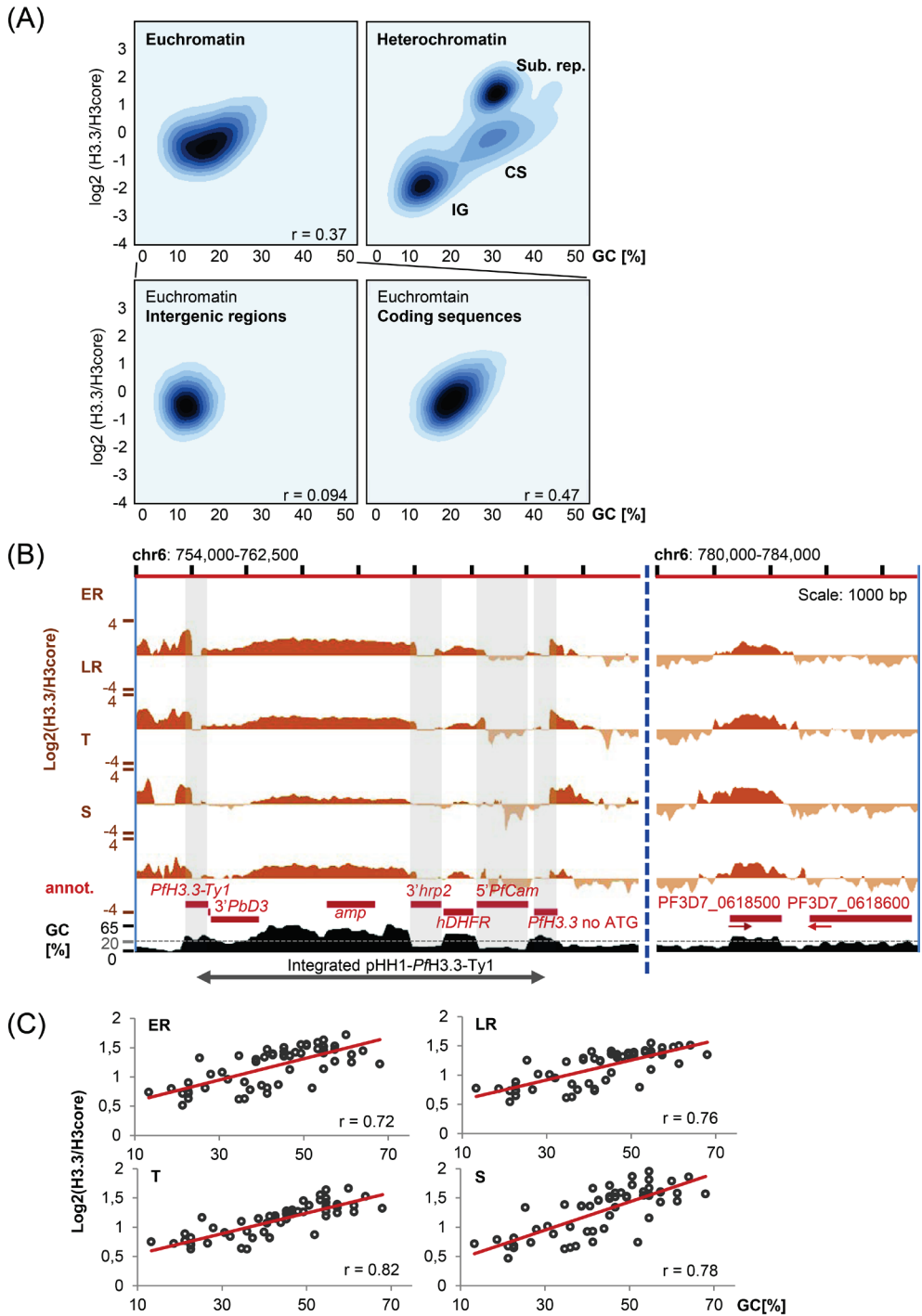
PfH3.3 is mainly enriched in euchromatic coding sequences and is generally absent in the heterochromatic domain with the exception of its enrichment over GC-rich subtelomeric repetitive sequences (Fig. 1). In these regions the occupancy of *PfH3.3* is largely invariable through the intraerythrocytic stages. Additionally, *PfH3.3* occupancy shows dynamic changes in euchromatic intergenic regions (Fig. 1). Centromeres are depleted of *PfH3.3* and *PfH3* (Fig. S4), as they are occupied by the third H3 variant, *PfCenH3* [13].

***PfH3.3* is demarcating the *P. falciparum* genome and exogenous DNA sequences in a GC-content coupled manner**

Visual inspection of the genome-wide *PfH3.3* profiles led to the observation that *PfH3.3* is mainly incorporated into GC-rich sequences (Fig. 1), including euchromatic coding sequences with an average GC-content of ~22% and the subtelomeric repetitive sequences with an GC-content ~32%. While at these GC-rich regions *PfH3.3* occupancy is invariable across the stages, AT-rich intergenic regions (average 14% GC) are only temporarily marked or completely depleted of *PfH3.3* (Fig. 1). To further investigate the correlation between GC-content and *PfH3.3* occupancy on a genome-wide level, we *in silico* subdivided the genome into 150 bp windows and calculated the GC percentage and *PfH3.3* coverage for each window. In line with our visual observation, GC-poor intergenic regions show the lowest *PfH3.3* coverage; whereas in euchromatic coding sequences *PfH3.3* levels exhibited clear correlation with the GC-content (Fig. 2A and Fig. S5A for data from all four stages).

Figure 2: *PfH3.3* localizes to GC-rich regions within the *P. falciparum* genome and exogenous DNA.

(A) Density plots depicting *PfH3.3* levels in relation to GC-content. *PfH3.3* levels and GC-content were calculated genome-wide per 150 bp windows. Data depicted are generated from schizont stages. IG: intergenic region, CS: coding sequence, Sub. Rep.: subtelomeric repetitive region. Pearson correlation values are displayed at the lower right corner of the graphs. (B) Screen shot from *PfH3.3* ChIP-seq ratio tracks encompassing the integrated pHH1-*PfH3.3*-Ty1 plasmid sequence (left) and the sequence of two nearby genes (right). Grey blocks indicate non-unique *P. falciparum* sequences within the pHH1-*PfH3.3*-Ty1 plasmid, which were excluded from the analysis in (C). (C) Scatter plot of GC percentage and *PfH3.3* occupancy over unique 75 bp windows of the integrated pHH1-*PfH3.3*-Ty1 plasmid. Stages: ER: early rings (10 hpi); LR: late rings (20 hpi), T: trophozoites (30 hpi), S: schizonts (40 hpi). Pearson correlation values are displayed at the lower right corner of the graphs.



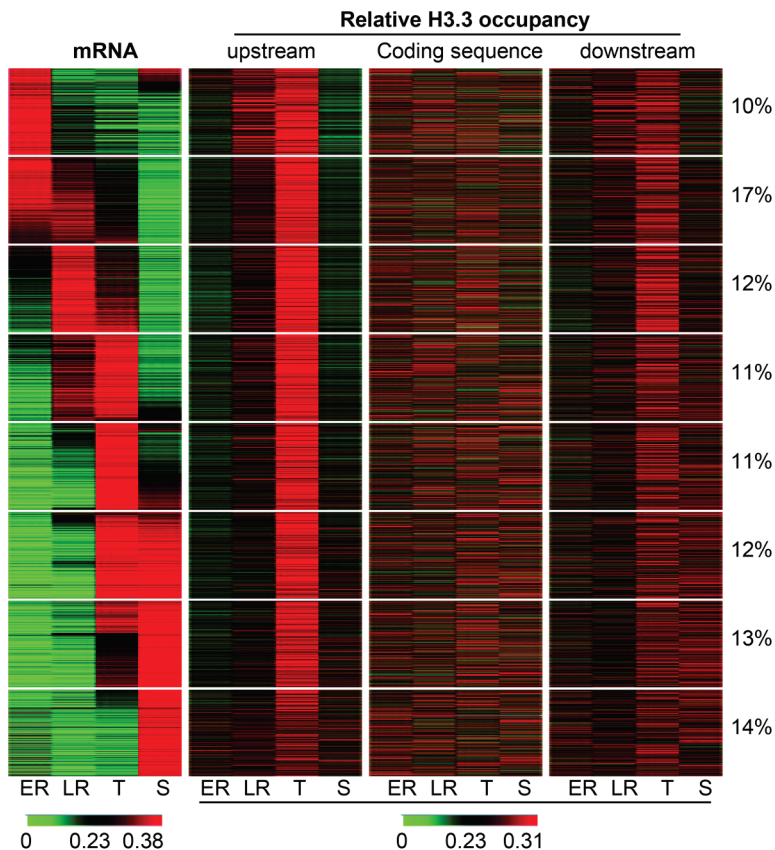


Figure 3: *PfH3.3* dynamically occupies intergenic regions predominantly during trophozoite stage.

Heatmap represents relative mRNA abundance of 3800 genes (heterochromatic and genes with low mRNA abundance have been excluded) and relative *PfH3.3* occupancy in upstream, coding and downstream region throughout intraerythrocytic development. Color-scale depicts the relative mRNA abundance in relation to the sum of the RPKM values in all four stages or the relative *PfH3.3* occupancy in proportion to the sum of occupancy values in all four stages (green – low, red – high). 8 groups of co-expressed genes were identified using k-means clustering. The percentage of genes belonging to each cluster is displayed next to the heatmap. Stages: ER: early rings (10 hpi), LR: late rings (20 hpi), T: trophozoites (30 hpi), S: schizonts (40 hpi).

Importantly, similar correlation was also observed in the ChIP-seq data obtained from the episomally tagged line (Fig. S5B). Subtelomeric GC-rich repetitive sequences display the highest *PfH3.3* levels. Thus, our genome-wide computational analysis is consistent with the initial observation.

Additionally, we tested whether *PfH3.3* is also incorporated into an exogenous DNA in a GC-dependent manner. Therefore, the non-*P. falciparum*-specific sequences within the integrated pHH1-Ty1-*PfH3.3* plasmid were examined for their GC-content and *PfH3.3*

occupancy. Also, here a clear correlation of incorporated *PfH3.3* and GC-content can be observed (Fig. 2B, C). As these exogenous sequences are unlikely to contain sequence-element for recruitment or detainment of *PfH3.3* or its chaperone, these findings support a base-content-driven genome marking by *PfH3.3*.

***PfH3.3* dynamically occupies intergenic regions predominantly during trophozoite development**

The dynamic occupancy of *H3.3* in some intergenic regions prompted us to test whether correlation exists between temporal gene activity and the dynamics of *PfH3.3* occupancy. Genes were clustered according to their relative mRNA abundance (matching RNA-seq data) during the intraerythrocytic cycle and matched with the respective *PfH3.3* occupancy values (in proportion of the sum of the occupancy values) for upstream regions, coding sequences and downstream regions (Fig. 3). Whereas *PfH3.3* coverage in coding sequences showed no obvious stage-dependent or transcription-coupled dynamics, mainly upstream regions but to some extent also downstream regions showed dynamic *PfH3.3* enrichment during the intraerythrocytic cycle. Interestingly, most intergenic regions have the highest relative *PfH3.3* occupancy during the trophozoite stage largely independent of their transcriptional dynamics.

To investigate *PfH3.3* enrichment in relation to gene expression level, we plotted the mRNA abundance (RPKM values) for each gene against *PfH3.3* occupancy in coding, upstream or downstream regions for each stage separately (Fig. S6). This analysis also did not reveal clear correlations between transcription and *PfH3.3* occupancy. Only in 5' intergenic regions and in particular at the trophozoite stage were moderate correlations observed.

In summary, *PfH3.3* is temporarily enriched in AT-rich, 5' intergenic regions mainly at the trophozoite stage. While this enrichment is somewhat more apparent for very highly expressed genes, and least apparent for (nearly) silent genes (Fig. S6), it does not exhibit clear correlation to gene expression dynamics (Fig. 3).

***PfH3.3* marks the promoter of the active and poised *var* gene**

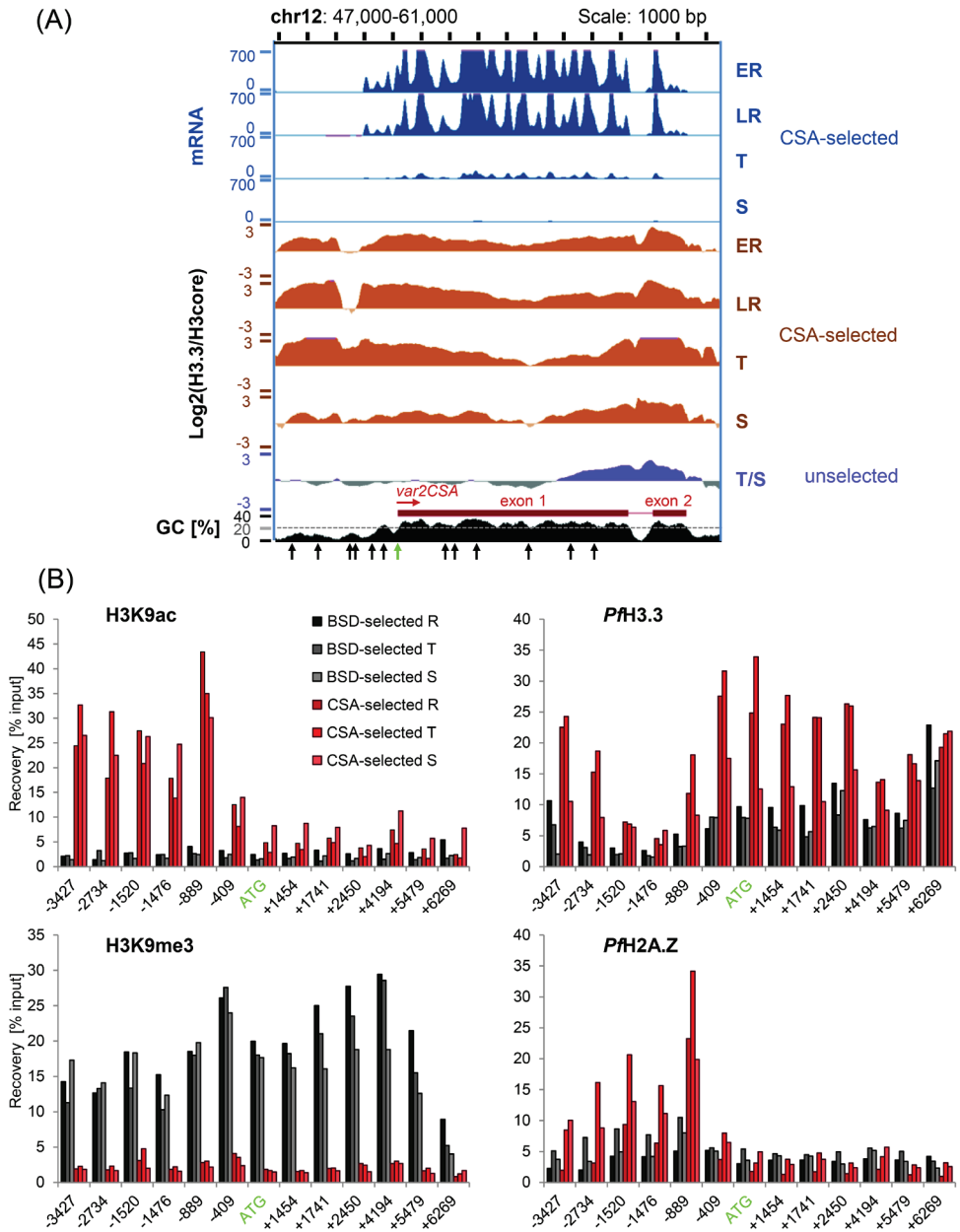
Persistence of *P. falciparum* within the human bloodstream primarily depends on the *var* multigene family, encoding for a highly variable cytoadherence protein called *P. falciparum* erythrocyte membrane protein 1 (*PfEMP1*) [52-55]. Only one of the ~60 *var* gene family members is expressed in any given parasite while other family members remain transcriptionally silenced. Importantly, *var* genes are expressed in a clonally variant fashion [56] as the same *var* gene is reactivated over several generation by a yet elusive mechanism of epigenetic memory [57, 58].

In order to investigate the role of *PfH3.3* in *var* gene regulation, we specifically looked at the *var2CSA* locus of CSA-selected (10, 20, 30, 40 hpi) (Fig. S7A) and non-selected (mix

of trophozoites and schizonts) parasites. Interestingly, *PfH3.3* is enriched in the promoter as well as in the coding region of the active as well as poised *var2CSA* gene (Fig. 4A) (Note that the *var* gene is only transcribed during rings stages while at trophozoites and schizonts stages it is poised, yet marked for reactivation in the next generation). The *var2CSA* gene in non-CSA selected parasites, however, is clearly less enriched in *PfH3.3* and only shows *PfH3.3* coverage at the *var2CSA* intron and the second exon (regions which are known to be transcribed from the *var* intronic promoter [55, 59-63]) (Fig. 4A). To ensure that this observation was not an artefact due to the comparison of two independently generated data sets (as the CSA-selected and non-selected parasites originate from independent experiments) we performed a ChIP-qPCR experiment that enabled us to directly compare the *var2CSA* locus of CSA-selected and non-selected parasites. First, a culture of endogenously *PfH3.3*-Ty1 expressing NF54-DCJ *P. falciparum* parasites was selected for *var2CSA* expression via chondroitin sulfate affinity purification (*var2CSA* positive line). Subsequently, half of the culture was treated with blasticidin (2.5 mM), which forced a switch to another *var* promoter (PF3D7_0223500) driving expression of the BSD resistance cassette [51]. Successful selection for parasites expressing *var2CSA* or PF3D7_0223500 was confirmed by RT-qPCR (Fig. S7). Subsequently, ChIP-qPCR was performed for the active histone modification H3K9ac, the inactive histone modification H3K9me3, the histone variant *PfH2A.Z* and for *PfH3.3* for three different time points (15, 25 and 40 hpi). The switch from *var2CSA* expression to PF3D7_0223500 was confirmed at the epigenetic level: CSA-selected parasites showed high H3K9ac recoveries for the whole promoter region of the *var2CSA* gene and low H3K9me3 recoveries for the promoter as well as coding sequences (Fig. 4B). BSD-selected parasites, however, showed the opposite pattern (Fig. 4B). As also reported earlier, only the CSA-selected parasite line showed high occupancy of *PfH2A.Z* close to the promoter start site 1500 bp upstream of the ATG [49]. Incorporation of *PfH3.3* into the promoter region of the *var2CSA* gene was clearly increased when the gene was active compared to its silenced state validating the observation made based on the ChIP-seq data. In summary, *PfH3.3* stably occupies the promoter region and coding sequence of the active *var* gene but is evidently less incorporated into the promoter and coding sequence of silenced *var* genes.

Figure 4: *PfH3.3* marks the promoter of the active and poised, but not the inactive *var* gene.

(A) RNA-seq and ChIP-seq ratio tracks displaying *PfH3.3* enrichment over H3core over the *var2CSA* locus. Data was obtained from *var2CSA*-selected parasites at four intraerythrocytic stages (ER, LR, T, S) as well as from unselected mixed stage parasites (mainly containing trophozoites and schizonts, T/S). Arrows indicate the approximate position of primers used in (B). (B) Bar graphs displaying ChIP-qPCR data for H3K9ac (euchromatic mark) and H3K9me3 (heterochromatic mark) marking as well as *PfH3.3* and *PfH2A.Z* occupancy over the *var2CSA* locus in CSA-selected (red) or - BSD-selected (grey) parasites. After affinity-based selection for *var2CSA*-expressing parasites



half of the parasite population was forced to express PF3D7_0223500 by blasticidin treatment (NF54-DCJ strain: first exon of PF3D7_0223500 is replaced by a blasticidin resistance cassette). Numbers indicate the position of the forward primer relative to the ATG. Stages (R: rings (15 hpi); T: trophozoites (25 hpi); S: schizonts (40 hpi)) are depicted by different shades.

Discussion

Here we present the first genome-wide analysis of histone variant H3.3 in the apicomplexan parasite, *P. falciparum*. Interestingly, *PfH3.3* has a distinct localization compared to other variants. While *PfH2A.Z* and *PfH2B.Z* [48] as well as *PfCenH3* [13] appear to occupy AT-rich sequences, *PfH3.3* is most commonly found in GC-rich coding sequences and the subtelomeric repetitive regions (Fig. 1) and only temporarily occupies AT-richer intergenic regions (Fig. 3). In many animals, H3.3 is also found in coding regions, where it is incorporated during transcriptional elongation [24, 25]. In *P. falciparum*, however, we could not observe a clear correlation between *PfH3.3* occupancy in euchromatic coding sequences and the steady-state mRNA level of the corresponding genes (Fig. 3, Fig. S6). Instead we find that most coding sequences are constantly marked by *PfH3.3* throughout the intraerythrocytic cycle and we observed a positive correlation between *PfH3.3* occupancy and the GC-content of the underlying DNA (Fig. 2A). This correlation is even more striking over exogenous DNA sequences that do not contain any *P. falciparum* “compatible” gene regulatory elements (Fig. 2B/C). Furthermore, although transcriptionally silent heterochromatic regions appear to have a generally lower *PfH3.3* occupancy, GC-richer coding sequences are yet more likely to contain some *PfH3.3* (Fig. 2A). These observations together suggest that *PfH3.3* incorporation to coding regions does not (necessarily) require transcriptional activity and *PfH3.3* occupancy is rather influenced by the GC-content of DNA sequence.

Next to constant occupancy of *PfH3.3* at euchromatic intergenic regions we observe clear *PfH3.3* marking at subtelomeric repeat sequences. In mammals, *PfH3.3* is also deposited at repressive chromatin regions and in mouse embryonic cells additionally at the telomeres [26, 28, 31, 64], suggesting that *PfH3.3* might serve a conserved telomere associated function. However, unlike in mouse embryonic stem cells [31, 64], *PfH3.3* occupancy seems to be uncoupled from K9me3 marking. In fact, *PfH3.3* appears to be specifically depleted from most heterochromatic regions - marked by K9me3 and HP1 in *P. falciparum* - except for the subtelomeric repetitive regions (Figs 1 and 2). Notably, these latter regions have the highest density of G quadruplex motifs in *P. falciparum* genome [65]. Since, earlier work in mouse established a potential link between G quadruplex sequences and H3.3 occupancy at the telomeres [66], it is conceivable that a similar link might underlie the subtelomeric H3.3 localization in *P. falciparum*. However, it is even more straightforward to think these regions have high *PfH3.3* occupancy simply because they have a distinctly higher GC content (Fig. 2A), providing further support for the above model.

In addition to the GC-based correlation of *PfH3.3* localization to coding and subtelomeric sequences, we observe dynamic incorporation of *PfH3.3* into AT-rich intergenic regions, almost exclusively at the trophozoite stages. This stage is peculiar in two ways: i) this is the stage when extensive replication of the parasites genome begins;

ii) this is the transcriptionally most active stage. Notably, in contrast to many other eukaryotic cells, cell division during the intraerythrocytic cycle of *P. falciparum* does not only double the DNA content, but rather results in 8 to 32 genome copies. Given that an enormous amount of histones needed to support the synthesis of the new chromatin it is conceivable that both H3 variants (*PfH3* and *PfH3.3*) need to be incorporated during replication to the newly synthesized chromatin, after which *PfH3.3* might be preferentially retained at GC-rich sequences. A similar phenomenon is observed in mouse spermatocytes, when almost all histones are replaced by protamine to form a highly compact paternal chromatin, nucleosomes that are retained at unmethylated CpG-rich sequences, primarily contain H3.3 [67]. Alternatively, the stage-specific incorporation of *PfH3.3* to promoter regions could be explained by temporal transcriptional activity, typical to intraerythrocytic stages of malaria parasites (Fig. 3). In fact, in many other organisms H2A.Z and H3.3 containing nucleosomes are located around the transcriptional start site (for review see ref. [68]). However, while our analysis revealed a moderate correlation between mRNA abundance and *PfH3.3* enrichment in promoter regions in particular at the trophozoite stage (Fig. S6) *PfH3.3* occupancy profiles in promoter regions are largely uncoupled from the dynamics of their transcriptional activity (Fig. 3). Accordingly, while we cannot completely exclude the possibility that *PfH3.3* is incorporated to some promoter regions in a transcription coupled manner, the extensive incorporation of *PfH3.3* to intergenic regions at the trophozoite stage is perhaps better explained by the replicative activity of this stage.

In mammalian cells, the different distribution patterns of H3.3 compared to H3 is achieved by histone- and site-specific chaperones. While H3.3 is deposited at euchromatic regions by the chaperone Histone regulator A (HIRA) replication-independently [20, 69], the death-associated protein (DAXX) together with the α -thalassemia and/or mental retardation X-linked syndrome protein (ATRX) mediates H3.3 deposition in pericentric heterochromatin and telomeric regions [26, 27]. H3 deposition, however, is mediated by the chaperone CAF-1 replication-dependently [20]. Recognition of H3 and H3.3 by these chaperons is mediated by amino acid residues at positions such as 87 and 90 [18, 19, 70], which also differ in *PfH3* and *PfH3.3* [18, 19, 42, 50, 70] and might support differential recognition by different chaperon proteins. Thus far, however, only one *P. falciparum* protein with weak homology to both CAF1 and HIRA has been identified (PF3D7_0501800). Therefore, additional experiments will have to be performed to determine whether this protein supports the incorporation of both *PfH3* and *PfH3.3* during replication or yet unidentified *PfH3/PfH3.3*-specific chaperons exist that ensure differential genomic incorporation of *PfH3* and *PfH3.3*.

Of particular interest is the specific *PfH3.3* incorporation to the promoter region of the active, but not inactive *var* gene (Fig. 4). This is similar to the earlier reported specific incorporation of the *PfH2A.Z/PfH2B.Z* double-variant nucleosomes around the

transcription start site of the active *var* gene during the ring stage, when this gene is being actively transcribed [49]. Furthermore, the active *var* promoter is marked by two active marks (H3K9ac and H3K4me3), while the poised *var* gene carries H3K9ac and H3K4me2 [58, 71, 72]. Interestingly, unlike *PfH2A.Z/PfH2B.Z*, *PfH3.3* is retained at this AT-rich *var* promoter even after active transcription has ceased (so called poised state). This raises the exciting possibility that *PfH3.3* may contribute to the epigenetic memory of *var* gene expression [57, 58], and enables the parasites to express the same *var* gene over several generations. In line with such hypothesis, nuclear transplantation experiments in *Xenopus laevis* showed that H3.3 mediates the epigenetic memory of certain genes; and H3.3 overexpression in nuclear transplant embryos even enhances this effect [73-75]. Furthermore, H3.3 enrichment at promoters has also been observed for inactive genes, which are potentially in a poised state [25, 76, 77] as we observe it for the poised *var2CSA* gene during the second half of the intraerythrocytic cycle. All in all, these observations made in other organisms support the idea of a potential role of *PfH3.3* in epigenetic memory in *P. falciparum*. In order to investigate the functional relevance of *PfH3.3* in parasite biology in general or in *var* gene regulation in particular, we made several attempts to tag *PfH3.3* with a AID-GFP-glmS [78, 79] or GFP-glmS [79] tag. Unfortunately, these attempts failed to result in viable parasite lines using these rather bulky conditional knock-down tags, hinting towards an essential function of *PfH3.3*. Accordingly, whether *PfH3.3* is critical for “epigenetic memory” of *var* gene expression and how it could exert such function will be the subject of future studies. In summary, we show that *PfH3.3* has a complex localization pattern that consists of conserved, but also parasite-specific features. It demarcates euchromatic coding and subtelomeric repetitive sequences in a GC-content coupled manner. Within the coding sequences it might promote the process of transcription by forming labile nucleosomes, as has been reported for other organisms. The strong marking of subtelomeric repetitive sequences resembles findings from higher eukaryotes, which could elude to a conserved, yet uncharacterized function. Excitingly, *PfH3.3* specifically marks the promoter region of the *var* gene in its active and poised, but not inactive state and could contribute to the process of antigenic variation which is critical to immune evasion of this deadly pathogen.

Acknowledgements

We are very grateful to : i) Hendrik G. Stunnenberg for advice and support at an early stage of this project; ii) Moritz Treeck and Tim-Wolf Gilberger for transfecting the Ty1-*PfH3.3* episomal construct to *P. falciparum* 3D7 parasites; iii) Sash Lopaticki and Alan Cowman for providing us with protocol and hands on experience for CSA selection of parasites; iv) Ron Dzikowski for providing us the NF54-DCJ parasite line and v) Andy Waters, Christine Clayton and David R. Westhead for fruitful advice throughout the project. We would also like to

thank Wieteke Hoeijmakers and Christa Toenhake for lively discussions and critical reading of the manuscript.

SF received her PhD fellowship from the European Community's Seventh Framework Program [grant number FP7/2007-2013] under grant agreements No 242095 and No ParaMet 290080. This work also received support from The Netherlands Organization for Scientific Research [NWO-Vidi 864.11.007] and the National Institutes of Health [EuPathDB-Driving Biological Project, subaward no 553539].

Author Contributions

S.A.F. and R.B designed and conceived the project. R.W.M.H. generated transgenic knock-down parasite lines and provided technical help. S.A.F. with R.B analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Methods

Generation of Ty1-tagged *PfH3.3* parasites

In order to profile *PfH3.3*, 3D7 *P. falciparum* parasites were transfected with a pARL-1a- plasmid [80] encoding a N-terminally Ty1-tagged version of *PfH3.3* (PF3D7_0617900) under the control of the chloroquine resistance transporter promoter (similar to pARL-1a-Ty1-H2A in ref. [47]) and cultured in presence of 40 nM WR99210. To generate *PfH3.3* profiles from endogenously expressed *PfH3.3*, NF54-DCJ *P. falciparum* parasites [51] were transfected with a pHH1-plasmid [81] encoding a C-terminally Ty1-tagged version of *PfH3.3* (PF3D7_0617900) and a CTG instead of an ATG as “start codon” (Fig. S1A). Parasites were cultured in presence of 2.5 nM WR99210. Single crossover integration of the plasmid resulted in parasites that express a fully functional *PfH3.3*-Ty1 under the endogenous promoter (ATG as start codon) and a non-functional untagged *PfH3.3* (CTG as start codon) (Fig. S1B). A clonal parasite line was obtained by limiting dilution. Correct integration of the pHH1-plasmid and absence of wild type parasites was confirmed by performing Southern Blot analysis (Fig. S1B, C). Expression of episomal and endogenously-expressed tagged *PfH3.3* was confirmed using the monoclonal anti-Ty1 antibody (BB2) [82] that exclusively recognizes the tagged *PfH3.3* protein on Western blot (Fig. S1D).

Parasite culture and collection of parasites

Parasites were cultured in standard RPMI medium supplemented with 10% human serum (AB), 0.2% NaHCO₃ and 2.5% human O⁺ red blood cells. Culturing occurred in 250 ml tissue culture flasks placed in candle jars or in an incubator with a gas composition of 3% O₂, 4% CO₂ and 93% N₂ and incubated at 37 °C. Parasites were synchronized with sorbitol treatments and Percoll gradient centrifugations as described previously [47]. During collections, medium was changed every 10 hours, but not less than 10 hours before collection to ensure optimal parasite growth. After

20 hours post invasion (hpi), medium volume was doubled. During each medium change parasite cultures were mixed to minimize growth differences within single culture flasks. For the generation of genome-wide *PfH3.3* profiles early ring (10 hpi), late ring (20 hpi), trophozoite (30 hpi) and schizont stages (40 hpi) were collected. For the ChIP-qPCR investigation of *PfH3.3 var2CSA*- or *PF3D7_0223500*-selected parasites, ring, trophozoite and schizont stages were collected at 15, 25 and 40 hpi. Parasite cultures used for RNA extraction or native ChIP were immediately placed and processed on ice (4 °C). Parasite cultures used for cross-linked ChIP were immediately treated with 1% formaldehyde (final concentration) and incubated at 37 °C for 10 min while shaking. Cross-linking reaction was quenched by adding 0.125 M glycine (final concentration). All subsequent steps were performed at 4 °C unless noted otherwise. To remove contaminating human white blood cells that would cause considerable background upon sequencing, all blood used for collection of sequence samples was filtered through Plasmodipure filters (EuroProxima). Ethic statement: Blood and serum was obtained from the local blood bank (Sanquin) with the full consent of healthy donors to use this material for malaria research.

var2CSA (PF3D7_1200600) and PF3D7_0223500 selection

Endogenously *PfH3.3-Ty1* expressing NF54-DCJ *P. falciparum* parasites were selected for *var2CSA* expression by affinity purification as in ref. [83]. For this purpose, petri dishes (150 × 15 mm, BD biosciences Falcon 351058) were coated with Chondroitin Sulfate A (0.05% CSA in PBS) overnight and subsequently blocked with 1% Casein/PBS solution for at least one hour and rinsed with RPMI twice. In the meanwhile, parasite cultures were centrifuged, resuspended in RPMI containing 10% human serum and transferred to the petri dishes. Subsequently, petri dishes containing the parasite culture were incubated for 30 min at 37 °C in a candle jar. Afterwards, unbound parasites and non-infected erythrocytes were removed by gentle RPMI washes. Bound parasites were extensively resuspended in complete medium, fresh blood was added, and parasites were cultured as described above. *var2CSA* panning was repeated several times to achieve an increasing proportion of *var2CSA*-expressing parasites. Subsequent switching to *PF3D7_0223500* expression, was enforced by culturing the parasites with 2.5 mM Blasticidin [51]. Successful *var* gene selection was confirmed via RNA extraction (RNeasy mini kit, Qiagen) followed by cDNA synthesis and subsequent quantitative real time PCR (qPCR) using *var* gene specific primers [71, 84] (Fig. S7A, B).

cDNA synthesis and qPCR

To synthesize the cDNA 0.5–5 µg of total RNA were mixed with random hexamer primers (0.5 µg, Roche), OligodT12-18 (0.5 µg, Invitrogen) and dNTPs (0.5 mM in the final volume of 20 µl, Invitrogen) and incubated for 5 min at 70 °C. Then first strand synthesis was performed for 1 hour at 42 °C in First Strand Buffer supplemented with DTT (10 mM), Superscript III (200 units) and RNasin Plus RNase inhibitor (40 units, Promega), after which superscript III was inactivated by incubation at 70 °C for 15 minutes. RT “minus” reactions (where Superscript III was replaced by water) were performed in all cases as negative control. Subsequent qPCR was performed using the CFX96 Real Time Systems C1000 Touch Thermal Cycler. *P. falciparum* genomic DNA served as standard (500 pg, 50 pg, 5 pg). SYBRgreen supermix (BioRad) and primers were mixed according to manufacturer’s instructions. qPCR was performed in 96 well plate format (BioRad) using the

following protocol: 95 °C for 3 min, (94 °C for 10 sec, 52 °C for 30 sec, 68 °C for 30 sec) 39x, 95 °C 1 min, 65 °C 1 min and a gradient from 65 °C to 94.5 °C with a 0.5 °C increase every 10 sec.

Western Blot analysis

Nuclear extracts from the 3D7 mother line (~40 ug), the 3D7 parasites expressing Ty1-*PfH3.3* episomally (~60 ug) and the NF54-DCJ parasites expressing *PfH3.3*-Ty1 under the endogenous promoter respectively, were separated on 18% SDS-PAGE and transferred to a protean nitrocellulose membrane (Whatman, 0.45 mm). The membrane was probed with mouse anti-Ty1 antibody (1:2000, BB2 antibody in ref. [82]) and rabbit anti-histone H3 antibody (1:3000, Abcam ab1791) and secondary goat anti-mouse IRDye680 (1:10000, LI-COR Biosciences 926-32220) and goat anti-rabbit IRDye 800CW (1:10000, LI-COR 926-32211). Fluorescence was measured on the Odyssey system (LI-COR Biosciences).

Southern Blot analysis

Genomic DNA was isolated from native nuclei using phenol-chloroform extraction as described previously [47]. 2.5 µg of genomic DNA and 0.5 ng of plasmid DNA were digested with EcoRI and XbaI (NEB) (Fig. S1B). A *PfH3.3* sequence specific hybridization probe was generated by PCR labelling using Digoxigenin-11-dUTP (DIG) (alkali-labile, Roche) according to manufacturer's instructions. The PCR reaction was performed using the Advantage[®] cDNA PCR Kit and Polymerase Mix (Clontech) and a DIG-dUTP to dTTP ratio of 1:3 under the following PCR conditions: 5 min 95 °C; 30 s 95 °C, 30 s 52 °C, 90 s 68 °C (35 cycles); 5 min 68 °C (Fig. S1B). Gel preparation, subsequent DNA transfer to a positively charged Nylon membrane (Hybond-N⁺, GE Healthcare Amersham), prehybridization (DIG Easy Hyb Kit from Roche, 65 °C) and probe-DNA hybridization (DIG Easy Hyb Kit from Roche, 65 °C) were all performed according to manufacturer's recommendations. The detection of the hybridized probe was performed using the DIG Wash and Block Buffer Set, the anti-Digoxigenin antibody (anti-Digoxigenin-AB conjugate Fab frag; 1:10 000 dilution) and the CDP-Star Kit (ready-to-use) from Roche according to manufacturer's instructions.

Chromatin immunoprecipitation

After erythrocyte and parasite lysis using 0.05% saponin and a hypotonic buffer (10 mM Tris pH 8.0, 3 mM MgCl₂, 0.2% NP-40, Roche Protease Inhibitor Cocktail) respectively, native nuclei (ChIP-seq of episomal Ty1-*PfH3.3* and *var* gene switch experiment, see below) or formaldehyde-cross-linked nuclei (ChIP-seq experiment) were separated from cell debris using a 0.25 M sucrose buffer cushion as described before [47]. Subsequent, cross-linked chromatin was prepared as described in ref. [85] with a few adjustments. In short, for generation of genome-wide *PfH3.3* profiles, enzymatic MNase digestion of cross-linked chromatin was performed using digestion buffer (50 mM Tris pH7.4, 4 mM MgCl₂, 1 mM CaCl₂, 0.075% NP40, Roche Protease Inhibitor Cocktail) with 0.5 U MNase (Worthington Biochemicals Corporation) in 150 µl aliquots for 6 min (10 hpi), 12 min (20 hpi) or 13 min (30 hpi & 40 hpi) at 37 °C. Optimal digestion times were empirically determined for each stage by test-digestion of a nuclear aliquot for varying times, to obtain primarily mono-nucleosomal DNA fragments. Digestion reactions were stopped by adding 150 µl quenching solution (2% Triton X100, 0.6% SDS, 300 mM NaCl, 6 mM EDTA, Roche Protease Inhibitor Cocktail) and placing the samples on ice. Subsequently, digested chromatin was

sonicated for 6×10 sec (setting low, BioruptorTM Next Gen, Diagenode) to free cross-linked chromatin from nuclear membranes. Approximately 200 ng of digested DNA-containing cross-linked chromatin was incubated in ChIP buffer (150 mM NaCl, 20 mM Tris pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.15% SDS, Roche Protease Inhibitor Cocktail) with 2 μ g of anti-Ty1, 1 μ g of anti-histone H3 (recognizes *PfH3* and *PfH3.3*, Abcam Ab1791) or 1 μ g polyclonal anti-IgG (Upstate 12-370) overnight at 4 °C while rotating, followed by the addition of 10 μ l ProtA and 10 μ l ProtG dynabeads (Life Technologies, 10008D and 10009D) and further incubation at 4 °C for 2 h. For the H3core antibody (capturing histone *PfH3* and *PfH3.3*) one ChIP reaction was performed, for *PfH3.3* (anti-Ty1) four ChIP reactions were performed in parallel per life cycle stage to ensure sufficient amount of DNA for subsequent ChIP-seq. After washing with three different buffers (buffer1, 1x: 20 mM Tris pH 8.0, 2 mM EDTA, 1% Triton-X100, 0.1% SDS, 150 mM NaCl; buffer 2, 2x: 20 mM Tris pH 8.0, 2 mM EDTA, 1% Triton-X100, 0.1% SDS, 500 mM NaCl and buffer 3, 2x: 10 mM Tris pH 8.0, 1 mM EDTA) immunoprecipitated chromatin was eluted using 1% SDS and 0.1 M NaHCO₃. Immunoprecipitated chromatin was decross-linked (1% SDS, 0.1 M NaHCO₃, 1 M NaCl) overnight at 45 °C shaking, followed by DNA isolation via QIAquick column purification (Qiagen).

To investigate *PfH3.3* and *PfH2A.Z* occupancy as well as H3K9ac and H3K9me3 marking of the *var2CSA* locus of *var2CSA*- or PF3D7_0223500-selected parasites, chromatin was extracted and native ChIPs were performed as described previously [47]. Native chromatin was digested with 0.5 U MNase in 150 μ l aliquots for 4 min (15 hpi *var2CSA*-selected parasites), 7 min (15 hpi, 25 hpi, 40 hpi PF3D7_0223500-selected parasites and 25 hpi *var2CSA*-selected parasites) and 10 min (40 hpi *var2CSA*-selected parasites) at 37 °C in a waterbath and regular tapping of the tubes (again optimal digestion times to primarily mono- nucleosomal fragments were determined empirically by a test-digestion of a nuclear-aliquot from each sample). Subsequent chromatin extraction was performed in low-salt buffer (buffer1: 50 mM Tris pH 7.4, 4 mM MgCl₂, 1 mM CaCl₂, Roche Protease Inhibitor Cocktail; buffer2: 10 mM Tris pH 7.4, 1 mM EDTA, Roche Protease Inhibitor Cocktail; buffer3: 1 mM Tris pH 7.4, 0.2 mM EDTA, Roche Protease Inhibitor Cocktail) and incubated with 1 volume ChIP incubation buffer (100 mM NaCl, 19.7 mM Tris pH 7.4, 6.25 mM EDTA, 1% Triton X-100, 0.1% SDS). For native ChIP reactions, native chromatin (containing about 200 ng DNA) from each time point of *var2CSA*- and PF3D7_0223500-selected parasites was incubated with 2 μ g of anti-Ty1 (*PfH3.3*), 0.5 μ l polyclonal H2A.Z [47], 1 μ g of polyclonal anti-H3K9/14ac antibody (Diagenode, C15410200), 1 μ g of polyclonal H3K9me3 (in house, corresponds to Merck Millipore 07-442) or 1 μ g polyclonal IgG (Upstate 12-370) overnight at 4 °C followed by the addition of 20 μ l ProtA/G dynabeads and further incubation at 4 °C for 2 h. 200 ng of native chromatin were kept as input. After washing with buffers three different buffers (buffer 1, 1x: 50 mM Tris pH 7.5, 10 mM EDTA, 100 mM NaCl, 0.5% Triton X-100, 0.05% SDS; buffer 2, 2x: 50 mM Tris pH 7.5, 10 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 0.05% SDS and buffer 3, 2x: 50 mM Tris pH 7.5; 10 mM EDTA, 250 mM NaCl) immunoprecipitated DNA was eluted and purified using PCR purification columns.

ChIP-qPCR

To test relative enrichment of *PfH3.3*-Ty1, H2A.Z, H3K9ac and H3K9me3 xChIP was performed as described above. qPCR (Bio-Rad MyIQ) on each ChIP sample was performed alongside with a 1:10

dilution series of the corresponding input DNA to calculate recovery as a percentage of input. Recovery values for the *var2CSA* locus were normalized across the different samples using primer sets for regions that are invariably marked with the histone mark/variant. For H3K9ac, PfH2A.Z and PfH3.3 primers for actin (PF3D7_1246200, coding sequence and -1100bp promoter region) and seryl-tRNA synthetase (PF3D7_0717700, coding sequence), HSP70 (PF3D7_0818900, coding sequence and -1000 promoter regions) and 18s RNA (PF3D7_1148600, coding sequence) were chosen. For H3K9me3 primers for bp the telomere repetitive regions, a rifin (PF3D7_0324500, -400bp promoter regions) and AP2G (PF3D7_1222600, coding sequence). Normalization factor was calculated by dividing the average of the recoveries from the control primers within each sample with the average of the average recovery values across the samples. Finally, all recovery values from the *var2CSA* specific primers have been multiplied by the normalization factors. The normalization factor scaled between 0.75 and 1.8.

ChIP-seq library preparation

For each sequencing library generated from endogenously PfH3.3-Ty1 expressing NF54-DCJ *P. falciparum* parasites 3 ng of ChIP DNA were end repaired, extended with 3' A-overhangs, and ligated to barcoded NextFlex adapters (Bio Scientific) [86]. Subsequent library amplification was performed under similar conditions as described previously [85] with a few adjustments: First, 2x KAPA HiFi HotStart ready-mix (KAPA Biosystems) and NextFlex primers were used for 4 cycles of PCR amplification (98 °C for 2 min; 4 cycles of 98 °C for 20 sec, 62 °C for 3 min; 62 °C for 5 min). Subsequently, amplified libraries were size selected for 270 bp (mono-nucleosomes + ~125 bp NextFlex adapter) using 2% E-Gel Size Select Agarose Gels (Thermo Fisher Scientific) and again amplified as described above for 9 cycles. To deplete adapter dimers and clean-up the DNA, libraries were subsequently purified with Agencourt AMPure XP bead (Beckman Coulter) (1:1 library beads ratio) before sequencing. Sequencing libraries generated from episomally Ty1-PfH3.3-expressing parasites were generated as described in refs [47, 86].

RNA-seq library preparation

The preparation of strand-specific RNA-seq libraries was performed as in ref. [85]. In short, after total RNA isolation (with on-column DNase treatment) and oligo-dT-selection [87] PolyA-selected RNA was further treated with DNase (TURBO DNase, Ambion) to remove remaining genomic DNA. Subsequently, 1.2 µg PolyA-selected total RNA equivalent (see ref. [87] for details) for 10 hpi and 2 µg for 20 hpi, 30 hpi and 40 hpi were fragmented by means of alkaline hydrolysis (5x fragmentation buffer: 200 mM Tris acetate pH 8.2, 500 mM potassium acetate, 150 mM magnesium acetate) for 2 min at 85 °C in 250 µl volume. RNA was precipitated overnight and checked for genomic DNA contamination using qPCR, followed by - if needed - another DNase treatment and RNA precipitation for clean-up. Remaining material was processed for strand-specific RNA-seq as in ref. [85]. To prevent unwanted DNA-dependent second strand cDNA synthesis - a major source of artificial antisense 'transcripts' [88] - 0.2 µg Actinomycin D was included into the first strand synthesis reaction, which was followed by a 15 min 70 °C enzyme deactivation step and QIAquick MinElute purification (Qiagen). During second strand synthesis replacement of dTTP with dUTPs resulted in incorporation of U-bases into the second cDNA strand. Subsequently, 2 ng double stranded cDNA were used for each sequencing library which

were prepared as described above with the exception that after adaptor ligation samples were treated with USER enzyme (New England Biolabs) for 15 min at 37 °C that resulted in dUTP-dependent second strand-specific degradation, followed by a first amplification cycle (4 cycles), size-selection for 300–400 bp fragments, a second amplification cycle (9 cycles) and Ampure Beads purification (see above).

High throughput sequencing

ChIP-seq and strand-specific RNA-seq libraries generated from endogenously *PfH3.3*-Ty1 expressing NF54-DCJ parasites were sequenced on a HiSeq 2000 system (Illumina) to obtain 100 bp single-end reads (TruSeq SE Cluster Kit v2). ChIP-seq libraries generated from episomal Ty1-*PfH3.3* expressing parasites were sequenced on a GAllx system (Illumina) to obtain 75 bp single-end reads (Standard Cluster Generation KitV4). 100 bp and 75 bp reads respectively were mapped against the *Plasmodium falciparum* genome assembly (PlasmoDB v6.1) for ChIP samples or against the *Plasmodium falciparum* annotated transcripts (PlasmoDB v6.1) for mRNA samples using BWA [89]. Single-end ChIP and RNA reads were filtered to mapping quality ≥ 15 and only uniquely mapped reads (between 13.5 and 23.5 million reads for ChIP-seq and 15.2 and 26.8 million reads for RNA-seq) were used for further analysis.

High-throughput sequencing data analysis

To visualize ChIP-seq and RNA-seq data in the UCSC browser, all libraries were normalized to the number of mapped reads per million (RPM). For log₂ ratio tracks *PfH3.3* values were divided through the respective *PfH3core* values and log₂ was calculated (“unmappable” regions containing the value 0 were discarded). Bedgraph.gz files were generated using bedtools v2.20.1. Genome browser tracks were set to ‘mean’ as windowing function and ‘16’ as smoothing window function.

To investigate the correlation of *PfH3.3* with the GC-content of the *Plasmodium falciparum* genome the genome was binned into 150 bp windows and the GC-content as well as *PfH3.3* and *PfH3core* tags were calculated in each window. *PfH3.3* tags were divided through *PfH3core* tags (windows containing the value 0 were discarded), log₂ transformed and plotted over the GC-content as a density plot using the python package seaborn 0.6.0 in ipython notebook (Enthought Canopy 1.5.5.3123). For more detailed analysis the 150 bp windows were intersected with different chromatin domains e.g. heterochromatic or euchromatic regions (at least 50% overlap) using bedtools v2.20.1 and depicted as described above.

In order to investigate the correlation of *PfH3.3* in upstream regions, downstream regions or coding sequences with mRNA abundance, the genome was subdivided as described [85]. In short, intergenic regions between convergent and divergent genes were split in half and each half was assigned to the nearest gene. Intergenic regions between genes with the same orientation were split in 3/4 and 1/4. The longer part was assigned to the upstream region of the nearest gene and the shorter part to the downstream region of the other gene. All upstream and downstream regions were shortened to a maximum of 3 kb as this covers the vast majority of upstream regions in full length but excludes some extremely long intergenic regions. *PfH3.3* and *PfH3core* tags were calculated within each upstream region, downstream region and coding sequence, normalized to the amount of reads per kb per million mapped reads (RPKM) and *PfH3.3*/*PfH3core* ratios (*PfH3.3*

occupancy) were determined. To assess RNA abundance of each gene, RNA-seq data were mapped against the annotated *P. falciparum* transcriptome, tags were counted for each transcript (excluding tRNAs, rRNAs, mitochondrial and apicoplast RNA) and RPKM values were calculated. To illustrate the dynamics of RNA expression during intraerythrocytic development, the relative RNA abundance was calculated by dividing the amount of tags per transcript of each intraerythrocytic stage through the sum of tags per transcript of all four investigated intraerythrocytic stages (Heterochromatic genes and 25% of euchromatic genes with the lowest expression has been excluded from this analysis). Subsequently, RNAs were clustered into eight groups according to their expression pattern throughout the intraerythrocytic cycle using k-means clustering (Multi Experiment Viewer 4.9.0). Clusters were visualized in a heatmap. After calculating the relative Pfh3.3 occupancy in upstream regions for all four intraerythrocytic stages as described above, the profile of Pfh3.3 occupancy in upstream regions was matched to the respective RNA expression profile. The same method was applied to coding sequences and downstream regions.

To illustrate the relationship between Pfh3.3 enrichment and gene expression level, RPKM values were log₁₀ transformed and plotted against log₂ transformed Pfh3.3 occupancy levels of matching upstream region, downstream region and coding sequence.

Data Availability

ChIP-seq and RNA-seq datasets will be available via GEO (accession number GSE80466) and PlasmoDB.

Reference

1. Ramachandran, S. & Henikoff, S. Replicating Nucleosomes. *Sci Adv* **1**, doi:10.1126/sciadv.1500587 (2015).
2. Gunjan, A., Paik, J. & Verreault, A. Regulation of histone synthesis and nucleosome assembly. *Biochimie* **87**, 625-635, doi:10.1016/j.biochi.2005.02.008 (2005).
3. Melters, D. P., Nye, J., Zhao, H. & Dalal, Y. Chromatin Dynamics in Vivo: A Game of Musical Chairs. *Genes (Basel)* **6**, 751-776, doi:10.3390/genes6030751 (2015).
4. Lawrence, M., Daujat, S. & Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet* **32**, 42-56, doi:10.1016/j.tig.2015.10.007 (2016).
5. Ammar, R. *et al.* Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* **1**, e00078, doi:10.7554/eLife.00078 (2012).
6. Talbert, P. B. *et al.* A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin* **5**, 7, doi:10.1186/1756-8935-5-7 (2012).
7. Waterborg, J. H. Evolution of histone H3: emergence of variants and conservation of post-translational modification sites. *Biochem Cell Biol* **90**, 79-95, doi:10.1139/o11-036 (2012).
8. Elsaesser, S. J., Goldberg, A. D. & Allis, C. D. New functions for an old variant: no substitute for histone H3.3. *Curr Opin Genet Dev* **20**, 110-117, doi:10.1016/j.gde.2010.01.003 (2010).
9. Cui, J. *et al.* Genome-wide identification, evolutionary, and expression analyses of histone H3 variants in plants. *Biomed Res Int* **2015**, 341598, doi:10.1155/2015/341598 (2015).
10. Falk, S. J. & Black, B. E. Centromeric chromatin and the pathway that drives its propagation. *Biochim Biophys Acta* **1819**, 313-321 (2013).
11. Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**, 264-275, doi:10.1038/nrm2861 (2010).

12. McKinley, K. L. & Cheeseman, I. M. The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol* **17**, 16-29, doi:10.1038/nrm.2015.5 (2016).
13. Hoeijmakers, W. A. *et al.* Plasmodium falciparum centromeres display a unique epigenetic makeup and cluster prior to and during schizogony. *Cell Microbiol* **14**, 1391-1401, doi:10.1111/j.1462-5822.2012.01803.x (2012).
14. Altheim, B. A. & Schultz, M. C. Histone modification governs the cell cycle regulation of a replication-independent chromatin assembly pathway in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **96**, 1345-1350 (1999).
15. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nat Struct Biol* **10**, 882-891, doi:10.1038/nsb996 (2003).
16. Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421-434, doi:10.1038/cr.2011.14 (2011).
17. Cui, B., Liu, Y. & Gorovsky, M. A. Deposition and function of histone H3 variants in *Tetrahymena thermophila*. *Mol Cell Biol* **26**, 7719-7730, doi:10.1128/MCB.01139-06 (2006).
18. Liu, C. P. *et al.* Structure of the variant histone H3.3-H4 heterodimer in complex with its chaperone DAXX. *Nat Struct Mol Biol* **19**, 1287-1292, doi:10.1038/nsmb.2439 (2012).
19. Elsässer, S. J. *et al.* DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. *Nature* **491**, 560-565, doi:10.1038/nature11608 (2012).
20. Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* **116**, 51-61 (2004).
21. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* **9**, 1191-1200 (2002).
22. Shu, H. *et al.* Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes. *Genome Biol* **15**, R62, doi:10.1186/gb-2014-15-4-r62 (2014).
23. Stroud, H. *et al.* Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **109**, 5370-5375, doi:10.1073/pnas.1203145109 (2012).
24. Schwartz, B. E. & Ahmad, K. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev* **19**, 804-814, doi:10.1101/gad.1259805 (2005).
25. Tamura, T. *et al.* Inducible deposition of the histone variant H3.3 in interferon-stimulated genes. *J Biol Chem* **284**, 12217-12225, doi:10.1074/jbc.M805651200 (2009).
26. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-691, doi:10.1016/j.cell.2010.01.003 (2010).
27. Drané, P., Ouararhni, K., Depaux, A., Shuaib, M. & Hamiche, A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* **24**, 1253-1265, doi:10.1101/gad.566910 (2010).
28. Santenard, A. *et al.* Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat Cell Biol* **12**, 853-862, doi:10.1038/ncb2089 (2010).
29. Wong, L. H. *et al.* Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. *Genome Res* **19**, 404-414, doi:10.1101/gr.084947.108 (2009).
30. Hake, S. B. *et al.* Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. *Proc Natl Acad Sci U S A* **102**, 6344-6349, doi:10.1073/pnas.0502413102 (2005).
31. Udugama, M. *et al.* Histone variant H3.3 provides the heterochromatic H3 lysine 9 trimethylation mark at telomeres. *Nucleic Acids Res*, doi:10.1093/nar/gkv847 (2015).
32. Jang, C. W., Shibata, Y., Starmer, J., Yee, D. & Magnuson, T. Histone H3.3 maintains genome integrity during mammalian development. *Genes Dev* **29**, 1377-1392, doi:10.1101/gad.264150.115 (2015).

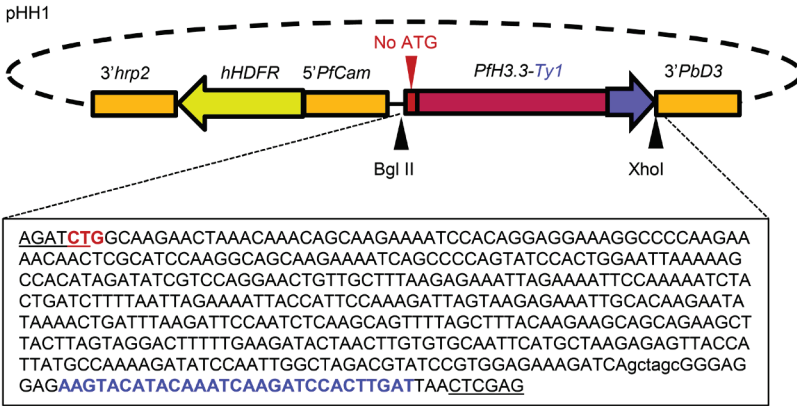
33. Wollmann, H. *et al.* Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. *PLoS Genet* **8**, e1002658, doi:10.1371/journal.pgen.1002658 (2012).
34. Jacob, Y. *et al.* Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. *Science* **343**, 1249-1253, doi:10.1126/science.1248357 (2014).
35. Sakai, A., Schwartz, B. E., Goldstein, S. & Ahmad, K. Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. *Curr Biol* **19**, 1816-1820, doi:10.1016/j.cub.2009.09.021 (2009).
36. Jin, C. & Felsenfeld, G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* **21**, 1519-1529, doi:10.1101/gad.1547707 (2007).
37. Hödl, M. & Basler, K. Transcription in the absence of histone H3.3. *Curr Biol* **19**, 1221-1226, doi:10.1016/j.cub.2009.05.048 (2009).
38. Couldrey, C., Carlton, M. B., Nolan, P. M., Colledge, W. H. & Evans, M. J. A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. *Hum Mol Genet* **8**, 2489-2495 (1999).
39. Schwartzenuber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226-231, doi:10.1038/nature10833 (2012).
40. Lan, F. & Shi, Y. Histone H3.3 and cancer: A potential reader connection. *Proc Natl Acad Sci U S A* **112**, 6814-6819, doi:10.1073/pnas.1418996111 (2015).
41. Hoeijmakers, W. A., Stunnenberg, H. G. & Bártfai, R. Placing the Plasmodium falciparum epigenome on the map. *Trends Parasitol* **28**, 486-495, doi:10.1016/j.pt.2012.08.006 (2012).
42. Miao, J., Fan, Q., Cui, L. & Li, J. The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation. *Gene* **369**, 53-65, doi:10.1016/j.gene.2005.10.022 (2006).
43. Flueck, C. *et al.* Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* **5**, e1000569, doi:10.1371/journal.ppat.1000569 (2009).
44. Lopez-Rubio, J. J., Mancio-Silva, L. & Scherf, A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5**, 179-190, doi:10.1016/j.chom.2008.12.012 (2009).
45. Pérez-Toledo, K. *et al.* Plasmodium falciparum heterochromatin protein 1 binds to trimethylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res* **37**, 2596-2606, doi:10.1093/nar/gkp115 (2009).
46. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498-511, doi:10.1038/nature01097 (2002).
47. Bártfai, R. *et al.* H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3. *PLoS Pathog* **6**, e1001223, doi:10.1371/journal.ppat.1001223 (2010).
48. Hoeijmakers, W. A. *et al.* H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the Plasmodium falciparum genome. *Mol Microbiol* **87**, 1061-1073, doi:10.1111/mmi.12151 (2013).
49. Petter, M. *et al.* H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite Plasmodium falciparum. *Mol Microbiol* **87**, 1167-1182, doi:10.1111/mmi.12154 (2013).
50. Sullivan, W. J. Histone H3 and H3.3 variants in the protozoan pathogens Plasmodium falciparum and Toxoplasma gondii. *DNA Seq* **14**, 227-231 (2003).
51. Dzikowski, R., Frank, M. & Deitsch, K. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathog* **2**, e22, doi:10.1371/journal.ppat.0020022 (2006).

52. Scherf, A., Lopez-Rubio, J. J. & Riviere, L. Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol* **62**, 445-470, doi:10.1146/annurev.micro.61.080706.093134 (2008).
53. Baruch, D. I. *et al.* 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 (1995).
54. Smith, J. D. *et al.* Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101-110 (1995).
55. Su, X. Z. *et al.* The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89-100 (1995).
56. Cortés, A., Crowley, V. M., Vaquero, A. & Voss, T. S. A view on the role of epigenetics in the biology of malaria parasites. *PLoS Pathog* **8**, e1002943, doi:10.1371/journal.ppat.1002943 (2012).
57. Voss, T. S., Bozdech, Z. & Bártfai, R. Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr Opin Microbiol* **20**, 88-95, doi:10.1016/j.mib.2014.05.007 (2014).
58. Chookajorn, T. *et al.* Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci U S A* **104**, 899-902, doi:10.1073/pnas.0609084103 (2007).
59. Epp, C., Li, F., Howitt, C. A., Chookajorn, T. & Deitsch, K. W. 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**, 116-127, doi:10.1261/rna.1080109 (2009).
60. Amit-Avraham, I. *et al.* Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **112**, E982-991, doi:10.1073/pnas.1420855112 (2015).
61. Calderwood, M. S., Gannoun-Zaki, L., Wellem, T. E. & Deitsch, K. W. *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**, 34125-34132, doi:10.1074/jbc.M213065200 (2003).
62. Broadbent, K. M. *et al.* A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol* **12**, R56, doi:10.1186/gb-2011-12-6-r56 (2011).
63. Siegel, T. N. *et al.* Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. *BMC Genomics* **15**, 150, doi:10.1186/1471-2164-15-150 (2014).
64. Voon, H. P. & Wong, L. H. New players in heterochromatin silencing: histone variant H3.3 and the ATRX/DAXX chaperone. *Nucleic Acids Res*, doi:10.1093/nar/gkw012 (2016).
65. Harris, L. M. & Merrick, C. J. G-quadruplexes in pathogens: a common route to virulence control? *PLoS Pathog* **11**, e1004562, doi:10.1371/journal.ppat.1004562 (2015).
66. Maizels, N. & Gray, L. T. The G4 genome. *PLoS Genet* **9**, e1003468, doi:10.1371/journal.pgen.1003468 (2013).
67. Erkek, S. *et al.* Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* **20**, 868-875, doi:10.1038/nsmb.2599 (2013).
68. Huang, C. & Zhu, B. H3.3 turnover: a mechanism to poise chromatin for transcription, or a response to open chromatin? *Bioessays* **36**, 579-584, doi:10.1002/bies.201400005 (2014).
69. Ray-Gallet, D. *et al.* HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Mol Cell* **9**, 1091-1100 (2002).

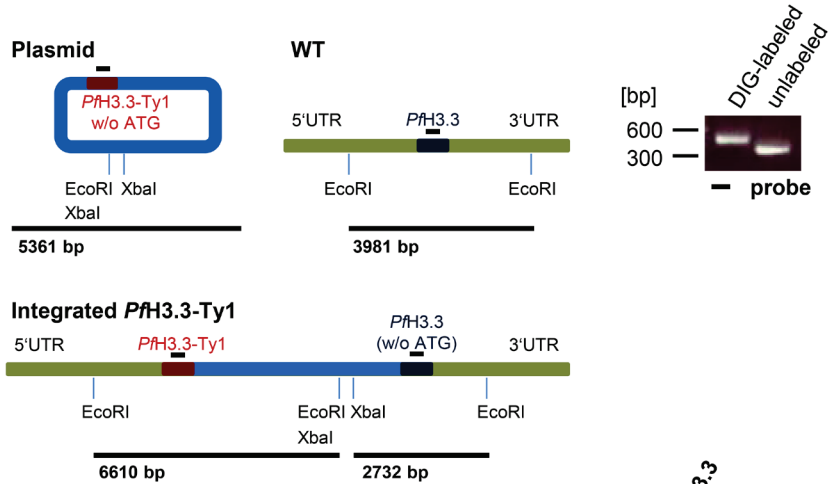
70. DeNizio, J. E., Elsässer, S. J. & Black, B. E. DAXX co-folds with H3.3/H4 using high local stability conferred by the H3.3 variant recognition residues. *Nucleic Acids Res* **42**, 4318-4331, doi:10.1093/nar/gku090 (2014).
71. Lopez-Rubio, J. J. *et al.* 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol* **66**, 1296-1305, doi:10.1111/j.1365-2958.2007.06009.x (2007).
72. Volz, J. *et al.* Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in *Plasmodium falciparum*. *Int J Parasitol* **40**, 109-121, doi:10.1016/j.ijpara.2009.09.002 (2010).
73. Ng, R. K. & Gurdon, J. B. Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer. *Proc Natl Acad Sci U S A* **102**, 1957-1962, doi:10.1073/pnas.0409813102 (2005).
74. Ng, R. K. & Gurdon, J. B. Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat Cell Biol* **10**, 102-109, doi:10.1038/ncb1674 (2008).
75. Ng, R. K. & Gurdon, J. B. Epigenetic inheritance of cell differentiation status. *Cell Cycle* **7**, 1173-1177 (2008).
76. Jin, C. & Felsenfeld, G. Distribution of histone H3.3 in hematopoietic cell lineages. *Proc Natl Acad Sci U S A* **103**, 574-579, doi:10.1073/pnas.0509974103 (2006).
77. Mito, Y., Henikoff, J. G. & Henikoff, S. Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**, 1408-1411, doi:10.1126/science.1134004 (2007).
78. Philip, N. & Waters, A. P. Conditional Degradation of *Plasmodium* Calcineurin Reveals Functions in Parasite Colonization of both Host and Vector. *Cell Host Microbe* **18**, 122-131, doi:10.1016/j.chom.2015.05.018 (2015).
79. Prommana, P. *et al.* Inducible knockdown of *Plasmodium* gene expression using the glmS ribozyme. *PLoS One* **8**, e73783, doi:10.1371/journal.pone.0073783 (2013).
80. Crabb, B. S. *et al.* Transfection of the human malaria parasite *Plasmodium falciparum*. *Methods Mol Biol* **270**, 263-276, doi:10.1385/1-59259-793-9:263 (2004).
81. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K. & Cowman, A. F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature* **403**, 906-909, doi:10.1038/35002615 (2000).
82. Brookman, J. L. *et al.* An immunological analysis of Ty1 virus-like particle structure. *Virology* **207**, 59-67, doi:10.1006/viro.1995.1051 (1995).
83. Noviyanti, R. *et al.* Multiple var gene transcripts are expressed in *Plasmodium falciparum* infected erythrocytes selected for adhesion. *Mol Biochem Parasitol* **114**, 227-237 (2001).
84. Salanti, A. *et al.* 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 (2003).
85. Kensche, P. R. *et al.* The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Res*, doi:10.1093/nar/gkv1214 (2015).
86. Hoeijmakers, W. A., Bártfai, R., François, K. J. & Stunnenberg, H. G. Linear amplification for deep sequencing. *Nat Protoc* **6**, 1026-1036, doi:10.1038/nprot.2011.345 (2011).
87. Hoeijmakers, W. A., Bártfai, R. & Stunnenberg, H. G. Transcriptome analysis using RNA-Seq. *Methods Mol Biol* **923**, 221-239, doi:10.1007/978-1-62703-026-7_15 (2013).
88. Perocchi, F., Xu, Z., Clauder-Münster, S. & Steinmetz, L. M. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res* **35**, e128, doi:10.1093/nar/gkm683 (2007).
89. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

Supplementary information

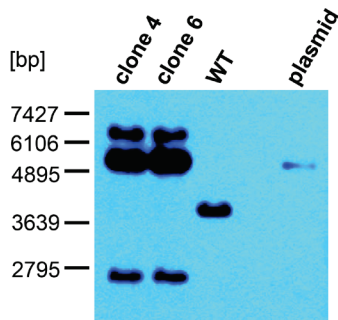
(A)



(B)



(C)



(D)

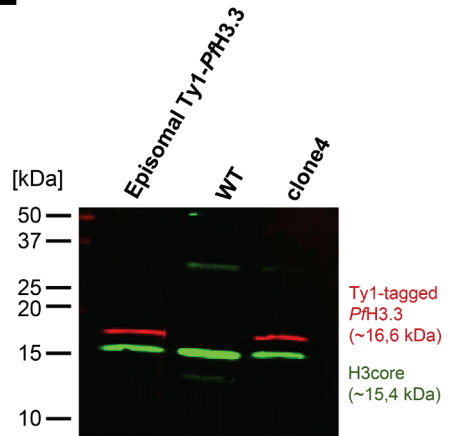
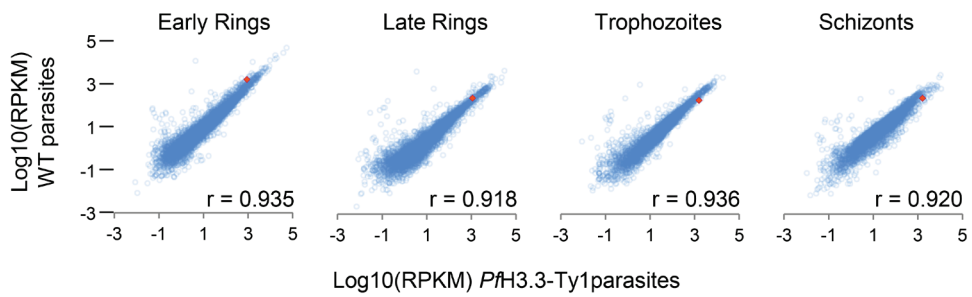


Figure S1: Generation of endogenously tagged *PfH3.3*-Ty1 parasite line.

(A) Map of the pHH1-*PfH3.3*-Ty1 integration vector. (B) Schematic representation of the fragments obtained from the i) integration vector ii) the wild type and iii) tagged *PfH3.3* locus after EcoRI/XbaI digestion. Gel picture shows the successful DIG labelling of 349bp DNA probe corresponding to the coding sequence of *PfH3.3*. (C) Southern Blot confirming successful integration of *PfH3.3*-Ty1 into the parasite's genome of two independent clones of the *P. falciparum* NF54- DCJ line. (D) Western Blot demonstrating the expression of Ty1-tagged *PfH3.3* both from an episomal construct as well as from the endogenous locus. The blot was stained with Ty1-tag specific antibody (BB2, red) and H3core antibody (green).

**Figure S2: Wild type and endogenously tagged *PfH3.3*-Ty1 parasites show a similar RNA expression pattern.**

RPKM values were calculated for wild type 3D7 85 and endogenously tagged *PfH3.3*-Ty1 DCJ parasites for all transcripts excluding tRNAs, rRNAs, mitochondrial and apicoplast RNA. Log10 transformed RPKM values are depicted for early rings, late rings, trophozoites and schizonts. Pearson correlation values are displayed at the lower right corner of the graphs. A red square represents *PfH3.3*.

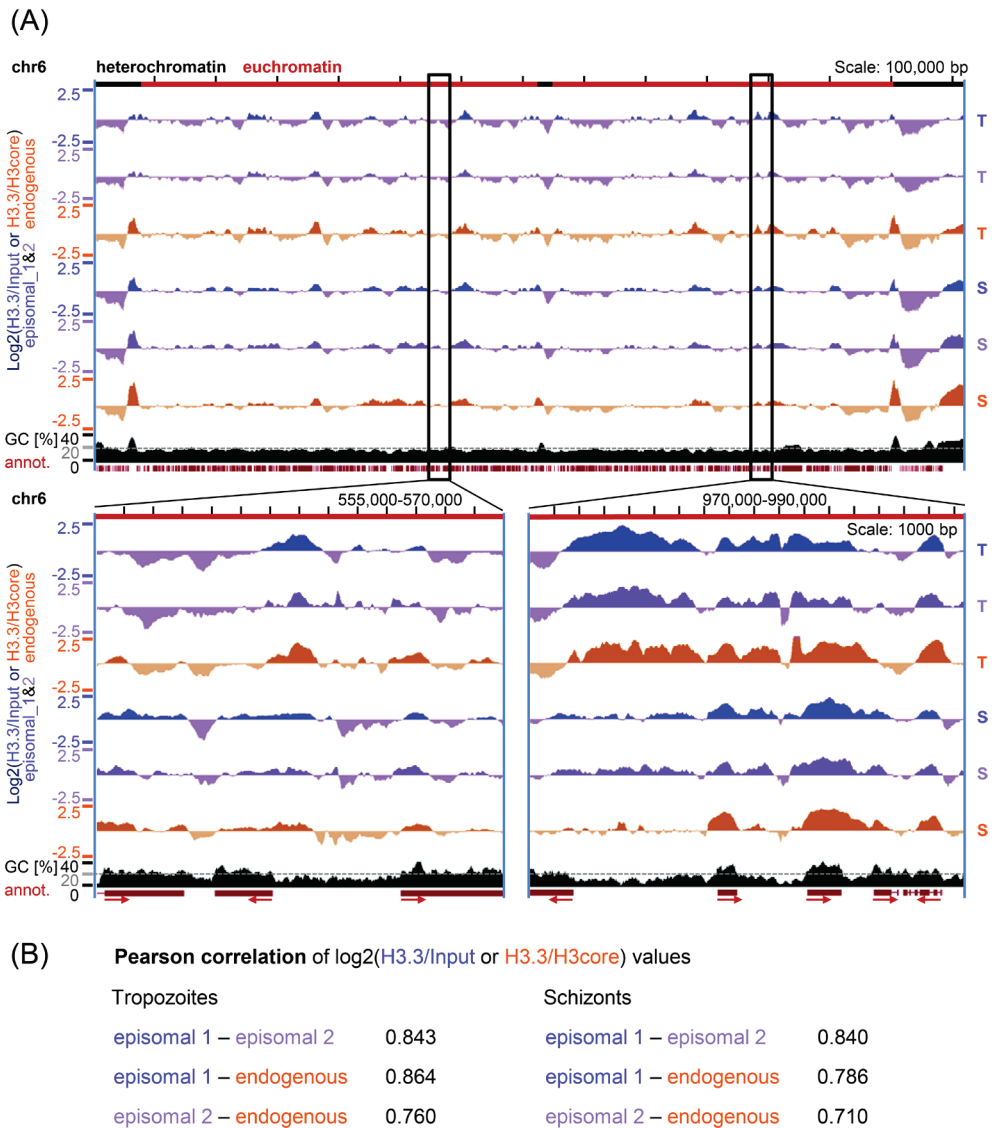


Figure S3: Ratio tracks of *Pf*H3.3-Ty1 ChIP-seq obtained from two different parasite lines at two stages of intraerythrocytic development.

(A) ChIP-seq ratio tracks of trophozoite (T) and schizont (S) stages are depicted in purple and violet for two biological replicates of the episomally Ty1-*Pf*H3.3 expressing 3D7 parasite line (N-terminally tagged *Pf*H3.3, native ChIP, linear amplification protocol, normalized over input) and in orange for endogenously *Pf*H3.3-Ty1 expressing NF54-DCJ parasites (C-terminally tagged *Pf*H3.3, cross-linked ChIP, KAPA amplification protocol, normalized over H3core values). (B) Pearson correlation values calculated from \log_2 transformed *Pf*H3.3/Input or *Pf*H3.3/H3core values for all three ChIPs-seq data sets in two intraerythrocytic stages.

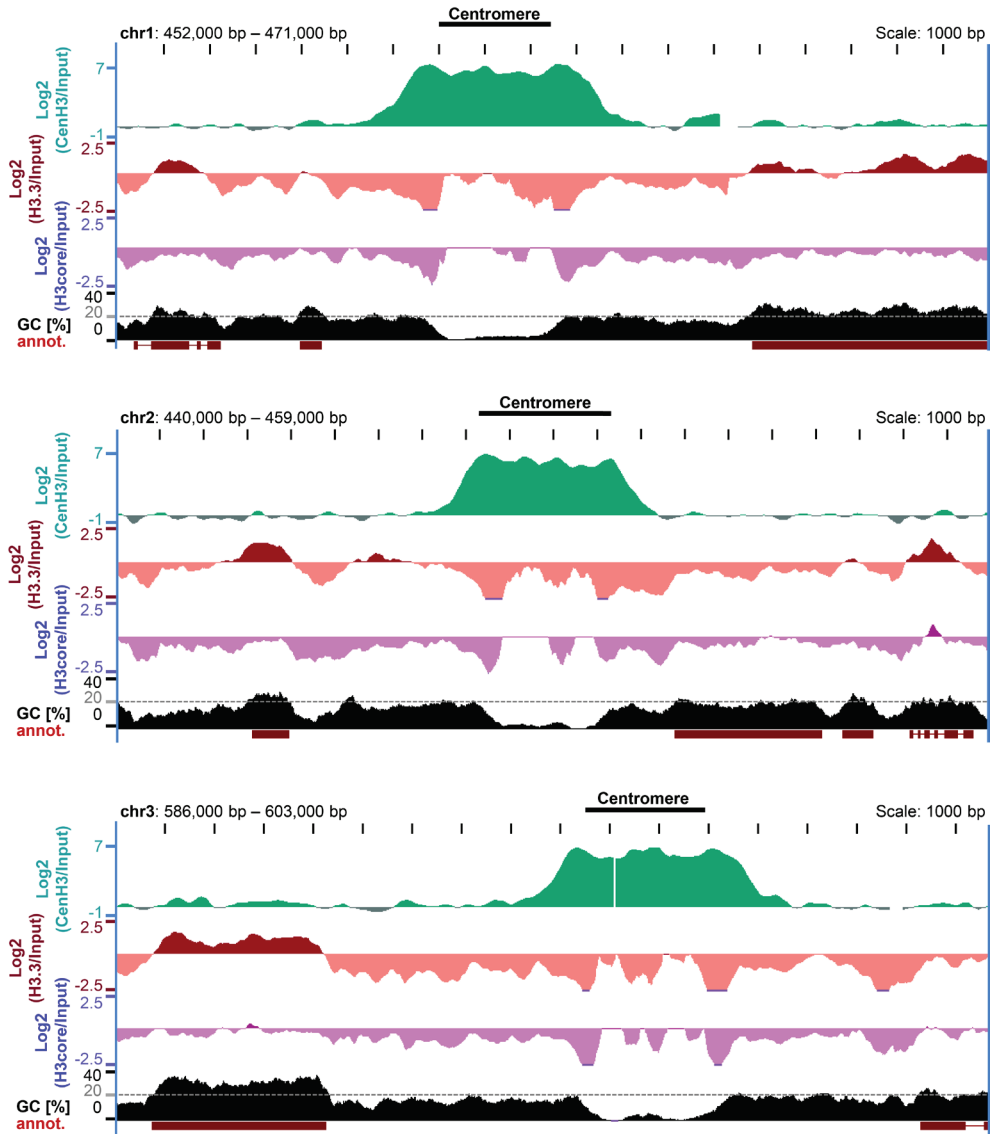


Figure S4: Centromere regions are *PfH3.3* depleted and marked by *PfCenH3*.

PfCenH3, but also *PfH3.3*-Ty1 and H3core ChIP-seq tracks were normalized over input since the H3core antibody only captures *PfH3* and *PfH3.3*. *PfCenH3* clearly marks centromeres (upper panel) whereas *PfH3.3* (middle panel) but also *PfH3* (lower panel) is depleted from these sites. Normalized ChIP-seq data are depicted from schizont stages (40hpi). Centromeres are indicated by black boxes.

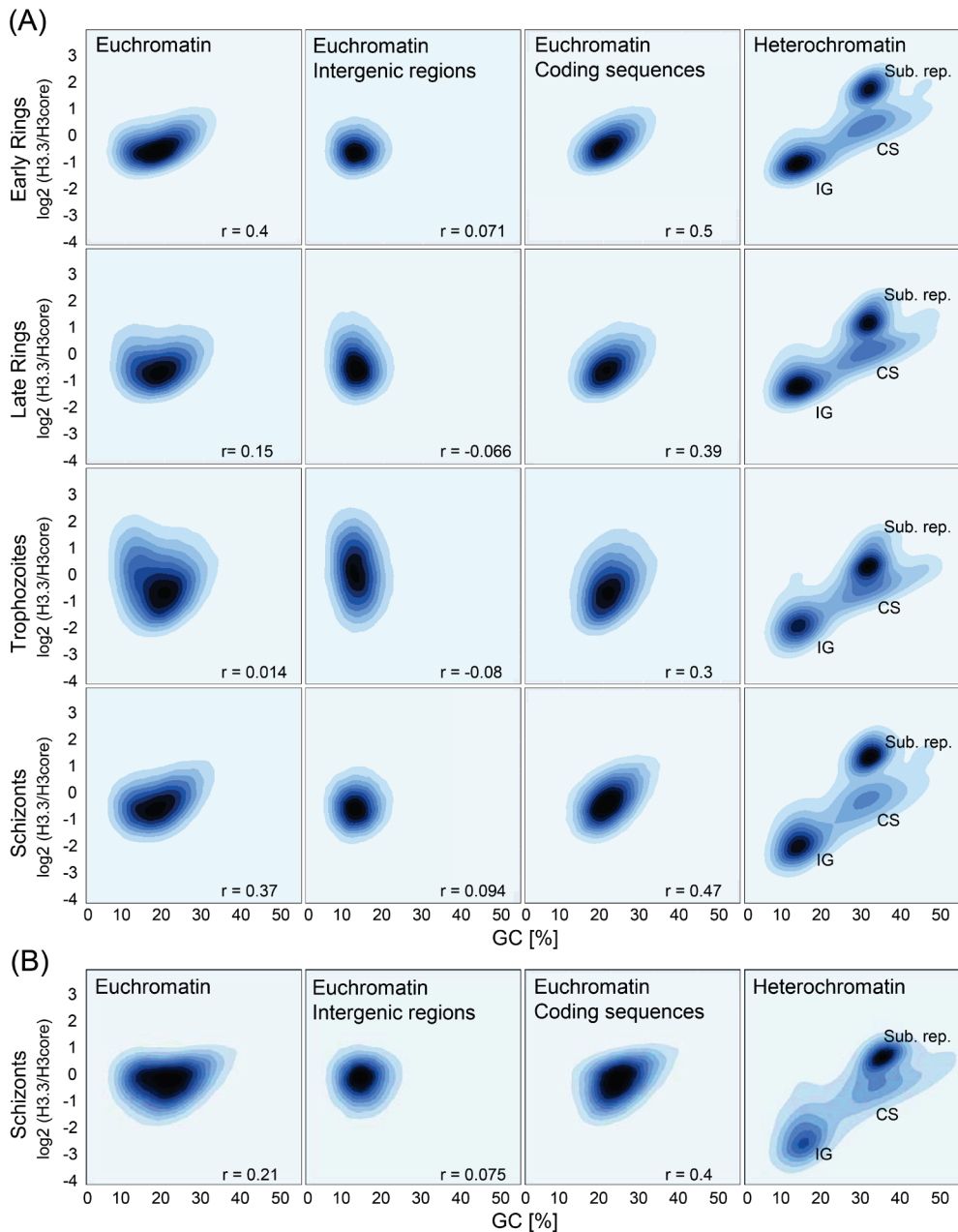


Figure S5: *PfH3.3* localizes to GC-rich regions within the *P. falciparum* genome throughout intraerythrocytic development.

Density plots depicting *PfH3.3* levels in relation to GC-content. *PfH3.3* levels and GC-content were calculated genome-wide per 150 bp windows. **(A)** Data depicted are generated from endogenously tagged *PfH3.3*-Ty1 DCJ parasites for four different intraerythrocytic stages: Early rings (10hpi), late rings (20hpi), trophozoites (30hpi) and schizonts (40hpi). **(B)** Data depicted are generated from

episomally Ty1-*PfH3.3* expressing parasites (see Fig. S3 “episomal 1”) at schizont stage. IG: intergenic region, CS: coding sequence, Sub. Rep.: subtelomeric repetitive region. Pearson correlation values are displayed at the lower right corner of the graphs.

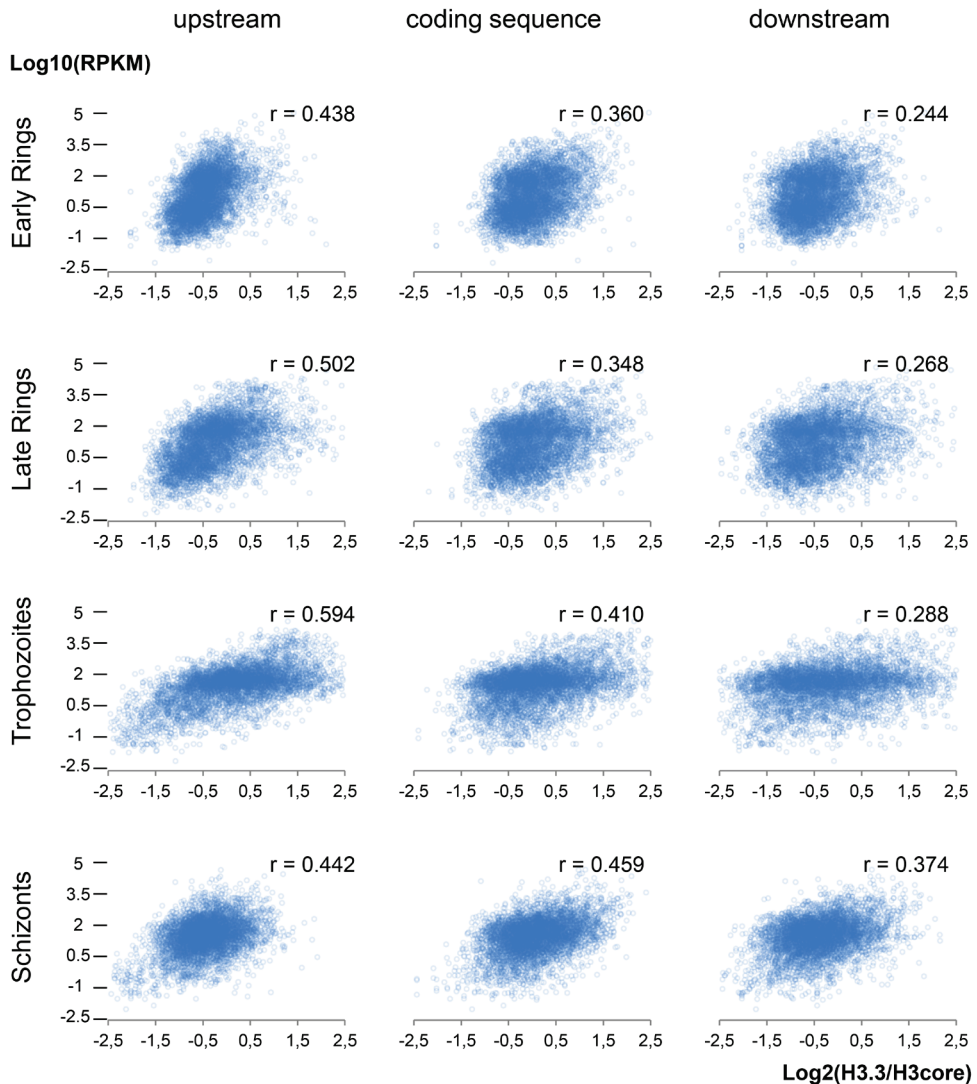
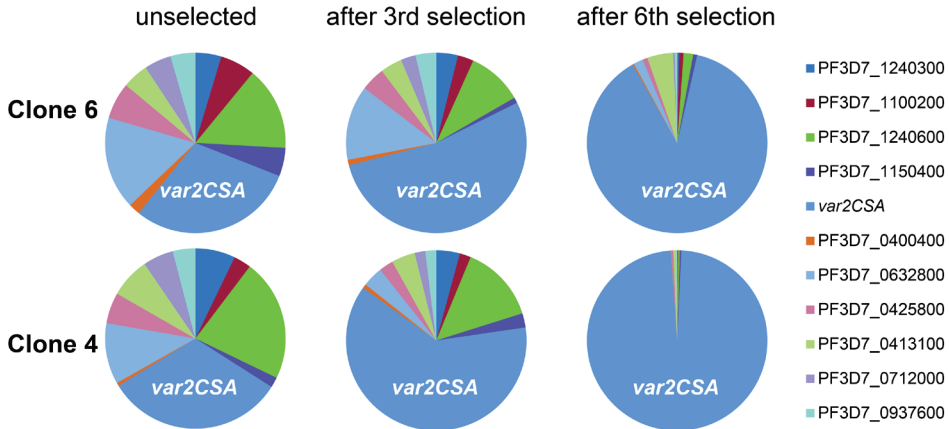


Figure S6: Correlation between steady state mRNA abundance and *PfH3.3* occupancy at four intraerythrocytic stages.

Scatter plots depicting \log_{10} transformed RPKM values from all transcripts excluding tRNAs, rRNAs, mitochondrial and apicoplast RNA in relation to \log_2 transformed *PfH3.3*/H3core ratios in upstream regions, coding sequences and downstream regions. RNA- and ChIP-seq analysis were performed on early rings, late rings, trophozoites and schizonts from the endogenously tagged *PfH3.3*-Ty1 DCJ parasite line. Pearson correlation values are displayed at the higher right corner of the graphs.

(A)
var2CSA (Pf3D7_1200600) selection for two clones



(B)
PF3D7_0223500 expression after Blasticidin treatment

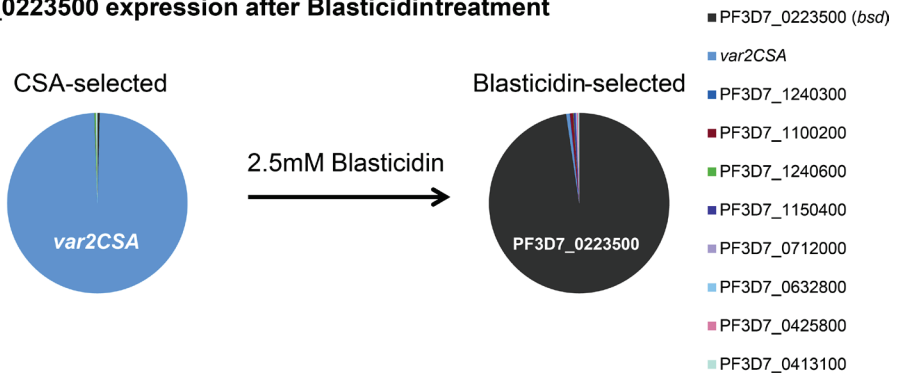


Figure S7: Endogenously *PfH3.3-Ty1* expressing NF54-DCJ *P. falciparum* parasites were successfully selected for *var2CSA* and PF3D7_223500 expression.

(A) Two independent, endogenously *PfH3.3-Ty1* expressing NF54-DCJ *P. falciparum* parasites clones were selected for *var2CSA* expression by repeated rounds of *var2CSA* panning. Successful *var* gene selection was confirmed via RNA extraction followed by cDNA synthesis and subsequent qPCR using one *var2CSA* specific primer pair and 10 primers pairs specific for 10 other *var* genes. (B) CSA-selected parasites were treated with Blasticidin. This selected for parasites expressing PF3D7_223500 since in this parasite line the PF3D7_223500 promoter drives the expression of the BSD resistance cassette (*bsd*). Successful *var* gene switch was confirmed via RNA extraction followed by cDNA synthesis and subsequent qPCR using one specific primer pair for *bsd*, one specific primer pair for *var2CSA* and primers pairs specific for 8 other *var* genes.

CHAPTER 4

Comparative heterochromatin profiling reveals conserved and unique epigenome signatures linked to adaptation and development of malaria parasites

Sabine A. Fraschka^{1,*}, Michael Filarsky^{2,3,*}, Regina Hoo⁴, Igor Niederwieser^{2,3}, Xue Yan Yam⁴, Nicolas M. B. Brancucci^{5,6}, Franziska Mohring⁷, Annals T. Mushunje⁴, Ximei Huang⁴, Peter R. Christensen⁸, Francois Nosten^{8,9}, Zbynek Bozdech⁴, Bruce Russell¹⁰, Robert W. Moon⁷, Matthias Marti^{5,6}, Peter R. Preiser⁴, Richárd Bártfai^{1,#}, Till S. Voss^{2,3,11,#}

Adapted from
Cell Host & Microbe 23, 407-420.e408 (2018)

¹Department of Molecular Biology, Faculty of Science, Radboud University, 6525 GA Nijmegen, the Netherlands

²Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, 4051 Basel, Switzerland

³University of Basel, 4001 Basel, Switzerland

⁴School of Biological Sciences, Nanyang Technological University, Singapore 637551, Singapore

⁵Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow G12 8QQ, UK

⁶Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, Boston MA 02155, USA

⁷Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

⁸Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot 63110, Thailand

⁹Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine Research building, University of Oxford Old Road campus, Oxford OX3 7FZ, UK

¹⁰Department of Microbiology and Immunology, University of Otago, Dunedin 9054, New Zealand

¹¹Lead Contact

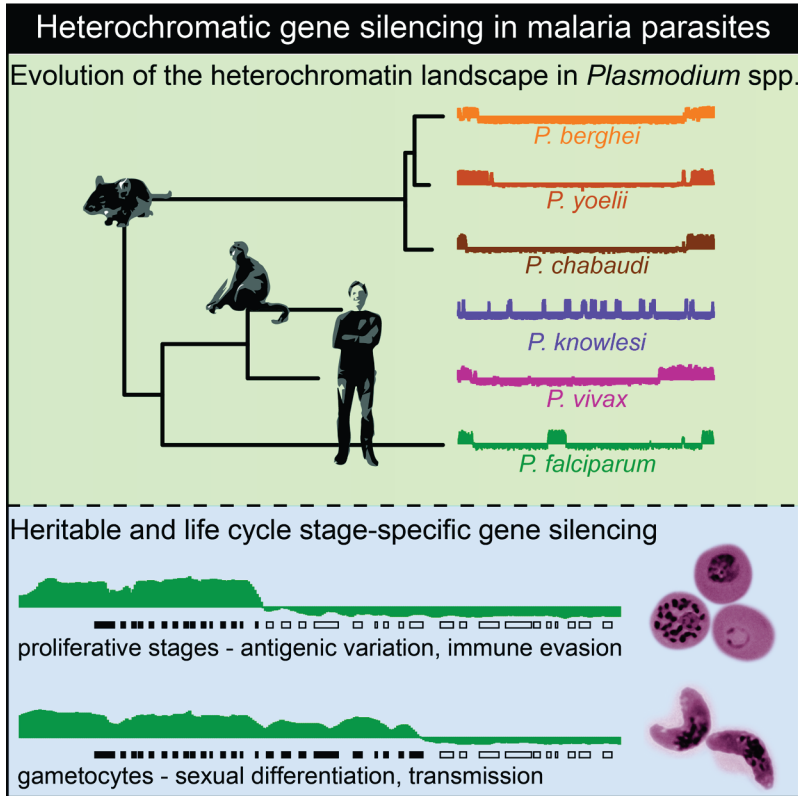
*These authors contributed equally to this work.

#Correspondence: till.voss@unibas.ch (T.S.V.), r.bartfai@science.ru.nl (R.B.)

Contribution:

Designed and performed ChIP-seq experiments, analysed these data, prepared illustrations and wrote the respective sections of the manuscript.

Graphical abstract



Summary

Heterochromatin-dependent gene silencing is central to the adaptation and survival of *Plasmodium falciparum* malaria parasites, allowing clonally variant gene expression during blood infection in humans. By assessing genome-wide heterochromatin protein 1 (HP1) occupancy, we present a comprehensive analysis of heterochromatin landscapes across different *Plasmodium* species, strains and life cycle stages. Common targets of epigenetic silencing include fast-evolving multi-gene families encoding surface antigens and a small set of conserved HP1-associated genes with regulatory potential. Many *P. falciparum* heterochromatic genes are marked in a strain-specific manner, increasing the parasite's adaptive capacity. Whereas heterochromatin is strictly maintained during mitotic proliferation of asexual blood stage parasites, substantial heterochromatin reorganization occurs in differentiating gametocytes and appears crucial for the activation of key gametocyte-specific genes and adaptation of erythrocyte remodeling machinery. Collectively, these findings provide a catalog of heterochromatic genes and reveal conserved and specialized features of epigenetic control across the genus *Plasmodium*.

Introduction

Malaria is caused by unicellular eukaryotes of the genus *Plasmodium* that belongs to an ancient group of obligate endoparasites known as Apicomplexa. The *Plasmodium* genus comprises a few hundred species infecting birds, reptiles or mammals and their radiation is estimated to have occurred about 130 million years ago (1). Members of the *Vinckeia* subgenus parasitize non-primate mammals, among which rodents and bats are the most abundant. This group includes parasites of rodents such as *P. berghei*, *P. yoelii*, *P. chabaudi* and *P. vinckei*, which serve as important models to interrogate *Plasmodium* biology. Parasites belonging to the subgenera *Plasmodium* and *Laverania* infect humans or other primates. Five species are known to naturally infect humans, namely *P. vivax*, *P. ovale*, *P. malariae*, *P. knowlesi* (all members of the *Plasmodium* clade) and *P. falciparum* (*Laverania* clade).

Malaria parasites of mammals are transmitted between their intermediate hosts by female *Anopheles* mosquitoes. Their life cycle is complex involving several stage transitions and replication phases as well as colonization of different cell types and tissues. In the bloodstream, parasites invade red blood cells (RBCs) and undergo intracellular multiplication via schizogony, which involves progression through the ring and trophozoite stages followed by multiple nuclear divisions before a single cytokinesis event leads to the production of up to 32 merozoites ready to invade other RBCs. Repeated rounds of these cycles are responsible for all malaria-related morbidity and mortality. For malaria transmission to occur, mosquitoes must ingest male and female gametocytes with their blood meal. These sexual precursors emerge at a low rate from the proliferating pool of blood stage parasites and are essential to complete sexual reproduction and subsequent sporozoite formation in the mosquito vector. Upon injection into the skin through a mosquito bite, sporozoites migrate to the liver, undergo intra-hepatic schizogony and release over 10'000 merozoites that commence blood stage infection.

Proteins involved in functions at the host-parasite interface have been key to the evolutionary success of malaria parasites (2). Genes encoding such factors comprise up to 15% of all parasite genes and belong to various dynamically evolving multi-gene families (3). Characteristic features of these gene families are that they (1) primarily encode proteins exported to the RBC; (2) display high levels of sequence polymorphism between paralogs and across strains, and substantial differences in copy number between species; (3) mostly locate to subtelomeric gene arrays (with the exception of *P. knowlesi* where they occur throughout the genome); and (4) are often species- or clade-specific (3, 4). A prime example of species-specific gene families is the 60-member *var* gene family in *P. falciparum*. Each *var* gene encodes a variant of the major surface antigen *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) that mediates adhesion of iRBCs to several host receptors (5). Members of gene families represented in multiple species include the *Plasmodium* interspersed repeat (*pir*) genes (6), *fam-a*, *-b*, *-c* genes (7), *Plasmodium* helical interspersed subtelomeric

(*phist*) genes (8, 9) or reticulocyte-binding-like (*rbl*) genes (10). Independent of their size and species distribution these gene families provide a fertile ground for genetic diversification and are a driving force of evolutionary adaptation.

A number of studies conducted in *P. falciparum* showed that these multi-gene families are located in heterochromatin (11-13). Heterochromatin is characterized by trimethylation of lysine 9 on histone H3 (H3K9me3) and the consequent binding of heterochromatin protein 1 (HP1), a conserved regulator of heterochromatin formation and heritable silencing (14). *P. falciparum* encodes a single HP1 protein termed PfHP1 (11, 15). In asexual blood stage parasites PfHP1/H3K9me3 demarcate large heterochromatic domains in all subtelomeric regions and in a few internal regions of some chromosomes (11-13). These heterochromatic domains are virtually confined to non-syntenic regions and include over 400 genes, almost all of which are members of multi-gene families (11). As a consequence, these genes are subject to clonally variant expression, providing the parasites with a strong potential for phenotypic diversification and rapid adaptation for instance through antigenic variation or expression of alternative invasion ligands or nutrient transporters (16, 17). In addition, a few single genes are also associated with PfHP1, some of which have orthologs in other *Plasmodium* spp. (11). One of these loci encodes the transcription factor AP2-G, the master regulator of gametocytogenesis (18, 19). PfHP1-dependent silencing of *pfap2-g* prevents sexual commitment, while activation of this locus triggers sexual conversion and subsequent gametocyte differentiation, thus facilitating parasite transmission to the mosquito vector (18-21).

These and other studies provided clear evidence that epigenetic regulation, particularly heterochromatin formation, is central to adaptation and survival of malaria parasites. To date, however, heterochromatin organization has almost exclusively been investigated in *P. falciparum* strain 3D7 blood stage schizonts. It is currently unknown whether the heterochromatin landscape differs between *P. falciparum* strains, whether other *Plasmodium* spp. display similar heterochromatin landscapes, or to what extent HP1 contributes to life cycle stage transitions and parasite differentiation.

Results

Conserved and species-specific aspects of the heterochromatin landscape across the *Plasmodium* genus

To investigate evolutionary aspects of heterochromatin organization we profiled genome-wide HP1 occupancy in multiple *Plasmodium* species by chromatin immunoprecipitation sequencing (ChIP-seq). For *P. falciparum* we used our recently generated polyclonal rabbit α -PfHP1 antibody (20). Guided by a phylogenetic tree constructed from HP1 orthologs (Figure S1) we generated additional polyclonal rabbit antibodies against PvHP1 (to study HP1 in *P. vivax* and *P. knowlesi*) and PbHP1 (to study HP1 in *P. berghei*, *P. chabaudi* and

P. yoelii). Immunofluorescence assays (IFAs) using these antibodies visualized punctate signals in the nuclei of all species that are reminiscent of the perinuclear chromosome end clusters observed in *P. falciparum* (20) (Figures 1A and S1). In Western blot analyses these antibodies detected a protein of the expected size of HP1 in each species (for *P. vivax* Western blot analysis was not performed due to lack of a suitable parasite sample) (Figure S1).

Using these antibodies, we mapped HP1 occupancy in schizonts of *P. falciparum* strain 3D7, *P. berghei* ANKA, *P. chabaudi chabaudi* AS, *P. yoelii yoelii* YM, *P. knowlesi* clone A1-C.1 (22) as well as *P. vivax* field isolates. In all species HP1 predominantly localizes to subtelomeric heterochromatic domains on all chromosomes and to a few internal regions on some chromosomes (Figures 1B and S2). The only exception is *P. knowlesi*, where subtelomeric occupancy is much less pronounced but numerous chromosome-internal HP1-demarcated domains are observed (Figures 1B and S2). We next calculated HP1 enrichment values for each gene and employed a binomial Gaussian mixture model to call HP1-associated genes with high confidence (Figure 1C and Table S1). HP1 occupancy in *P. falciparum* is largely restricted to the *var*, *rif*, *stevor*, *phist*, *pfmc-2tm* and other gene families encoding known or predicted exported proteins in accord with a previous report (11). In *P. vivax* most HP1-occupied genes belong to the *vir* family and members of the *cir*, *bir* and *yir* families make up the majority of HP1-associated genes in *P. chabaudi*, *P. berghei*, and *P. yoelii*, respectively. The dispersed HP1-demarcated domains in *P. knowlesi* capture the *kir* and *SICAv* families and the interstitial telomere repeat sequences (ITSs) that are linked to these loci throughout the genome (4) (Figure 1C, S3 and Table S1).

Most other HP1-associated genes in *P. vivax*, *P. knowlesi* and the three rodent-infecting species are members of gene families encoding other known or predicted exported proteins including *phist*, *stp1*, *fam-a*, *fam-b* and *fam-c* genes (Figure 1C, Table S1) (3). Moreover, several species possess small heterochromatic gene families involved in RBC invasion such as the *pc235*, *pb235* and *py235* genes encoding rho-try proteins (*P. chabaudi*, *P. berghei*, *P. yoelii*) (23), or in metabolism such as *lpl* genes encoding lysophospholipases (*P. falciparum*, *P. vivax*, *P. knowlesi*, *P. chabaudi*, *P. yoelii*) and *acs* genes encoding acyl-CoA synthetases (*P. falciparum*, *P. chabaudi*).

All species also contain a few HP1-associated genes encoding proteins involved in regulation of gene expression, vesicular transport, cell division, RBC invasion, sexual development or transmission (summarized in the category “other” (Figure 1C and Table S1). Notably, while the multi-gene families have no or limited orthology, most of these genes have orthologs including some with conserved synteny across species. The extent of HP1 enrichment at these loci varied across species and most were bound by HP1 only in one species (Tables S1 and S2). However, six conserved syntenic orthologs were associated with HP1 in more than one species (Figure 2A).

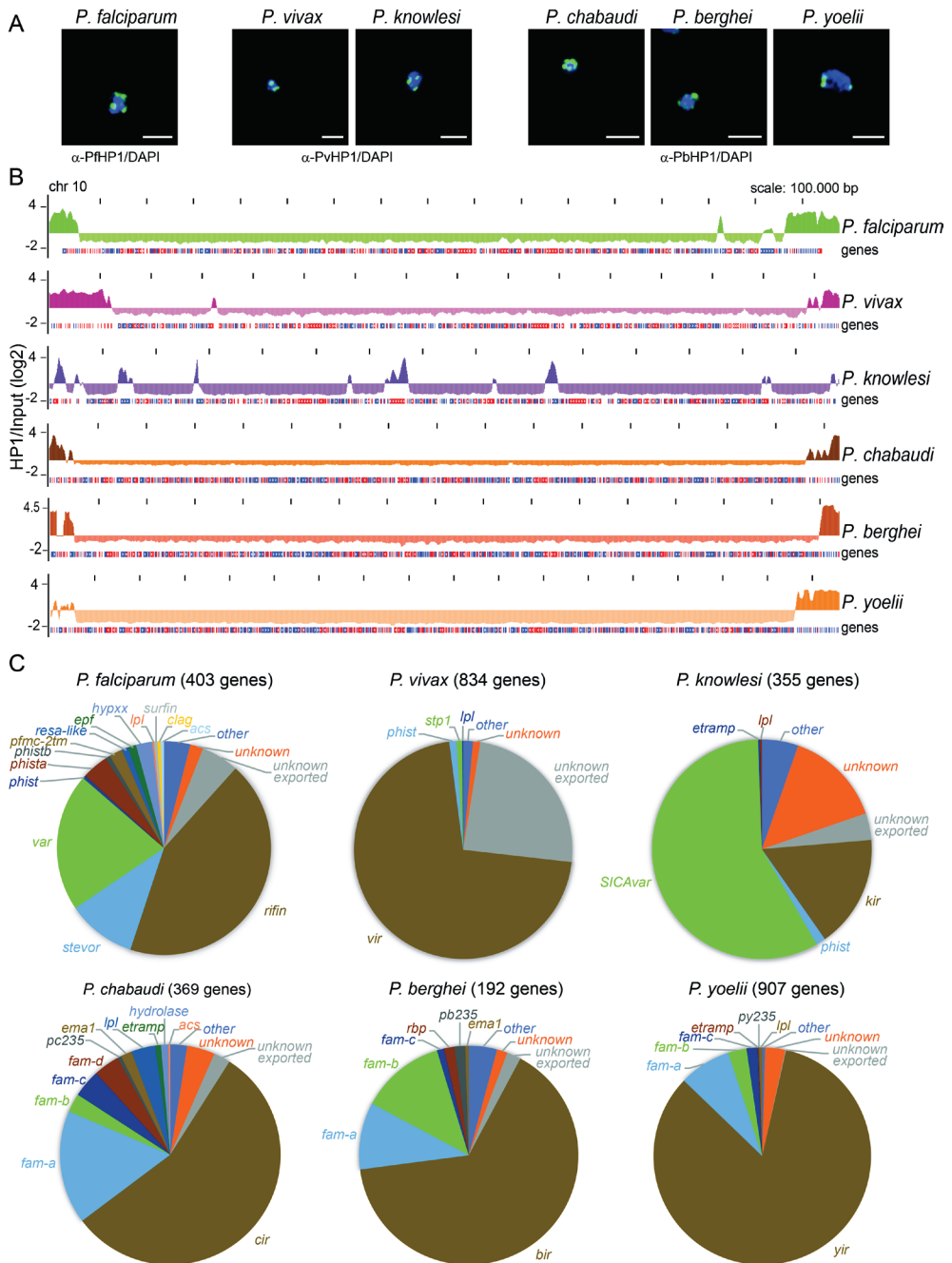


Figure 1. HP1 localisation and genome-wide HP1 occupancy in six different *Plasmodium* species.

(A) IFAs showing HP1 localisation (green) in *P. falciparum* (α -PfHP1 antibodies), *P. vivax* and *P. knowlesi* (α -PvHP1 antibodies), and *P. chabaudi*, *P. berghei*, *P. yoelii* (α -PbHP1 antibodies) trophozoites. Nuclei were stained with DAPI (blue). Scale bar, 2.5 μ m. **(B)** Log₂-transformed

ChIP/Input ratio tracks from schizont stages of six *Plasmodium* species. Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes. **(C)** Relative composition of heterochromatic genes in six *Plasmodium* species, classified into multi-gene families or groups of “unknown”, “unknown exported” and “other”. Numbers indicate the total number of high-confidence heterochromatic genes. See also Figures S1, S2 and S3 and Tables S1 and S2.

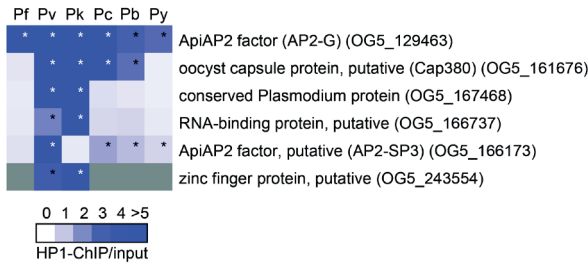
Four of them encode putative transcriptional or post-transcriptional regulators of gene expression, namely the ApiAP2 TFs AP2-G (18, 19) and AP2-SP3/AP2-Tel (24, 25), an RNA-binding protein and a CCCH-type zinc finger (ZnF) protein that is only conserved in *P. vivax* and *P. knowlesi*. Interestingly, *ap2-g* was the only gene with clear HP1 enrichment in all species (Figure 2B), underscoring its crucial role in controlling the switch to sexual differentiation (18-21). *ap2-sp3/ap2-tel* was bound by HP1 in *P. vivax* and the three species infecting rodents.

The two genes encoding the ZnF and RNA-binding proteins and a gene encoding a conserved *Plasmodium* protein of unknown function were significantly enriched only in the *P. vivax/knowlesi* clade (Figure 2A). *cap380*, encoding an oocyst capsule protein essential for oocyst development in *P. berghei* (26), is associated with HP1 in *P. vivax*, *P. knowlesi*, *P. berghei* and *P. chabaudi*, partially marked in *P. falciparum*, but not in *P. yoelii* (Figure 2C). In summary, in all *Plasmodium* species examined most HP1-enriched genes belong to species-, clade- or pan-specific multi-gene families with documented or probable functions in antigenic variation, immune evasion or host cell invasion. In addition, each species contains a small number of HP1-associated single copy genes, many of which are conserved in other *Plasmodium* spp. and have known or predicted roles in fundamental parasite biology.

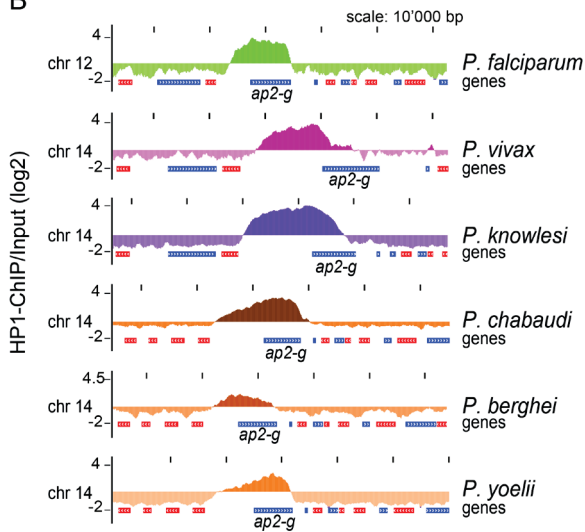
***P. knowlesi* parasites proliferating in macaque or human RBCs display altered PkHP1 occupancy at several loci**

P. knowlesi parasites have been adapted to continuous *in vitro* culture in human RBCs (22, 27). We reasoned that the adaptation to growth in human RBCs may have involved epigenetic changes. We therefore compared the PkHP1 binding profiles of *P. knowlesi* clones A1-C.1 (see above) and A1-H.1, which have been adapted to long-term *in vitro* culture in *M. fascicularis* and human RBCs, respectively (22). Only twelve genes differentially marked by PkHP1 between the two clones (≥ 2.5 -fold change in PkHP1 occupancy) (Table S3). The six loci with higher PkHP1 enrichment in human RBC-adapted parasites encode a KIR protein, a lysophospholipase, a PHIST protein, and three tryptophan-rich antigens (TRAGs) (Figure 3A, Table S3). The six genes with reduced PkHP1 occupancy in human RBC-adapted parasites encode two members of the SICAvAr family, a protein of unknown function, a predicted exported protein, the secreted ookinete protein PSOP7, and a putative histone RNA hairpin-binding protein (Figure 3B, Table S3).

A



B



C

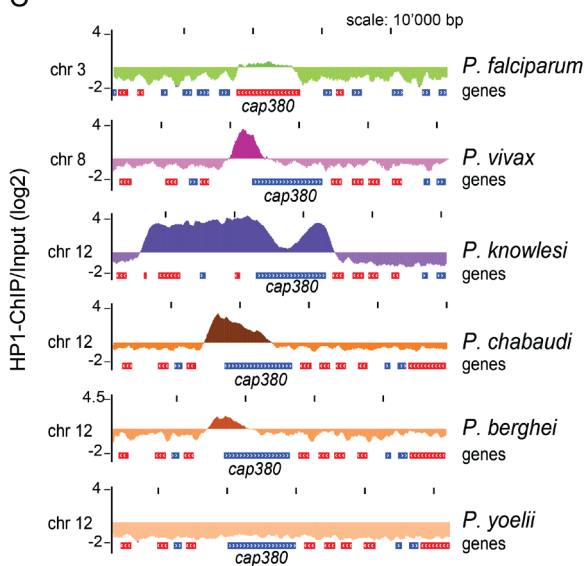


Figure 2. Conserved single copy genes associated with HP1 in more than one species.

(A) HP1 enrichment values for conserved syntenic orthologs in six *Plasmodium* species. Asterisks denote high confidence heterochromatic genes ($p > 0.99999$). **(B)** HP1 enrichment tracks over the *ap2-g* locus in six *Plasmodium* species. Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes. **(C)** HP1 enrichment tracks over the *cap380* locus in six *Plasmodium* species. See also Tables S1 and S2.

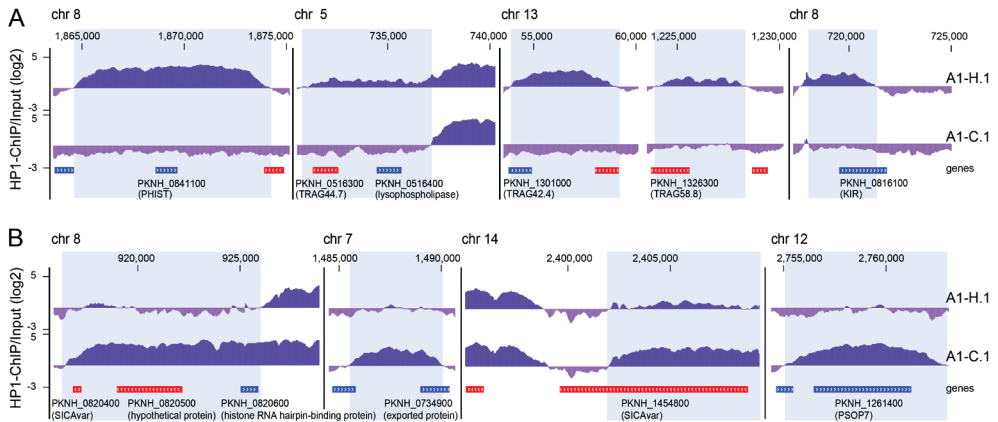


Figure 3. Genes differentially marked by PkHP1 in *P. knowlesi* parasites invading human or macaque RBCs.

(A) HP1 enrichment tracks over the six genes showing >2.5-fold increased HP1 occupancy in *P. knowlesi* A1-H.1 compared to A1-C.1. Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes. **(B)** HP1 enrichment tracks over the six genes showing >2.5-fold increased HP1 occupancy in *P. knowlesi* clone A1-C.1 compared to A1-H.1. See also Table S3.

Noteworthy, TRAG proteins interact with RBC receptors and have proposed roles in invasion (28, 29) and PHIST proteins play central roles in RBC remodelling (9). Furthermore, the *P. knowlesi* gene encoding a *Plasmodium* exported protein of unknown function (PKNH_0734900) has an ortholog in *P. vivax*, a parasite that naturally infects humans. Hence, these epigenetic changes may indeed represent signatures of positive selection during adaptation but replicate *in vitro* selection experiments and further characterization of candidates are needed to test this intriguing possibility experimentally.

Heterochromatin organization is variable between different *P. falciparum* strains

We next profiled HP1 occupancy in *P. falciparum* schizont stages of strain NF54 (30), the NF54-derived clone 3D7 (31), and the recently culture-adapted Ghanaian strain Pf2004 (32, 33) and Cambodian strain NF135 (34). To allow direct comparison of PfHP1 occupancy, all ChIP-seq reads were mapped against the 3D7 genome (PlasmoDB v26). The four strains displayed largely similar heterochromatin organization but distinct PfHP1 occupancy was still evident, predominantly at the border of heterochromatic domains (Figure 4A). Importantly, mapping the Pf2004 ChIP-seq reads against the matching Pf2004 genome revealed that many changes in PfHP1 occupancy occurred in syntenic regions (Figure S4), demonstrating that differences at heterochromatin borders are not solely due to genetic rearrangements.

To identify genes with altered PfHP1 occupancy ChIP/input enrichment values were z-score transformed and grouped using k-means clustering (Figure 4B and Table S4). Of all

heterochromatic genes (clusters 1 to 11; 279 genes) one third displayed variable PfHP1 occupancy across strains (clusters 5 to 11; 88 genes). Interestingly, most of these genes localize close to heterochromatin boundaries (Figure 4B) and show variation in expression between laboratory lines (16) (Table S4). While most *var*, *stevor* and *rifin* genes were stably marked by PfHP1, members of other gene families such as *phist*, *fikk* or *surfin* and genes encoding unknown exported proteins were overrepresented among the variably marked genes (Figures 4C and 4D, Table S4). Pseudogenes were also more abundant in this class, suggesting they may provide “buffer zones” for heterochromatin reorganization. Variable PfHP1 occupancy was also observed for most PfHP1-associated single copy genes and small gene families (category “other”) (Figure 4D). This set includes genes encoding proteins implicated in erythrocyte invasion (*eba-181*) (35), RBC remodeling in gametocytes (*geco*) (36), mosquito midgut invasion (*warp*) (37), sporozoite maturation or egress (*ccp1*, *crmp4*) (38, 39), or liver stage development (*Isa1*) (40). Notably, however, four such genes (*ap2-g*, *clag3.2*, *dblmsp2* and another gene encoding a DBL-domain containing protein) showed stable PfHP1 enrichment in all strains (Figure 4C), suggesting stable heterochromatin inheritance at these loci provides a selective growth advantage *in vitro*. Indeed, depletion of HP1 from the *pfap2-g* locus leads to cell cycle exit and sexual differentiation (20). *dblmsp2* encodes a putative invasion factor expressed only in a small fraction of schizonts (41).

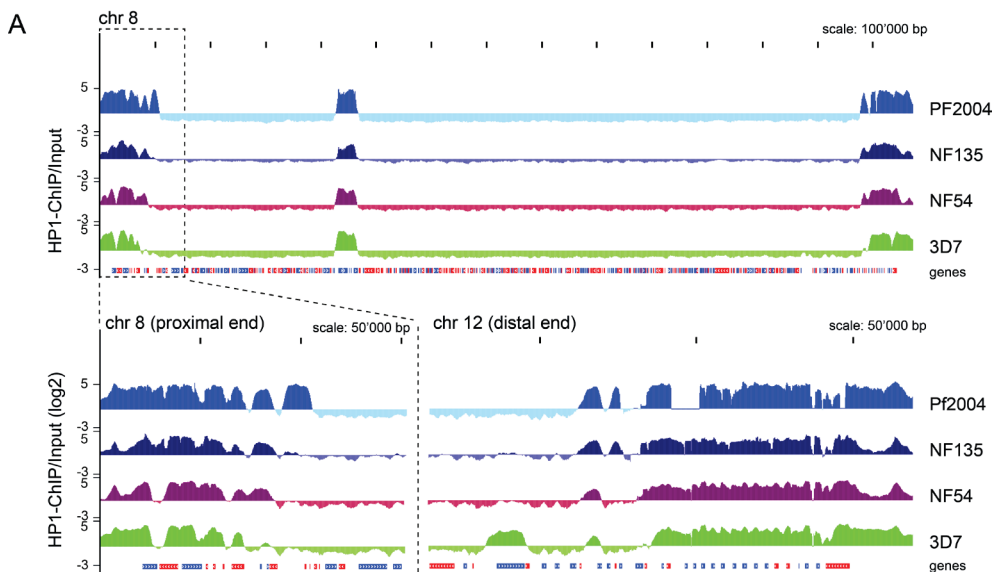
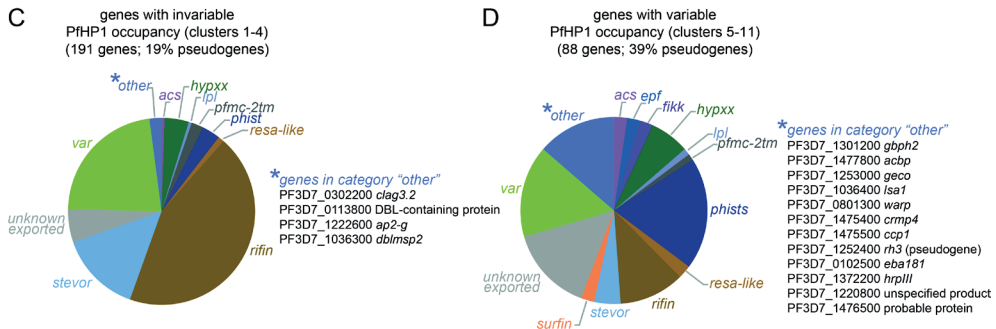
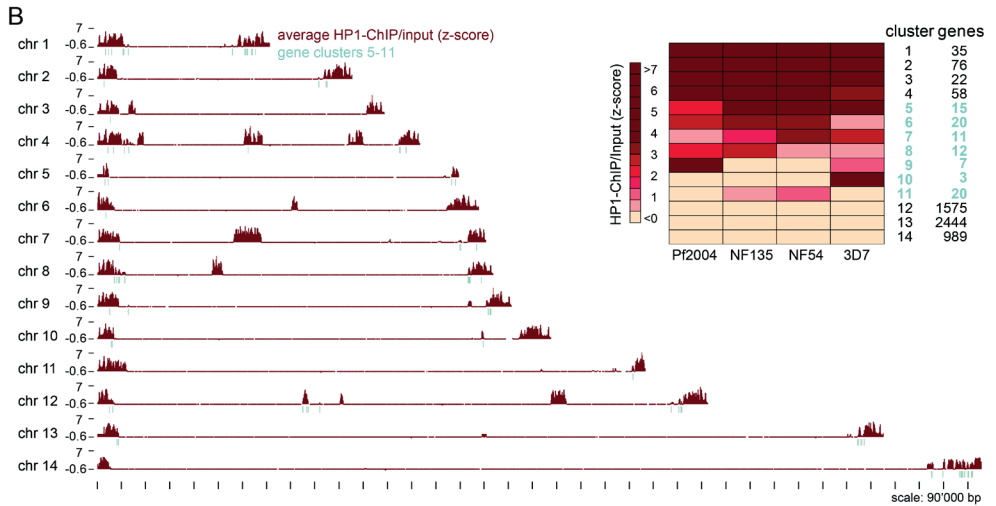


Figure 4. Strain-specific differences in heterochromatin organization in *P. falciparum* schizonts. (A) Log₂-transformed ChIP/Input ratio tracks from *P. falciparum* strains Pf2004, NF135, NF54 and 3D7 schizonts. Chromosome 8 and zoom-ins of its proximal end and the distal end of chromosome 12 are depicted as representative examples. Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes.



(B) Heatmap based on k-means clustering of z-score-transformed ChIP/Input ratios calculated for each gene. Clusters containing genes with variable HP1 occupancy are marked in turquoise. Chromosome maps depict the position of variably marked genes (turquoise) in relation to HP1-demarcated heterochromatin (brown tracks; average z-score-transformed ChIP/Input ratios across the strains calculated in 1000 bp windows). **(C)** Relative composition of invariably marked heterochromatic genes (clusters 1 to 4), classified into multi-gene families or groups of “unknown exported” and “other”. **(D)** Relative composition of variably marked heterochromatic genes (clusters 5 to 11), classified into multi-gene families and groups of “unknown exported” and “other”. See also Figure S4 and Table S4.

clag3.2 and its paralog clag3.1 encode related variants of the surface transport channel PSAC (42); parasites express either one of the two variants and some preferentially express clag3.1 *in vitro* (43, 44). Collectively, these observations highlight a high degree of variability in heterochromatin organization that likely contributes to phenotypic variation of malaria parasites. Interestingly, by analyzing gene expression data from field isolates (45) we found that variably marked heterochromatic genes display a significantly higher degree of expression variation compared to euchromatic genes and, to a lesser extent, also to invariably marked genes, suggesting that this relation may be relevant *in vivo* (Figure S4 and Table S4).

Heterochromatin organization is invariable between different stages of asexual intra-erythrocytic development

To assess whether and to what extent PfHP1-dependent gene expression contributes to the regulation of gene expression during the intra-erythrocytic developmental cycle (IDC) we mapped PfHP1 occupancy in *P. falciparum* 3D7 ring stages, trophozoites and schizonts. The profiles were highly similar in all three stages (Figure 5A). PfHP1 enrichment values of individual genes were highly correlated and we did not identify any genes with significantly altered PfHP1 occupancy across the IDC (Figure 5B and Table S5). Comparison of our data with a RNA-seq dataset (46) confirmed that most PfHP1-associated genes are expressed at low levels during the IDC (11) and that most clonally variant genes are PfHP1 target genes (16) (Figure 5C). Interestingly, genes with lower PfHP1 occupancy showed somewhat higher expression and this set includes many experimentally confirmed clonally variant genes (16).

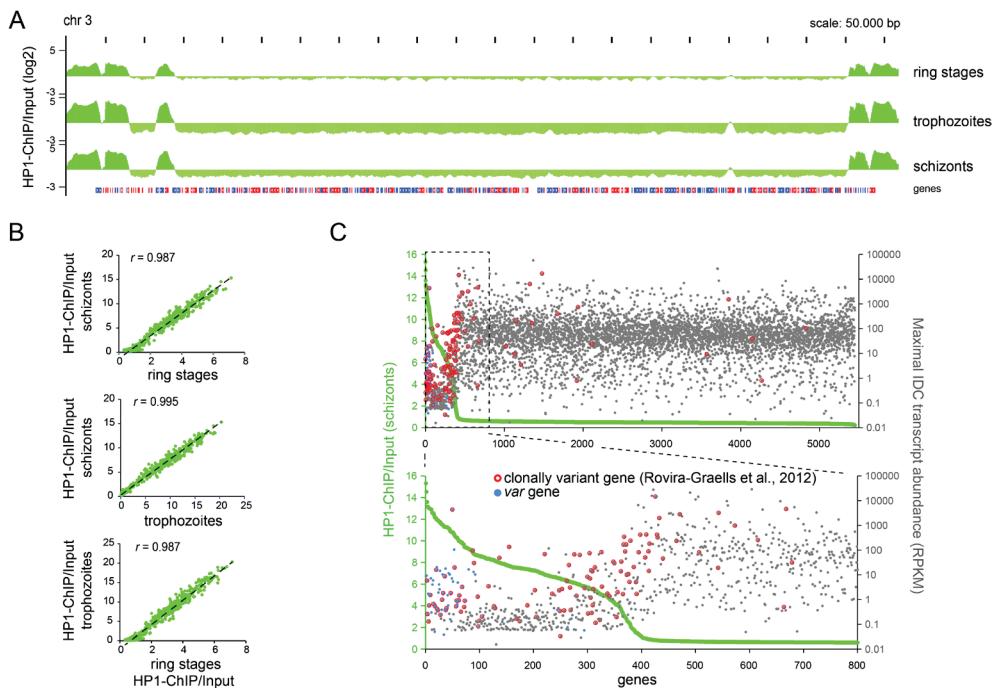


Figure 5. Genome-wide PfHP1 localization is invariable across the IDC.

(A) Log₂-transformed ChIP/Input ratio tracks from *P. falciparum* 3D7 ring, trophozoite and schizont stages. Chromosome 3 is depicted as a representative example. Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes. **(B)** Pairwise comparisons of PfHP1 coverage of individual genes between the three IDC stages. r , Pearson correlation values. **(C)** Scatter plots displaying for each gene the maximum transcript level during the IDC (46) (grey dots) in relation to HP1 occupancy in schizonts (green dots). Genes were sorted according to HP1 occupancy. Genes with clonally variant expression (16) are marked with a red circle. *var* genes are indicated as blue dots. See also Table S5.

var genes appear to be special in this regard since they show moderate expression despite high PfHP1 occupancy levels. In summary, these results reveal that PfHP1-mediated silencing does not contribute in any major way to the temporal regulation of gene expression during the IDC.

The switch from asexual proliferation to sexual differentiation in *P. falciparum* is accompanied by marked changes in the heterochromatin landscape

We performed PfHP1 ChIP-seq experiments on Pf2004 schizonts, stage II/III gametocytes and stage IV/V gametocytes. Mapping the PfHP1 ChIP-seq reads against both the 3D7 and Pf2004 reference genomes highlighted clear differences in PfHP1 occupancy between asexual and sexual stages that were particularly evident from the expansion of subtelomeric heterochromatic domains in gametocytes (Figures 6A, 6B and S5).

Calculation of PfHP1 enrichment values followed by z-score transformation and k-means clustering identified 104 genes with altered PfHP1 occupancy between schizonts and gametocytes (clusters 5 to 8) (Figure 6B and Table S6). Of these, only 15 genes showed reduced PfHP1 occupancy in gametocytes (cluster 5). This set includes *pfap2-g* (18), the gametocyte-specific gene *pfgeco* (36) and seven additional genes at the distal end of chromosome 14 that include five known markers of early gametocytogenesis (*pfg14_744*, *pfg14_748*, PF3D7_1476600, PF3D7_1477400, *gexp17*) (47-49) (Figures 6A, 6B and S5, Table S6). Clusters 6 to 8 contain 89 genes specifically bound by PfHP1 in gametocytes (Figures 6A, 6B and S6, Table S6). Intriguingly, this set is enriched for genes encoding proteins implicated in RBC remodeling. Of particular interest is the subtelomeric region at the left arm of chromosome 2 where the heterochromatic domain is extended by almost 50 kb in gametocytes. This differentially marked region includes three genes encoding proteins involved in knob formation, namely the knob-associated heat shock protein 40 (KAHsp40) (50), PfEMP3 (51) and the knob-associated histidine-rich protein (KAHRP) (52). In addition, five members of the *fikk* family, which encode exported serine-threonine protein kinases implicated in host cell remodeling (53, 54), are enriched in PfHP1 in gametocytes. Increased PfHP1 occupancy is also observed at the gene encoding MESA, an exported protein of unknown function that binds to the RBC membrane skeleton protein 4.1 (55) and at 15 *phist* genes.

In summary, these data suggest that heterochromatin remodeling contributes in a major way to the establishment of a gametocyte-specific transcriptional program. It should be noted, however, that the majority of genes differentially expressed between asexual and sexual blood stages are not marked by HP1 in either stage (11, 56), suggesting that sequence-specific transcription factors such as AP2-G and AP2-G2 (18, 19, 57) are the main drivers of stage-specific gene expression during sexual differentiation.

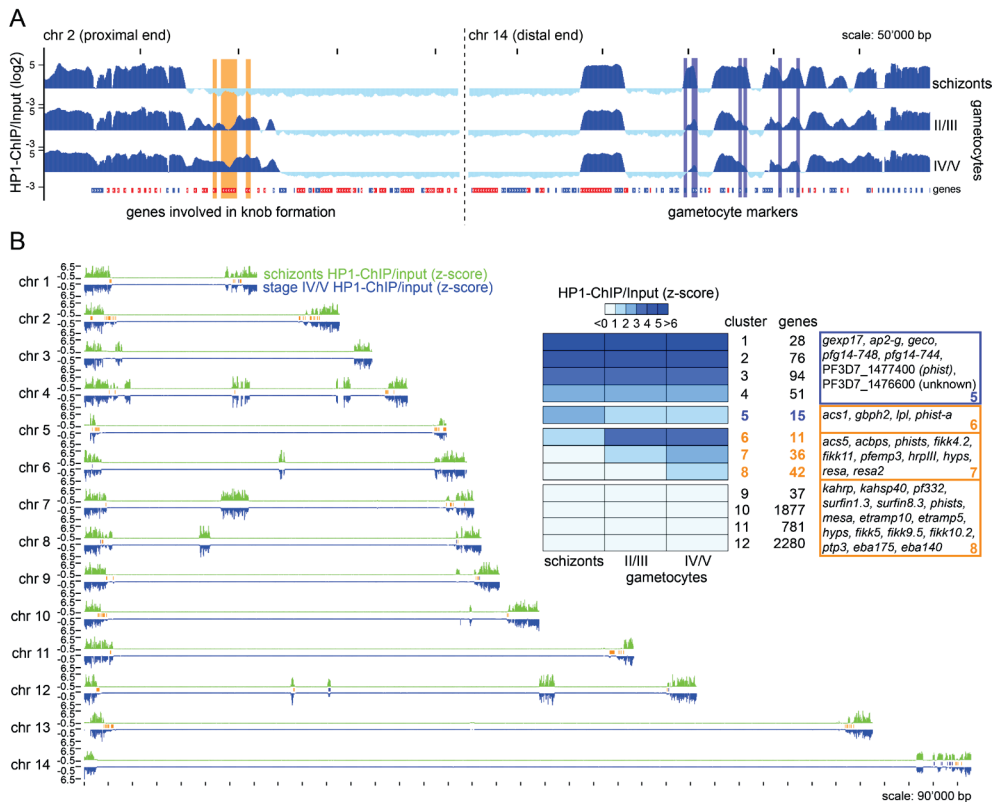


Figure 6. Differences in heterochromatin organization between asexual and sexual *P. falciparum* blood stage parasites.

(A) Log₂-transformed ChIP-seq ChIP/Input ratio tracks from *P. falciparum* Pf2004 schizonts and stage II/III and stage IV/V gametocytes. The proximal end of chromosome 2 is depicted as an example for an expanded heterochromatic domain in gametocytes. Genes involved in knob formation (*kahsp40*, PF3D7_0201800; *pfemp3*, PF3D7_0201900; *kahrp*, PF3D7_0202000) are marked in orange. Early gametocyte markers PF3D7_1476500, PF3D7_1476600, PF3D7_1477300 (*pfg14_744/phist*), PF3D7_1477400 (*phist*), PF3D7_1477700 (*pfg14_748/phista*), PF3D7_1478000 (*gexp17*) at the distal end of chromosome 14 have reduced PfHP1 occupancy in gametocytes and are marked in purple.

(B) Heatmap based on k-means clustering of z-score-transformed ChIP/Input ratios calculated for each gene. Examples of genes with reduced (cluster 5) or increased (clusters 6 to 8) HP1 occupancy in gametocytes are highlighted in purple or orange, respectively. Chromosome maps depict the position of genes with reduced (purple) or increased (orange) HP1 occupancy in gametocytes in relation to PfHP1-demarcated heterochromatin (green tracks and blue inverted tracks are z-score-transformed ChIP/Input ratios in schizonts and stage IV/V gametocytes, respectively, calculated in 1000 bp windows). See also Figure S5 and Table S6.

Discussion

We show that heterochromatin formation at chromosome ends and their perinuclear clustering is a conserved feature of chromatin organization across the *Plasmodium* genus. Furthermore, in all six species examined heterochromatin primarily embeds members of the various species-, clade- or pan-specific multi-gene families with known or predicted roles in antigenic variation and other host-parasite interrelations, independent of chromosomal location. While in *P. berghei*, *P. chabaudi*, *P. yoelii* and *P. vivax* heterochromatin is mostly confined to chromosome ends, *P. falciparum* features some additional intra-chromosomal heterochromatic islands, and in *P. knowlesi* chromosome-internal HP1-demarcated domains are scattered throughout the genome. These differences in heterochromatin distribution mirror the differences in the genome-wide localization of gene families between the species. These observations, in particular the intriguing association of PkHP1 with the numerous individual *kir* and *SICAvar* loci (Figure S3), lends support to the idea that unknown DNA elements linked to sequences of gene family members may be directly involved in the formation and/or local containment of heterochromatin. The ITS elements found at *kir* and *SICAvar* loci are interesting first candidates to be tested for such putative functions.

Some conserved single copy genes are subject to HP1-dependent gene silencing in at least one of the species analyzed. Heterochromatinization of these genes may be used to prevent their expression during certain phases of the life cycle and/or to express them in a clonally variant manner to facilitate alternative phenotypes conducive to parasite adaptation. Even though some of these genes display only subtle HP1 enrichment and may represent false positive hits, this list constitutes a valuable resource for the exploration of genus-, clade- or species-specific heterochromatic genes with functions in key biological processes (Table S2). Here, we focused our attention on the six syntenic orthologs that are heterochromatic in more than one species. Remarkably, next to *ap2-g* this set includes three other genes with likely roles in regulating gene expression. Of these, AP2-SP3/AP2-Tel that has recently been studied in *P. falciparum* and *P. berghei*. In *P. falciparum*, this factor binds the telomeric tract (25) and disruption of *ap2-sp3/ap2-tel* impairs parasite proliferation (58). In *P. berghei*, however, AP2-SP3/AP2-Tel is dispensable for intra-erythrocytic growth and sexual development but essential for sporozoite maturation (24). These rather conflicting findings may be explained by functional divergence of AP2-SP3/AP2-Tel in different malaria parasites. Consistent with this hypothesis we found that *ap2-sp3/ap2-tel* is marked by HP1 in *P. vivax* and the three species infecting rodents, but not in *P. falciparum* and *P. knowlesi*. Another interesting HP1 target gene encodes a putative CCCH-type ZnF protein in *P. vivax* (PVP01_0604500) and *P. knowlesi* (PKNH_0603500). Proteins carrying these domains typically bind RNA and control gene expression by regulating mRNA turnover (59). Hence, this factor may act in a similar way to regulate important processes specifically in the *P. vivax* / *P. knowlesi* clade and it will be interesting to know in which life cycle stage(s) this

may take effect.

A previous study reported a substantial degree of variegated gene expression in *P. falciparum* and showed that most genes affected are located in heterochromatin (16). Their results also suggested that the transcriptional states of individual genes are stably inherited during *in vitro* culture. Our results are consistent with these findings. First, up to one third of all heterochromatic genes showed variable PfHP1 occupancy between strains, which likely contributes to differential gene expression. Second, we observed a complete lack of variation in PfHP1 occupancy between the different IDC stages and found only twelve differentially marked genes between two clones of *P. knowlesi* that have been cultured independently in RBCs from two different hosts for over 200 generations (22). Together, these findings suggest that heterochromatin is faithfully maintained and that heritable changes occur rather infrequently during asexual proliferation *in vitro*.

In many multicellular eukaryotes epigenetic mechanisms are employed in a developmental context to progressively silence groups of genes no longer required in differentiated cells (60). We found that in a somewhat analogous fashion many genes display altered HP1 occupancy between asexually reproducing and sexually differentiating parasites. Importantly, since these cell populations were generated from the same strain in one continuous *in vitro* culture experiment, the observed changes directly reflect the dynamics of heterochromatin re-structuring associated with the cell fate switch. Besides *ap2-g*, a few other early gametocyte-specific genes showed reduced HP1 occupancy already in stage II/III gametocytes, suggesting their de-repression occurred alongside that of *ap2-g* during sexual commitment or in the subsequent gametocyte ring stages. On the contrary, a larger group of genes devoid of HP1 in schizonts became heterochromatinized during gametocyte differentiation. Strikingly, many of these genes encode RBC remodelling factors. The most compelling example is related to the knob structures, parasite-induced aggregates underneath the erythrocyte membrane that are crucial for the PfEMP1-dependent adherence of iRBCs to endothelial cells and their consequent sequestration in the microvasculature (61). Although stage I to IV gametocytes also sequester, primarily in the bone marrow (62), their cytoadhesive properties are markedly different and reflected in the absence of knobs in gametocyte-infected erythrocytes (63, 64). Our findings suggest that the mechanism responsible for preventing expression of these structures in gametocytes is based on HP1-dependent silencing of *kahrp* and other genes linked to knob formation. Given that many additional genes implicated in host cell remodeling also become associated with HP1 in gametocytes, we speculate that *P. falciparum* gametocytes use heterochromatin spreading as a general mechanism to inactivate host cell remodeling machinery that is crucial for the survival of asexual parasites but incompatible with the distinct biology of differentiating gametocytes.

Qualitative comparison between our data and genome-wide H3K9me3 ChIP-seq profiles obtained from *P. falciparum* oocysts and salivary gland sporozoites (65) suggests that

further expansion of heterochromatic domains in these life cycle stages might lead to silencing of yet another set of genes during development in the mosquito (Figure S5). Collectively, our findings reveal that distinct changes in heterochromatin organization accompany developmental stage transitions during parasite transmission, reflecting the different biology, environmental niches, and requirements for rapid adaptive responses associated with each life cycle stage. Such silencing mechanisms must be reversed at some point to enable re-expression of affected genes during the life cycle stages where their expression is required. Two recent studies provided evidence that epigenetic reprogramming during mosquito passage may reset virulence gene expression in *P. chabaudi* (66, 67). Based on our results HP1 likely plays a central role in such a process, and it will be interesting to build on the tools and knowledge generated here to investigate this intriguing possibility in more detail.

In conclusion, we demonstrate that the HP1-dependent silencing of genes implicated in antigenic variation, invasion, or sexual conversion is evolutionary conserved in malaria parasites. We further identify a number of genes that are marked by HP1 specifically in one or a few species only. These may play crucial roles in the adaptive control of species- or clade-specific processes. Our results also reveal that gametocyte differentiation is accompanied by changes in heterochromatin distribution that potentially affect the expression of more than 100 genes. This raises the exciting possibility that despite their large evolutionary distance malaria parasites employ a similar strategy as metazoans to regulate expression of cell type-specific genes via heterochromatinization.

Acknowledgments

We thank Neil Almond and Jo Hall for the supply of *M. fascicularis* blood samples, Sylwia Boltryk for preparing Pf2004 gDNA for PacBio sequencing, Marga van de Vegte-Bolmer and Robert Sauerwein for providing the NF54 and NF135 strains, Pascal Maeser for generating the HP1 phylogenetic tree, Natalie Hofmann for providing a thin smear of a *P. vivax*-infected blood sample and Simon van Heeringen, Wout Megchelenbrink and Lei Zhu for help with bioinformatic analyses. We thank Thomas D. Otto, the Pf3K consortium (www.malariagen.net/projects/pf3k) and the Wellcome Trust Sanger Institute to grant access to the Pf2004 sequence and annotation. We thank the staff and patients attending the Mae Sot Malaria Clinic and Clinics associated with the Shoklo Malaria Research Unit (SMRU). SMRU is part of the Mahidol-Oxford University Research Unit supported by the Wellcome Trust of Great Britain. S.F. received a PhD fellowship from the European Community's Seventh Framework Program under grant agreements No 242095 and No ParaMet 290080. N.M.B.B. received a PostDoc.Mobility fellowship from the Swiss National Science Foundation (P300PA_160975). A.T.M and X.H. acknowledge support through an NTU Graduate Research Scholarship. R.W.M. and F.M. are supported by an MRC Career Development Award funded by the UK MRC and UK Department for International

Development. B.R. is supported by a University of Otago (New Zealand) Start-up grant and the Marsden Fund (17-UOO-241). M.M. is supported by a career development award from the Burroughs Wellcome Fund. P.R.P. is supported by the Singapore Ministry of Education Academic Research Fund Tier 2 (MOE2012-T2-2-093), the Singapore Ministry of Health's National Medical Research Council under its Individual Research Grant NMRC/1292/2011 and A*STAR-UK MRC Joint Grant (10/1/22/24/630). R.B. received support from The Netherlands Organization for Scientific Research (NWO-Vidi 864.11.007). T.S.V. is supported by grants from the Swiss National Science Foundation (31003A_163258, 31003A_143916,).

Author contributions

S.A.F. and M.F. designed and performed experiments, analysed data, prepared illustrations and wrote the paper. I.N. produced affinity-purified α -PbHP1 and α -PvHP1 antibodies. N.M.B.B. produced crosslinked Pf2004 samples and M.M. supervised these experiments. X.Y.Y., A.T.M. and X.H. produced crosslinked *P. yoelii*, *P. chabaudi* and *P. berghei* samples and carried out IFAs and Western blots. R.H. carried out data analysis, prepared illustrations and wrote parts of the paper. B.R. and P.R.C. conducted experimental work related to the collection and processing of *P. vivax* blood samples. F.N. provided the diagnostic, clinical and ethical oversight for the collection of *P. vivax* from malaria patients. R.W.M. and F.M. produced crosslinked *P. knowlesi* samples and carried out IFAs and Western blots. Z.B. provided conceptual advice. P.R.P. designed experiments, analysed data and wrote parts of the manuscript. R.B. and T.S.V. conceived and supervised this study, analysed data, prepared illustrations and wrote the manuscript.

Declaration of interests

The authors declare no competing interests

Star methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Till S. Voss (till.voss@swisstp.ch).

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|--|-----------------|
| Antibodies | | |
| Antibodies | | |
| Rabbit polyclonal anti- <i>Plasmodium falciparum</i> HP1 | Brancucci et al., 2014 | N/A |
| Rabbit polyclonal anti- <i>Plasmodium berghei</i> HP1 | This paper | N/A |
| Rabbit polyclonal anti- <i>Plasmodium vivax</i> HP1 | This paper | N/A |
| Alexa Fluor 488-conjugated Goat anti-Rabbit IgG | ImmunoJackson | Cat#111-545-003 |
| Alexa Fluor 488-conjugated Goat anti-Rabbit IgG | Thermo Fisher Scientific | Cat#A-11008 |
| Bacterial and Virus Strains | | |
| <i>E. coli</i> Rosetta 2 (DE3) | EMD Millipore | Cat#71397-3 |
| Biological Samples | | |
| <i>Plasmodium vivax</i> clinical isolates | Shoklo Malaria Research Unit, Thailand | N/A |
| <i>Plasmodium vivax</i> thin blood smear | Swiss TPH / PNGIMR | N/A |
| Chemicals, Peptides, and Recombinant Proteins | | |
| <i>Plasmodium berghei</i> HP1 | This paper | N/A |
| <i>Plasmodium vivax</i> HP1 | This paper | N/A |
| Dynabeads Protein A | Thermo Fisher Scientific | Cat#10008D |
| Dynabeads Protein G | Thermo Fisher Scientific | Cat#10009D |
| Nycodenz | Axis-Shield | Cat#1002424 |
| Formaldehyde solution 36.5-38% | Sigma-Aldrich | Cat#F8775 |
| ProLong Antifade mountant with DAPI | Thermo Fisher Scientific | Cat#P36931 |
| VECTASHIELD mounting medium containing DAPI | Vector Laboratories | Cat#H-1200 |
| cOmplete™, Mini Protease Inhibitor Cocktail | Sigma-Aldrich | Cat#11836153001 |
| Critical Commercial Assays | | |
| NEXTflex® ChIP-Seq Barcodes - 24 | Bio Scientific | Cat#NOVA-514122 |
| KAPA HiFi HotStart ready mix | KAPA Biosystems | Cat#KM2602 |
| Agencourt AMPure XP | Beckman Coulter | Cat#A63880 |
| NextSeq 500/550 High Output v2 kit (75 cycles) | Illumina | Cat#FC-404-2005 |
| Deposited Data | | |

| | | |
|---|---|---|
| ChIP-Seq data | This paper | GEO: GSE102695 |
| Pf2004 draft genome assembly | Pf3K consortium and the Wellcome Trust Sanger Institute | ftp://ftp.sanger.ac.uk/pub/project/pathogens/Plasmodium/falciparum/PF3K/SecondSetReferenceGenomes/DraftAnnotation/Pf2004/ . |
| Experimental Models: Cell Lines | | |
| Parasite strain: <i>Plasmodium falciparum</i> 3D7 | Walliker et al., 1987 | Alan Cowman, WEHI, Melbourne, Australia |
| Parasite strain: <i>Plasmodium falciparum</i> Pf2004/164tdT | Brancucci et al., 2015 | N/A |
| Parasite strain: <i>Plasmodium falciparum</i> NF54 | Deleamarre and van der Kaay, 1979 | Robert Sauerwein, Radboudumc, Nijmegen, NL |
| Parasite strain: <i>Plasmodium falciparum</i> NF135 | Teirlinck et al., 2013 | Robert Sauerwein, Radboudumc, Nijmegen, NL |
| Parasite strain: <i>Plasmodium knowlesi</i> A1-H.1 | Moon et al., 2013 | Mike Blackman, Francis Crick Institute, London, UK |
| Parasite strain: <i>Plasmodium knowlesi</i> A1-C.1 | Moon et al., 2013 | Mike Blackman, Francis Crick Institute, London, UK |
| Parasite strain: <i>Plasmodium berghei</i> ANKA | BEI Resources (MR4) | N/A |
| Parasite strain: <i>Plasmodium chabaudi chabaudi</i> AS | Spence et al., 2011 | Jean Langhorne, Francis Crick Institute, London |
| Parasite strain: <i>Plasmodium yoelii yoelii</i> YM | BEI Resources (MR4) | N/A |
| Experimental Models: Organisms/Strains | | |
| Mouse: BALB/c (female) | In Vivos Pte Ltd. Singapore | NA |
| Oligonucleotides | | |
| Pb_F: aaaagatttcacatgacaggatcagatg | This paper | N/A |
| Pb_R: ttcctcgagcaccgttctatatctaagtc | This paper | N/A |
| SUMO_F: tttcatatgcatcatcatcatcacgggtcggactcagaagtcaatc | This paper | N/A |
| SUMO_R: cctaggatccggcgccaccaatctgttctctgtg | This paper | N/A |
| Pv_F: actggatccgatgaagagtttgaatagg | This paper | N/A |
| Pv_R: tgtgctcgactactaggccgttcggtatcg | This paper | N/A |
| Recombinant DNA | | |
| pET20b(+) | EMD Millipore | Cat#69739-3 |
| pET_PbHP1-6xHis | This paper | N/A |
| pETA-HS | This paper | N/A |
| pETA-HS-PvHP1 | This paper | N/A |
| pETA-Strep | This paper | N/A |
| pETA_Strep-PvHP1-6xHis | This paper | N/A |
| Software and Algorithms | | |

| | | |
|---|------------------------------|---|
| Clustal Omega | Sievers et al., 2011 | https://www.ebi.ac.uk/Tools/msa/clustalo/ |
| MEGA7 | Kumar et al., 2016 | http://www.megasoftware.net/download_form |
| Olympus DP manager software (v2.2.1.195) | Olympus | N/A |
| ImageJ | Schneider et al., 2012 | https://imagej.net/Downloads |
| Nikon Elements Advanced Research | Nikon | N/A |
| Leica IM1000 software | Leica | N/A |
| Fiji | Schinderlin et al., 2012 | https://imagej.net/Fiji |
| jvenn | Bardou et al., 2014 | http://jvenn.toulouse.inra.fr/app/index.html |
| BWA (v0.7.12-r1039) | Li and Durbin, 2009 | https://insidedna.me/tool_page_assets/pdf_manual/bwa.pdf |
| SAMtools (v1.2) | Li and Handsaker et al, 2009 | http://www.htslib.org/download/ |
| BEDTools (v2.20.1) | Quinlan and Hall, 2010 | http://bedtools.readthedocs.io/en/latest/content/installation.html |
| SignalMap Software v2.0 | Roche | http://sequencing.roche.com/en/products-solutions/by-category/target-enrichment/software/signal-map-software.html |
| UCSC Genome Browser | UCSC Genome Browser | https://genome-store.ucsc.edu/ |
| Rstudio (v3.3.2) | RStudio | https://www.rstudio.com/products/rstudio/download/ |
| Other | | |
| Plasmidpure filters | EuroProxima | Cat#8011Filter25u |
| HisTrap HP | GE Healthcare | Cat#17-5248-01 |
| HiTrap Protein A HP | GE Healthcare | Cat#17-0403-01 |
| HiTrap NHS activated HP column | GE Healthcare | Cat#17-0716-01 |
| StrepTrap HP | GE Healthcare | Cat#28-9136-30 |
| Amicon Ultra Centrifugal Filter 10KDa | EMD Millipore | Cat#UFC801024 |
| E-Gel Size Select agarose gel | Thermo Fisher Scientific | Cat#G661012 |
| Non-woven fabric filters | ZXBio.net | N/A |
| Glycerolyte Solution 57 | Fenwal | Cat#FWL4A7831 |
| BD Vacutainer™ Plastic Blood Collection Tubes with Sodium Heparin: Conventional Stopper | Fisher Scientific | Cat#BD367874 |
| BD Vacutainer™ Plastic Blood Collection Tubes with K2EDTA: Tube Stopper | Fisher Scientific | Cat#BD367844 |

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse model

Mice used in this study (BALB/c mice; age 6-8 weeks; weight 25-30 g) were maintained in accordance with the NAELAR (National Advisory Committee for Laboratory Animal Research) guidelines under the Animal & Birds (Care and Use of Animals for Scientific Purposes) Rules of Singapore with approval from the Institutional Animal Care and Use Committee (IACUC) of Nanyang Technological University (NTU) of Singapore (Approval number: ARFSBS/NIE A002). All animals used in this study were obtained from InVivos Pte Ltd and subsequently housed under SPF conditions at NTU. *P. berghei*, *P. chabaudi* and *P. yoelii* infections for obtaining parasites for chromatin extraction were performed on female BALB/c mice (age 6-8 weeks; weight 25-30g). Mice were infected with an initial inoculum of 5×10^5 parasites and were exsanguinated by cardiac puncture when parasitaemia levels reached between 10-20%.

P. falciparum parasites

P. falciparum parasites were cultured at 37°C at 5% haematocrit based on the original protocol published by Trager and Jensen (68). Growth synchronization was achieved by repeated sorbitol treatments (69). 3D7 parasites were cultivated with AB⁺ human RBCs in RPMI 1640/25 mM Hepes standard culture medium supplemented with 0.5% Albumax II. NF54 and NF135 parasites were cultivated with O⁺ human RBCs in RPMI 1640/25 mM Hepes standard culture medium supplemented with 10 % human serum. Pf2004/164tdTom parasites (33) were cultivated with AB⁺ human RBCs in RPMI 1640/25 mM Hepes standard culture medium supplemented with 10 % human serum. Pf2004/164tdTom gametocytes were generated by inducing sexual commitment as described (33). After re-invasion cultures were treated with 50mM N-acetylglucosamine (70) for three consecutive days to eliminate asexual parasites.

P. knowlesi parasites

P. knowlesi A1-H.1 parasites were grown at 37°C in O⁺ human RBCs obtained from the United Kingdom National Blood Transfusion Service. *P. knowlesi* A1-C.1 parasites were grown in *M. fascicularis* blood provided by NIBSC (UK), which was collected by venous puncture into K2 EDTA BD Vacutainers (Fisher Scientific) as described previously (22). Samples of *M. fascicularis* blood used for parasite culture were provided by the National Institute for Biological Standards and Control. The rationale and procedures for venepuncture and blood sample collection were reviewed by the local Animal Welfare and Ethical Review Body (the Institutional Review Board) of the National Institute for Biological Standards and Control and performed under licence (PPL70/8506) granted by the United Kingdom Home Office as governed by United Kingdom law under the Animals (Scientific Procedures) Act 1986. Animals were handled in strict accordance with the “Code of Practice Part 1 for the housing and care of animals (21/03/05)” available at <https://www.gov.uk/research-and-testing-using-animals>. The work also met the National Centre for the Replacement Refinement and Reduction of Animals in Research (NC3Rs) guidelines on primate accommodation, care, and use (<https://www.nc3rs.org.uk/non-human-primate-accommodation-care-and-use>), which exceed the legal minimum standards required by the United Kingdom Animals (Scientific Procedures) Act 1986, associated Codes of Practice, and the US Institute for Laboratory Animal Research Guide. Parasite cultures were synchronised by centrifugation through a density cushion of Nycodenz (Axis-Shield) as

previously described (22).

P. vivax parasites

The clinical *P. vivax* isolates examined in this study were collected under the following approved ethical guidelines and protocols: OXTREC 45-09 and OXTREC 17-11 (Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, Oxford, UK), MUTM 2008-215 from the Ethics committee of the Faculty of Tropical Medicine (Mahidol University, Bangkok, Thailand) and MRAC No. 16.01 from the Medical Research Advisory Committee of Papua New Guinea. To obtain *P. vivax* parasites for chromatin extraction eight clinical isolates were collected from malaria patients attending clinics run by the Shoklo Malaria Research Unit, Mae Sot, Thailand. Five milliliters of whole blood were collected in lithium heparin collection tubes by venepuncture from each patient using Sodium Heparin BD Vacutainers (Fisher Scientific). After leukocyte depletion using non-woven fabric filters (ZXBio.net) these samples were cryopreserved in glycerolyte 57 solution (Fenwal) and stored in liquid nitrogen (71). Frozen samples were thawed using a series of NaCl gradients (12%, 1.6% and 0.9%) and matured *ex vivo* for 40 hours as described (71, 72).

METHOD DETAILS

P. falciparum sample collection and chromatin preparation

Chromatin from 3D7 ring stages (8-16 hpi), trophozoites (32-30 hpi) or schizonts (40-48 hpi) (approximately $0.75-1.5 \times 10^9$ parasites each) and from Pf2004/T164dTom schizonts, stage II/III gametocytes (day four after re-invasion) or stage IV/V gametocytes (day nine after re-invasion) (approximately 2×10^8 parasites each) was prepared by crosslinking cultures with 1% formaldehyde (Sigma-Aldrich) for 15 min at 37°C. Crosslinking reactions were quenched by 0.125 M glycine. Nuclei were isolated by releasing parasites from iRBCs using 0.05% saponin followed by lysis in CLB (20 mM HEPES, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.65% NP-40, 1mM DTT, 1x protease inhibitor (Sigma-Adrich), pH 7.9). Nuclei were washed and snap-frozen in CLB supplemented with 50% glycerol. Chromatin from NF54 and NF135 schizont stages (approximately $2-4 \times 10^9$ parasites each) was prepared by passing the cultures through Plasmodipure filters (EuroProxima) to remove white blood cells prior to formaldehyde crosslinking for 10 min at 37°C and quenching in 0.125 M glycine. Nuclei were isolated by releasing parasites from iRBCs using 0.05% saponin followed by gentle homogenisation (pestle B, 15 strokes) in CLB2 (10 mM Tris-HCl, 3 mM MgCl₂, 0.2% NP40, 1x protease inhibitor (Sigma-Adrich), pH 8.0) and centrifugation through a 0.25 M sucrose cushion (in CLB2) at 2000 rpm for 10 min at 4°C. Nuclei were snap-frozen in CLB2 supplemented with 20% glycerol. Frozen nuclei were thawed and resuspended in sonication buffer (50mM Tris-HCl, 1% SDS, 10mM EDTA, 1x protease inhibitor (Sigma-Adrich), pH 8.0) and sonicated for 20-24 cycles of 30 sec ON/30 sec OFF (setting high, Bioruptor™ Next Gen, Diagenode). Chromatin fragment sizes ranged from 100-600 bp as determined by de-crosslinking a 50 µl aliquot and running the purified DNA on a 1% agarose gel.

P. knowlesi sample collection and chromatin preparation

Cultures containing a schizont parasitaemia of around 5% were passed through Plasmodipure filters (EuroProxima) to remove white blood cells prior to crosslinking for 10 min at 37°C in 1% formaldehyde (Sigma-Aldrich) and quenching in 0.125 M glycine. Red blood cells were removed by lysis on ice for 10 min with 0.15% saponin/PBS before washing the pellet in PBS and snap-freezing the parasite pellets

in liquid nitrogen. Nuclei were isolated by lysis in CLB, washed and aliquots corresponding to approximately 1×10^9 nuclei were snap-frozen in CLB supplemented with 50% glycerol. Preparation of sheared chromatin was performed as described above for *P. falciparum*.

P. berghei, *P. chabaudi* and *P. yoelii* sample collection and chromatin preparation

For each of the three parasite species, whole blood of infected mice containing approximately 5×10^9 schizonts (three mice for *P. yoelii yoelii* YM, four mice for *P. chabaudi chabaudi* AS, five mice for *P. berghei* ANKA) was diluted in standard *P. falciparum* culture medium and passed through a Plasmodipure filter (EuroProxima) to remove white blood cells. The purified RBCs were collected by centrifugation at 2'000 rpm for 5 min, resuspended in 30 ml culture medium and crosslinked at 37°C for 10 min in presence of 1% formaldehyde (Sigma-Aldrich). Crosslinking reactions were quenched by 0.125 M glycine. The crosslinked RBCs suspension was split into three equal aliquots, centrifuged at 2'000 rpm for 5 min, supernatants were removed and the RBC pellets snap-frozen in liquid nitrogen. Nuclei were isolated by releasing parasites from iRBCs using 0.05% saponin followed by lysis in CLB (20 mM Hepes, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.65% NP-40, 1mM DTT, 1x protease inhibitor (Sigma-Adrich), pH 7.9). Again, nuclei were washed and aliquots corresponding to approximately 1×10^9 nuclei were snap-frozen in CLB supplemented with 50% glycerol. Preparation of sheared chromatin was performed as described above for *P. falciparum*.

P. vivax sample collection and chromatin preparation

The *P. vivax ex vivo* schizont cultures were crosslinked at 37°C for 10 min in presence of 1% formaldehyde (Sigma-Aldrich) and subsequently the reactions were quenched by 0.125 M glycine. The crosslinked RBCs were centrifuged at 200 g for 5 min, supernatants were removed and the RBC pellets snap-frozen in liquid nitrogen. The eight samples were thawed and pooled and nuclei isolated by releasing parasites from iRBCs using 0.05% saponin followed by lysis in CLB. Nuclei were washed and snap-frozen in CLB supplemented with 50% glycerol. Preparation of sheared chromatin was performed as described above for *P. falciparum*.

Phylogenetic analysis of *Plasmodium* HP1 orthologs

Protein sequences of *Plasmodium* HP1 orthologs were downloaded from PlasmoDB v33 and used to perform a multiple sequence alignment using Clustal Omega (73) with default parameters. Phylogenetic tree construction was done with MEGA7 (74) using the Neighbor-joining method and 1'000 bootstrap replicates.

Generation and affinity purification of α -PbHP1 and α -PvHP1 antibodies

All recombinant proteins were expressed in Rosetta2(DE3) cells (EMD Millipore) using auto-induction (75). The sequence encoding PbHP1 was amplified from gDNA using primers Pb_F (aaaagattcatatgacagatcagatg) and Pb_R (ttccctcgagaccgttctatatctaagtc) and cloned into pET20b(+) (EMD Millipore) using *Nde*I and *Xho*I restriction sites in order to express PbHP1 fused to a C-terminal 6xHis tag (pET_PbHP1-6xHis). Recombinant PbHP1-6xHis was purified using a HisTrap HP column (GE Healthcare) using buffer NiB (50 mM H₃PO₄, 0.5 M NaCl, 20 mM imidazole, pH 7.4) supplemented with 8 M urea for lysis, binding and washing, and buffer NiE (50mM H₃PO₄, 0.5 M NaCl, 225 mM imidazole, pH 7.4) containing 8 M urea for elution. The elution was diluted 1:4 with H₂O and the eluted proteins were precipitated with trichloroacetic acid (TCA). PvHP1 was expressed as an N-terminally tagged

6xHis-SUMO fusion protein (HS-PvHP1). The parental expression vector pETA-HS was generated by introducing a sequence encoding a fusion tag consisting of a 6xHis stretch followed by *Saccharomyces cerevisiae* SUMO, amplified from gDNA using primers SUMO_F (tttcatatgcatcatcatcatcatcacgggtcggactcagaagtcaatc) and SUMO_R (cctaggatccggcggccaccaatctgttctctgtg) between the *NdeI* and *BamHI* sites of pET20(b)+ (EMD Millipore), yielding a vector similar to the one described by Malakhov and colleagues (76). The *pvhp1* insert was cloned into pETA-HS by ligating a *BamHI/XhoI*-digested PCR product amplified from gDNA using primers Pv_F (actggatccgatgaagagtttgaatagg) and Pv_R (tgtgctcgcgactactaggccgttcggtatcg) (pETA-HS-PvHP1). HS-PvHP1 was purified using a HisTrap HP column (GE Healthcare) and buffers NiB for lysis, binding and washing, and buffer NiE for elution. The elution was subject to buffer exchange with NiB supplemented with 1 mM Tris-(2-carboxyethyl)-phosphin (TCEP) and the HS-PvHP1 fusion protein was digested using recombinant SUMO protease (L403-K621 of *S. cerevisiae* ULP1; GB1-UPL1-6xHis). The cleaved tag and SUMO protease were subtracted using a HisTrap HP column (GE Healthcare) and the purified untagged PvHP1 protein was precipitated with TCA.

Purified recombinant PbHP1-6xHis and untagged PvHP1 were used to immunize rabbits (Pacific Immunology). Total rabbit IgG from anti-PbHP1 and anti-PvHP1 immune sera were purified using HiTrap Protein A HP columns (GE Healthcare) using a mild arginine elution method similar to the one described by Arakawa and colleagues (77), with the exception that a linear combined pH and arginine gradient elution was used instead of stepwise elution. Rabbit sera were diluted 1:3 using buffer IgGA (750 mM L-arginine, 150 mM H₃PO₄, 150 mM citric acid, pH 7.3), bound to 5 ml HiTrap Protein A HP columns (GE Healthcare) and washed with five column volumes (CVs) of buffer IgGA. Antibodies were eluted using a linear gradient (eight CVs) of buffers IgGA to IgGB (2 M L-arginine, 150 mM H₃PO₄, 150 mM citric acid, adjusted to pH 3.7 using HCl). The antibodies eluted in a symmetrical peak (maximum at 1.4 M arginine and pH 5). Purified antibodies were subject to buffer exchange with PBS.

Affinity purification of α -PbHP1 antibodies was done as previously described for α -PfHP1 antibodies (20) with the exception that the PbHP1-6xHis antigen was bound to the nickel column in buffer NiB containing 2 M urea. For the PvHP1 antigen, we first generated the parental pETA_Strep vector facilitating expression of N-terminally Strep(II)-tagged and C-terminally 6xHis-tagged fusion proteins by replacing the *NdeI/BamHI* fragment in pET20b(+) with an annealed double-stranded oligonucleotide (Strep_F (tatggtagctggagccaccgcagttcgaaaaag) and Strep_R (gatcctttttcgaactcgggtggctccagctagcca)) encoding Met-Ala-Ser-Strep(II). Next, the same PCR product that was used to generate HS_PvHP1 (see above) was cloned into the pETA_Strep vector using *BamHI* and *XhoI* to obtain pETA_Strep-PvHP1-6xHis. The Strep(II)-PvHP1-6xHis fusion protein was purified using a HisTrap HP column (GE Healthcare) and the same buffers as used for purification of HS-PvHP1 (see above) followed by a StrepTrap HP column (GE Healthcare) using buffer NiB containing 1 mM EDTA for washing or 2.5 mM desthiobiotin for elution. The purified protein was subject to buffer exchange with coupling buffer (0.2 M NaHCO₃, 0.5 M NaCl, pH 8.3). Next, the protein was coupled to a HisTrap NHS-activated HP column (GE Healthcare) following the supplier's instructions. α -PvHP1 IgG was diluted 1:5 in buffer IgGA and bound to the Strep(II)-PvHP1-6xHis column. The column was washed with 20 CVs of buffer IgGA and eluted with buffer IgGB. Both, α -PvHP1 and α -PbHP1 antibodies were finally subject to buffer exchange with PGS (20 mM H₃PO₄, 30 mM KOH, 25% glycerol, 250 mM Na₂SO₄, pH 6.8) and concentrated to 0.4-0.7 mg/ml using an Amicon Ultra spin filter with a 10K cutoff (EMD Millipore).

Fluorescence microscopy

IFAs for *P. falciparum* were performed as described previously (20). IFAs for *P. berghei*, *P. chabaudi* and *P. yoelii* were performed with acetone:methanol (9:1)-fixed cells using rabbit α -PbHP1 (1:250) and Alexa Fluor 488-conjugated α -rabbit IgG (1:500) (ImmunoJackson). Slides were viewed under Olympus IX71 fluorescence microscope using a 100x oil immersion objective and equipped with an Olympus DP30BW camera. Images were acquired via the Olympus DP manager software (v2.2.1.195) and processed using ImageJ (v1.440) (78). IFAs for *P. knowlesi* were performed using blood smears fixed with 4% paraformaldehyde for 30 min followed by three washes in PBS and permeabilisation in 0.1% Triton-X100 for 10 min. Slides were blocked overnight at 4°C in 3% BSA/PBS and then labelled with rabbit α -PvHP1 (1:600) and Alexa Fluor 488-conjugated α -rabbit IgG (1:5'000) (Thermo Fisher Scientific). The smear was mounted in ProLong Antifade mountant with DAPI (Thermo Fisher Scientific). Slides were viewed with a Nikon Ti E inverted microscope using a 100x oil immersion objective and imaged with an ORCA Flash 4.0 CMOS camera (Hamamatsu). Images were acquired and processed using the Nikon Elements Advanced Research software package. IFAs for *P. vivax* were performed using methanol-fixed thin blood smears. Slides were blocked using 3% BSA/PBS and then labeled with α -PvHP1 (1:500) and Alexa Fluor 488- conjugated α -rabbit IgG (1:500) (Thermo Fisher Scientific) antibodies in 3% BSA/PBS. Slides were mounted using VECTASHIELD mounting medium containing DAPI (Vector Laboratories). Images were taken at 100-fold magnification on a Leica DM 5000B microscope with a Leica DFC 345 FX camera, acquired via the Leica IM1000 software, and processed using Fiji (79). For each experiment, images were acquired and processed with identical settings.

Western Blot

P. berghei, *P. chabaudi* and *P. yoelii* schizonts were enriched using a 50-60% Histodenz (Sigma-Aldrich) gradient. Parasites were released from iRBCs by saponin lysis, resuspended in Urea extraction buffer (EMD Millipore) and separated by SDS-PAGE. Proteins were detected using rabbit α -PbHP1 (1:2'000) antibodies. *P. knowlesi* A1-H.1 schizonts were enriched on a density cushion of 55% Nycodenz (Axis-Shield) stock solution (27.6% (wt/vol) Nycodenz in 10 mM Hepes, pH 7.0) in RPMI-1640 medium. RBCs were lysed using 0.15% saponin/PBS and the resultant parasite pellet was diluted 1:100 in Urea extraction buffer (40 mM Tris-HCl, 1 mM EDTA, 8 M Urea, 5% SDS, 1x protease inhibitor cocktail (Sigma-Adrich), 1% β -mercaptoethanol), mixed with SDS sample buffer and separated by SDS-PAGE alongside a similarly treated uninfected RBC control. PkHP1 was detected with rabbit α -PvHP1 antibodies (1:5'000).

Chromatin immunoprecipitation

For each ChIP reaction, sonicated chromatin containing 500 ng of DNA was incubated in incubation buffer (0.75% SDS, 5% Triton-X100, 750 mM NaCl, 5 mM EDTA, 2.5 mM EGTA, 100 mM Hepes, pH 7.4) with either 1 μ g rabbit α -PfHP1 (for *P. falciparum*), 1 μ g rabbit α -PvHP1 (for *P. vivax* and *P. knowlesi*) or 1 μ g rabbit α -PbHP1 (for *P. berghei*, *P. chabaudi* and *P. yoelii*) as well as 10 μ l protA and 10 μ l protG Dynabeads suspension (Thermo Fisher Scientific).

For each sample four ChIP reactions were prepared and incubated overnight at 4°C while rotating. Beads were washed twice with wash buffer 1 (0.1% SDS, 0.1% DOC, 1% Triton-X100, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes, pH 7.4), once with wash buffer 2 (0.1% SDS, 0.1% DOC, 1% Triton-X100, 500 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes, pH 7.4), once with wash buffer

3 (250 mM LiCl, 0.5% DOC, 0.5% NP-40, 1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes, pH 7.4) and twice with wash buffer 4 (1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes, pH 7.4). Each wash was performed for 5 min at 4°C while rotating. Subsequently, immunoprecipitated chromatin was eluted in elution buffer (1% SDS, 0.1M NaHCO₃) at room temperature. The eluted chromatin samples and the corresponding input samples (sonicated input chromatin containing 500 ng DNA) were de-crosslinked in 1% SDS/0.1 M NaHCO₃/1 M NaCl at 45°C overnight while shaking. For each parasite strain or species the separate ChIP samples were combined and the DNA was purified using QIAquick MinElute PCR columns (Qiagen).

High throughput sequencing

For each sequencing library 2-10 ng of ChIP or input DNA were end-repaired, extended with 3' A-overhangs and ligated to barcoded NextFlex adapters (Bio Scientific) as described previously (80). Libraries were amplified (98°C for 2 min; four cycles 98°C for 20 sec, 62°C for 3 min; 62°C for 5 min) using KAPA HiFi HotStart ready mix (KAPA Biosystems) and NextFlex primer mix (Bio Scientific) as described (46). 225-325 bp fragments (including the 125 bp NextFlex adapter) were size-selected using a 2% E-Gel Size Select agarose gel (Thermo Fisher Scientific) and amplified by PCR for eight or ten cycles (Table S7) under the same condition as described above. Library purification and removal of adapter dimers was performed with Agencourt AMPure XP beads in a 1:1 library:beads ratio (Beckman Coulter). ChIP-seq libraries were sequenced for 75 bp single-end reads using the NextSeq 500/550 High Output v2 kit (Illumina) on the Illumina NextSeq 500 system.

QUANTIFICATION AND STATISTICAL ANALYSIS

High throughput sequencing data analysis

Using BWA samse (v0.7.12-r1039) (81) sequencing reads were mapped against the respective reference genomes available on PlasmoDB v26, namely *P. berghei* ANKA, *P. chabaudi chabaudi*, *P. yoelii yoelii* YM, *P. falciparum* 3D7, *P. knowlesi* H and *P. vivax* P01 (PlasmoDB v29). Reads from the *P. falciparum* Pf2004 ChIP-seq libraries were additionally mapped against a Pf2004 genome assembly, which was obtained after long-read PacBio sequencing in the framework of the Pf3K reference project. Mapped reads were filtered to mapping quality ≥ 15 (SAMtools v1.2) (82) and only uniquely mapped reads (3.4-22 million reads for α -HP1 ChIP samples and 6.5-55 million reads for input samples) were used for further analysis (Table S7). ChIP-seq data were visualized in the UCSC Genome browser (<https://genome-store.ucsc.edu/>). All libraries were normalized to the number of mapped reads per million (RPM) and bedgraph files were generated using BEDtools (v2.20.1) (83). For log₂ ratio tracks α -HP1 ChIP values were divided by input values and log₂-transformed using BEDtools (v2.20.1) (83). Within the UCSC genome browser tracks were smoothed and the windowing function was set as 'mean'.

To calculate the HP1 coverage for individual genes, tags were counted in a 1000 bp window (ATG \pm 500 bp) for each coding sequence and offset by +1 to avoid division by zero while calculating fold changes in coverage. α -HP1 ChIP-seq and input tag counts were normalized to the number of reads per kb per million mapped reads (RPKM). ChIP-seq enrichment values were calculated as α -HP1 ChIP [RPKM]/input [RPKM]. Genes encoded by the mitochondrial or apicoplast genomes and nuclear genes with low mappability (input RPKM < 5) were excluded from downstream analysis.

To visualize the genome-wide HP1 coverage in schizont stages of the six *Plasmodium* species

(*P. berghei*, *P. yoelii*, *P. chaubaudi*, *P. knowlesi*, *P. vivax*, *P. falciparum*) the respective reference genomes were divided into 1000 bp windows using BEDtools (v2.20.1) (83). For each window ChIP-seq enrichment values were calculated as described above, log₂-transformed and visualized using the software SignalMap v2.0 (www.sequencing.roche.com). Windows with less than five tag counts in the ChIP-seq and/or the input sample were set to '0' and defined as regions with low mappability. To identify and compare the sets of heterochromatic genes across the six *Plasmodium* species ChIP-seq enrichment values were calculated and assigned to either a 'heterochromatic' or 'euchromatic' compartment. To do so, we fitted a bivariate Gaussian mixture model to the data and calculated the probabilities (p) for genes to belong to either one of the two compartments using the modelling tool 'normalmixEM' from the R package 'mixtools'. For further analysis genes with $p > 0.99999$ for the 'heterochromatic' compartment were considered high confidence heterochromatic genes. Genes with $0.99999 > p > 0.95$ were considered potential heterochromatic genes and genes with $p < 0.95$ were placed in the 'euchromatic' compartment. To look for orthologs and syntenic orthologs, high confidence heterochromatic genes ($p > 0.99999$) for each species were imputed into PlasmoDB v33. For instance, the orthologs (syntenic/non-syntenic) for *P. falciparum* heterochromatic genes (403 genes) were transformed into orthologs of *P. berghei* ANKA, *P. yoelii yoelii* YM, *P. chaubaudi chaubaudi*, *P. knowlesi* strain H and *P. vivax* P01 using the function 'transform by orthology' in PlasmoDB. Similarly, the orthologs for heterochromatic genes in *P. vivax* (834 genes), *P. knowlesi* (355 genes), *P. berghei* (192 genes), *P. chaubaudi* (369 genes) and *P. yoelii* (907 genes) were individually transformed into orthologs of the five other *Plasmodium* species. Ortholog sets (syntenic/non-syntenic) among the species were identified using jvenn (Bardou et al., 2014) and assigned according to the species identifier pf, pv, pk, pb, pc or py (Tables S1 and S2).

To investigate the association between heterochromatic region and distribution of *kir* and *SICAvar* genes as well as interstitial telomere repeat sequences (ITs) PkHP1 coverage in *P. knowlesi* schizont stages was depicted as described above. Coding sequences of *kir*/*kir*-like and *SICAvar* genes were depicted according to their genomic coordinates within the *P. knowlesi* H genome (PlasmoDB v26). ITs were identified by searching the *P. knowlesi* genome for occurrences of GGGTTTA or GGGTTCA repeats on both strands using regular expression. The number of these sequences were counted at every 100 bp window and windows with three or more hits were considered (imperfect repeats with mismatches were not considered). To compare PkHP1 occupancy between two clones of *P. knowlesi* (A1-H.1 and A1-C.1) the ratio between PkHP1 occupancy values (ChIP/Input) were calculated for each gene. Based on visual inspection of the UCSC Genome browser tracks we considered 2.5-fold difference as a marked and likely influential change in PkHP1 occupancy. Genes with low mappability (input RPKM < 5) in at least one of the clones were excluded from downstream analyses. To allow direct comparison of HP1 gene coverage across different *P. falciparum* strains (3D7, Pf2004, NF135 and NF54) PfHP1 ChIP-seq reads from all four strains were mapped against the *P. falciparum* 3D7 reference genome (PlasmoDB v26). ChIP-seq reads from strain Pf2004 schizonts were additionally mapped against the Pf2004 reference genome (Table S7). Note that although matching reference genomes do exist for NF54 and NF135 (Plas_falc_NF54_v1 and Plas_falc_NF135_5_C10_v1; <http://protists.ensembl.org>) the assembly and annotation of these reference genomes are fragmented and not sufficiently informative for analysis of heterochromatin organization. Genes with low mappability (input RPKM < 5) in at least one of the strains were excluded from downstream analyses. For the remaining genes ChIP-seq enrichment values were z-score transformed, k-means

clustered and depicted as heatmap using the R package 'pheatmap'. To visualize the average HP1 occupancy for all strains investigated in this study (Figure 4B), the 3D7 reference genome was divided into 1000 bp windows using BEDtools (v2.20.1) (83). For each window ChIP-seq enrichment values were calculated as described above, z-score transformed, averaged across the strains and visualized using the software SignalMap v2.0 (www.sequencing.roche.com). Coding sequences of the genes in k-means clusters 5 to 11 were depicted according to their location within the *P. falciparum* 3D7 genome (PlasmoDB v26).

For the comparison of intra-erythrocytic stages, ChIP-seq enrichment values were calculated using ChIP RPKM values of ring stages (8-16 hpi), trophozoites (24-32 hpi) or schizonts (40-48 hpi) and input RPKM values obtained from the schizont sample. The relation between PfHP1 gene coverage and transcript abundance was visualized by sorting PfHP1 ChIP-seq enrichment values for each gene in schizont stages from high to low, plotted against the corresponding transcript abundance value (RPKM). For the latter, directional RNA-seq data from eight intra-erythrocytic stages (46) were aligned against the annotated *P. falciparum* 3D7 transcriptome (PlasmoDB v26) and filtered for uniquely mapped reads and mapping quality ≥ 15 . Reads were separated according to the strand they mapped to (sense strand FLAG16; antisense strand FLAG0). Reads aligning to the sense strand were used for further analysis. For each transcript (excluding mitochondrial RNA and apicoplast RNA) tags were counted, offset by +1 and normalized to the number of reads per kb per million mapped reads. For each gene the maximum transcript abundance value (RPKM) observed during intra-erythrocytic development was plotted in the scatter plot and *var* genes as well as genes displaying clonally variant expression (16) were specifically highlighted.

To assess differences in PfHP1 occupancy between Pf2004 schizonts, stage II/III gametocytes and stage IV/V gametocytes ChIP-seq reads were mapped against the *P. falciparum* 3D7 genome (PlasmoDB v26) and the *P. falciparum* Pf2004 reference genome (Table S7). Genes with low mappability (input RPKM < 5) in at least one of the stages were excluded from downstream analysis. For the remaining genes ChIP-seq enrichment values were z-score transformed, k-means clustered and depicted as a heatmap using the R package 'pheatmap'. To visualize the genome-wide PfHP1 occupancy for schizont stages and stage IV/V gametocytes (Figure 6C) the *P. falciparum* 3D7 reference genome was divided into 1000 bp windows using BEDtools (v2.20.1) (83). For each window ChIP-seq enrichment values were calculated as described above, z-score transformed and visualized using the software SignalMap v2.0 (www.sequencing.roche.com). Coding sequences of the genes in k-means clusters 5 and 6 to 8 were depicted according to their location within the *P. falciparum* 3D7 genome (PlasmoDB v26). Additionally, we visually compared our Pf2004 PfHP1 ChIP-seq data with the midgut oocyst and salivary gland sporozoite H3K9me3 ChIP-seq datasets generated by Gómez-Díaz and colleagues (65) using the UCSC genome browser (<https://genome-store.ucsc.edu/>). Midgut oocyst and salivary gland sporozoite H3K9me3 ChIP-seq and input data (GEO accession numbers GSM1981878, GSM1981880, GSM1981883, GSM1981885) were aligned to the *P. falciparum* 3D7 reference genome (PlasmoDB v26) (Table S7) and processed as described above to generate bedgraph \log_2 H3K9me3-ChIP/Input ratio files.

Data availability

The ChIP-seq data have been deposited in the Gene Expression Omnibus repository under ID code GSE102695. The sequence and annotation of the Pf2004 genome is available at

<ftp://ftp.sanger.ac.uk/pub/project/pathogens/Plasmodium/falciparum/PF3K/SecondSetReferenceGenomes/DraftAnnotation/Pf2004/>.

References

1. S. L. Perkins, Malaria's many mates: past, present, and future of the systematics of the order Haemosporida. *J. Parasitol.* **100**, 11-25 (2014).
2. L. S. Swapna, J. Parkinson, Genomics of apicomplexan parasites. *Crit Rev. Biochem. Mol. Biol.* **52**, 254-273 (2017).
3. A. J. Reid, Large, rapidly evolving gene families are at the forefront of host-parasite interactions in Apicomplexa. *Parasitology* **142 Suppl 1**, S57-S70 (2015).
4. A. Pain *et al.*, The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**, 799-803 (2008).
5. J. D. Smith, J. A. Rowe, M. K. Higgins, T. Lavstsen, Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. *Cell Microbiol.* **15**, 1976-1983 (2013).
6. D. Cunningham *et al.*, The *pir* multigene family of *Plasmodium*: antigenic variation and beyond. *Mol. Biochem. Parasitol.* **170**, 65-73 (2010).
7. T. D. Otto *et al.*, A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol.* **12**, 86- (2014).
8. T. J. Sargeant *et al.*, Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* **7**, R12- (2006).
9. J. D. Warncke, I. Vakonakis, H. P. Beck, *Plasmodium* Helical Interspersed Subtelomeric (PHIST) Proteins, at the Center of Host Cell Remodeling. *Microbiol. Mol. Biol. Rev.* **80**, 905-927 (2016).
10. K. Gunalan *et al.*, The role of the reticulocyte-binding-like protein homologues of *Plasmodium* in erythrocyte sensing and invasion. *Cell Microbiol.* **15**, 35-44 (2013).
11. C. Flueck *et al.*, *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog.* **5**, e1000569- (2009).
12. J. J. Lopez-Rubio, L. Mancio-Silva, A. Scherf, Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host. Microbe* **5**, 179-190 (2009).
13. A. M. Salcedo-Amaya *et al.*, Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A* **106**, 9655-9660 (2009).
14. G. Lomberk, L. Wallrath, R. Urrutia, The Heterochromatin Protein 1 family. *Genome Biol.* **7**, 228- (2006).
15. K. Perez-Toledo *et al.*, *Plasmodium falciparum* heterochromatin protein 1 binds to trimethylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res.* **37**, 2596-2606 (2009).
16. N. Rovira-Graells *et al.*, Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res* **22**, 925-938 (2012).
17. T. S. Voss, Z. Bozdech, R. Bartfai, Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr. Opin. Microbiol.* **20**, 88-95 (2014).
18. B. F. Kafsack *et al.*, A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* **507**, 248-252 (2014).
19. A. Sinha *et al.*, A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* **507**, 253-257 (2014).

20. N. M. Brancucci *et al.*, Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host. Microbe* **16**, 165-176 (2014).
21. B. I. Coleman *et al.*, A *Plasmodium falciparum* histone deacetylase regulates antigenic variation and gametocyte conversion. *Cell Host. Microbe* **16**, 177-186 (2014).
22. R. W. Moon *et al.*, Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc. Natl. Acad. Sci. U. S. A* **110**, 531-536 (2013).
23. J. Iyer *et al.*, Invasion of host cells by malaria parasites: a tale of two protein families. *Mol. Microbiol.* **65**, 231-249 (2007).
24. K. Modrzynska *et al.*, A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the *Plasmodium* Life Cycle. *Cell Host. Microbe* **21**, 11-22 (2017).
25. M. Sierra-Miranda *et al.*, PfAP2Tel, harbouring a non-canonical DNA-binding AP2 domain, binds to *Plasmodium falciparum* telomeres. *Cell Microbiol.* **19**, (2017).
26. P. Srinivasan, H. Fujioka, M. Jacobs-Lorena, PbCap380, a novel oocyst capsule protein, is essential for malaria parasite survival in the mosquito. *Cell Microbiol.* **10**, 1304-1312 (2008).
27. C. Lim *et al.*, Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat. Commun.* **4**, 1638- (2013).
28. K. Tyagi *et al.*, Recognition of Human Erythrocyte Receptors by the Tryptophan-Rich Antigens of Monkey Malaria Parasite *Plasmodium knowlesi*. *PLoS. ONE.* **10**, e0138691- (2015).
29. M. Zeeshan *et al.*, Host-parasite interaction: selective Pv-fam-a family proteins of *Plasmodium vivax* bind to a restricted number of human erythrocyte receptors. *J. Infect. Dis.* **211**, 1111-1120 (2015).
30. B. J. Delemarre, H. J. van der Kaay, [Tropical malaria contracted the natural way in the Netherlands]. *Ned. Tijdschr. Geneesk.* **123**, 1981-1982 (1979).
31. D. Walliker *et al.*, Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661-1666 (1987).
32. S. R. Elliott *et al.*, Antibody recognition of heterologous variant surface antigens after a single *Plasmodium falciparum* infection in previously naive adults. *Am J Trop Med Hyg* **76**, 860-864 (2007).
33. N. M. Brancucci *et al.*, An assay to probe *Plasmodium falciparum* growth, transmission stage formation and early gametocyte development. *Nat. Protoc.* **10**, 1131-1142 (2015).
34. A. C. Teirlinck *et al.*, NF135.C10: a new *Plasmodium falciparum* clone for controlled human malaria infections. *J. Infect. Dis.* **207**, 656-660 (2013).
35. T. W. Gilberger *et al.*, A novel erythrocyte binding antigen-175 paralogue from *Plasmodium falciparum* defines a new trypsin-resistant receptor on human erythrocytes. *J. Biol. Chem.* **278**, 14480-14486 (2003).
36. B. J. Morahan *et al.*, Functional analysis of the exported type IV HSP40 protein PfGECO in *Plasmodium falciparum* gametocytes. *Eukaryot. Cell* **10**, 1492-1503 (2011).
37. M. Yuda *et al.*, von Willebrand Factor A domain-related protein, a novel microneme protein of the malaria ookinete highly conserved throughout *Plasmodium* parasites. *Mol. Biochem. Parasitol.* **116**, 65-72 (2001).
38. N. Simon *et al.*, Sexual stage adhesion proteins form multi-protein complexes in the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **284**, 14537-14546 (2009).
39. B. Douradinha *et al.*, *Plasmodium* Cysteine Repeat Modular Proteins 3 and 4 are essential for malaria parasite transmission from the mosquito to the host. *Malar. J.* **10**, 71- (2011).
40. S. A. Mikolajczak *et al.*, Disruption of the *Plasmodium falciparum* liver-stage antigen-1 locus causes a differentiation defect in late liver-stage parasites. *Cell Microbiol.* **13**, 1250-1260 (2011).

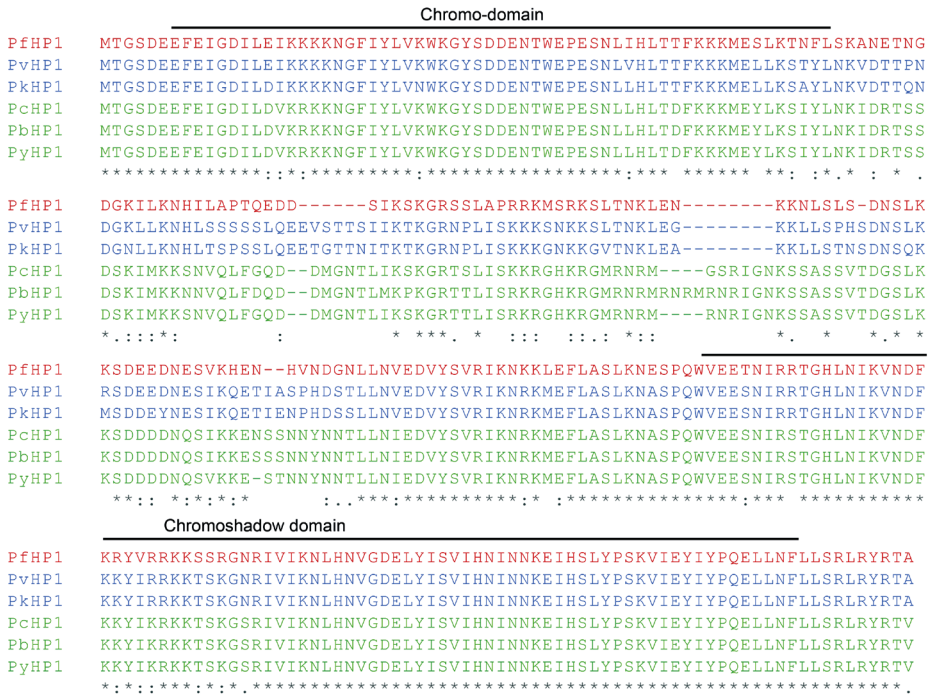
41. A. Amambua-Ngwa *et al.*, Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS. Genet.* **8**, e1002992- (2012).
42. W. Nguitragool *et al.*, Malaria parasite clag3 genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell* **145**, 665-677 (2011).
43. A. Cortes *et al.*, Epigenetic silencing of Plasmodium falciparum genes linked to erythrocyte invasion. *PLoS Pathog.* **3**, e107- (2007).
44. C. A. Comeaux *et al.*, Functional analysis of epigenetic regulation of tandem RhopH1/clag genes reveals a role in Plasmodium falciparum growth. *Mol. Microbiol.* **80**, 378-390 (2011).
45. S. Mok *et al.*, Drug resistance. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* **347**, 431-435 (2015).
46. P. R. Kensche *et al.*, The nucleosome landscape of Plasmodium falciparum reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Res.* **44**, 2110-2124 (2016).
47. S. Eksi *et al.*, Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in Plasmodium falciparum. *Mol. Biochem. Parasitol.* **143**, 90-99 (2005).
48. F. Silvestrini *et al.*, Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics* **9**, 1437-1448 (2010).
49. S. Eksi *et al.*, Plasmodium falciparum gametocyte development 1 (Pfgdv1) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS. Pathog.* **8**, e1002964- (2012).
50. P. Acharya, S. Chaubey, M. Grover, U. Tatu, An exported heat shock protein 40 associates with pathogenesis-related knobs in Plasmodium falciparum infected erythrocytes. *PLoS. ONE.* **7**, e44605- (2012).
51. B. L. Pasloske *et al.*, Cloning and characterization of a Plasmodium falciparum gene encoding a novel high-molecular weight host membrane-associated protein, PfEMP3. *Mol. Biochem. Parasitol.* **59**, 59-72 (1993).
52. L. G. Pologe, J. V. Ravetch, A chromosomal rearrangement in a P. falciparum histidine-rich protein gene is associated with the knobless phenotype. *Nature* **322**, 474-477 (1986).
53. M. C. Nunes, J. P. Goldring, C. Doerig, A. Scherf, 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 (2007).
54. L. M. Kats *et al.*, An exported kinase (FIKK4.2) that mediates virulence-associated changes in Plasmodium falciparum-infected red blood cells. *Int. J. Parasitol.* **44**, 319-328 (2014).
55. K. L. Waller *et al.*, Mature parasite-infected erythrocyte surface antigen (MESA) of Plasmodium falciparum binds to the 30-kDa domain of protein 4.1 in malaria-infected red blood cells. *Blood* **102**, 1911-1914 (2003).
56. J. A. Young *et al.*, The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol. Biochem. Parasitol.* **143**, 67-79 (2005).
57. M. Yuda, S. Iwanaga, I. Kaneko, T. Kato, Global transcriptional repression: An initial and essential step for Plasmodium sexual development. *Proc. Natl. Acad. Sci. U. S. A* **112**, 12824-12829 (2015).
58. B. Balu, N. Singh, S. P. Maher, J. H. Adams, A genetic screen for attenuated growth identifies genes crucial for intraerythrocytic development of Plasmodium falciparum. *PLoS. ONE.* **5**, e13282- (2010).
59. M. Fu, P. J. Blackshear, RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins. *Nat. Rev. Immunol.* **17**, 130-143 (2017).

60. J. S. Becker, D. Nicetto, K. S. Zaret, H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* **32**, 29-41 (2016).
61. J. A. Boddey, A. F. Cowman, Plasmodium nesting: remaking the erythrocyte from the inside out. *Annu. Rev. Microbiol.* **67**, 243-269 (2013).
62. R. Joice *et al.*, Plasmodium falciparum transmission stages accumulate in the human bone marrow. *Sci. Transl. Med.* **6**, 244re5- (2014).
63. R. E. Sinden, Gametocytogenesis of Plasmodium falciparum in vitro: an electron microscopic study. *Parasitology* **84**, 1-11 (1982).
64. M. Tiburcio *et al.*, Early gametocytes of the malaria parasite Plasmodium falciparum specifically remodel the adhesive properties of infected erythrocyte surface. *Cell Microbiol.* **15**, 647-659 (2013).
65. E. Gomez-Diaz *et al.*, Epigenetic regulation of Plasmodium falciparum clonally variant gene expression during development in Anopheles gambiae. *Sci. Rep.* **7**, 40655- (2017).
66. T. Brugat *et al.*, Antibody-independent mechanisms regulate the establishment of chronic Plasmodium infection. *Nat. Microbiol.* **2**, 16276- (2017).
67. P. J. Spence *et al.*, Vector transmission regulates immune control of Plasmodium virulence. *Nature* **498**, 228-231 (2013).
68. W. Trager, J. B. Jensen, Cultivation of malarial parasites. *Nature* **273**, 621-622 (1978).
69. C. Lambros, J. P. Vanderberg, Synchronization of Plasmodium falciparum erythrocytic stages in culture. *J. Parasitol.* **65**, 418-420 (1979).
70. Q. L. Fivelman *et al.*, Improved synchronous production of Plasmodium falciparum gametocytes in vitro. *Mol. Biochem. Parasitol.* **154**, 119-123 (2007).
71. C. Borlon *et al.*, Cryopreserved Plasmodium vivax and cord blood reticulocytes can be used for invasion and short term culture. *Int. J. Parasitol.* **42**, 155-160 (2012).
72. B. Russell *et al.*, A reliable ex vivo invasion assay of human reticulocytes by Plasmodium vivax. *Blood* **118**, e74-e81 (2011).
73. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539- (2011).
74. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).
75. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207-234 (2005).
76. M. P. Malakhov *et al.*, SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5**, 75-86 (2004).
77. T. Arakawa *et al.*, Elution of antibodies from a Protein-A column by aqueous arginine solutions. *Protein Expr. Purif.* **36**, 244-248 (2004).
78. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675 (2012).
79. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682 (2012).
80. W. A. Hoesjmakers, R. Bartfai, K. J. Francoijs, H. G. Stunnenberg, Linear amplification for deep sequencing. *Nat. Protoc.* **6**, 1026-1036 (2011).
81. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**, 1754-1760 (2009).
82. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* **25**, 2078-2079 (2009).
83. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* **26**, 841-842 (2010).

Supplementary information

SUPPLEMENTARY FIGURES

A



B

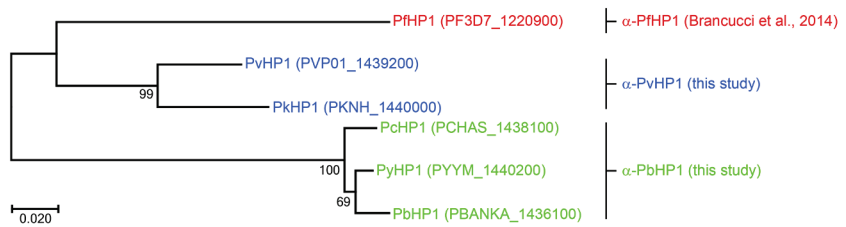


Figure S1. Phylogenetic analysis of HP1 orthologs in *Plasmodium* species (related to Figures 1 and S2, and STAR Methods).

(A) Multiple sequence alignment of *Plasmodium* HP1 orthologs. Conserved residues are marked by asterisks. The positions of the PfHP1 chromo- and chromoshadow domains (11) are indicated above the alignment. (B) Phylogenetic tree of *Plasmodium* HP1 orthologs. Bootstrap values are indicated at the nodes (based on 1'000 replicates). Antibodies used to immune-precipitate HP1-associated chromatin from the six *Plasmodium* species for ChIP-seq experiments are shown on the right.

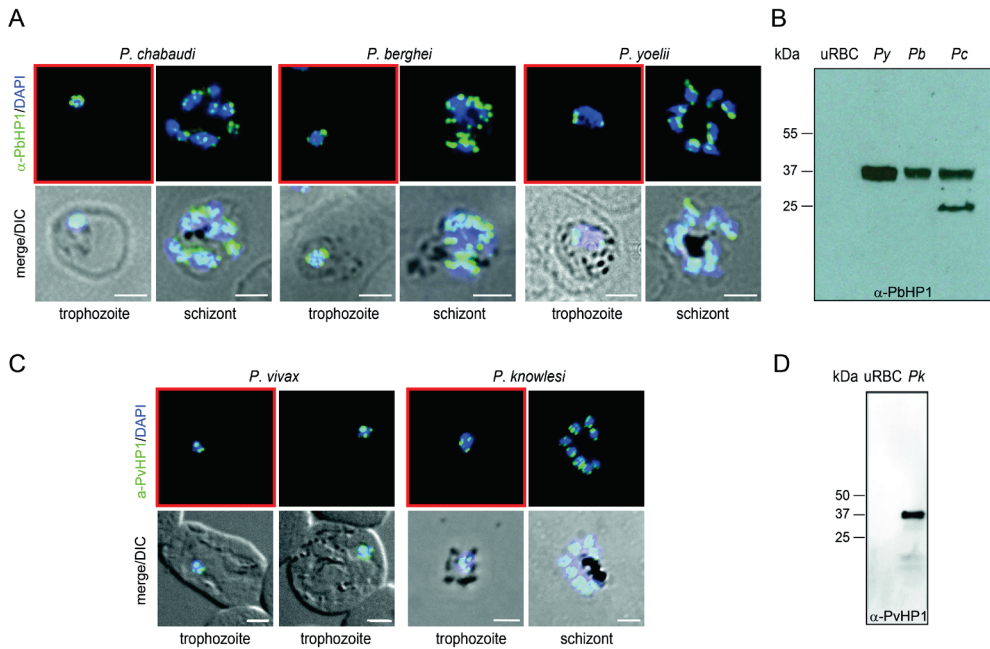


Figure S2. Detection of HP1 by IFAs and Western blot analysis (related to Figures 1 and S1, and STAR Methods).

(A) Representative IFA images obtained with α -PbHP1 antibodies showing perinuclear HP1 localisation (green) in *P. chabaudi*, *P. berghei* and *P. yoelii* trophozoite and schizont stage parasites. Nuclei were stained with DAPI (blue). DIC, differential interference contrast. Scale bar, 2.5 μ m. The red-framed images are identical to those presented in Figure 1A. **(B)** α -PbHP1 Western blot analysis detects a single protein at sizes expected for PyHP1, PbHP1 and PchHP1 (~32-33 kDa) in whole cell lysates generated from *P. yoelii* (lane 2), *P. berghei* (lane 3) and *P. chabaudi* (lane 4), respectively, but not in uninfected RBCs (lane 1; negative control). The lower band in lane 4 may represent a proteolytic degradation product of PchHP1. Generation of the polyclonal affinity-purified α -PbHP1 antibodies is explained in detail in the Experimental Procedures section. **(C)** Representative IFA images obtained with α -PvHP1 antibodies showing perinuclear HP1 localisation (green) in *P. vivax* trophozoites (note that schizont stages were not observed in this sample) and *P. knowlesi* trophozoite and schizont stage parasites. Nuclei were stained with DAPI (blue). DIC, differential interference contrast. Scale bar, 2.5 μ m. The red-framed images are identical to those presented in Figure 1A. **(D)** α -PvHP1 Western blot analysis detects a single protein at the expected size for PkHP1 (~32 kDa) in whole cell lysates generated from *P. knowlesi* (lane 2), but not in uninfected RBCs (lane 1; negative control). Generation of the polyclonal affinity-purified α -PvHP1 antibodies is explained in detail in the Experimental Procedures section.

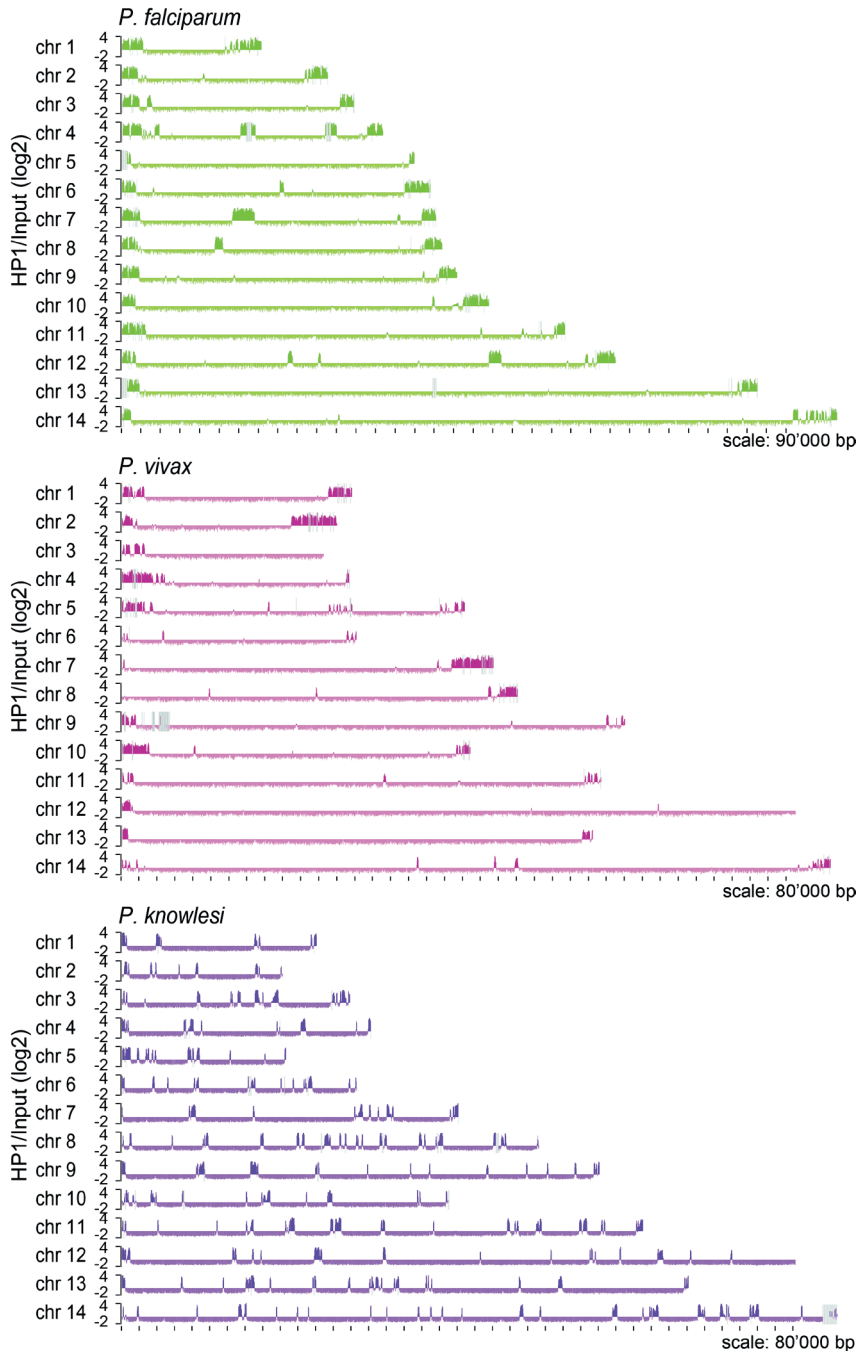
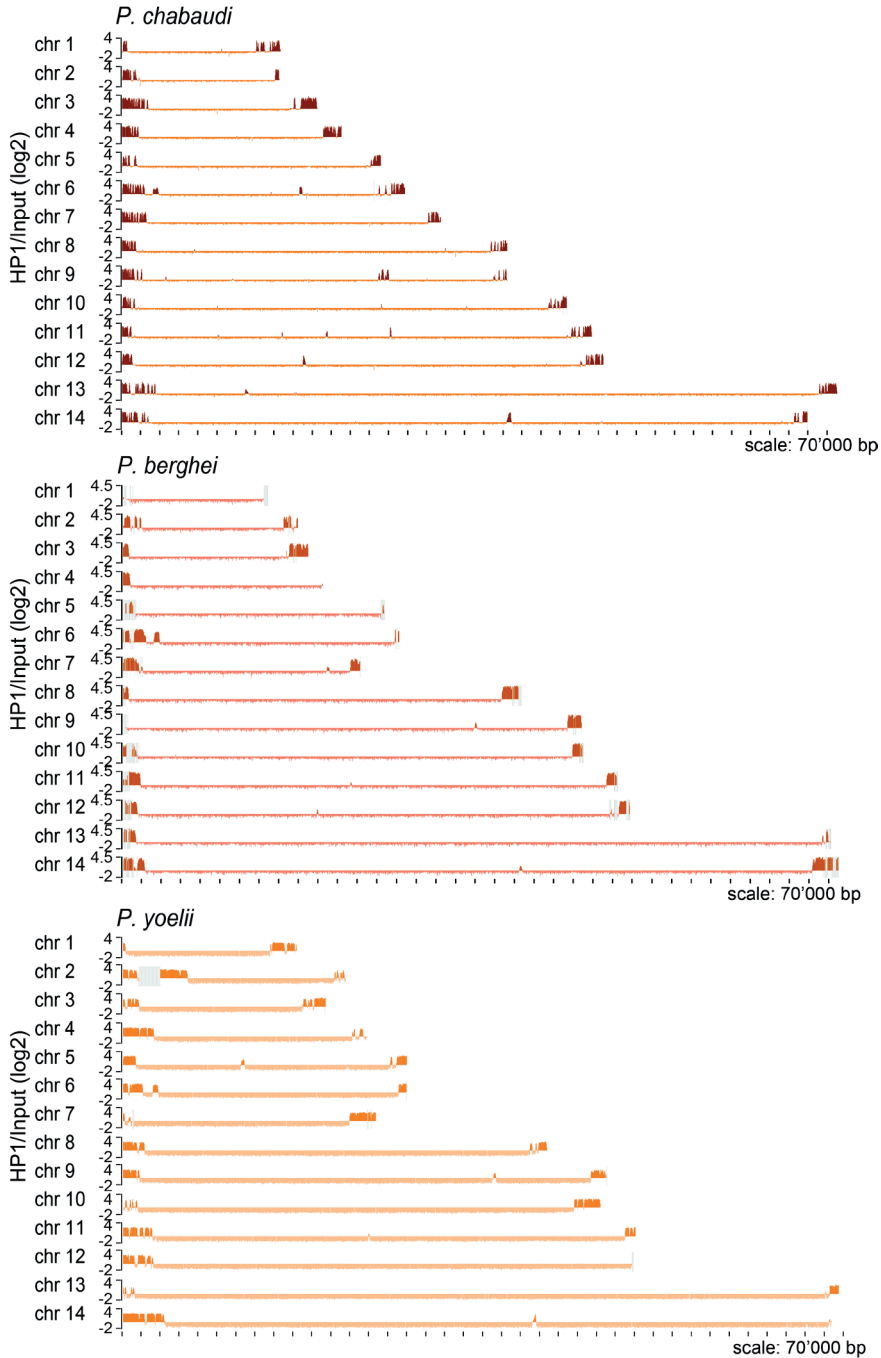


Figure S3: Genome-wide overviews of HP1 localization in schizont stages of six different *Plasmodium* species (related to Figure 1).

Chromosome maps depicting HP1-demarcated heterochromatin over all 14 chromosomes of schizont



stages from *P. falciparum* (3D7), *P. vivax*, *P. knowlesi* (A1-C.1) *P. chabaudi*, *P. berghei* and *P. yoelii*. Log₂-transformed HP1-ChIP/Input ratios were calculated in 1000 bp windows. Regions with low mappability are highlighted by grey bars.

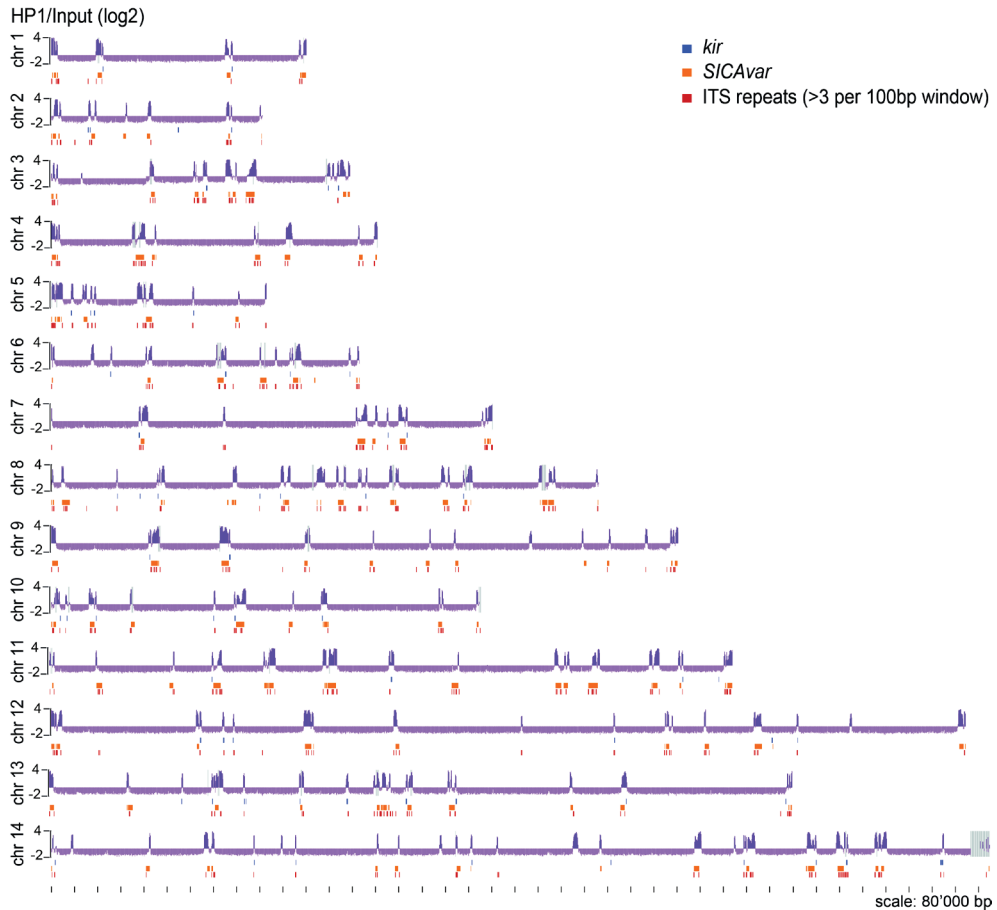
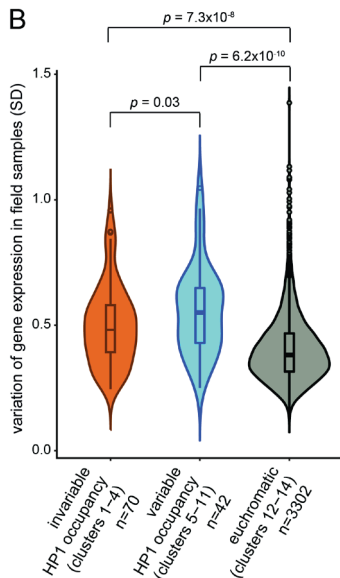
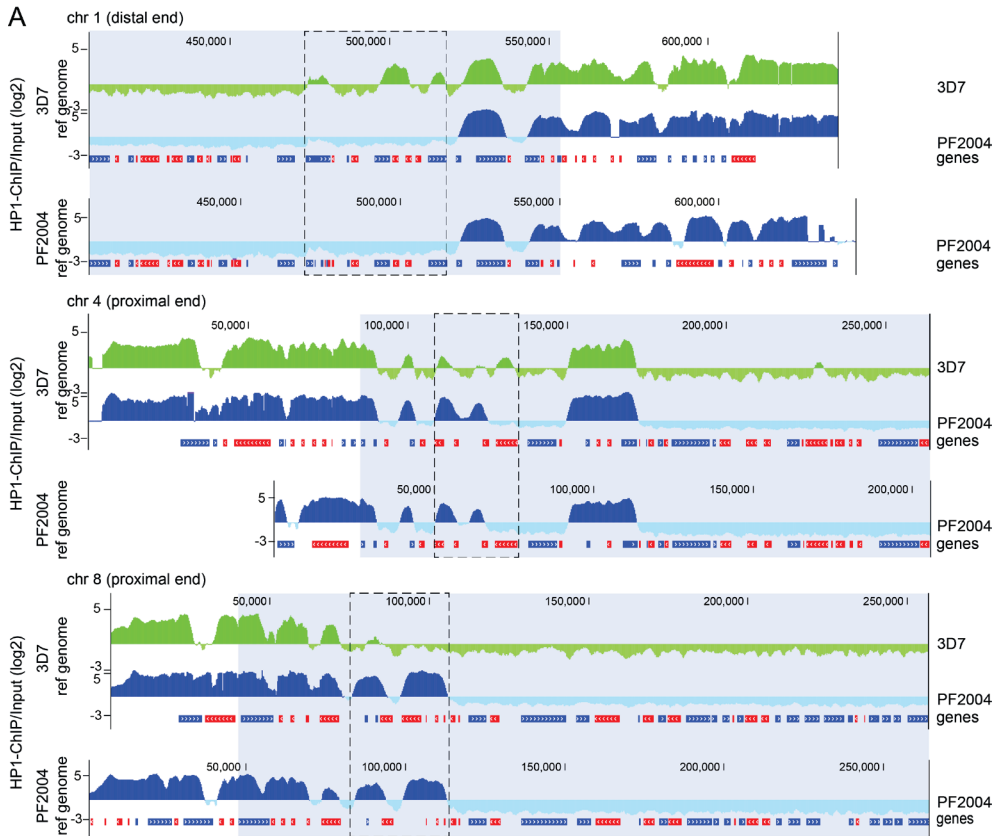


Figure S4: Genome-wide association of *SICAvAr* genes, *kir* genes and ITS elements with heterochromatic regions in *P. knowlesi* (related to Figure 1).

Chromosome maps depicting HP1-demarcated heterochromatin on all 14 chromosomes in *P. knowlesi* (A1-C.1). Log₂-transformed HP1-ChIP/Input ratios were calculated in 1000 bp windows. Coding sequences of *kir*/*kir*-like genes and *SICAvAr* genes are depicted in blue and orange, respectively. 100 bp windows containing three or more interstitial telomere repeat sequences (ITSs) are marked in red. Regions with low mappability are highlighted by grey bars.

Figure S5. Strain-specific differences in heterochromatin organization in syntenic regions of the genome (related to Figure 4 and Table S4).

(A) Log₂-transformed PfHP1-ChIP/Input ratio tracks generated from *P. falciparum* Pf2004 and 3D7



schizont stage parasites. The upper tracks were generated by mapping 3D7 (green) and Pf2004 (blue) ChIP-seq reads against the 3D7 reference genome (PlasmoDB v26). The Pf2004 ChIP-seq data were also mapped against the matching *P. falciparum* Pf2004 genome (lower tracks). Syntenic regions are marked in grey and regions with changing PfHP1 occupancy within these regions are framed. Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes. **(B)** Violin plot of gene expression variation in the heterochromatic and euchromatic groups of genes. Y-axis: standard deviation (SD) of relative gene expression levels determined by microarray analysis in 549 *P. falciparum* field samples (45) (only group A isolates, which all consisted of young ring stage parasites and displayed consistently low gametocyte densities (45), were included in this analysis). X-axis: with invariable HP1 occupancy (clusters 1-4) and variable HP1 occupancy (clusters 5-11) are shown in brown and blue, respectively. Euchromatic genes (clusters 12-14) are shown in grey. *p* values of Wilcoxon Rank Sum tests are shown on top for each pair of groups.

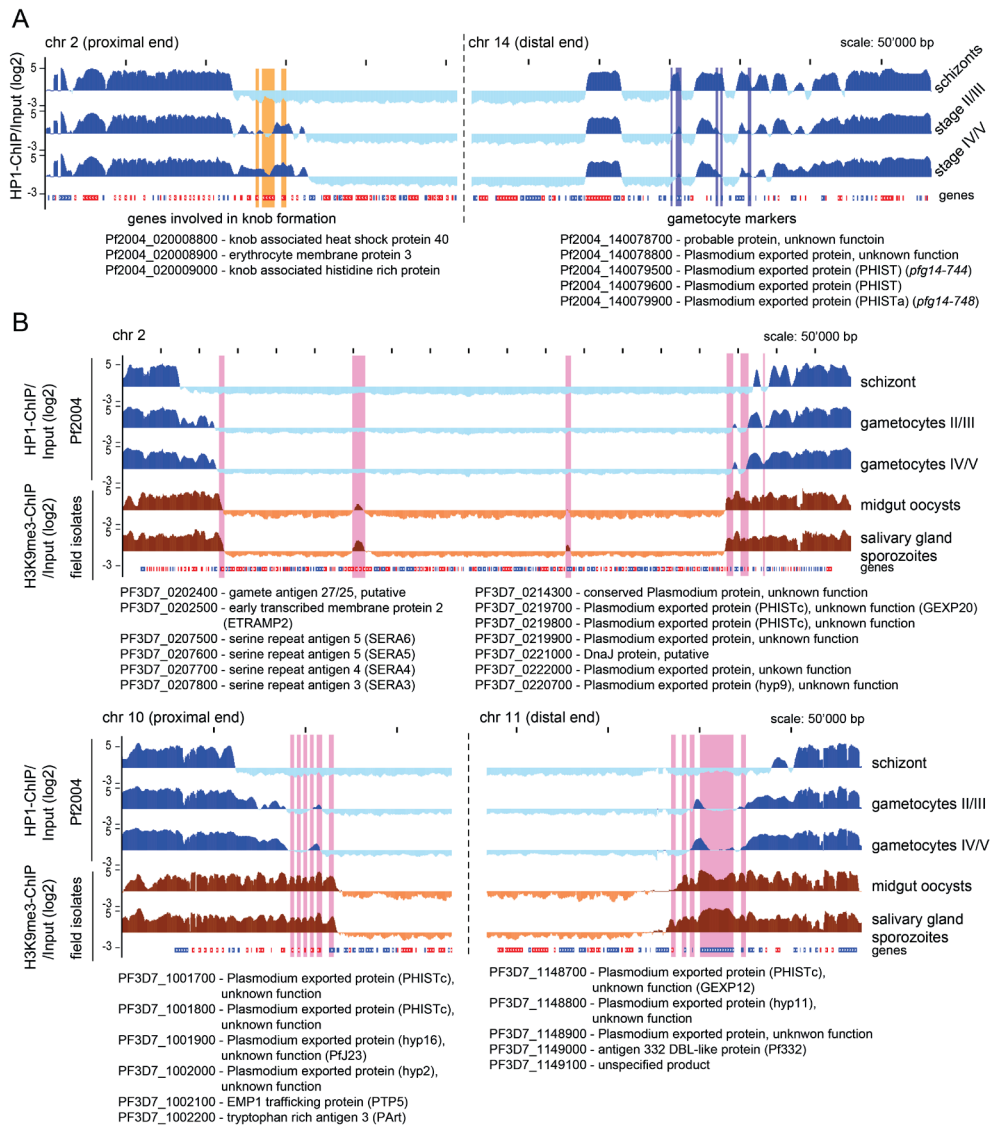


Figure S6. Altered heterochromatin organization in *P. falciparum* gametocytes, midgut oocysts and salivary gland sporozoites (related to Figure 6 and Table S6).

(A) Log₂-transformed PfHP1-ChIP/Input ratio tracks generated from *P. falciparum* Pf2004 schizonts and stage II/III and IV/V gametocytes. ChIP-seq reads were aligned against the Pf2004 genome and depicted as in Figure 6A. Genes involved in knob formation (orange) and early gametocyte markers (purple) are highlighted and their Pf2004 Gene IDs are shown. Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes. **(B)** Log₂-transformed PfHP1-ChIP/Input ratio tracks from different *P. falciparum* life cycle stages mapped against the 3D7 reference genome. The heterochromatic domains on chromosome 2 and zoom-in views of the proximal and distal ends of chromosome 10 and 11, respectively, are depicted as representative examples. PfHP1-ChIP/Input ratio tracks of Pf2004 schizonts, stage II/III and IV/V gametocytes are depicted in blue. H3K9me3-

ChIP/Input ratio tracks of midgut oocysts and salivary gland sporozoites from *P. falciparum* field isolates (65) are depicted in brown. Genes differentially marked in sporozoites compared to gametocytes are highlighted in pink (the corresponding Gene IDs are listed below the tracks). Coding sequences are shown as blue (sense strand) and red (antisense strand) boxes.

SUPPLEMENTARY TABLES

Table S1-S7, please download online.

Table S1. Genome-wide HP1 occupancy in six *Plasmodium* species (related to Figures 1 and 2, and Tables S2, S3, S4, S5 and S7).

ChIP/input enrichment values were calculated over a 1000 bp window (ATG \pm 500 bp) for each coding sequence of the respective reference genomes in *P. falciparum*, *P. vivax*, *P. knowlesi*, *P. chabaudi*, *P. berghei* and *P. yoelii* schizont stages. Column A: chromosome number. Columns B-C: Nucleotide positions at the start and end of each 1000 bp window. Columns D-E: Gene IDs and annotations. Columns F-G: RPKM values for the α -ChIP and input samples. Column H: HP1 enrichment values (ChIP/input ratio). Columns I-J: Probability values (p) for genes to belong to either the euchromatic or the heterochromatic group of genes. Column K: Heterochromatic gene family classification. Columns L-M: *Plasmodium* species encoding an ortholog or syntenic ortholog. Columns N-O: Orthologs are heterochromatic in multiple species or a single species only. High-confidence heterochromatic genes ($p > 0.99999$) and potential heterochromatic genes ($0.99999 > p > 0.95$) are highlighted in orange and light blue, respectively. Genes with very low mappability (RPKM value < 5 in at least one input sample) are highlighted in grey and were excluded from downstream analyses. This file contains six worksheets. (xlsx format).

Table S2. List of all HP1-enriched genes that have orthologs in at least one other *Plasmodium* species (related to Figures 1 and 2, and Tables S1 and S7).

Columns A-B: Gene IDs and annotations. Column C: HP1 enrichment values (ChIP/input ratio). Column D: Heterochromatic gene family classification. Columns E-F: *Plasmodium* species encoding an ortholog or syntenic ortholog. Column G: Number of *Plasmodium* species in which the orthologous gene is HP1-enriched. Columns H-AB: Gene IDs of orthologs in *P. falciparum*, *P. vivax*, *P. knowlesi*, *P. chabaudi*, *P. berghei* and *P. yoelii*. HP1-enriched orthologs are highlighted in orange. (xlsx format).

Table S3. Genome-wide PkHP1 occupancy in *P. knowlesi* clones A1-H.1 and A1-C.1 (related to Figure 3, and Tables S1 and S7).

ChIP/input enrichment values were calculated over a 1000 bp window (ATG \pm 500 bp) for each coding sequence of the *P. knowlesi* reference genome in *P. knowlesi* clones A1-H.1 and A1-C.1 schizont stages. Column A: chromosome number. Columns B-C: Nucleotide positions at the start and end of each 1000 bp window. Columns D-E: Gene IDs and annotations. Columns F-I: RPKM values for the α -ChIP and input samples for each of the two clones. Columns J-K: PkHP1 enrichment values (ChIP/input ratio) for each of the two clones. Column L: fold change in PkHP1 occupancy in A1-H.1 compared to A1-C.1. Genes with ≥ 2.5 -fold higher or lower PkHP1 occupancy in A1-H.1 are highlighted in light or dark orange, respectively. The ChIP-seq values for *P. knowlesi* clone A1-C.1 are identical to those listed in Table S1. Genes with very low mappability (RPKM value < 5 in at least one input sample) are highlighted in grey and were excluded from downstream analyses. (xlsx format).

Table S4. Genome-wide HP1 occupancy in four different *P. falciparum* strains (related to Figures 4 and S4, and Tables S1 and S7).

ChIP/input enrichment values were calculated over a 1000 bp window (ATG \pm 500 bp) for each coding sequence of the *P. falciparum* 3D7 reference genome in schizont stages of *P. falciparum* strains Pf2004, NF135, NF54, and 3D7. Column A: chromosome number. Columns B-C: Nucleotide positions at the start and end of each 1000 bp window. Columns D-E: Gene IDs and annotations. Columns F-M: RPKM values for the α -ChIP and input samples for each of the four strains. Columns N-Q: PfHP1 enrichment values (ChIP/input ratio) for each of the four strains. Columns R-U: z-score-transformed ChIP enrichment values for each of the four strains. Column V: k-means cluster numbers (1-14). Column W: gene family classification. Column X: heterochromatic pseudogenes. Column Y: genes previously shown to be variably expressed *in vitro* (16). Column Z: degree of variation in gene expression (SD) in field samples (45). Genes with variable PfHP1 occupancy between the strains are marked in blue (k-means clusters 5 to 11). Genes with very low mappability (RPKM value < 5 in at least one input sample) are highlighted in grey and were excluded from downstream analyses. The ChIP-seq values for *P. falciparum* 3D7 schizonts are identical to those listed in Tables S1 and S5. (xlsx format).

Table S5. Genome-wide PfHP1 occupancy in ring stages, trophozoites and schizonts (related to Figure 5, and Tables S1 and S7).

ChIP/input enrichment values were calculated over a 1000 bp window (ATG \pm 500 bp) for each coding sequence of the *P. falciparum* 3D7 reference genome in ring stages, trophozoites and schizonts stages of *P. falciparum* 3D7 parasites. Column A: chromosome number. Columns B-C: Nucleotide positions at the start and end of each 1000 bp window. Columns D-E: Gene IDs and annotations. Columns F-H: RPKM values for thea-ChIP samples for each of the three IDC stages. Column I: RPKM values for the input sample from schizonts. Columns J-L: PfHP1 enrichment values (ChIP/input ratio) for each of the three IDC stages. Column M: peak transcript abundance during the IDC (RNA-seq RPKM values from (46)). Column N: genes previously shown to be variably expressed in 3D7 parasites (16). Genes with very low mappability (RPKM value < 5 in the input sample) are highlighted in grey and were excluded from downstream analyses. The ChIP-seq values for *P. falciparum* 3D7 schizonts are identical to those listed in Tables S1 and S4. (xlsx format).

Table S6. Genome-wide HP1 occupancy in *P. falciparum* schizonts and gametocytes (related to Figures 6 and S5, and Tables S4 and S7).

ChIP/input enrichment values were calculated over a 1000 bp window (ATG \pm 500 bp) for each coding sequence of the *P. falciparum* 3D7 reference genome in schizont stages, stage II/III and stage IV/V gametocytes of *P. falciparum* strain Pf2004. Column A: chromosome number. Columns B-C: Nucleotide positions at the start and end of each 1000 bp window. Columns D-E: Gene IDs and annotations. Columns F-K: RPKM values for the α -ChIP and input samples for each of the three stages. Columns L-N: PfHP1 enrichment values (ChIP/input ratio) for each of the three stages. Columns O-Q: z-score-transformed ChIP enrichment values for each of the three stages. Column R: k-means cluster numbers (1-12). Genes with lower or higher PfHP1 occupancy in gametocytes compared to schizonts are marked in green (k-means cluster 5) or blue (k-means clusters 6 to 8), respectively. Genes with very low mappability (RPKM value < 5 in at least one input sample) are highlighted in grey and were excluded from downstream analyses. The ChIP-seq values for *P. falciparum* Pf2004 schizonts are identical to those listed in Table S4. (xlsx format).

Table S7. ChIP-seq libraries (related to Tables S1 to S6, and STAR Methods).

Column A: Information on the species, strains, clones and stages of *Plasmodium* parasite samples harvested for ChIP-seq experiments. Column B: DNA sample type used for library preparation. Column C: Antibodies used in the ChIP experiment. Column D: Number of PCR cycles used in the first and

second library amplification. Columns E-F: Number of reads before filtering (total) and after filtering for uniquely mapping reads (unique). Column G: Versions of the reference genomes used for read mapping.

CHAPTER 5

GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing

Michael Filarsky,^{1,2} Sabine A. Fraschka,³ Igor Niederwieser,^{1,2} Nicolas M.B. Brancucci,^{1,2} Eilidh Carrington,^{1,2} Elvira Carrió,^{1,2} Suzette Moes,⁴ Paul Jenoe,⁴ Richárd Bártfai,³ Till S. Voss^{1,2,*}

Published in
Science 359, 1259 (2018)

¹Swiss Tropical and Public Health Institute, Basel 4051, Switzerland

²University of Basel, Basel 4003, Switzerland

³Department of Molecular Biology, Radboud University, Nijmegen 6525GA, The Netherlands

⁴Biozentrum, University of Basel, Basel 4056, Switzerland

*Corresponding author. Email: till.voss@unibas.ch

Contribution:

Designed, performed and analyzed ChIP-Seq and RNA-seq experiments.

This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in *Science* on Vol. 359, Issue 6381, pp. 1259-1263 on 16 Mar 2018, DOI: [10.1126/science.aan6042](https://doi.org/10.1126/science.aan6042).

Malaria is caused by *Plasmodium* parasites that proliferate in the bloodstream. During each replication cycle some parasites differentiate into gametocytes, the only forms able to infect the mosquito vector and transmit malaria. Sexual commitment is triggered by activation of AP2-G, the master transcriptional regulator of gametocytogenesis. Heterochromatin protein 1 (HP1)-dependent silencing of *ap2-g* prevents sexual conversion in proliferating parasites. Here, we identified *Plasmodium falciparum* gametocyte development 1 (GDV1) as an upstream activator of sexual commitment. We found that GDV1 targeted heterochromatin and triggered HP1 eviction thus de-repressing *ap2-g*. Expression of GDV1 was responsive to environmental triggers of sexual conversion and controlled via a *gdv1* antisense RNA. Hence, GDV1 appears to act as cell fate decision factor that induces sexual differentiation by antagonizing HP1-dependent gene silencing.

Heterochromatin protein 1 (HP1) is a conserved regulator of heterochromatin formation, heritable gene silencing and variegated gene expression (1). In *Plasmodium falciparum*, HP1-dependent clonally variant expression allows parasites to adapt rapidly to environmental challenges encountered during infection (2-4). For example, immune evasion via antigenic variation of *var*/PfEMP1 is the hallmark of *Plasmodium* survival. Other processes, such as expression of red blood cell (RBC) invasion ligands or nutrient transporters, are similarly regulated in this parasite (4). Most clonally variant genes cluster in subtelomeric domains but some also occur in chromosome-internal heterochromatic regions. In addition, HP1 forms microdomains at some euchromatic genes (2). One of these encodes the transcription factor AP2-G that is required for sexual conversion and differentiation (2, 5-7). HP1-dependent regulation of *ap2-g* controls the rate at which parasites commit to sexual differentiation (7).

To explore the mechanisms regulating HP1 occupancy in *P. falciparum* we identified HP1-interacting proteins by liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of native HP1 complexes that were purified by co-immunoprecipitation (co-IP) from parasites expressing GFP-tagged HP1 (7) (Fig. 1A and Table S3). Interestingly, we consistently observed GDV1 among the potential HP1 interaction partners (Table S1). GDV1 is a nuclear protein implicated in sexual commitment and early gametocytogenesis but its exact function remains unknown (8). We therefore created a parasite line for the conditional expression of fluorescently labelled ectopic GDV1 (GDV1-GFP-DD) (Fig. 1B). Proteins tagged with the immunophilin protein-folding chaperone FKBP destabilisation domain (DD) are proteolytically degraded unless cells are cultured in presence of Shield-1, a small molecule ligand stabiliser (9, 10). Thus, GDV1-GFP-DD is barely detectable in parasites cultured in absence of Shield-1 (3D7/GDV1-GFP-DD^{OFF}), but its expression is markedly induced in parasites grown in presence of Shield-1 (3D7/GDV1-GFP-DD^{ON}) (Fig. 1, B and C). In agreement with the co-IP results GDV1-GFP-DD co-localizes with HP1 at the nuclear periphery (Fig. 1C and Fig. S1). Furthermore, we found that recombinant HP1 and

GDV1 formed a complex (Fig. S1) and that HP1 co-purified with GDV1-GFP-DD in reverse co-IPs (Fig. 1D, Tables S2 and S4). The chromodomain-helicase-DNA-binding protein 1 (CHD1) and a protein of unknown function (PF3D7_1451200) also consistently co-purified with both HP1 and GDV1-GFP-DD (Tables S1 and S2). Given that CHD1 plays important roles in cell fate decision and heterochromatin remodelling in other organisms (11, 12) and that GDV1 is implicated in gametocytogenesis (8) it appears that this putative regulatory complex may function in activating sexual commitment.

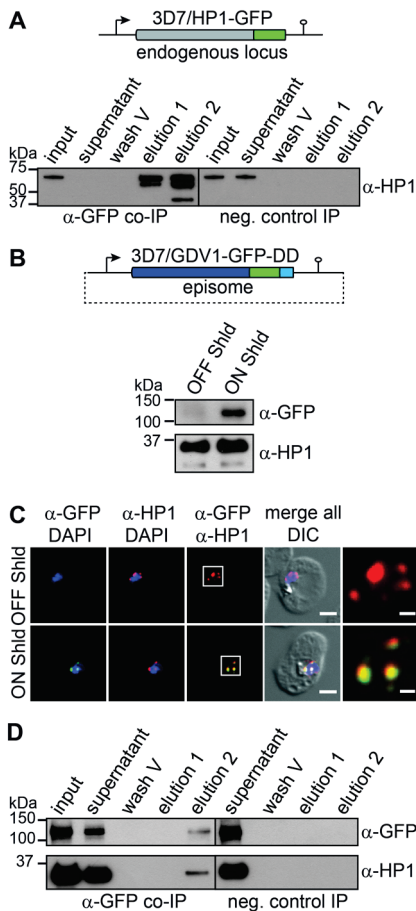


Fig. 1. GDV1 interacts with HP1.

(A) Endogenous *hp1* locus in 3D7/HP1-GFP parasites and α -HP1 Western blots of the α -HP1-GFP co-IP and negative control samples. Results are representative of three biological replicates. (B) *gdv1-gfp-dd* expression plasmid and α -GFP Western blots of 3D7/GDV1-GFP-DD^{OFF} and 3D7/GDV1-GFP-DD^{ON} parasites. α -HP1 antibodies served as loading control. (C) GDV1-GFP-DD/HP1 co-localisation IFAs in 3D7/GDV1-GFP-DD^{OFF} and 3D7/GDV1-GFP-DD^{ON} trophozoites (24-32 hpi). DIC, differential interference contrast. Scale bar, 2.5 μ m (0.5 μ m for the magnified views in the rightmost images). Results are representative of three biological replicates. (D) α -GFP and α -HP1 Western blots of the α -GDV1-GFP-DD co-IP and negative control samples. Results are representative of three biological replicates.

Malaria parasites proliferate by iterative rounds of intra-erythrocytic replication through schizogony, merozoite release and RBC re-invasion. The decision to enter gametocytogenesis is made in the cell cycle prior to sexual differentiation; sexually committed schizonts release merozoites that invade RBCs and differentiate all into either female or male gametocytes (13, 14) (Fig. 2A).

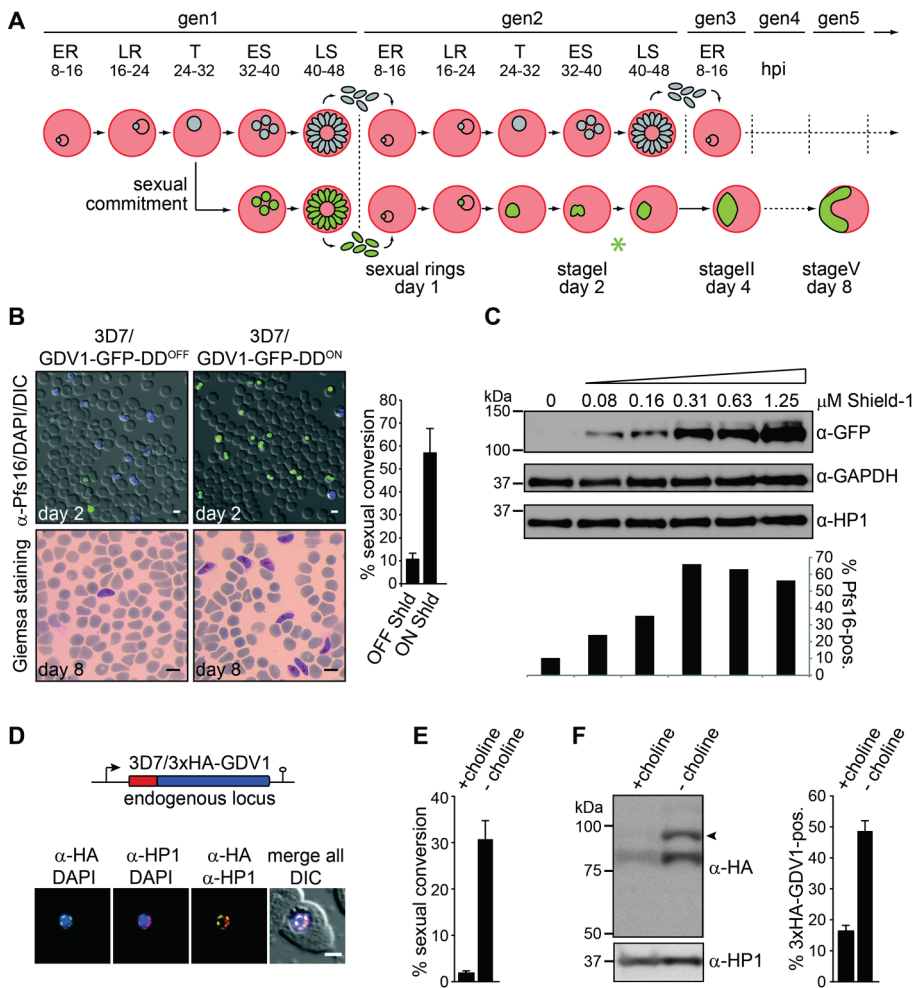
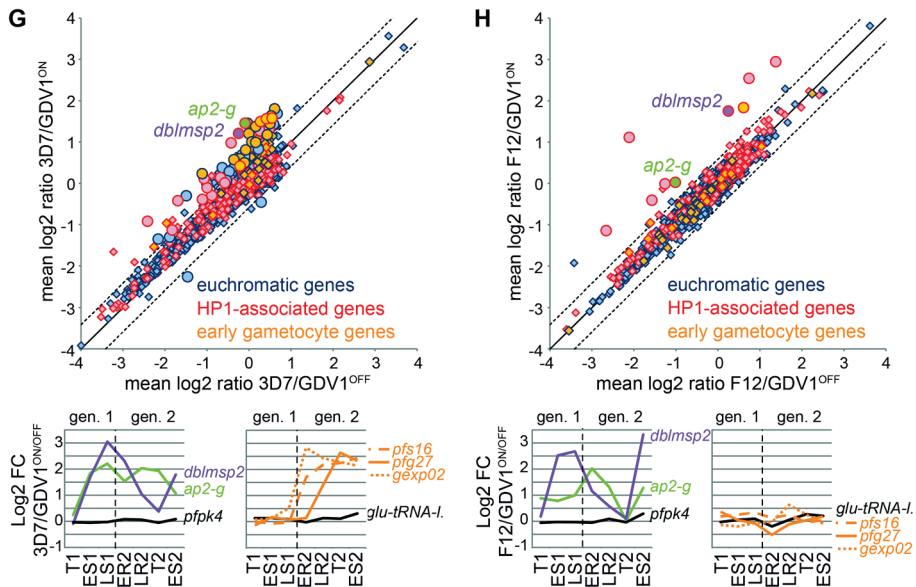


Fig. 2. GDV1 induces sexual commitment and differentiation.

(A) Schematic illustrating the iterative cycles of schizogony and RBC re-invasion (top) or sexual commitment, RBC re-invasion and gametocyte differentiation (bottom). ER/LR, early/late ring stages; T, trophozoites; ES/LS, early/late schizonts; gen, generation; hpi, hours post-invasion; asterisk, time point of α -Pfs16 IFAs. **(B)** Top panel: α -Pfs16 IFAs identifying stage I gametocytes. Quantification of Pfs16-positive parasites is shown at the right (results are the mean of three biological replicates (200 infected RBCs counted per sample); error bars indicate SD). Bottom panel: Giemsa-stained blood smears showing stage V gametocytes. Scale bars, 5 μ m. **(C)** Western blot showing GDV1-GFP-DD expression in presence of increasing Shield-1 concentrations. α -GAPDH and α -HP1 antibodies served as loading controls. Percentages of Pfs16-positive parasites are shown at the bottom (400 infected RBCs counted per sample). **(D)** Endogenous *gdv1* locus in 3D7/3xHA-GDV1 parasites and 3xHA-GDV1/HP1 co-localisation IFAs in trophozoites (24-32 hpi). Scale bar, 2.5 μ m. **(E)** Sexual conversion rates in 3D7/3xHA-GDV1 parasites cultured in presence or absence of choline (results are the mean of three biological replicates (>190 infected RBCs counted per sample); error bars indicate SD). **(F)** Left panel: Western blot showing 3xHA-GDV1 expression levels in 3D7/3xHA-GDV1 parasites cultured in presence or absence of choline. α -HP1 antibodies served as loading control. Right panel: Percentages



of 3xHA-GDV1-positive parasites in presence or absence of choline (results are the mean of three biological replicates (>100 infected RBCs counted per sample); error bars indicate SD). **(G, H)** Comparison of mean expression levels of all genes in 3D7/GDV1-GFP-DD^{ON} versus 3D7/GDV1-GFP-DD^{OFF} (G) and F12/GDV1-GFP-DD^{ON} versus F12/GDV1-GFP-DD^{OFF} parasites (H). Significantly deregulated genes are indicated by circles (mean fold change cut-off >1.5; q-value (fdr) cut-off <0.15). Known early gametocyte markers (7, 8, 17) are labelled orange. Line graphs show fold changes in expression across seven consecutive TPs. *pfs16/pfg27/gexp02*, early gametocyte markers (15, 22, 32); *pk4* (PF3D7_0628200)/*glu-tRNA-I*. (PF3D7_1331700), control genes (7).

To test if GDV1 triggers sexual commitment, 3D7/GDV1-GFP-DD^{OFF} parasites were split and cultured in the absence or presence of Shield-1. After re-invasion, stage I gametocytes were quantified by immuno-fluorescence assays (IFA) using antibodies against the gametocyte marker Pfs16 (15). Strikingly, the 3D7/GDV1-GFP-DD^{ON} population displayed a sexual conversion rate of 57.2% (+/- 10.0 SD) compared to 11.0% (+/- 2.4 SD) in 3D7/GDV1-GFP-DD^{OFF} parasites, and these gametocytes differentiated normally into both male and female gametocytes and showed a typical female-biased sex ratio (Fig. 2B and Fig. S2). Moreover, Shield-1 titration revealed a positive correlation between ectopic GDV1-GFP-DD expression levels and sexual conversion rates (Fig. 2C). To test if endogenous GDV1 levels similarly correlate with gametocyte conversion we used CRISPR/Cas9-based gene editing to append a triple hemagglutinin (HA) tag to the N-terminus of GDV1 (3D7/3xHA-GDV1) (Fig. S3). Endogenous 3xHA-GDV1 co-localised with HP1 as expected (Fig. 2D and Fig. S3) but was only expressed in some parasites. We next quantified 3xHA-GDV1 expression under conditions that either suppress or favour sexual conversion. To this end, we made use of the recent discovery of choline as an inhibitor of sexual commitment (16). 3D7/3xHA-GDV1 parasites

cultured in the presence or absence of 2 mM choline displayed sexual commitment rates of 1.8% (+/- 0.3 SD) or 30.9% (+/- 3.8 SD), respectively (Fig. 2E). Strikingly, parasites cultured in the absence of choline showed markedly increased 3xHA-GDV1 expression levels (Fig. 2F). This was accounted for by a higher proportion of 3xHA-GDV1-positive cells (48.6% (+/- 3.4 SD) in absence compared to 16.4% (+/- 1.8 SD) in presence of choline) (Fig. 2F) and comparatively higher 3xHA-GDV1 expression levels in individual 3xHA-GDV1-positive parasites (Fig. S3). Together, these results show that GDV1 activates sexual conversion in a dose-dependent manner and that endogenous GDV1 expression can be induced by environmental signals triggering sexual commitment.

We next performed comparative transcriptome analyses using two-colour microarrays. 3D7/GDV1-GFP-DD^{OFF} ring stage parasites were split, cultured separately in absence or presence of Shield-1 and total RNA was harvested at seven paired time points spanning the remaining 24 hours of generation 1 (24-32 hours post-invasion (hpi); 32-40 hpi, 40-48 hpi) and the first 40 hours after re-invasion in generation 2 (8-16 hpi, 16-24 hpi, 24-32 hpi, 32-40 hpi) (Fig. 2A). As expected, GDV1-GFP-DD expression triggered a transcriptional response characteristic of sexual commitment and early differentiation. This was evident from the induction of *ap2-g* in generation 1, followed by activation of early gametocyte markers (5, 7, 8, 17) after re-invasion (Fig. 2G, Fig. S4 and Table S5). In F12 parasites, a 3D7-derived gametocyte-deficient clone carrying a loss-of-function mutation in *ap2-g* (5, 18), GDV1-GFP-DD expression still activated *ap2-g* but failed to launch a sexual differentiation response (Fig. 2H and Table S6). Next to *ap2-g* only eight other genes were significantly induced in F12/GDV1-GFP-DD^{ON} parasites, all of which are marked by HP1. This set included *dblmsp2*, which was also induced in 3D7/GDV1-GFP-DD^{ON} parasites (Fig. 2, G and H). Given that DBLMSP2 is a merozoite surface antigen expressed only in a small subpopulation of schizonts (19, 20) the GDV1-dependent activation of the *dblmsp2* locus suggests it may be expressed specifically in sexually committed schizonts. In summary, these findings show that GDV1 is an upstream activator of sexual commitment and likely triggers this process by antagonising HP1-dependent silencing of *ap2-g*.

To test if GDV1 associates with heterochromatin *in vivo* we conducted comparative ChIP-seq experiments. 3D7/GDV1-GFP-DD^{OFF} parasites were split at 28-34 hpi, cultured in parallel in the absence or presence of Shield-1 and paired chromatin samples were harvested two (30-36 hpi), six (34-40 hpi) and ten (38-44 hpi) hours after Shield-1 addition. We found that (1) GDV1-GFP-DD associates specifically with heterochromatin throughout the genome (Fig. 3A, Fig. S5, Table S7); (2) GDV1-GFP-DD occupancy was markedly higher in 3D7/GDV1-GFP-DD^{ON} compared to 3D7/GDV1-GFP-DD^{OFF} parasites (Fig. S5, Table S7); and (3) GDV1-GFP-DD occupancy is highly correlated with that of HP1 (Fig. 3B). Moreover, GDV1-GFP-DD occupancy peaked six hours post-induction and decreased substantially thereafter (Fig. 3A, Fig. S5, Table S7). This drop in GDV1-GFP-DD signal coincided with a reduced HP1 occupancy over heterochromatic genes in 3D7/GDV1-GFP-DD^{ON} compared to 3D7/GDV1-GFP-DD^{OFF}

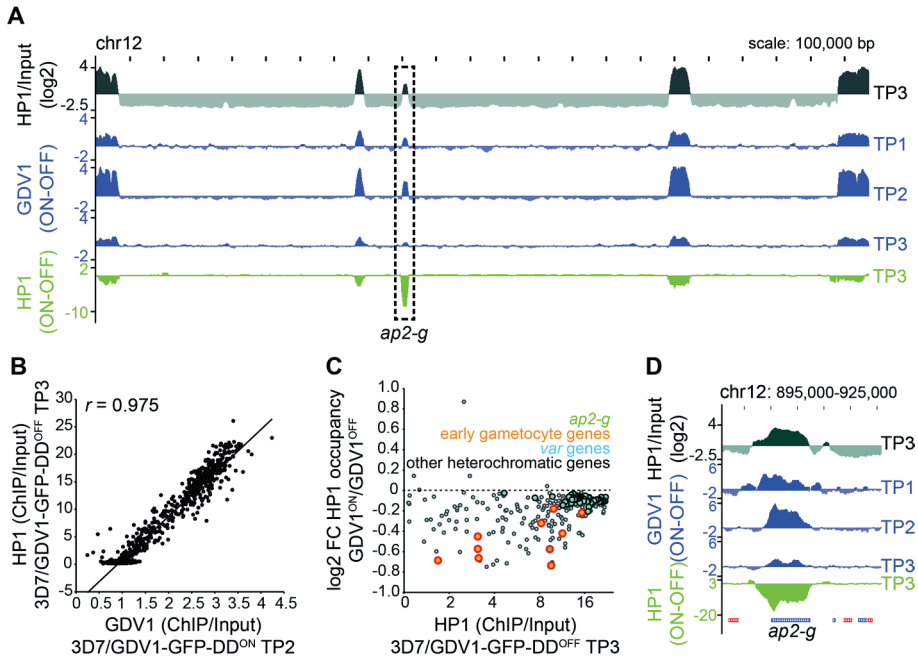


Fig. 3. GDV1 associates with heterochromatin throughout the genome and triggers HP1 removal at *ap2-g*.

(A) HP1 over input ratio track from 3D7/GDV1-GFP-DD^{OFF} schizonts (38-44 hpi, TP3) (grey). ChIP-seq subtraction tracks display relative enrichment of GDV1-GFP-DD in 3D7/GDV1-GFP-DD^{ON} schizonts two (30-36 hpi, TP1), six (34-40 hpi, TP2) and ten (38-44 hpi, TP3) hours after Shield-1 addition (blue), and relative depletion of HP1 in 3D7/GDV1-GFP-DD^{ON} parasites at TP3 (green). (B) Correlation between GDV1-DD-GFP enrichment in 3D7/GDV1-GFP-DD^{ON} (34-40 hpi, TP2) and HP1 enrichment in 3D7/GDV1-GFP-DD^{OFF} schizonts at each coding region. r , Pearson correlation coefficient. (C) Fold change in HP1 enrichment upon GDV1-GFP-DD overexpression in relation to HP1 enrichment in 3D7/GDV1-GFP-DD^{OFF} schizonts for each heterochromatic gene. (D) Zoom-in view of the enrichment/subtraction tracks at the *ap2-g* locus.

parasites (Fig. 3, A and C, Table S7). While the vast majority of heterochromatic loci, in particular those displaying high HP1 occupancy such as *var* genes, displayed marginally decreased HP1 levels, some genes exhibited as much as 40% reduced HP1 occupancy (Fig. 3C and Table S7). This group of genes includes *ap2-g* and most known HP1-associated early gametocyte markers including *geco* (21), *pfgexp17* (22) and *pfg14_748* (8, 17) (Fig. 3, C and D, Table S7). These data are consistent with the microarray results, where GDV1-GFP-DD expression activated *ap2-g* and early gametocyte genes but had no effect on the expression of the bulk of heterochromatic genes (Fig. 2, G and H).

Of note, given the 50-60% sexual conversion rate observed for 3D7/GDV1-GFP-DD^{ON} parasites (see above), a 30-40% reduction in HP1 occupancy indicates that HP1 may be depleted at these loci specifically in sexually committed parasites but single cell approaches

are required to confirm this hypothesis. Overall, we suggest that GDV1 destabilises heterochromatin and thus allows specific transcription factors to activate expression of *ap2-g* and other gametocyte-specific heterochromatic genes, and this may play an important role in the positive auto-regulatory feedback loop proposed to reinforce AP2-G expression in committed parasites (5, 6, 23). How GDV1 achieves specificity in unlocking specific HP1-associated genes despite binding heterochromatin genome-wide is a challenging question to be addressed in the future.

Since GDV1 activates sexual commitment, the question arises of how parasites limit GDV1 expression to prevent sexual conversion in asexual schizonts. A recent study identified a multi-exon long non-coding *gdv1* antisense RNA (asRNA) that initiates downstream of the *gdv1* locus and overlaps with the ATG start codon of *gdv1* (24), which is a hallmark feature of regulatory asRNAs (25). To investigate if the *gdv1* asRNA participates in regulating sexual commitment we created a *gdv1* asRNA loss-of-function mutant in F12 parasites (F12/*gdv1*-askO) (Fig. 4A and Fig. S6). Strand-specific RNA-seq analysis identified a small set of genes that were consistently differentially expressed between F12/*gdv1*-askO and F12 wild-type parasites (17 up- and 23 down-regulated genes) (Fig. 4B, Table S8). Strikingly, and similar to F12 parasites expressing ectopic GDV1-GFP-DD (Fig. 2H), *ap2-g*, *dblmsp2* and two early gametocyte genes (*pfg14_748*, PF3D7_1477400) (8, 17) were markedly induced in F12/*gdv1*-askO parasites, and all except one up-regulated gene are HP1-associated genes (Fig. 4B, Fig. S6, Table S8). *gdv1* sense transcripts were slightly increased in the F12/*gdv1*-askO population, while *gdv1* antisense transcripts were undetectable as expected (Fig. 4, B and C, Fig. S6, Table S8).

These results indicated that the *gdv1* asRNA acts as a negative regulator of GDV1 expression. To confirm this hypothesis, we tagged endogenous GDV1 in these parasites (F12/3xHA-GDV1/*gdv1*-askO) and observed that indeed almost all parasites expressed 3xHA-GDV1 (96.7% +/- 2.5 SD) (Fig. S7). Lastly, we show that deletion of the *gdv1* asRNA locus in a conditional AP2-G mutant resulted in a markedly increased production of gametocytes (Fig. S8 and supplementary text). Together, these findings demonstrate a central role for the *gdv1*-asRNA in regulating GDV1-dependent activation of sexual commitment. We anticipate this mechanism likely involves inhibiting GDV1 expression by interference with *gdv1* mRNA transcription, stability or translation, similar to asRNA-mediated gene regulation in other organisms (26). In conclusion, we identified GDV1-mediated heterochromatin destabilisation as an epigenetic control strategy regulating sexual cell fate decision in malaria parasites. Our discovery of the *gdv1*-asRNA as a negative regulator of sexual commitment is reminiscent of lncRNA-mediated control of gametogenesis in yeasts (27, 28). In *S. cerevisiae*, nutritional stress triggers gametogenesis by activating the transcriptional regulator Inducer of Meiosis 1 (IME1) (28). A lncRNA in the *ime1* promoter and antisense transcription of *ime4* are key factors in preventing

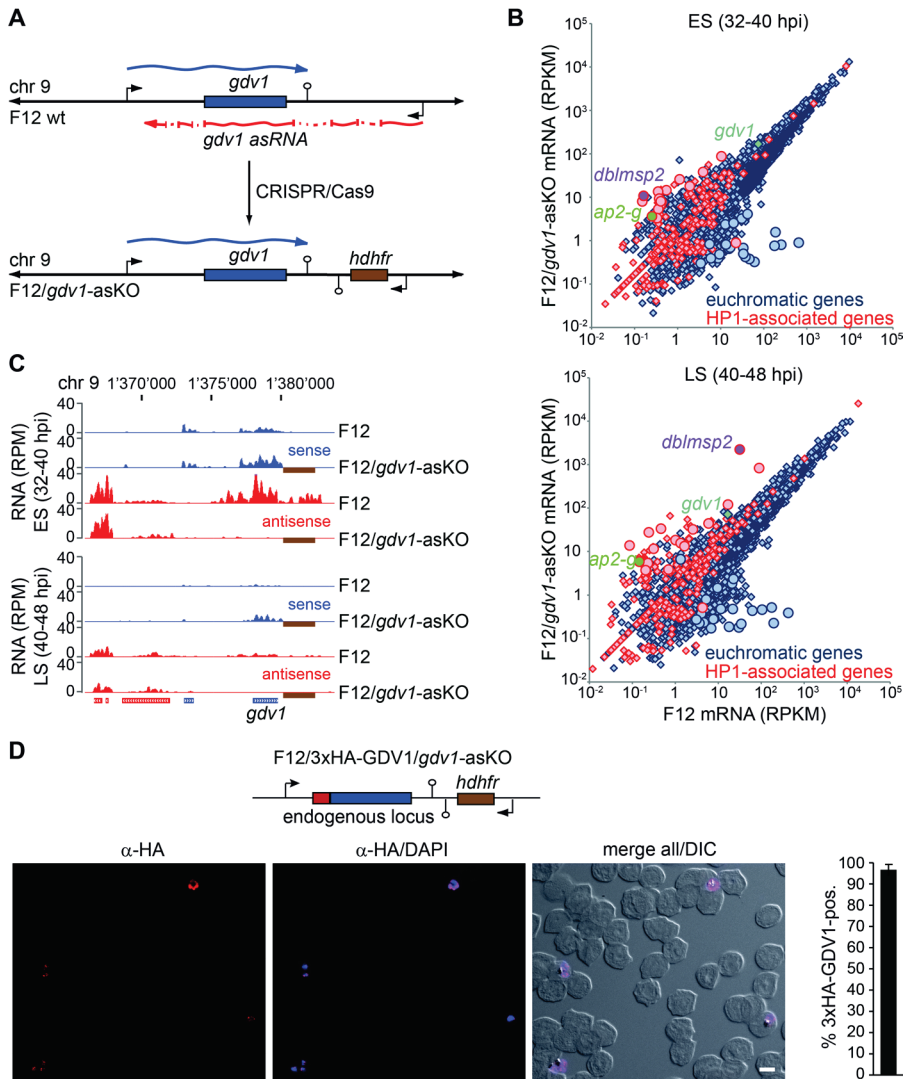


Fig. 4. A *gdv1* antisense RNA antagonises GDV1-dependent sexual commitment.

(A) *gdv1* locus in F12 wild-type and F12/*gdv1*-asKO parasites. The *gdv1* sense transcript (blue), five-exon *gdv1*-asRNA (24) (red) and *hdhfr* resistance marker (brown) are highlighted. (B) Comparison of gene expression levels in F12 wild-type and F12/*gdv1*-asKO early (ES) and late (LS) schizonts. Genes de-regulated > 5-fold in both TPs are indicated by circles. (C) UCSC genome browser screenshots of RNA-seq coverage plots over the *gdv1* locus in F12 wild-type and F12/*gdv1*-asKO early (ES) and late (LS) schizonts. The *hdhfr* resistance cassette downstream of the *gdv1* locus in F12/*gdv1*-asKO parasites and absent in the 3D7 reference genome is indicated by a brown box. (D) Endogenous *gdv1* locus in F12/3xHA-GDV1/*gdv1*-asKO parasites and α -HA overview IFA in early schizonts (ES, 32-40 hpi). DIC, differential interference contrast. Scale bar, 5 μ m. Percentage of 3xHA-GDV1-positive parasites is shown at the right (results are the mean of three biological replicates (100 infected RBCs counted per sample); error bars indicate SD).

IME1 expression under non-inducing conditions (29, 30). These parallels raise the exciting possibility that evolutionary divergent unicellular eukaryotes may employ a conceptually similar regulatory logic to control entry into the sexual phases of their life cycles. Our study contributes to understanding the molecular pathway underlying the formation of malaria transmission stages and provides an experimental tool for the mass production of gametocytes to support the development of transmission-blocking interventions.

References

1. S. H. Kwon, J. L. Workman, The heterochromatin protein 1 (HP1) family: put away a bias toward HP1. *Mol. Cells* **26**, 217-227 (2008).
2. C. Flueck *et al.*, Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS. Pathog.* **5**, e1000569- (2009).
3. N. Rovira-Graells *et al.*, Transcriptional variation in the malaria parasite Plasmodium falciparum. *Genome Res* **22**, 925-938 (2012).
4. T. S. Voss, Z. Bozdech, R. Bartfai, Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr. Opin. Microbiol.* **20**, 88-95 (2014).
5. B. F. Kafsack *et al.*, A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* **507**, 248-252 (2014).
6. A. Sinha *et al.*, A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. *Nature* **507**, 253-257 (2014).
7. N. M. Brancucci *et al.*, Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host. Microbe* **16**, 165-176 (2014).
8. S. Eksi *et al.*, Plasmodium falciparum gametocyte development 1 (Pfgdv1) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS. Pathog.* **8**, e1002964- (2012).
9. L. A. Banaszynski *et al.*, A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004 (2006).
10. C. M. Armstrong, D. E. Goldberg, An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. *Nat. Methods* **4**, 1007-1009 (2007).
11. A. Gaspar-Maia *et al.*, Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**, 863-868 (2009).
12. L. Bugga, I. E. McDaniel, L. Engie, J. A. Armstrong, The Drosophila melanogaster CHD1 chromatin remodeling factor modulates global chromosome structure and counteracts HP1a and H3K9me2. *PLoS. ONE.* **8**, e59496- (2013).
13. M. C. Bruce, P. Alano, S. Duthie, R. Carter, Commitment of the malaria parasite Plasmodium falciparum to sexual and asexual development. *Parasitology* **100 Pt 2**, 191-200 (1990).
14. F. Silvestrini, P. Alano, J. L. Williams, Commitment to the production of male and female gametocytes in the human malaria parasite Plasmodium falciparum. *Parasitology* **121 Pt 5**, 465-471 (2000).
15. M. C. Bruce *et al.*, Cellular location and temporal expression of the Plasmodium falciparum sexual stage antigen Pfs16. *Mol. Biochem. Parasitol.* **65**, 11-22 (1994).
16. N. M. B. Brancucci *et al.*, Lysophosphatidylcholine Regulates Sexual Stage Differentiation in the Human Malaria Parasite Plasmodium falciparum. *Cell* **171**, (2017).
17. S. Eksi *et al.*, Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in Plasmodium falciparum. *Mol. Biochem. Parasitol.* **143**, 90-99 (2005).

18. P. Alano *et al.*, Plasmodium falciparum: parasites defective in early stages of gametocytogenesis. *Exp. Parasitol.* **81**, 227-235 (1995).
19. J. A. Pearce *et al.*, Characterisation of two novel proteins from the asexual stage of Plasmodium falciparum, H101 and H103. *Mol. Biochem. Parasitol.* **139**, 141-151 (2005).
20. A. Amambua-Ngwa *et al.*, Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS. Genet.* **8**, e1002992- (2012).
21. B. J. Morahan *et al.*, Functional analysis of the exported type IV HSP40 protein PfGECO in Plasmodium falciparum gametocytes. *Eukaryot. Cell* **10**, 1492-1503 (2011).
22. F. Silvestrini *et al.*, Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics* **9**, 1437-1448 (2010).
23. A. Poran *et al.*, Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* **551**, 95-99 (2017).
24. K. M. Broadbent *et al.*, Strand-specific RNA sequencing in Plasmodium falciparum malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC. Genomics* **16**, 454- (2015).
25. F. Huber *et al.*, Protein Abundance Control by Non-coding Antisense Transcription. *Cell Rep.* **15**, 2625-2636 (2016).
26. V. Pelechano, L. M. Steinmetz, Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880-893 (2013).
27. E. Hiriart, A. Verdell, Long noncoding RNA-based chromatin control of germ cell differentiation: a yeast perspective. *Chromosome Res.* **21**, 653-663 (2013).
28. F. J. van Werven, A. Amon, Regulation of entry into gametogenesis. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **366**, 3521-3531 (2011).
29. B. Gelfand *et al.*, Regulated antisense transcription controls expression of cell-type-specific genes in yeast. *Mol. Cell Biol.* **31**, 1701-1709 (2011).
30. F. J. van Werven *et al.*, Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150**, 1170-1181 (2012).
31. R. Edgar, M. Domrachev, A. E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210 (2002).
32. P. Alano, S. Premawansa, M. C. Bruce, R. Carter, A stage specific gene expressed at the onset of gametocytogenesis in Plasmodium falciparum. *Mol. Biochem. Parasitol.* **46**, 81-88 (1991).
33. W. Trager, J. B. Jensen, Cultivation of malarial parasites. *Nature* **273**, 621-622 (1978).
34. C. Lambros, J. P. Vanderberg, Synchronization of Plasmodium falciparum erythrocytic stages in culture. *J. Parasitol.* **65**, 418-420 (1979).
35. K. Witmer *et al.*, Analysis of subtelomeric virulence gene families in Plasmodium falciparum by comparative transcriptional profiling. *Mol. Microbiol.* **84**, 243-259 (2012).
36. T. S. Voss *et al.*, A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. *Nature* **439**, 1004-1008 (2006).
37. M. Ghorbal *et al.*, Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32**, 819-821 (2014).
38. T. G. Montague *et al.*, CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401-W407 (2014).
39. K. Labun *et al.*, CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44**, W272-W276 (2016).
40. J. A. Thomas *et al.*, Development and Application of a Simple Plaque Assay for the Human Malaria Parasite Plasmodium falciparum. *PLoS. ONE.* **11**, e0157873- (2016).
41. T. S. Voss *et al.*, Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters. *Mol Microbiol* **48**, 1593-1607 (2003).

42. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682 (2012).
43. C. A. Daubenberg *et al.*, The N¹-terminal domain of glyceraldehyde-3-phosphate dehydrogenase of the apicomplexan *Plasmodium falciparum* mediates GTPase Rab2-dependent recruitment to membranes. *Biol. Chem.* **384**, 1227-1237 (2003).
44. M. P. Malakhov *et al.*, SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5**, 75-86 (2004).
45. C. Flueck *et al.*, A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS. Pathog.* **6**, e1000784- (2010).
46. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207-234 (2005).
47. T. Ponnudurai, A. H. Lensen, J. F. Meis, J. H. Meuwissen, Synchronization of *Plasmodium falciparum* gametocytes using an automated suspension culture system. *Parasitology* **93 (Pt 2)**, 263-274 (1986).
48. R. Carter, P. M. Graves, "Gametocytes" in *Malaria: Principles and Practice of Malariology*, W. H. Wernsdorfer, I. McGregor, Eds. (Churchill Livingstone, Edinburgh, 1988), vol. 1, pp. 253-305.
49. W. A. Hoeijmakers, R. Bartfai, K. J. Francoijs, H. G. Stunnenberg, Linear amplification for deep sequencing. *Nat Protoc* **6**, 1026-1036 (2011).
50. P. R. Kensche *et al.*, The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Res.* **44**, 2110-2124 (2016).
51. Z. Bozdech *et al.*, The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS. Biol.* **1**, E5- (2003).
52. H. J. Painter, L. M. Altenhofen, B. F. Kafsack, M. Llinas, Whole-genome analysis of *Plasmodium* spp. Utilizing a new agilent technologies DNA microarray platform. *Methods Mol. Biol.* **923**, 213-219 (2013).
53. V. G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A* **98**, 5116-5121 (2001).
54. A. J. Saldanha, Java Treeview--extensible visualization of microarray data. *Bioinformatics.* **20**, 3246-3248 (2004).
55. W. A. Hoeijmakers, R. Bartfai, H. G. Stunnenberg, Transcriptome analysis using RNA-Seq. *Methods Mol. Biol.* **923**, 221-239 (2013).
56. P. Prommana *et al.*, Inducible knockdown of *Plasmodium* gene expression using the glmS ribozyme. *PLoS. ONE.* **8**, e73783- (2013).
57. P. Y. Watson, M. J. Fedor, The glmS riboswitch integrates signals from activating and inhibitory metabolites *in vivo*. *Nat. Struct. Mol. Biol.* **18**, 359-363 (2011).
58. W. C. Winkler *et al.*, Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286 (2004).

Acknowledgements

We are grateful to M. van de Vegte-Bolmer and R. Sauerwein for determining gametocyte sex ratios and providing α -Pfs16 antibodies, to D. Richard for providing the pL6-3HA_glmS-246 plasmid and to T. Haefliger for technical assistance. This work was supported by the Swiss National Science Foundation (grant numbers 31003A_143916, 31003A_163258, BSCGI0_157729) and the Netherlands Organization for Scientific Research (NWO-Vidi 864.11.007). All data and code to understand and assess the conclusions of this research are available in the main text, supplementary materials and via the following repository: Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) (31) accession GSE95549

(microarray data) and GSE94901 (ChIP-seq and RNA-seq data). M.F. designed and performed experiments, analysed data, prepared illustrations and wrote the paper. S.A.F. performed and analysed ChIP-Seq and RNA-seq experiments. I.N. designed and cloned CRISPR/Cas9 mother plasmids and performed experiments involving recombinant proteins. N.M.B.B. performed experiments related to the 3D7/3xHA-GDV1 and F12/3xHA-GDV1/*gdv1*-asKO parasites. E. Carrington performed and analysed RT-qPCR experiments. E. Carrió performed experiments involving 3D7/AP2-G-GFP-DDgImS parasites. S.M. performed LC-MS/MS experiments. P.J. provided conceptual advice. P.J., R.B and T.S.V. provided resources. R.B. designed, supervised and analysed experiments. T.S.V. conceived of the study and designed, supervised, and analysed experiments and wrote the paper. All authors contributed to editing of the manuscript.

Materials and Methods

Parasite culture, transfection constructs and transgenic cell lines

P. falciparum 3D7 and F12 parasites were cultured as described (33). Growth synchronization was achieved by repeated sorbitol treatments (34). To generate parasite lines 3D7/GDV1-GFP-DD and F12/GDV1-GFP-DD expressing ectopic GDV1-GFP-DD, the *gdv1* coding sequence (primers *gdv1*_pHGFPDD_BamHI and *gdv1*_pHGFPDD_NotI) was cloned in frame with *gfp-dd* into pHcamGFP-DD using *Bam*HI and *Sal*I. pHcamGFP-DD was obtained by inserting a PCR fragment encoding the FKBP destabilizing domain (DD) in frame with *gfp* into *Spe*I/*Sal*I-digested pHcamGFP (35). 3D7 and F12 parasites were transfected with pHcamGDV1-GFP-DD and selected with 4nM WR99210 (WR) as described (36). Induction of GDV1-GFP-DD expression was achieved by addition of 312.5 nM Shield-1 to the culture medium unless otherwise indicated.

We generated three new mother plasmids (p_gC, pH_gC and pB_gC) for CRISPR/Cas9-based genome engineering in *P. falciparum* using the following cloning steps. First, plasmid pgRNA was obtained by Gibson assembly joining the *Nco*I/*Pvu*II fragment containing the plasmid backbone and U6 single guide RNA (sgRNA) expression cassette of pL-6_eGFP (37) with a PCR fragment containing a *Hind*III-*Sal*I-*Eco*RI-*Bam*HI multi-cloning site (MCS) amplified from pET-32a(+) using primers PCR1_F and PCR1_R. In a second step, the *Bsa*I site in the *amp*^R gene in pgRNA was eliminated using a Gibson assembly reaction by replacing the 207 bp *Bsa*I/*Fsp*I *amp*^R fragment with a corresponding 207 bp PCR fragment obtained with primers PCR2_F and PCR2_R, the latter of which introduced a silent G to A mutation in the *Bsa*I recognition sequence (GGTCTC to GATCTC). This plasmid was named pgRNA2. In a third step, the *Btg*ZI recognition site in the U6 sgRNA expression cassette in pgRNA2 was replaced with a *Bsa*I recognition sequence to obtain pgRNA3. To achieve this, the *Btg*ZI-digested pgRNA2 plasmid and a 56 bp fragment obtained by annealing the two complementary oligonucleotides *Bsa*I_F and *Bsa*I_R were joined by Gibson assembly. In a fourth step, a Cas9 expression cassette was introduced into pgRNA3 using Gibson assembly to obtain the mother plasmid p_gC. Here, *Nco*I-digested pgRNA3 was joined with three PCR fragments amplified from pUF1_Cas9 (37) encoding (1) the *P. falciparum* *hsp86* promoter (primers PCR4a_F and PCR4a_R), (2) *Streptococcus pyogenes* Cas9 (primers PCR4b_F and PCR4b_R), and (3) the *P. berghei* *dhfr-ts* terminator (primers PCR4c_F and PCR4c_R). In the fifth step, a human dihydrofolate reductase (hDHFR) resistance cassette was introduced into p_gC by Gibson

assembly to obtain the mother plasmid pH_{gC}. This was achieved by assembling the *Bam*HI-digested p_{gC} with three PCR fragments encoding (1) the *P. falciparum cam* promoter amplified from 3D7 gDNA (primers PCR5a_F and PCR5b_R), (2) the *hdhfr* gene amplified from pL6_eGFP (primers PCR5b_F and PCR5b_R), and (3) the *P. falciparum hrp2* terminator amplified from pL6_eGFP (primers PCR5c_F and PCR5c_R). The mother plasmid pB_{gC} is a modified version of pH_{gC} where the hDHFR cassette is replaced with a blasticidin deaminase (BSD) resistance cassette using Gibson assembly. This was achieved by assembling *Bam*HI/*Xho*I-digested pH_{gC} with three PCR fragments representing (1) the *cam* promoter amplified from pH_{gC} with primers B1F and B1R; (2) the *bsd* gene amplified from pBcam_3xHA (2) using primers B2F and B2R; and (3) the *hrp2* terminator amplified from pH_{gC} using primers B3F and B3R. We also generated pHF_{gC} and pBF_{gC}, modified versions of pH_{gC} and pB_{gC}, respectively, expressing hDHFR fused to the bifunctional negative selection marker yeast cytosine deaminase/uridyl phosphoribosyl transferase (hDHFR-yFCU) instead of hDHFR alone. For pHF_{gC}, pH_{gC} was digested with *Xho*I and *Xba*I (cuts within *hdhfr*) and assembled with four PCR products: (1) the 3' end of *hdhfr* starting at the *Xba*I site amplified from pH_{gC} using primers HF1F and HF1R; (2) the 5' part of *yfcu* up to the internal *Bam*HI site amplified from pL6-eGFP (37) using primers HF2F and HF2R; (3) the 3' part of *yfcu* starting at the internal *Bam*HI site amplified from pL6-eGFP (37) using primers HF3F and HF3R (note that the separation of the *yfcu* sequence into two parts was required to eliminate the internal *Bam*HI site); and (4) the *hrp2* terminator amplified from pH_{gC} using primers HF4F and B3R. For pBF_{gC}, the *Bam*HI/*Xho*I-digested pB_{gC} vector was assembled with two PCR products: (1) the *cam* promoter and *bsd* coding sequence amplified from pB_{gC} using primers B1F and BF1R; and (2) the *yfcu* coding sequence and *hrp2* terminator amplified from pHF_{gC} with primers HF2F and B3R. The mother plasmid p_{gC} serves as a template to generate CRISPR/Cas9 knockout plasmids (single plasmid approach), and the pH_{gC}, pB_{gC}, pHF_{gC} and pBF_{gC} mother plasmids are used together with a donor plasmid for gene tagging.

To generate 3D7/AP2-G-GFP-DDgImS parasites we constructed the CRISPR/Cas9 transfection vector pH_{gC}-*ap2g*-3'-*gfp*-*dd*-*gImS* allowing for marker-free genome editing using a single plasmid. We first generated plasmid pH_{gC}-*ap2g*-3' by inserting two annealed complementary oligonucleotides (*sgt*_{*ap2g3'*}_F, *sgt*_{*ap2g3'*}_R) encoding the sgRNA target sequence *sgt*_{*ap2-g*}-3' and appropriate single-stranded overhangs into the sgRNA expression cassette in pH_{gC} using *Bsa*I digestion and T4 DNA ligase. The sgRNA target sequence *sgt*_{*ap2g3'*} (agttatagggaaattcaaa) is positioned 60 bp downstream of the *ap2-g* coding sequence and was identified using CHOPCHOP (38, 39). We then generated a second plasmid carrying an *ap2g-gfp-dd-gImS* donor assembly (pD_{*ap2g-gfp-dd-gImS*}) using two subsequent cloning steps. First, plasmid pD_{*ap2g-gfp*} was generated using Gibson assembly joining four PCR fragments encoding (1) the plasmid backbone amplified from pUC19 using primers PCRA_F and PCRA_R, (2) a 393 bp *ap2-g* 5' homology box (HB) (spanning bps 6904-7296 of the *ap2-g* gene) amplified from 3D7 gDNA using primers PCRB_F and PCRB_R, (3) the *gfp* coding sequence (plus N-terminal GSAG linker) amplified from pH-GFP (7) using primers PCRC_F and PCRC_R, and (4) a 670 bp *ap2-g* 3' HB (spanning the region 60-730 bp downstream of the *ap2-g* gene) amplified from 3D7 gDNA using primers PCRD_F and PCRD_R. Second, pD_{*ap2g-gfp-dd-gImS*} was constructed using Gibson assembly of three PCR fragments representing (1) the *ap2-g* 5' HB/pUC19 backbone/*ap2-g* 3' HB assembly amplified from pD_{*ap2g-gfp*} using primers 1F and 1R, (2) the *gfp-dd* sequence amplified from pH-GFP-DD (7) using primers 2F and 2R, and (3) the *gImS*-246 sequence amplified from pL6-3HA_{*gImS*}-246 (kind gift from Dave Richard) using primers 3F and 3R. The final CRISPR/Cas9

transfection vector pH_gC-*ap2g-3'-gfp-dd-glmS* was then obtained by introducing the *EcoRI/HindIII*-digested *ap2g-gfp-dd-glmS* donor assembly amplified from pD_*ap2g-gfp-dd-glmS* (primers EcoF and HindR) into *EcoRI/HindIII*-digested pH_gC-*ap2g-3'*. The pH_gC-*ap2g-3'-gfp-dd-glmS* construct was transfected into 3D7 parasites and 3D7/AP2-G-GFP-DDglmS parasites were selected with 4 nM WR and 2.5 mM (D)- α -glucosamine (GlcN) (Sigma Aldrich) for six days and then maintained in absence of WR until a stably propagating parasite population was established. Successful editing of the *ap2-g* locus was confirmed by PCR on gDNA. Stabilisation of AP2-G-GFP-DD expression was achieved by removing GlcN and adding 1350 nM Shield-1 to the culture medium.

To disrupt the antisense RNA locus downstream of the *gdv1* coding sequence in F12 wild-type and 3D7/AP2-G-GFP-DDglmS parasites we generated the CRISPR/Cas9 knockout vectors p_gCH-*gdv1*-askO and p_gCB-*gdv1*-askO, respectively. First, p_gC-*gdv1*-askO-pre was created by inserting a *hdhfr* resistance cassette flanked by the 5' and 3' HBs into p_gC using Gibson assembly. This was achieved by assembling *EcoRI/HindIII*-digested p_gC with three PCR fragments encoding (1) a 440 bp 5' HB amplified from 3D7 gDNA (spanning the last 93 bp of the *gdv1* coding sequence and 347 bp of the downstream sequence; primers p_gC-*gdv1*-askO_5'HB_for and p_gC-*gdv1*-askO_5'HB_rev), (2) a *hdhfr* resistance cassette amplified from pH_gC (primers p_gCH-*gdv1*-askO_5'HB_for and p_gCH-*gdv1*-askO_5'HB_rev), and (3) a 753 bp 3' HB amplified from 3D7 gDNA (spanning the region 3,554-4,307 bp downstream of the *gdv1* coding sequence; primers p_gCH-*gdv1*-askO_3'HB_for and p_gCH-*gdv1*-askO_3'HB_rev). To generate the final p_gCH-*gdv1*-askO construct, the two complementary oligonucleotides (sgt_*gdv1*-3'_F and sgt_*gdv1*-3'_R) encoding the sgRNA target sequence sgt_*gdv1*-3' and appropriate single-stranded overhangs were annealed and inserted into the sgRNA expression cassette using *BsaI*-digested p_gC-*gdv1*-askO-pre and T4 DNA ligase. The sgRNA target sequence sgt_*gdv1*-3' (tgtattactaaagaatgaaa) is positioned 1,639 bp downstream of the *gdv1* coding sequence and was identified using CHOPCHOP (38, 39). For the p_gCB-*gdv1*-askO construct, p_gCH-*gdv1*-askO was cut with *XhoI* and *EcoRI* to release the *hdhfr* resistance cassette and the 3' HB. Gibson assembly was then used to insert PCR fragments encoding the blasticidin deaminase (*bsd*) resistance cassette amplified from pB_gC (primers p_gCB-*gdv1*-askO_for and p_gCB-*gdv1*-askO_rev) and the identical 753 bp 3'HB amplified from 3D7 gDNA (primers p_gCB-*gdv1*-askO_3'HB_for and p_gCB-*gdv1*-askO_3'HB_rev). The p_gCH-*gdv1*-askO construct was transfected into F12 parasites and F12/*gdv1*-askO parasites were selected with 4 nM WR. The p_gCB-*gdv1*-askO construct was transfected into 3D7/AP2-G-GFP-DDglmS parasites grown in presence of 2.5 mM GlcN and 3D7/AP2-G-GFP-DDglmS/*gdv1*-askO parasites were selected with 2.5 μ g/ml blasticidin-S-HCl (BSD). Successful disruption of the *gdv1*-asRNA locus and replacement with the *hdhfr* or *bsd* resistance cassettes was confirmed by PCR on gDNA. Stabilisation of AP2-G-GFP-DD expression was achieved by removing GlcN and adding 1350 nM Shield-1 to the culture medium.

To create the 3D7/3xHA-GDV1 and F12/3xHA-GDV1/*gdv1*-askO parasite lines we prepared the CRISPR/Cas9 transfection vectors pB_gC-*gdv1*-5' and the donor plasmid pFDon_3xha-*gdv1*. pFDon contains an expression cassette for *yfcu* under control of the *hsp86* promoter and was cloned by Gibson assembly of four PCR fragments: (1) the pUC18 *E. coli* plasmid backbone amplified with primers D1F and D1R; (2) the *hsp86* promoter amplified from pHF_gC using primers D2F and D2R; (3) the *yfcu* coding sequence and *hrp2* terminator amplified from pHF_gC using primers D3F and D3R; and (4) the MCS amplified from pHF_gC using primers D4F and D4R. To obtain pFDon_3xha-*gdv1*, three PCR fragments were inserted into the pFDon plasmid by Gibson assembly: (1) a 5' HB spanning 504 bp of

the *gdv1* upstream region (-505 to -1) amplified from 3D7 gDNA (primers pFDon_GDV1_HA_for and pFDon_GDV1_HA_rev); (2) a 93 bp fragment coding for a 3xHA-tag amplified from pBcam_3xHA (2) (primers pFDon_GDV1_HA_HAtag_for and pFDon_GDV1_HA_HAtag_rev); and (3) a 3' HB spanning 488 bp of the *gdv1* coding region (+3 to +491) amplified from 3D7 gDNA. The 3' HB fragment was created in a two-step process. First, a fragment containing three mismatch mutations in the sgRNA target sequence *sgt_gdv1-5'* (tgaaagtaaaaggaatagt to tgaaGgtTaaaggCaatagt) was amplified from 3D7 gDNA (primers pFDon_GDV1_HA_3'HB_mut_for and pFDon_GDV1_HA_3'HB_mut_rev). This sequence was then used as template for a second round of amplification adding overhangs for Gibson assembly (primers pFDon_GDV1_HA_3'HB_for and pFDon_GDV1_HA_3'HB_rev). To create the pB_gC-*gdv1-5'* CRISPR/Cas9 plasmid two complementary oligonucleotides (*sgt_gdv1-5'_F* and *sgt_gdv1-5'_R*) encoding the sgRNA target sequence *sgt_gdv1-5'* (tgaaagtaaaaggaatagt; positioned 28 bp downstream of the start codon of the *gdv1* coding sequence) were annealed and inserted into the *BsaI* site in pB_gC. Equal amounts of both plasmids were transfected into 3D7 wild-type and F12/*gdv1*-asKO parasites and 3D7/3xHA-GDV1 and F12/3xHA-GDV1/*gdv1*-asKO parasites were selected with 5 µg/ml BSD for 10 days and then maintained in absence of drug pressure until stably propagating parasite populations were established. Successful editing of the *gdv1* locus was confirmed by PCR on gDNA. The 3D7/3xHA-GDV1 line was cloned out as described (40). All oligonucleotide sequences used for cloning and confirmation of CRISPR/Cas9-based gene editing by PCR are provided in Table S9.

Preparation of parasite nuclear extracts

Nuclei were isolated as described (41). Prior to nuclear protein extraction, nuclei were treated with DNase I and RNase A (20mM Hepes pH 7.4, 10mM NaCl, 5mM MgCl₂, 1mM CaCl₂, 0.1% NP-40, 1x protease inhibitor (Roche Diagnostics), 1x PhosSTOP (Roche Diagnostics), 0.5 U/ml DNase I (Roche Diagnostics), 1mg RNase A (Roth)) in a 250 ml reaction for 5min at 37°C. Reactions were adjusted to 500 ml using 2x high salt extraction buffer (20mM Hepes pH 7.4, 1M NaCl, 6mM EDTA, 2mM EGTA, 2mM TCEP, 1x protease inhibitor (Roche Diagnostics), 1x PhosSTOP (Roche Diagnostics), 50% Glycerol, 0.1%NP-40) and nuclear proteins were extracted in 2.5 pellet volumes high salt extraction buffer (20mM Hepes pH 7.4, 500mM NaCl, 3mM EDTA, 1mM EGTA, 1mM TCEP, 1x protease inhibitor (Roche Diagnostics), 1x PhosSTOP (Roche Diagnostics), 25% Glycerol, 0.1%NP-40) containing 0.02mg/ml H3K9me3 peptides (Diagenode) and incubated at 4°C for 1 h. After centrifugation at 20800g at 4°C for 30min supernatants were immediately used for co-IP experiments.

Co-Immunoprecipitation (co-IP)

High salt nuclear extract from 3D7/HP1-GFP or 3D7/GDV1-GFP-DD^{ON} schizonts was diluted to 250mM NaCl using dilution buffer (20mM Hepes pH 7.4, 1mM EDTA, 1mM TCEP, 1x protease inhibitor (Roche Diagnostics), 1x PhosSTOP (Roche Diagnostics), 25% Glycerol) and incubated with negative control agarose beads (Chromotek) for 30 min at 4°C under constant agitation for lysate pre-clearing. The supernatant was split and incubated with GFP-Trap agarose beads (Chromotek) or negative control agarose beads (Chromotek) for 1 h at 4°C under constant agitation. The supernatant was recovered and beads were washed four times in wash buffer 1 (20mM Hepes pH 7.4, 250mM NaCl, 1mM EDTA, 1mM TCEP, 0.05% NP-40) and once in wash buffer 2 (20mM Hepes pH 7.4, 250mM NaCl, 1mM EDTA, 1mM TCEP). Bound proteins were eluted for 1 h at RT under constant agitation using 2M arginine

buffer (2M arginine, pH 4) followed by a second elution in 2% SDS buffer (2% SDS, 20mM Tris-HCl, pH 8) for 10 min at 95°C. All centrifugation steps were performed at 4°C for 3 min at 500g. Samples were analysed by Western blot using mouse mAb a-GFP (Roche Diagnostics #11814460001) or rabbit a-PfHP1 (7) antibodies.

Capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS)

For the HP1 co-IP experiments, both the 2M arginine elution and the 2% SDS elution were analysed by LC-MS/MS. For the GDV1-GFP-DD co-IP experiments only the 2% SDS elution samples were analysed. Proteins in the 2M arginine elution buffer were reduced with 10 mM DTT at 37°C for 1 hr and alkylated with 50 mM iodo-acetamide for 15 min at room temperature. Proteins were digested with 250 ng endoproteinase LysC (Wako, Neuss, Germany) for two hours at 37°C followed by 500 ng trypsin (Worthington, Lakewood, NJ, USA) overnight. The digest was stopped with 1% TFA and desalted on a microspin column (The Nest Group, Southborough, MA, USA) according to the manufacturer's recommendations. Proteins eluted with 2% SDS were precipitated with 10% TCA, washed twice in ice-cold acetone and air-dried. The protein pellets were dissolved in 30 ml 100 mM Tris-HCl (pH 8.0)/6M Urea, containing 10 mM DTT and reduced, alkylated and digested with 250 ng endoproteinase LysC as above. The urea concentration was diluted to 2 M with 100 mM Tris-HCl (pH 8.0) and the sample was further digested with 500 ng trypsin overnight at 37°C. The digest was acidified with 1% TFA and the sample was desalted on a MicroSpin cartridge.

The peptides were dried in a SpeedVac and dissolved in 40 mL 0.1% formic acid and analysed by capillary liquid chromatography tandem MS (LC-MS/MS) using a home-packed separating column (0.075 mm x 15 cm) packed with Reprosil C18 reverse-phase material (2.4 mm particle size, Dr. Maisch, Ammerbuch-Entringen, Germany). The column was connected on line to an Orbitrap Elite FT hybrid instrument (Thermo Scientific, Reinach, Switzerland). The solvents used for peptide separation were 0.1% formic acid in water/0.005% TFA (solvent A) and 0.1% formic acid/0.005% TFA and 80% acetonitrile in water (solvent B). 2 ml of peptide digest were injected with a Proxeon nLC capillary pump (Thermo Scientific) set to 0.3 ml/min. A linear gradient from 0 to 40% solvent B in solvent A in 95 min was delivered with the nano pump at a flow rate of 300 nl/min. After 95 min the percentage of solvent B was increased to 75% in ten minutes. The eluting peptides were ionized at 2.5 kV. The mass spectrometer was operated in data-dependent mode. The precursor scan was done in the Orbitrap set to 60,000 resolution, while the fragment ions were mass analyzed in the LTQ instrument. A top twenty method was run so that the twenty most intense precursors were selected for fragmentation. The MS/MS spectra were searched against a combined *P. falciparum* (www.plasmoDB.org; release 11.1)/human annotated protein database using Proteome Discoverer 1.4 (Thermo Scientific, Reinach, Switzerland) using the two search engines Mascot and SequestHT (Tables S3 and S4). For the search, oxidized methionine and N-terminal protein acetylation were used as variable modifications. The identifications were filtered to a peptide false discovery rate of 1%.

Fluorescence microscopy

IFAs were performed with methanol-fixed or 4% formaldehyde/0.0075% glutaraldehyde-fixed cells using the following primary antibodies: mouse mAb a-GFP (Roche Diagnostics #11814460001), 1:100; rat mAb a-HA (3F10, Roche Diagnostics #11867423001), 1:100; rabbit a-PfHP1 (7), 1:200; mouse mAb a-Pfs16 (kind gift from Robert W. Sauerwein), 1:500. Secondary antibodies: Alexa Fluor 568-

conjugated a-rabbit IgG (Molecular Probes #A11011), 1:500; Alexa Fluor 488-conjugated a-rabbit IgG (Molecular Probes #A11008), 1:500; Alexa Fluor 488-conjugated a-mouse IgG (Molecular Probes #A11001), 1:500; Alexa Fluor 488-conjugated a-rat IgG (Molecular Probes #A11006), 1:250; Alexa Fluor 568-conjugated a-rat IgG (Molecular Probes #A11077), 1:250. Nuclei were stained with DAPI. Images were taken on a Leica DM 5000B microscope with a Leica DFC 345 FX camera, acquired via the Leica IM1000 software, and processed using Fiji (42). For each experiment, images were acquired and processed with identical settings.

Western Blot

Parasites were released from infected RBCs by saponin lysis, resuspended in 8M Urea, 5%SDS, 40mM Tris pH 6.8, 1% β -Mercaptoethanol, 1x protease inhibitor (Roche Diagnostics) and separated on NuPage 4-12% Bis-Tris gels (Novex). Proteins were detected with mouse mAb a-GFP (Roche Diagnostics #11814460001), 1:500; rabbit a-PfHP1 (7), 1:5'000; mouse mAb a-PfGAPDH (43), 1:20'000; rat mAb a-HA (3F10, Roche Diagnostics #11867423001), 1:10'000; mouse mAb a-6xHis (R&D Systems #MAB050R), 1:5'000.

In vitro protein-protein interaction experiments

A plasmid vector with the following expression cassette was constructed for co-expression of HP1 and GDV1: T7 promoter/*strep(II)*-*hp1*/T7 promoter/*6xhis*-*sumo*-*gdv1*/T7 terminator (pStrep-HP1_HS-GDV1). We used pETA-HS and pETA-Strep as parental vectors, which we previously constructed for the expression of N-terminal 6xHis-SUMO-tagged and N-terminally Strep(II)-tagged fusion proteins, respectively. pETA-HS was obtained by introducing a sequence encoding a 6xHis stretch followed by *Saccharomyces cerevisiae* SUMO (amplified from gDNA using primers SUMO_F and SUMO_R between the *NdeI* and *BamHI* sites of pET20(b)+ (Novagen), yielding a vector similar to the one described elsewhere (44). pETA-Strep was generated by replacing the *NdeI/BamHI* fragment in pET20b(+) with an annealed double-stranded oligonucleotide (Strep_F and Strep_R) encoding Met-Ala-Ser-Strep(II). The co-expression plasmid pStrep-HP1_HS-GDV1 was cloned by Gibson assembly of four PCR products representing (1) the backbone of pETA-Strep amplified using the primers SF and SR; (2) the *hp1* coding sequence (cds) starting at +4 amplified from 3D7 gDNA using primers HP1F and HP1R; (3) the T7 promoter and the sequence encoding the 6xHis-SUMO tag amplified from pETA-HS with primers HSF and HSR; and (4) the *gdv1* cds starting at +4 amplified from 3D7 gDNA using primers GDV1F and GDV1R. A control vector with the same expression setup was made by replacing the *gdv1* cds with a fragment encoding amino acids M173 to N311 of the ApiAP2 DNA-binding protein SIP2 (45). This fragment was first ligated into pETA-HS (resulting in pETA-HS-SIP2) using the *SfoI* and *XhoI* restriction sites and the *XhoI*-digested *sip2* PCR product amplified from 3D7 gDNA with primers SIP2F and SIP2R. Next, the pStrep-HP1_HS-SIP2 co-expression vector was Gibson assembled using two PCR products representing (1) the pStrep-HP1_HS-GDV1 vector omitting the *gdv1* cds (amplified using primers SHHSF and SHHSR); and (2) the *sip2* cds fragment amplified from pETA-HS-SIP2 using primers SIP2F2 and SIP2R2. After cloning in *E. coli* DH5alpha and DNA sequencing, the plasmids were transformed into Rosetta2(DE3) cells (Novagen) for recombinant protein expression.

Recombinant proteins were expressed by auto-induction at 25°C (46). Once the cultures were grown to saturation, the bacteria were spun down and pellets frozen. Stock solutions for chromatography were 3 M arginine phosphate (3 M L-arginine, 1.6 M H₃PO₄, pH 7.3), 4 M imidazole phosphate (4 M

imidazole, 0.888 M H_3PO_4), 4 M ammonium phosphate (2.8 M $(\text{NH}_4)_2\text{HPO}_4$, 1.2 M $(\text{NH}_4)\text{H}_2\text{PO}_4$). Recombinant proteins were purified using the following strategy: (1) nickel affinity purification: bacterial pellets were thawed, resuspended in buffer Ni-A (200 mM arginine phosphate, 500 mM NaCl, 20 mM imidazole phosphate, 5 mM 6-aminocaproic acid, 0.5 mM TCEP, pH 7.2) and lysed by sonication. The lysate was cleared by centrifugation and loaded on a 5 ml HisTrap HP column (GE-Healthcare). The column was washed with 15 column volumes (CV) of buffer Ni-A and the proteins were eluted using five CV-long linear gradient of Ni-A to Ni-B (Ni-A with 400 mM imidazole phosphate). Both the Strep(II)-HP1/HisSUMO-GDV1 complex and HisSUMO-SIP2 eluted in the same range of the gradient. (2) StrepTactin affinity purification: the nickel column elution obtained from the GDV1/HP1 lysate was loaded on a 5 ml StrepTrap HP column (GE Healthcare), washed with ten CV of buffer Strep-A (200 mM arginine phosphate, 300 mM ammonium phosphate, 30 mM NaOH, 5 mM 6-aminocaproic acid, 1 mM EDTA, 0.5 mM TCEP, pH 7.4) and eluted using buffer Strep-B (400 mM arginine phosphate, 300 mM ammonium phosphate, 5 mM 6-aminocaproic acid, 1 mM EDTA, 2.5 mM desthiobiotin, 0.5 mM TCEP, pH 7.4). To purify Strep(II)-HP1 without interaction partner, the nickel column flow through obtained from the SIP2-HP1 control lysate was used and purified using the same conditions. (3) Hydrophobic interaction chromatography: the salt concentration of the StrepTactin column elutions were adjusted to 0.75 M ammonium phosphate and loaded on a 1 ml Phenyl HP column (GE Healthcare). The column was washed using ten CV of buffer HIC-A (400 mM arginine phosphate, 750 mM ammonium phosphate, 5 mM 6-aminocaproic acid, 1 mM EDTA, 0.5 mM TCEP, pH 7.2) and the proteins were eluted using a 25 CV-long linear gradient of buffer HIC-A to HIC-B (400 mM arginine phosphate, 5 mM 6-aminocaproic acid, 1 mM EDTA, 50 % (w/v) ethylene glycol, 0.5 mM TCEP, pH 7.2). The Strep(II)-HP1/HisSUMO-GDV1 complex was concentrated using an Amicon spin filter (Millipore) with a 10 K cut-off. Protein samples were analysed by SDS/polyacrylamide gel electrophoresis followed by Coomassie staining or Western blot analysis.

All oligonucleotide sequences used for cloning of expression vectors are provided in Table S9.

Gametocyte conversion, differentiation and sex ratios

To quantify sexual conversion rates of 3D7/GDV1-GFP-DD parasites, synchronous 3D7/GDV1-GFP-DD^{OFF} cultures were split at 8-16 hours post-invasion and one half was maintained in the absence of Shield-1 and one half was cultured in the presence of Shield-1 (3D7/GDV1-GFP-DD^{ON}). After re-invasion, parasites were fixed at 38-46 hpi either with methanol (Fig. 2B) or 4% formaldehyde/0.0075% glutaraldehyde (Fig. S2) and a-Pfs16 IFAs combined with DAPI staining was performed to quantify sexual conversion rates (proportion of stage I gametocytes among all infected RBCs). To assess gametocyte differentiation, cultures were treated with 50 mM N-acetyl-D-glucosamine (GlcNAc) for the first four days after re-invasion to eliminate asexual parasites (47) and then with normal culture medium for another 4-6 days when mature stage V gametocytes were observed in Giemsa-stained thin blood smears.

To quantify male and female gametocytes synchronous 3D7/GDV1-GFP-DD^{OFF} cultures were split at 8-16 hours post-invasion and one half was maintained in the absence of Shield-1 while the other half was cultured in the presence of Shield-1 (3D7/GDV1-GFP-DD^{ON}). After re-invasion, parasites were treated with 50 mM GlcNAc as above. Gametocytaemia was determined four days after re-invasion by visual inspection of Giemsa-stained blood smears and used to calculate the sexual conversion rate (gametocytaemia observed on day four in relation to the total parasitaemia observed on day one after

re-invasion). Eleven days after re-invasion male and female gametocytes were quantified by inspection of Giemsa-stained blood smears as described elsewhere (48) (for each population three individual blood smears were scored blindly).

To assess 3xHA-GDV1 expression under growth conditions that either suppress or induce sexual commitment, 3D7/3xHA-GDV1 parasites were split at 24-30 hours post-invasion and pulsed with defined minimal fatty acid medium (RPMI-1640 medium, 25 mM HEPES, 24 mM sodium bicarbonate, 100 mM hypoxanthine, 0.39% fatty acid-free BSA (Sigma-Aldrich), 30 mM oleic acid, 30 mM palmitic acid), either supplemented with 2 mM choline (suppresses sexual commitment) or lacking choline (induces sexual commitment), respectively (16). Twelve hours after pulsing relative 3xHA-GDV1 expression levels were determined by Western blot and IFAs were performed to determine the proportion of parasites expressing 3xHA-GDV1 in haemozoin-positive parasites (1n-4n). Sexual conversion rates were determined as described above by visual inspection of Giemsa-stained blood smears after GlcNAc treatment.

To quantify sexual conversion rates of 3D7/AP2-G-GFP-DDgImS and 3D7/AP2-G-GFP-DDgImS/*gdv1*-askO parasites, synchronous cultures were split at 24-32 hours post-invasion and one half was maintained in the presence of glucosamine/absence of Shield-1 and one half was cultured in the absence of glucosamine/presence of Shield-1. Sexual conversion rates were determined as described above by a-Pfs16 IFAs on methanol-fixed thin smears.

Chromatin immunoprecipitations, high throughput sequencing and data analysis

3D7/GDV1-GFP-DD^{OFF} parasites were synchronized twice 18 hours apart to obtain a six hour growth window (18-24 hpi). At 28-34 hpi cultures were split and one half was maintained in the absence of Shield-1 and one half was cultured in the presence of Shield-1 (3D7/GDV1-GFP-DD^{ON}). Paired samples of 3D7/GDV1-GFP-DD^{OFF} and 3D7/GDV1-GFP-DD^{ON} parasites were harvested at three consecutive time points (TPs) two (30-36 hpi, TP1), six (34-40 hpi, TP2) and ten (38-44 hpi, TP3) hours after addition of Shield-1 by crosslinking with 1% formaldehyde for 15 min at 37°C. The crosslinking reaction was quenched by 0.125 M glycine. Nuclei were isolated by releasing parasites from infected RBCs using 0.05% saponin followed by lysis in CLB (20 mM Hepes, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.65% NP-40, 1 mM DTT, 1x protease inhibitor (Roche Diagnostics)). Nuclei were washed and snap-frozen in CLB supplemented with 50% glycerol. Nuclei were resuspended in sonication buffer (50 mM Tris pH 8, 1% SDS, 10 mM EDTA, 1x protease inhibitor (Roche Diagnostics)) and sonicated for 20 cycles of 30 sec ON/30 sec OFF (setting high, Bioruptor™ Next Gen, Diagenode). Fragment sizes ranged from 100-600 bp as determined by de-crosslinking a 50 ml aliquot and running the purified DNA on a 1% agarose gel. ChIPs were performed under rotation overnight at 4°C in incubation buffer (5% Triton-X-100, 750 mM NaCl, 5 mM EDTA, 2.5 mM EGTA, 100 mM Hepes pH 7.4) by combining sonicated chromatin (500 ng DNA content) with either 1 mg mouse mAb a-GFP (Roche Diagnostics #11814460001) or 1 mg rabbit a-PfHP1 (7) in the presence of 10 ml protA and 10 ml protG Dynabeads (Life Technologies, #10008D and #10009D). Beads were washed for 5 min at 4°C while rotating with 400 ml of the following wash buffers: 2x wash buffer 1 (0.1% SDS, 0.1% DOC, 1% Triton-X100, 1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes pH 7.4), 1x wash buffer 2 (0.1% SDS, 0.1% DOC, 1% Triton-X100, 500 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes pH 7.4), 1x wash buffer 3 (250 mM LiCl, 0.5% DOC, 0.5% NP-40, 1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes pH 7.4), 2x wash buffer 4 (1 mM EDTA, 0.5 mM EGTA, 20 mM Hepes pH 7.4). The immunoprecipitated chromatin was eluted in 200 ml elution buffer (1%

SDS, 0.1 M NaHCO₃) while rotating for 20 min at RT and de-crosslinked (1% SDS, 0.1 M NaHCO₃, 1 M NaCl) in a 45°C shaking heat-block overnight. In parallel, 30 ml of sonicated input chromatin was de-crosslinked under the same conditions. The DNA was purified using QIAquick MinElute PCR columns (Qiagen) and eight separate a-GFP ChIPs or four separate a-PfHP1 ChIPs were combined.

To prepare sequencing libraries 1 ng of input DNA, 1 ng a-GFP ChIP DNA or 5 ng a-HP1 ChIP DNA were end repaired, extended with 3' A-overhangs and ligated to barcoded NextFlex adapters (Bio Scientific, #514122) as described (49). Libraries were amplified using an optimized KAPA protocol (50) using KAPA HiFi HotStart ready mix (KAPA Biosystems, KM2602), NextFlex primer mix (Bio Scientific, #514122) and the following PCR program: 98°C for 2 min; four cycles of 98°C for 20 sec, 62°C for 3 min; 62°C for 5 min. Amplified libraries were size-selected for 225-325 bp (mono-nucleosomes + 125 bp NextFlex adapter) using 2% E-Gel Size Select agarose gels (Invitrogen, #G6610-02) and amplified by PCR for ten cycles using the above conditions. Libraries were purified and adapter dimers removed by Agencourt AMPure XP beads purification using a 1:1 library:beads ratio (Beckman Coulter, #A63880). ChIP-seq libraries were sequenced on the Illumina NextSeq 500 system to generate 75 bp single-end reads (NextSeq 500/550 High Output v2 kit). Reads were mapped against the *P. falciparum* 3D7 reference genome from PlasmoDB v26 (www.plasmodb.org) using BWA samse (v0.7.12-r1039) and filtered to mapping quality ≥ 15 (SAM tools v1.2). Only uniquely mapped reads (between 3.7 and 11.5 million reads) were used for further analysis.

To visualize ChIP-seq data in the UCSC Genome browser, all libraries were normalized to the amount of mapped reads per million (RPM). Bedgraph files were generated using bedtools (v2.20.1). For log₂ ratio tracks HP1 or GDV1-GFP-DD ChIP values were divided by input values and log₂-transformed. For subtraction tracks ChIP values obtained from 3D7/GDV1-GFP-DD^{OFF} parasites were subtracted from those obtained from 3D7/GDV1-GFP-DD^{ON} parasites. Within the UCSC genome browser tracks were smoothed (10-16) and the windowing function was set as 'mean'. To assess HP1 and GDV1-GFP-DD coverage of individual genes, tags were counted for each coding region and offset by +1 to avoid division by zero while calculating fold change in coverage. Coding region counts were normalized to the number of reads per kb per million mapped reads (RPKM). Genes with very low mapability (input RPKM < 5) were excluded from further analysis. ChIP-seq enrichment values (ChIP/input) were calculated over each coding region.

Microarray experiments and data analysis

3D7/GDV1-GFP-DD^{OFF} parasites were synchronized twice 16 hours apart to obtain an eight hour growth window (16-24 hpi). After re-invasion parasites were synchronised again at 0-8 hpi. At 4-12 hpi the culture was split and one half was maintained in the absence of Shield-1 and one half was cultured in the presence of Shield-1 (3D7/GDV1-GFP-DD^{ON}). For both populations, total RNA was isolated in the same cell cycle (generation 1) from trophozoites (T1; 24-32 hpi), early schizonts (ES1; 32-40 hpi), late schizonts (LS1; 40-48 hpi), and after re-invasion (generation 2) from early ring stages (ER2; 8-16 hpi), late ring stages (LR2; 16-24 hpi), trophozoites (T2; 24-32 hpi), early schizonts (ES2; 32-40 hpi). Corresponding samples of F12/GDV1-GFP-DD^{OFF} and F12/GDV1-GFP-DD^{ON} parasites were prepared accordingly. RNA extraction and cDNA synthesis was carried out as described (51). Cy5-labelled test cDNAs were hybridised against Cy3-labelled cDNA reference pool consisting of total RNA obtained from 3D7 wild-type parasites harvested at five consecutive TPs along intra-erythrocytic parasite development (7). Equal amounts of Cy5- and Cy3-labelled samples were hybridised on a

P. falciparum 8x15K Agilent gene expression microarray (GEO platform ID GPL15130) for 16 hours at 65°C in an Agilent hybridisation oven (G2545A) (52). Slides were scanned using the GenePix scanner 4000B and GenePix pro 6.0 software (Molecular Devices). The raw microarray data representing relative transcript abundance ratios between each test sample and the reference pool (Cy5/Cy3 log2 ratios) were subjected to lowess normalization and background filtering as implemented by the Acuity 4.0 program (Molecular Devices). Flagged features and features with either Cy3 or Cy5 intensities lower than two-fold the background were discarded. Log2 ratios for multiple probes per gene were averaged. Transcripts showing expression values at all seven consecutive paired time points were included for downstream analysis. We performed paired two-class Significance Analysis of Microarrays (SAM) (53) to identify genes with significant differential expression between 3D7/GDV1-GFP-DD^{OFF} and 3D7/GDV1-GFP-DD^{ON} or F12/GDV1-GFP-DD^{OFF} and F12/GDV1-GFP-DD^{ON} parasites, respectively (q-value (fdr) cut-off <0.15; mean fold change cut-off >1.5). Heatmaps were generated using Java Treeview (54). Scatterplots were generated using Microsoft Excel. The processed microarray data are listed in Table S5 (3D7/GDV1-GFP-DD) and S6 (F12/GDV1-GFP-DD).

RNA-seq library preparation, high throughput sequencing and data analysis

F12 and F12/*gdv1*-asKO parasites were synchronized twice 16 hours apart to obtain an eight hour growth window. Parasites were harvested at 32-40 hpi and at 40-48 hpi. Total RNA was isolated using Ribozol (Amresco) according to the manufacturer's manual and further purified using the RNeasy Plus Mini Kit (Qiagen) for removal of gDNA. Residual gDNA was digested with TURBO DNA-free™ DNaseI (Ambion) and the RNA was quantified by NanoDrop. Total RNA was polyA-selected using the Oligotex mRNA Mini Kit (Qiagen, #70022) according to manufacturer's instructions. Subsequently, 2 mg of polyA-selected total RNA equivalent were fragmented by alkaline hydrolysis (5x fragmentation buffer: 200 mM Tris-acetate pH 8.2, 500 mM potassium acetate, 150 mM magnesium acetate) for 1 min 45 sec at 85°C in a 150 ml volume and precipitated as previously described (55), followed by two TURBO DNase treatments (Ambion, #AM2238) to remove possible remnants of genomic DNA. Further library preparation for strand-specific RNA-seq was carried out as described (50). In short, first-strand cDNA synthesis was performed with AT-corrected random N9 primers (76% AT) in presence of 0.2 mg Actinomycin D (Thermo 446 Fisher Scientific) to prevent DNA-dependent second-strand cDNA synthesis. To maintain the strand-specific information during second-strand synthesis dTTPs were replaced with dUTPs. Resulting double-stranded cDNAs (1.9-3.2 ng) were end-repaired and extended with 3' A-overhangs. After ligation of NextFlex adapters (Bio Scientific, #514122), libraries were treated with USER enzyme (NEB, #M5505L) to specifically degrade the dUTP-containing second strand and subsequently amplified by PCR (98°C for 2 min; 4 cycles of 98°C for 20 sec, 62°C for 3 min; 62°C for 5 min) using KAPA HiFi HotStart ready mix (KAPA Biosystems, #KM2602) and NEXTflex primer mix (Bio Scientific, #514122). Next, libraries were size-selected for 300-400 bp fragments using 2% E-Gel Size Select agarose gels (Invitrogen, #G6610-02) and amplified for another 12 cycles using the same conditions as above. Subsequent depletion of adapter dimers and library purification was performed using Agencourt AMPure XP beads (Beckman Coulter, #A63880) in a 1:1 library beads ratio.

Strand-specific RNA-seq libraries were sequenced on the Illumina NextSeq 500 system to generate 75 bp single-end reads (TruSeq SR Cluster Kit v2), which were mapped against the *P. falciparum* 3D7 transcriptome from PlasmoDB v26 using BWA samse (version 0.7.12-r1039). Reads were filtered for mapping quality ≥15 (samtools version 1.2) and only uniquely mapped reads (between 1.6 and 4.7

million reads) were used for further analysis. After separating the reads according to the strand they mapped to (sense strand FLAG16, antisense strand FLAG0), sense and antisense transcript abundance was assessed by counting the respective tags for each transcript except for transcripts originating from the mitochondrial and apicoplast genomes. Resulting tag counts were offset by +1 and normalized to the number of reads per kb per million mapped reads (RPKM). RPKM ratios were calculated when at least one of the two samples per time point had a RPKM value ≥ 3 . To visualize RNA-seq data in the UCSC Genome browser, 75 bp reads were additionally mapped against the annotated *P. falciparum* 3D7 genome (PlasmoDB v26), filtered as described above and split into reads mapping to the sense strand (FLAG16) and antisense strand (FLAG0), respectively. RNA-seq libraries were normalized to the number of mapped reads per million (RPM) and bedgraph files were generated (bedtools v2.20.1). After uploading the tracks to the UCSC genome browser, tracks were smoothed (5) and the windowing function was set as 'mean'.

Reverse transcription quantitative PCR (RT-qPCR)

F12 and F12/*gdv1*-asKO parasites were synchronized twice 16 hours apart to obtain an eight hour growth window. Parasites were harvested at 40-48 hpi. Total RNA was isolated using Ribozol (Amresco) according to the manufacturer's manual and further purified using the RNeasy Plus Mini Kit (Qiagen) for removal of gDNA. Residual gDNA was digested with TURBO DNA-free™ DNaseI (Ambion) and the RNA was quantified by NanoDrop. All samples were tested negative for contaminating gDNA by qPCR. 2 mg total RNA was reverse-transcribed using the RETROscript Kit (Ambion). qPCR reactions were performed at primer concentrations of 0.5 mM using Power SYBR® Green PCR Master Mix (Applied Biosystems) on a StepOnePlus™ Real-Time PCR System (Applied Biosystems) in a reaction volume of 20 ml. All reactions were run in triplicate yielding virtually identical Ct values. Cycling conditions were: 95°C for 10 min, followed by 40 cycles of 95°C/15 sec and 60°C/1 min. Product-specific amplification was ensured by melting curve analysis for each reaction. Relative transcript levels were calculated by normalization against the house-keeping gene encoding eukaryotic translation initiation factor 2-alpha kinase (*pk4*, PF3D7_0628200). All primer sequences are listed in Table S9.

Supplementary materials

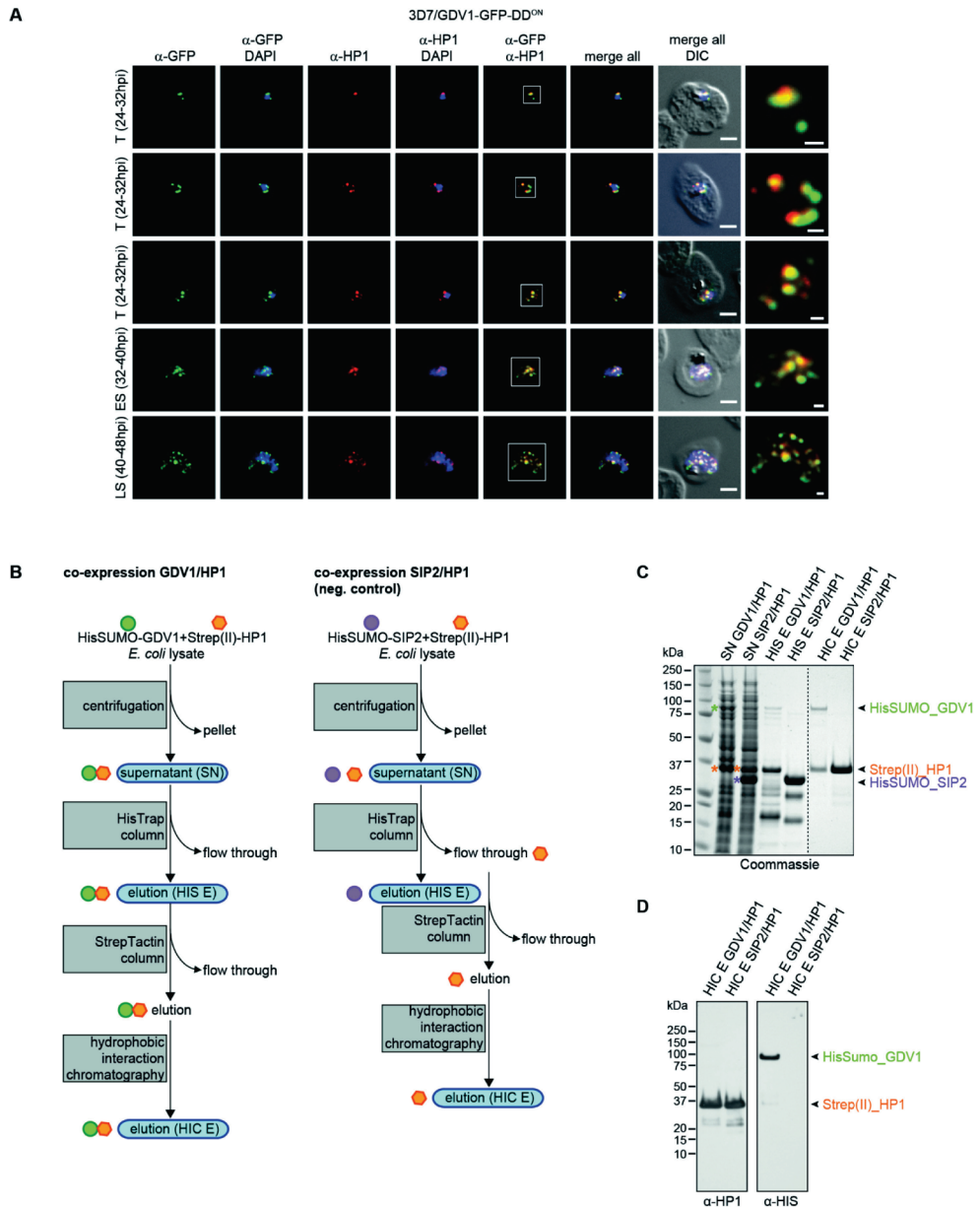
SUPPLEMENTARY TEXT

Deletion of the *gdv1* asRNA locus in F12 parasites resulted in a marked increase in *ap2-g* transcript levels (Fig. 4B, Fig. S6 and Table S8). Hence, to be able to knock-out the *gdv1* asRNA in 3D7 parasites we first had to modify the *ap2-g* locus such that AP2-G expression is prevented under default culture conditions. To achieve this in a most robust way we placed expression of AP2-G under control of two conditional expression systems. Using a single CRISPR/Cas9-based gene editing step we tagged AP2-G with a C-terminal GFP-DD tag and at the same time inserted a *gImS* riboswitch element into the 3' untranslated region (Fig. S8). The *gImS* riboswitch allows regulating gene expression at the level of the mRNA; in presence of the amino sugar glucosamine (GlcN) the *gImS* ribozyme structure mediates self-cleavage of the mRNA leading to its degradation (56-58). As expected, parasites did not express AP2-G-GFP-DD and failed to produce gametocytes when cultured in absence of Shield-1 and presence of GlcN. In presence of Shield-1 and absence of GlcN AP2-G-GFP-DD expression was detectable by IFA in

the nuclei of some parasites and gametocytes were produced at a rate of 11.0% (+/- 0.3 SD) (Fig. S8). Having established this conditional AP2-G loss-of-function mutant we next deleted the *gdv1* asRNA locus in these parasites (3D7/AP2-G-GFP-DDgImS/*gdv1*-asKO). 3D7/AP2-G-GFP-DDgImS/*gdv1*-asKO parasites still failed to produce gametocytes when AP2-G-GFP-DD expression was prevented in absence of Shield-1 and presence of GlcN. However, compared to the 3D7/AP2-G-GFP-DDgImS control line 3D7/AP2-G-GFP-DDgImS/*gdv1*-asKO parasites displayed a substantially increased sexual conversion rate of 39.3% (+/- 2.8 SD) when AP2-G-GFP-DD expression was stabilised, showing that the *gdv1*-asRNA antagonises the GDV1-dependent induction of gametocyte conversion.

SUPPLEMENTARY FIGURES

Fig. S1. GDV1 co-localises and interacts with HP1. (A) a-GFP (green) and a-HP1 (red) co-localisation IFAs in 3D7/GDV1-GFP-DD^{ON} trophozoites (T, 24-32 hpi), early schizonts (ES, 32-40 hpi) and late schizonts (LS, 40-48 hpi). Scale bar, 2.5 mm. White frames refer to the magnified views in the rightmost images (scale bar, 0.5 mm). DIC, differential interference contrast. Results are representative of three biological replicate experiments. **(B)** Flow chart of the purification approach used to demonstrate direct protein-protein interaction between GDV1 and HP1 co-expressed in *E.coli* as HisSumo-tagged GDV1 and Strep(II)-tagged HP1. Co-expression of a HisSUMO-tagged N-terminal fragment of the heterochromatin-associated ApiAP2 factor PfsIP2 (45) and Strep(II)-tagged HP1 was used as negative control. Samples highlighted in blue are shown in the Coomassie-stained polyacrylamide gel shown in panel C. **(C)** Coomassie-stained SDS-polyacrylamide gel. Lane 1: protein size standard. Lanes 2, 3: supernatants (SN) of bacterial lysates show co-expression of the respective fusion proteins GDV1/HP1 (lane 2) and SIP2/HP1 (lane 3) (asterisks). Lanes 4, 5: elution samples obtained after Ni²⁺ affinity purification (HIS E) of the GDV1/HP1 SN (lane 4) or the SIP2/HP1 SN



(lane 5). Strep(II)-tagged HP1 co-purified specifically with HisSUMO-GDV1 (lane 4) but not with HisSUMO-SIP2 (lane 5). Lanes 6, 7: the Ni²⁺ column elution of the GDV1/HP1 sample (containing both HisSUMO-GDV1 and Strep(II)-HP1) and Ni²⁺ column flow through of the SIP2/HP1 negative control sample (containing Strep(II)-HP1) were further purified using StrepTactin affinity purification followed by hydrophobic interaction chromatography (HIC). Analysis of the HIC eluates (HIC E) show that HisSUMO-GDV1 co-purified with Strep(II)-HP1 (lane 6). **(D)** α -HP1 and α -HIS Western blot confirm the identities of HisSUMO-GDV1 and Strep(II)-HP1 in the HIC elution samples.

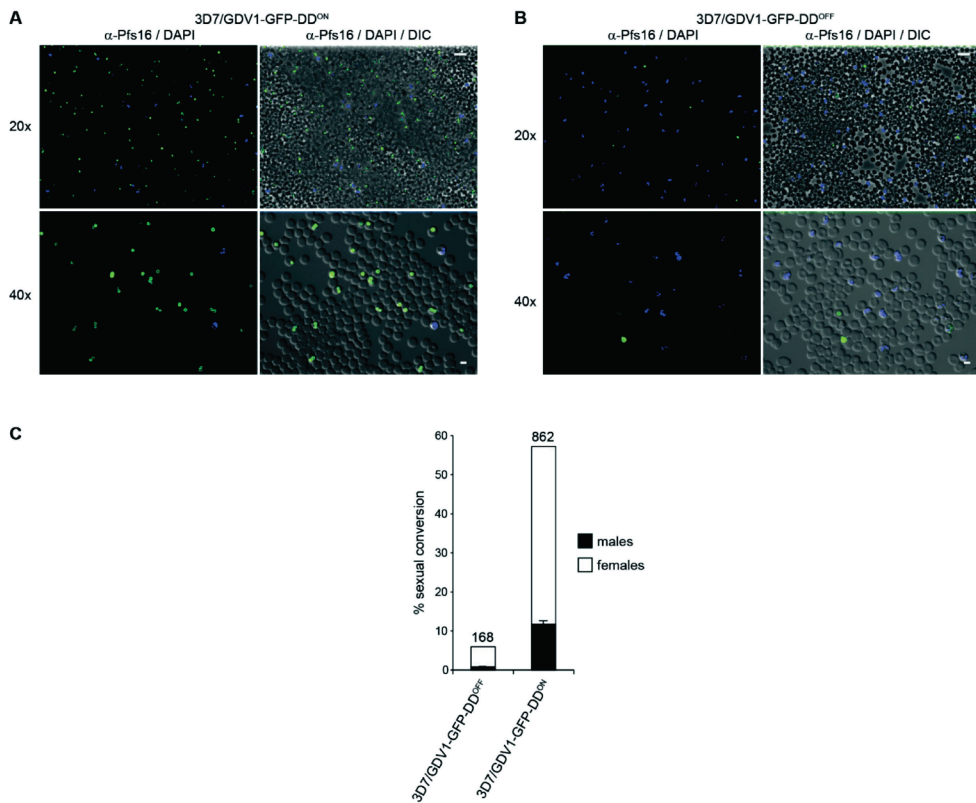
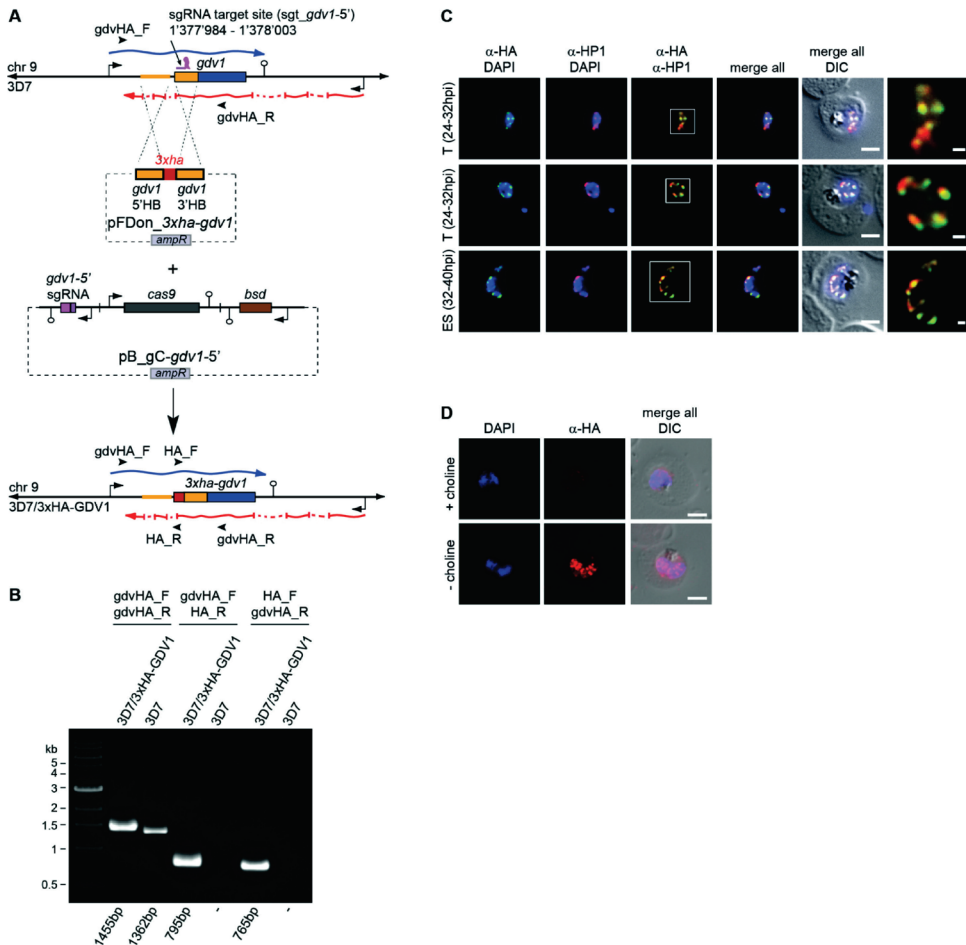


Fig. S2. GDV1 over-expression induces sexual commitment.

(A, B) Overview images of IFAs performed on formaldehyde/glutaraldehyde-fixed cultures 38-46 hours after re-invasion (day 2) using α-Pfs16 antibodies, highlighting the large proportion of stage I gametocytes (green) in 3D7/GDV1-GFP-DD^{ON} (A) compared to 3D7/GDV1-GFP-DD^{OFF} (B) parasites. Nuclei were stained with DAPI. Scale bars, 15 mm (20x magnification) and 5 mm (40x magnification). (C) Proportion of male and female gametocytes in 3D7/GDV1-GFP-DD^{OFF} and 3D7/GDV1-GFP-DD^{ON} populations. Results represent the mean of three technical replicates (error bars represent SD). Total number of scored gametocytes is shown above the bars.

Fig. S3. Generation of the 3D7/3xHA-GDV1 line using CRISPR/Cas9 and expression of endogenous 3xHA-GDV1.

(A) Schematic maps of the *gdv1* locus (PF3D7_0935400) in 3D7 parasites (top), the pFDon_3xha-gdv1 donor and pB_gC-gdv1-5' CRISPR/Cas9 transfection vectors (center), and the modified *gdv1* locus after CRISPR/Cas9-based genome editing in 3D7/3xHA-GDV1 parasites (bottom). The *gdv1* sense



transcript and multi-exon antisense transcript (24) are indicated by undulated arrows in blue and red, respectively. The nucleotide positions of the *sgt_gdv1-5'* sgRNA target sequence is indicated (chromosome 9 coordinates). The pFDon_3xha-gdv1 donor plasmid contains a 3xha sequence (red) flanked on either side by a homology box (HB) (orange) for homology-directed repair. The pB_gC-gdv1-5' plasmid contains expression cassettes for SpCas9 (grey), the sgRNA (purple) and the *bsd* resistance marker (brown). Successful gene editing results in the expression of a N-terminally tagged 3xHA-GDV1. PCR primer binding sites used to confirm successful gene editing are indicated by arrowheads. **(B)** PCR on gDNA from 3D7/3xHA-GDV1 and 3D7 wild-type parasites. Primers *gdvHA_F* and *gdvHA_R* bind to chromosomal sequences outside the HBs and amplify a 1455 bp or 1362 bp fragment from the edited or wild-type *gdv1* locus, respectively. The *gdvHA_F/HA_R* and *HA_F/gdvHA_R* primer combinations are specific for the edited locus and amplify 795 bp and 765 bp fragments, respectively. **(C)** a-HA (green) and a-HP1 (red) co-localisation IFAs in 3D7/3xHA-GDV1 trophozoites (T, 24-32 hpi) and early schizonts (ES, 32-40 hpi). Scale bar, 2.5 mm. White frames refer to the magnified views in the rightmost images (scale bar, 0.5 mm). DIC, differential interference contrast. Results are representative of two biological replicate experiments. **(D)** Representative a-HA (red) IFAs of 3D7/3xHA-GDV1 parasites cultured in presence or absence of 2 mM choline. Nuclei were stained with DAPI. Scale bar, 2.5 mm. DIC, differential interference contrast.

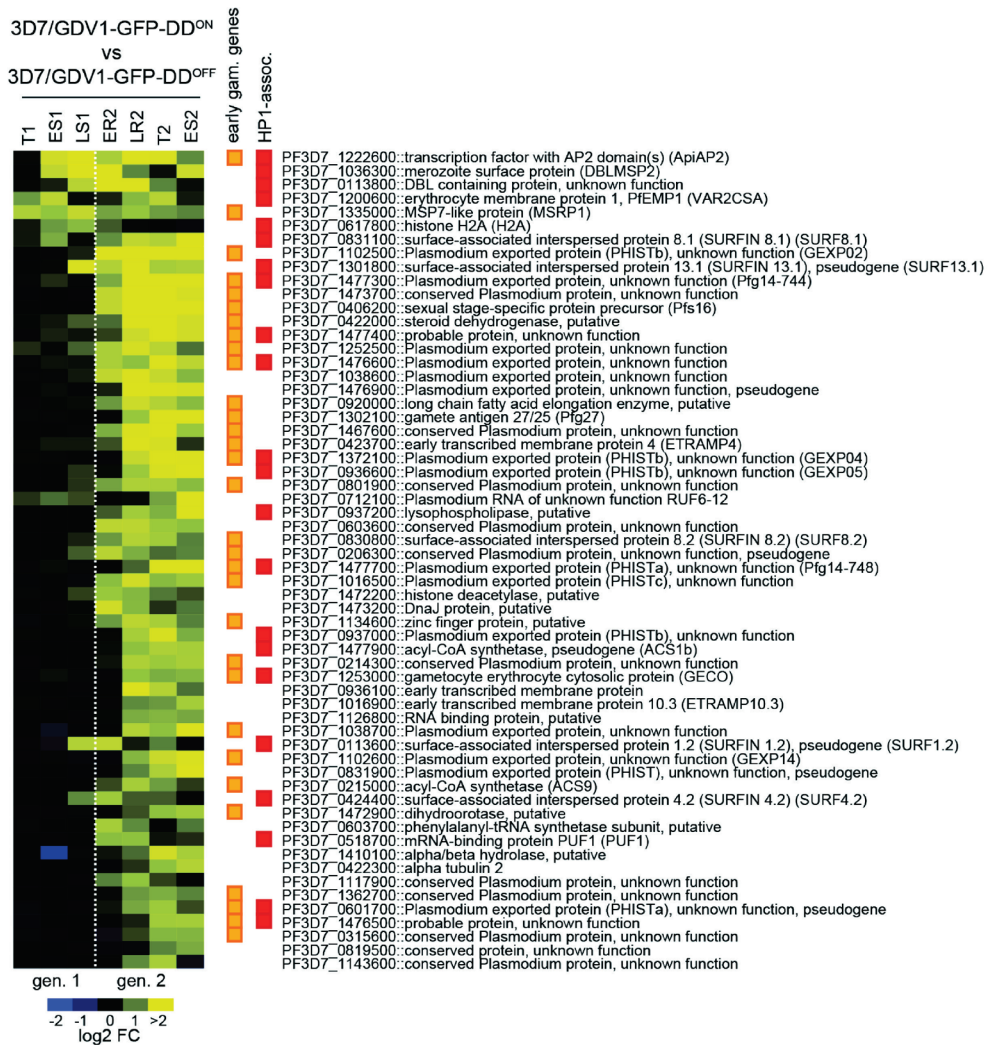


Fig. S4. Induction of GDV1-GFP-DD expression triggers a gametocyte-specific transcriptional response.

The heatmap shows the temporal gene induction profiles of all genes with (1) higher transcript levels in 3D7/GDV1-GFP-DD^{ON} compared to the 3D7 reference pool (positive log₂ Cy5/Cy3 ratio), and (2) greater than 2-fold higher transcript levels in 3D7/GDV1-GFP-DD^{ON} vs 3D7/GDV1-GFP-DD^{OFF} (log₂ fold change >1) in any two of the seven consecutive TPs. HP1-associated genes identified previously (2) and in this study (Table S7) are highlighted by red boxes. Early gametocyte genes identified in previous studies (7, 8, 17) are highlighted by orange boxes. FC, fold change; gen., generation. T1/ES1/LS1, trophozoites, early schizonts, late schizonts in generation 1; ER2/LR2/T2/ES2, early ring stages, late ring stages, trophozoites, early schizonts in generation 2.

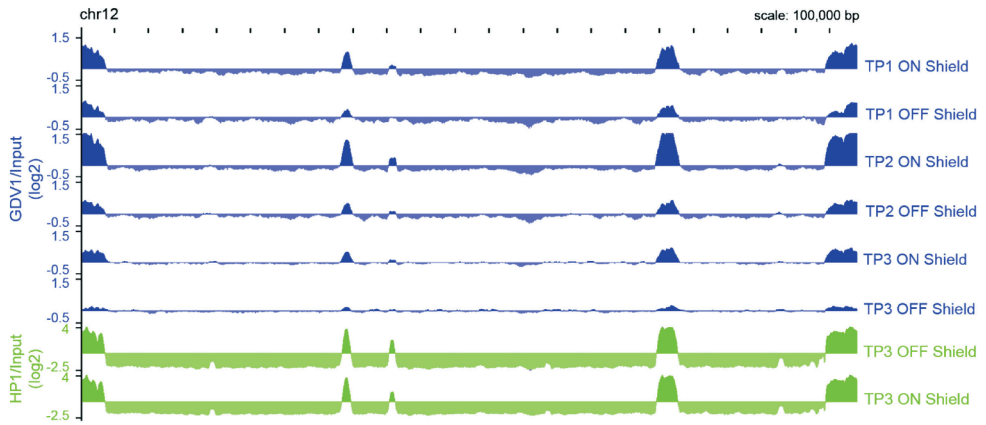


Fig. S5. Increased GDV1 occupancy at heterochromatic domains at all three TPs in 3D7/GDV1-GFP-DD^{ON} compared to 3D7/GDV1-GFP-DD^{OFF} parasites.

ChIP-seq ratio tracks display relative enrichment (ChIP/input) of GDV1-GFP-DD (blue) or HP1 (green) along chromosome 12 in 3D7/GDV1-GFP-DD^{ON} parasites cultured in the presence of Shield-1 for two (TP1, 30-36 hpi), six (TP2, 34-40 hpi) or ten hours (TP3, 38-44 hpi) and in the matching control populations grown in the absence of Shield-1 (3D7/GDV1-GFP-DD^{OFF}).

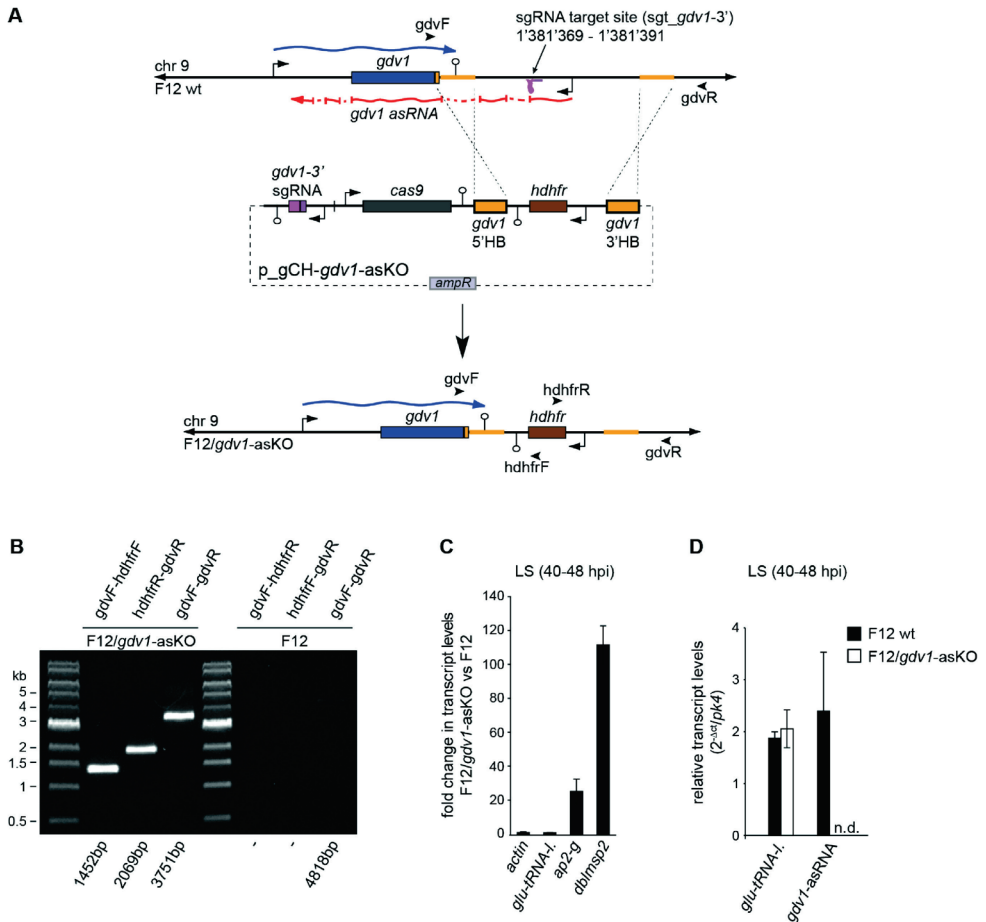


Fig. S6. Generation of the F12/*gdv1*-asKO parasite line using CRISPR/Cas9.

(A) Schematic maps of the *gdv1* locus (PF3D7_0935400) in F12 parasites (top), the p_gCH-*gdv1*-asKO CRISPR/Cas9 transfection vector (center), and the modified *gdv1* locus after CRISPR/Cas9-based genome editing in F12/*gdv1*-asKO parasites (bottom). The *gdv1* sense transcript and multi-exon antisense transcript (24) are indicated by undulated arrows in blue and red, respectively. The nucleotide positions of the *sgt_gdv1-3'* sgRNA target sequence is indicated (chromosome 9 coordinates). The p_gCH-*gdv1*-asKO plasmid contains expression cassettes for SpCas9 (grey), the sgRNA (purple) and the *hdhfr* resistance marker (brown) flanked by two homology boxes (HB) (orange) for homology-directed repair. PCR primer binding sites used to confirm successful gene editing are indicated by arrowheads. **(B)** PCR on gDNA from F12/*gdv1*-asKO and F12 wild-type parasites. Primers *gdvF* and *gdvR* bind to chromosomal sequences outside the HBs and amplify a 3751 bp or 4818 bp fragment from the edited or wild-type *gdv1* locus, respectively. The *gdvF*/*hdhfrF* and *hdhfrR*/*gdvR* primer combinations are specific for the edited locus and amplify 1452 bp and 2069 bp fragments, respectively. **(C)** RT-qPCR data confirming the increase in *ap2-g* and *dblmsp2* transcript levels in F12/*gdv1*-asKO compared to F12 wild-type parasites (results are the mean of three biological replicates; error bars represent SD). *actin* (PF3D7_1264200)/*glul-tRNA-I* (PF3D7_1331700), control genes. **(D)** RT-qPCR showing the absence of *gdv1* asRNA in F12/*gdv1*-asKO parasites. Values represent

expression levels relative to the house-keeping gene encoding eukaryotic translation initiation factor 2-alpha kinase (*pk4*, PF3D7_0628200). *glu-tRNA-l*. (PF3D7_1331700), control gene. Results are the mean of three biological replicates (error bars represent SD). n.d., not detectable.

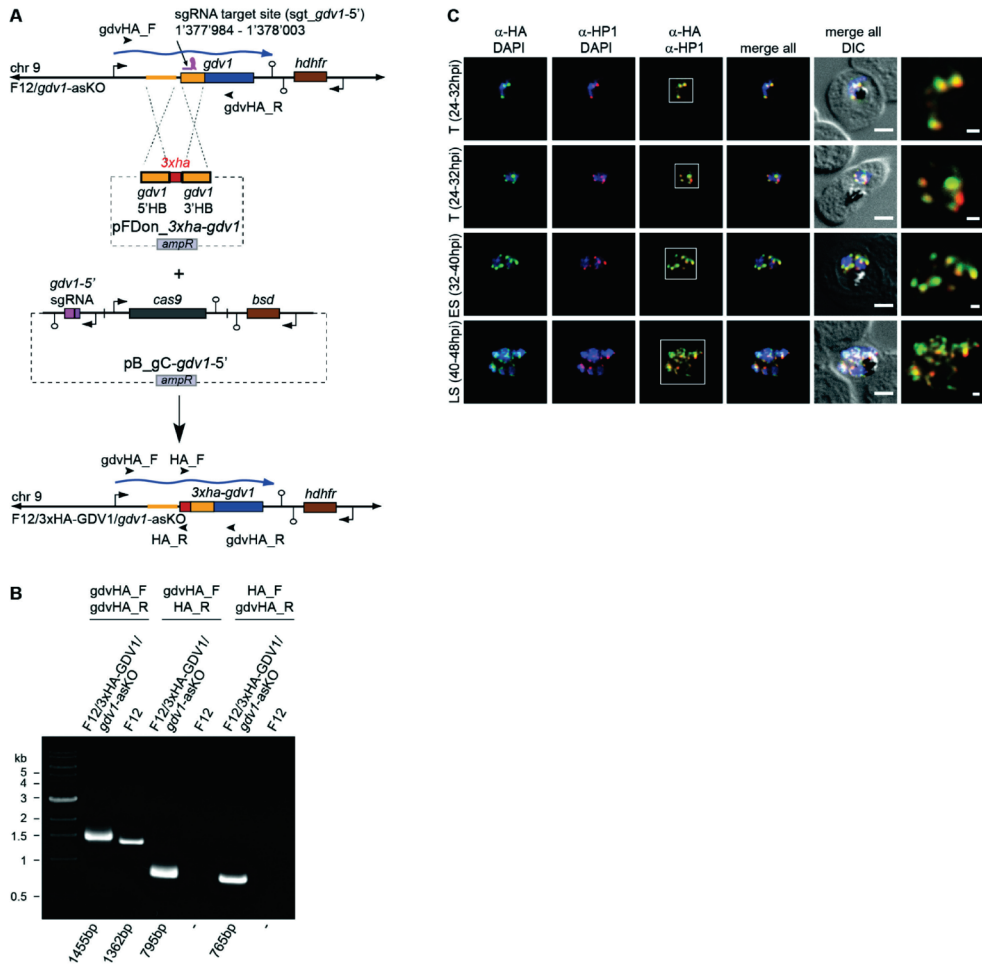


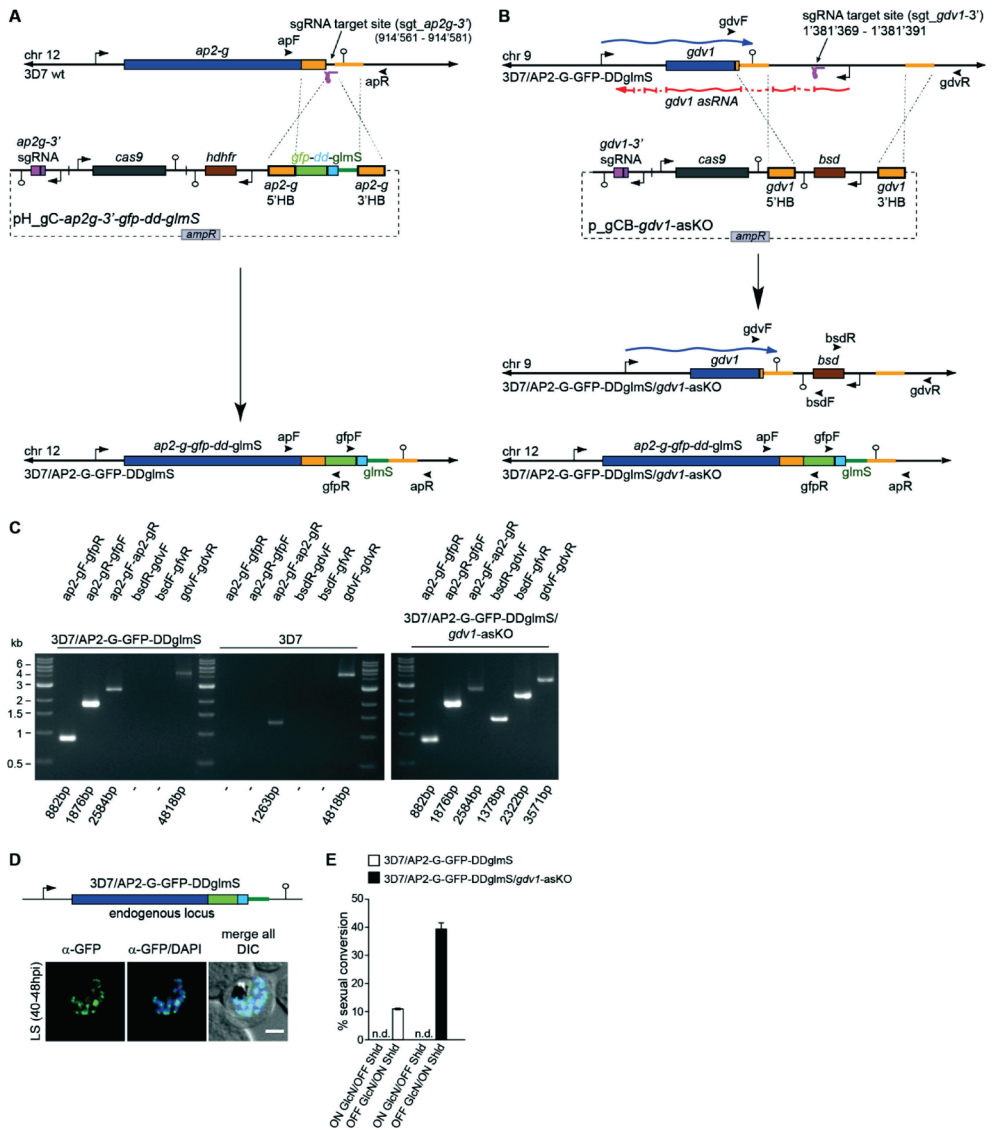
Fig. S7. Generation of the F12/3xHA-GDV1/*gdv1*-asKO line using CRISPR/Cas9 and expression of endogenous 3xHA-GDV1.

(A) Schematic maps of the *gdv1* locus (PF3D7_0935400) in F12/*gdv1*-asKO parasites (top), the pFDon_3*xha-gdv1* donor and pB_gC-*gdv1-5'* CRISPR/Cas9 transfection vectors (center), and the modified *gdv1* locus after CRISPR/Cas9-based genome editing in F12/3xHA-GDV1/*gdv1*-asKO parasites (bottom). The *gdv1* sense transcript is indicated by an undulated arrow in blue. The nucleotide positions of the *sgt_gdv1-5'* sgRNA target sequence is indicated (chromosome 9 coordinates). The pFDon_3*xha-gdv1* donor plasmid contains a 3*xha* sequence (red) flanked on either side by a homology box (HB) (orange) for homology-directed repair. The pB_gC-*gdv1-5'* plasmid contains expression cassettes for SpCas9 (grey), the sgRNA (purple) and the *bsd* resistance marker

(brown). Successful gene editing results in the expression of a N-terminally tagged 3xHA-GDV1. PCR primer binding sites used to confirm successful gene editing are indicated by arrowheads. **(B)** PCR on gDNA from F12/3xHA-GDV1/*gdv1*-asKO and F12 wild-type parasites. Primers *gdvHA_F* and *gdvHA_R* bind to chromosomal sequences outside the HBs and amplify a 1455 bp or 1362 bp fragment from the edited or wild-type *gdv1* locus, respectively. The *gdvHA_F/HA_R* and *HA_F/gdvHA_R* primer combinations are specific for the edited locus and amplify 795 bp and 765 bp fragments, respectively. **(C)** a-HA (green) and a-HP1 (red) co-localisation IFAs in F12/3xHA-GDV1/*gdv1*-asKO trophozoites (T, 24-32 hpi), early schizonts (ES, 32-40 hpi) and late schizonts (LS, 40-48 hpi). Scale bar, 2.5 mm. White frames refer to the magnified views in the rightmost images (scale bar, 0.5 mm). DIC, differential interference contrast. Results are representative of two biological replicate experiments.

Fig. S8. Generation of the 3D7/AP2-G-GFP-DDgImS and 3D7/AP2-G-GFP-DDgImS/*gdv1*-asKO parasite lines using CRISPR/Cas9.

(A) Schematic maps of the *ap2-g* locus (PF3D7_1222600) in 3D7 parasites (top), the pD_ *ap2-g-gfp-dd-gImS* donor and pH_ *gC-ap2g-3'* CRISPR/Cas9 transfection vectors (center), and the modified *ap2-g* locus after CRISPR/Cas9-based genome editing in 3D7/AP2-G-GFP-DDgImS parasites (bottom). The nucleotide positions of the *sgt_ap2g-3'* sgRNA target sequence is indicated (chromosome 12 coordinates). The pD_ *ap2-g-gfp-dd-gImS* donor plasmid contains a *gfp-dd-gImS* fusion sequence flanked on either side by a homology box (HB) (orange) for homology-directed repair. The pH_ *gC-ap2g-3'* plasmid contains expression cassettes for SpCas9 (grey), the sgRNA (purple) and the *hdhfr* resistance marker (brown). Successful gene editing results in the expression of a C-terminally tagged AP2-G-GFP-DD fusion protein that is under control of the *gImS* riboswitch system (56). PCR primer binding sites used to confirm successful gene editing are indicated by arrowheads. **(B)** Schematic maps of the *gdv1* locus (PF3D7_0935400) in 3D7/AP2-G-GFP-DDgImS parasites (top), the p_ *gCB-gdv1*-asKO CRISPR/Cas9 transfection vector (center), and the modified *gdv1* and *ap2-g* loci after CRISPR/Cas9-based genome editing in 3D7/AP2-G-GFP-DDgImS/*gdv1*-asKO parasites (bottom). The *gdv1* sense transcript and multi-exon antisense transcript (24) are indicated by undulated arrows in blue and red, respectively. The nucleotide positions of the *sgt_gdv1-3'* sgRNA target sequence is indicated (chromosome 9 coordinates). The p_ *gCB-gdv1*-asKO plasmid contains expression cassettes for SpCas9 (grey), the sgRNA (purple) and the *bsd* resistance marker (brown) flanked by two homology boxes (HB) (orange) for homology-directed repair. PCR primer binding sites used to confirm successful gene editing are indicated by arrowheads. **(C)** PCR on gDNA from 3D7/AP2-G-GFP-DDgImS/*gdv1*-asKO, 3D7/AP2-G-GFP-DDgImS and 3D7 wild-type parasites. Primers *apF* and *apR* bind to chromosomal sequences outside the *ap2-g* HBs and amplify a 2584 bp or 1263 bp fragment from the edited or wild-type *ap2-g* locus, respectively. Primers *gdvF* and *gdvR* bind to chromosomal sequences outside the *gdv1* HBs and amplify a 3751 bp or 4818 bp fragment from the edited or wild-type *gdv1* locus,



respectively. The apF/gfpR and apR/gfpF primer combinations are specific for the edited *ap2-g* locus and amplify 882 bp and 1876 bp fragments, respectively. The bsdR/gdvF and bsdF/gdvR primer combinations are specific for the edited *gdv1* locus and amplify 1378 bp and 2322 bp fragments, respectively. **(D)** Endogenous *ap2-g* locus in the 3D7/AP2-G-GFP-DDglmS line and a-GFP IFA showing AP2-G-GFP-DD expression in a late schizont (LS, 40-48 hpi) from a 3D7/AP2-G-GFP-DDglmS culture grown in presence of Shield-1 and absence of glucosamine (GlcN). DIC, differential interference contrast. Scale bar, 2.5 μ m. Results are representative of three biological replicates. **(E)** Sexual conversion rates in 3D7/AP2-G-GFP-DDglmS and 3D7/AP2-G-GFP-DDglmS/*gdv1*-asKO parasites cultured either in presence of GlcN/absence of Shield-1 (AP2-G-GFP-DD expression prevented) or absence of GlcN/presence of Shield-1 (AP2-G-GFP-DD expression permitted) (results are the mean of three biological replicates (>200 infected RBCs counted per sample); error bars indicate SD).

Supplementary tables

Table S1. Potential HP1-interacting proteins identified by LC-MS/MS in the eluates of three independent Co-IP experiments. Gene IDs and annotations have been retrieved from PlasmoDB (www.plasmodb.org). This table lists all proteins detected in all three HP1-GFP Co-IP eluates (Co-IP Exp#1-3) and undetected in all negative control eluates. Note that for each experiment the sequence coverage and number of unique peptides indicated derive from either the 2 M arginine or subsequent 2% SDS elution samples, whichever displayed the higher number of unique tryptic peptides detected (see also Table S3). Cov, % sequence coverage; Pep, number of unique tryptic peptides detected.

| GeneID | Description | Co-IP Exp#1 | | Co-IP Exp#2 | | Co-IP Exp#3 | |
|---------------|--|-------------|-----|-------------|-----|-------------|-----|
| | | Cov | Pep | Cov | Pep | Cov | Pep |
| PF3D7_1220900 | heterochromatin protein 1 (HP1) | 77.82 | 30 | 65.79 | 23 | 71.43 | 24 |
| PF3D7_0935400 | gametocyte development protein 1 (GDV1) | 43.91 | 19 | 9.68 | 6 | 22.20 | 10 |
| PF3D7_0818200 | 14-3-3 protein (14-3-3I) | 43.89 | 7 | 17.94 | 3 | 10.31 | 3 |
| PF3D7_1451200 | conserved Plasmodium protein, unknown function | 17.42 | 24 | 14.10 | 14 | 15.43 | 21 |
| PF3D7_1441400 | FACT complex subunit SSRP1, putative (FACT-S) | 11.46 | 5 | 9.49 | 4 | 7.91 | 3 |
| PF3D7_0517400 | FACT complex subunit SPT16, putative (FACT-L) | 1.14 | 1 | 8.41 | 8 | 3.24 | 3 |
| PF3D7_1023900 | chromodomain-helicase-DNA-binding protein 1 homolog, putative (CHD1) | 0.45 | 1 | 0.45 | 1 | 5.26 | 12 |
| PF3D7_1357400 | conserved Plasmodium protein, unknown function | 1.70 | 2 | 3.56 | 3 | 3.48 | 3 |
| PF3D7_1456000 | transcription factor with AP2 domain(s) (ApiAP2) | 2.91 | 4 | 0.80 | 1 | 3.20 | 3 |

Table S2. Potential GDV1-interacting proteins identified by LC-MS/MS in the eluates of three independent Co-IP experiments. Gene IDs and annotations have been retrieved from PlasmoDB (www.plasmodb.org). This table lists all proteins detected in all three GDV1-GFP-DD Co-IP eluates (Co-IP Exp#1-3) and undetected in all negative control eluates (see also Table S4). Cov, % sequence coverage; Pep, number of unique tryptic peptides detected.

| GeneID | Description | Co-IP Exp#1 | | Co-IP Exp#2 | | Co-IP Exp#3 | |
|---------------|--|-------------|-----|-------------|-----|-------------|-----|
| | | Cov | Pep | Cov | Pep | Cov | Pep |
| PF3D7_0505800 | small ubiquitin-related modifier, putative (SUMO) | 54 | 5 | 46 | 4 | 37 | 2 |
| PF3D7_1220900 | heterochromatin protein 1 (HP1) | 50.38 | 14 | 28.95 | 9 | 23.31 | 6 |
| PF3D7_0935400 | gametocyte development protein 1 (GDV1) | 32.89 | 19 | 23.54 | 10 | 18.36 | 9 |
| PF3D7_1211800 | polyubiquitin (PfpUB) | 44.62 | 1 | 21 | 1 | 21 | 1 |
| PF3D7_1451200 | conserved Plasmodium protein, unknown function | 15.82 | 24 | 20.81 | 29 | 5.72 | 7 |
| PF3D7_1023900 | chromodomain-helicase-DNA-binding protein 1 homolog, putative (CHD1) | 10.16 | 30 | 8.74 | 23 | 5.44 | 13 |

| | | | | | | | |
|---------------|------------------------------------|------|---|------|---|------|---|
| PF3D7_1027800 | 60S ribosomal protein L3, putative | 3.11 | 1 | 3.11 | 1 | 3.11 | 1 |
| PF3D7_1447000 | 40S ribosomal protein S2, putative | 2.94 | 1 | 3.31 | 1 | 3.31 | 1 |

Table S3-S8, please download online.

Table S3. Mass spectrometry results of the HP1-GFP co-IP experiments. This file lists all proteins and peptides identified by Mascot and Sequest HT searches of the results obtained from LC-MS/MS analysis of protein elutions 1 (E1_Arg) and 2 (E2_SDS) obtained after HP1-GFP co-IP (HP1 Co-IP) or negative control IPs (neg. control). The experiments have been performed in three biological replicates (#1-3). The file contains 12 worksheets. (xlsx format).

Table S4. Mass spectrometry results of the GDV1-GFP-DD co-IP experiments. This file lists all proteins and peptides identified by Mascot and Sequest HT searches of the results obtained from LC-MS/MS analysis of protein elutions (E_SDS) obtained after GDV1-GFP co-IP (GDV1 Co-IP) or negative control IPs (neg. control). The experiments have been performed in three biological replicates (#1-3). The file contains 6 worksheets. (xlsx format).

Table S5. Processed microarray data obtained from 3D7/GDV1-GFP-DD parasites. Columns A-B: Gene ID and gene annotation. Column C: HP1 target genes previously identified by ChIP-on-chip (2) and in this study (Table S7). Column D: known early gametocyte markers (7, 8, 17). Columns E-K: Cy5/Cy3 log2 ratios for all transcripts and seven TPs harvested from 3D7/GDV1-GFP-DD^{OFF} parasites. Columns L-R: Cy5/Cy3 log2 ratios for all transcripts and seven TPs harvested from 3D7/GDV1-GFP-DD^{ON} parasites. Column S-Y: fold change in gene expression (log2) in 3D7/GDV1-GFP-DD^{ON} compared to 3D7/GDV1-GFP-DD^{OFF} parasites for each of the seven paired TPs. Column Z: mean fold change in gene expression in 3D7/GDV1-GFP-DD^{ON} compared to 3D7/GDV1-GFP-DD^{OFF} parasites across all seven TPs. Column AA: SAM significance q-values (% fdr). Column AB: significantly up-regulated (u) or down-regulated (d) genes. T1/ES1/LS1, trophozoites, early schizonts, late schizonts in generation 1; ER2/LR2/T2/ES2, early ring stages, late ring stages, trophozoites, early schizonts in generation 2. (xlsx format).

Table S6. Processed microarray data obtained with F12/GDV1-GFP-DD parasites. Columns A-B: Gene ID and gene annotation. Column C: HP1 target genes previously identified by ChIP-on-chip (2) and in this study (Table S7). Column D: known early gametocyte markers (7, 8, 17). Columns E-K: Cy5/Cy3 log2 ratios for all transcripts and seven TPs harvested from F12/GDV1-GFP-DD^{OFF} parasites. Columns L-R: Cy5/Cy3 log2 ratios for all transcripts and seven TPs harvested from F12/GDV1-GFP-DD^{ON} parasites. Column S-Y: fold change in gene expression (log2) in F12/GDV1-GFP-DD^{ON} compared to F12/GDV1-GFP-DD^{OFF} parasites for each of the seven paired TPs. Column Z: mean fold change in gene expression in F12/GDV1-GFP-DD^{ON} compared to F12/GDV1-GFP-DD^{OFF} parasites across all seven TPs. Column AA: SAM significance q-values (% fdr). Column AB: significantly up-regulated (u) or down-regulated (d) genes. T1/ES1/LS1, trophozoites, early schizonts, late schizonts in generation 1; ER2/LR2/T2/ES2, early ring stages, late ring stages, trophozoites, early schizonts in generation 2. (xlsx format).

Table S7. GDV1-GFP-DD and HP1 ChIP-seq enrichment values. ChIP/input enrichment values were calculated over each coding region of the Pf_3D7_v3 reference genome in 3D7/GDV1-GFP-DD^{ON} and 3D7/GDV1-GFP-DD^{OFF} parasites at TPs 1-3 (30-36 hpi, 34-40 hpi, 38-44 hpi). Columns A-B: Gene ID and gene annotation. Column C: chromosome number. Columns D-E: Nucleotide positions at the start and end of each gene. Column F: HP1-associated genes (HP1 ChIP/input >1 in 3D7/GDV1-GFP-DD^{OFF}

control parasites at TP3). Column G: known early gametocyte markers (7, 8, 17). Columns H-M: GDV1-GFP-DD ChIP/input values for each population and TP. Columns N-P: fold change in GDV1-GFP-DD occupancy between parasites cultured in the presence or absence of Shield-1 for each TP. Columns Q-R: HP1 ChIP/input values for both populations at TP3. Column S: log₂ fold change in HP1 occupancy between parasites cultured in the presence or absence of Shield-1 for TP3. (xlsx format).

Table S8. Processed RNA-seq data of sense transcripts obtained from F12 wild-type and F12/*gdv1*-askO parasites. Columns A-B: Gene ID and gene annotations. Column C: HP1 target genes previously identified by ChIP-on-chip (2) and in this study (Table S7). Columns D-E: RPKM values for early schizonts (32-40 hpi) of F12 wild-type and F12/*gdv1*-askO parasites. Column F: fold change (FC) in transcript abundance in F12/*gdv1*-askO compared to F12 wild-type early schizonts (32-40 hpi). Columns G-H: RPKM values for late schizonts (40-48 hpi) of F12 wild-type and F12/*gdv1*-askO parasites. Column I: fold change (FC) in transcript abundance in F12/*gdv1*-askO compared to F12 wild-type late schizonts (40-48 hpi). Column J: Genes consistently up-regulated (up) or down-regulated (down) in both time points in F12/*gdv1*-askO compared to F12 wild-type parasites. Genes with RPKM values < 3 in both samples per time point are highlighted in grey and were excluded from downstream analyses. (xlsx format)

Table S9. Oligonucleotides used in this study.

CHAPTER 6

Chromatin accessibility-based characterization of the gene regulatory network underlying *Plasmodium falciparum* blood stage development

Christa Geeke Toenhake^{1,5}, Sabine Anne-Kristin Fraschka^{1,5}, Mahalingam Shanmugiah Vijayabaskar^{3,4}, David Robert Westhead³, Simon Jan van Heeringen² and Richárd Bártfai^{1,6,#}

Adapted from
Cell Host & Microbe 23, 557-569.e559

¹Radboud University, Faculty of Science, Department of Molecular Biology, Nijmegen, The Netherlands

²Radboud University, Faculty of Science, Department of Molecular Developmental Biology, Nijmegen, The Netherlands

³School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom

⁴Current address: Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom

⁵These authors contributed equally to this work.

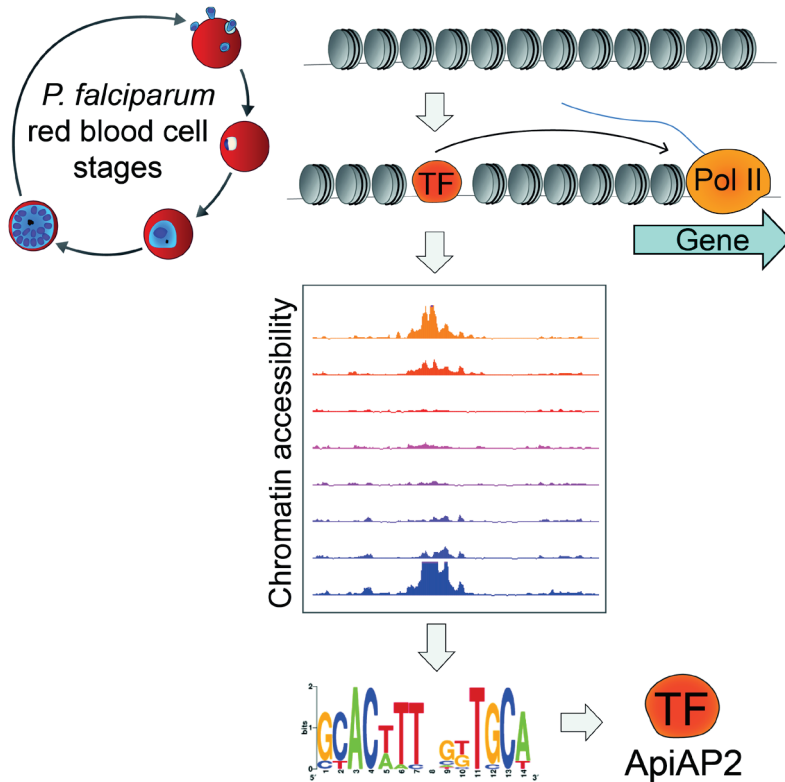
⁶Lead Contact

Correspondence: r.bartfai@science.ru.nl

Contribution:

Designed and performed RNA-seq and reporter gene experiments, analysed these data, prepared the respective illustrations and wrote the respective sections of the manuscript.

Graphical abstract



Summary

Underlying the development of malaria parasites within erythrocytes, and the resulting pathogenicity, is a hardwired program that secures proper timing of gene transcription and production of functionally relevant proteins. How stage-specific gene expression is orchestrated *in vivo* remains unclear. Here, using the assay for transposase accessible chromatin-sequencing (ATAC-seq), we identified ~4000 regulatory regions in *P. falciparum* intraerythrocytic stages. The vast majority of these sites is located within 2 kb upstream of transcribed genes and their chromatin accessibility pattern correlates positively with abundance of the respective mRNA transcript. Importantly, these regions are sufficient to drive stage-specific reporter gene expression and DNA motifs enriched in stage-specific sets of regulatory regions interact with members of the *P. falciparum* AP2 transcription factor family. Collectively, this study provides initial insights into the *in vivo* gene regulatory network of *P. falciparum* intraerythrocytic stages and should serve as a valuable resource for future studies.

Introduction

Malaria, caused by infection with parasites of the *Plasmodium* genus, remains a major health and economic burden [1]. The parasite's life cycle is intriguingly complex, requiring adaptation to several different host cell environments and transmission between the human host and the mosquito vector. The approximately 48 hr intraerythrocytic development of *P. falciparum* is responsible for most disease symptoms. It involves the invasion, remodelling, consumption, and rupture of human red blood cells while the parasite replicates by schizogony, giving rise to 16-32 new parasites [2]. Underlying this development and the pathogenicity of the parasite is a gene expression program that secures proper timing of gene transcription and production of functionally relevant proteins. However, despite being a fundamental eukaryotic process and a potential target of drug-based intervention, our understanding of gene expression regulation in *Plasmodium* is still in its infancy [3].

During the intraerythrocytic development cycle (IDC), the majority of genes is transcribed in a "just-in-time" manner, with peak mRNA abundances correlating with the need for the products they encode for [4]. Although post-transcriptional and translational control mechanisms operate in this stage as well [5, 6], the initial production of mRNAs, dictated by transcriptional and epigenetic mechanisms, remains a major and rate limiting step in the gene expression process during the blood-stage cycle. In *P. falciparum*, epigenetic regulation of gene expression is most evident in heterochromatin-mediated gene silencing of, for example, antigenic variation genes, the selection of erythrocyte invasion pathways, and control of gametocyte conversion rate (for review see [7]). This type of regulation is, however, limited to genes located in subtelomeric regions and a few chromosome-internal heterochromatic islands [8, 9], while the largest part of the parasite genome is in a transcriptionally permissive, euchromatic state.

These observations collectively point to an important role for transcriptional control mechanisms in stage-specific gene expression regulation, including the action of *trans*-acting transcription factors (TFs) that bind to specific DNA sequences and stimulate or inhibit the assembly and/or activity of the RNA polymerase II pre-initiation complex. Such sequence-specific TFs are, however, relatively low in numbers, constituting roughly 1% of all protein-coding genes [10, 11] compared to ~3% in yeast or 6% in human. Despite general scarcity of sequence-specific TFs, the relevance of the Apicomplexan AP2 family of TFs in *Plasmodium* has become evident over the past decade, mainly through the use of knock-out or knock-down experiments [12-21]. While these functional genomic approaches have been very powerful to dissect the function of TFs outside of the IDC, they could only suggest the essentiality of AP2 factors during the IDC. Furthermore, rather little is known about DNA elements that act in concert with these specific TFs. Most of our current understanding on *cis*-regulatory DNA elements stems from deletion analyses of promoters (e.g. [22-24]), *in silico* DNA motif predictions (e.g. [25-28]), and protein binding microarray studies

defining the *in vitro* sequence preference of recombinant ApiAP2 domains (Campbell et al., 2010). Although these studies have certainly been valuable and some of the DNA motif predictions could indeed be confirmed by ChIP-sequencing experiments (Kaneko et al., 2015; Santos et al., 2017), we still lack an accurate, genome-wide overview of *cis*-regulatory DNA elements and their activity *in vivo*.

The binding of specific *trans*-factors to the DNA is associated with the eviction and/or destabilization of nucleosomes thereby creating a more “accessible” chromatin environment. As a first attempt to explore open chromatin structures in *P. falciparum*, formaldehyde-assisted isolation of regulatory elements (FAIRE-seq) has been employed [29]. While this study reported increased accessibility at active promoter regions, the resolution of the data was not sufficient to improve the identification of regulatory elements. In a previous study, we applied MNase-sequencing to profile the nucleosome landscape and provided proof of principle that the chromatin environment of a predicted regulatory element is depleted of nucleosomes and this signature could be used to predict active regulatory elements [30]. As a completion of these efforts, here we set out to identify gene regulatory elements *in vivo* and on a genome-wide scale by directly profiling chromatin accessibility using the assay for transposase accessible chromatin-sequencing (ATAC-seq [31]). We combined ATAC-seq and directional RNA-sequencing (RNA-seq) on eight tightly synchronized *P. falciparum* IDC stages to profile gene regulatory events. Furthermore, we combined bioinformatics, biochemical, and reporter gene assays to characterize these *cis*-regulatory elements and their interactions with transcription factors. Collectively, this study represents a major step towards dissection of the transcriptional regulation network of this deadly pathogen and provides a valuable resource for future studies aiming to characterize or use gene regulatory elements.

Results

ATAC-seq identifies accessible chromatin regions in the AT-rich *Plasmodium* genome

To identify and profile transcription factor binding events, we performed ATAC-seq on synchronized *P. falciparum* 3D7 parasites at eight consecutive time points during their IDC (from 5 to 40 hr post invasion [hpi]). Considerable signal was detected in coding sequences (Figure S1A, purple track “t40 all”) and in subtelomeric regions of the genome (data not shown). We reasoned that this was likely due to the sequence bias of the enzyme [32], in combination with the distinctly higher GC-content of these sequences as compared to the AT-rich intergenic regions [33]. To correct for such biases as well as biases introduced during library preparation and sequencing, we performed the same assay on naked, genomic DNA (gDNA). This control library also showed a distinctly higher read count in the GC-richer coding sequences (Figure S1A, bottom gray track “gDNA all”) and subtelomeric regions.

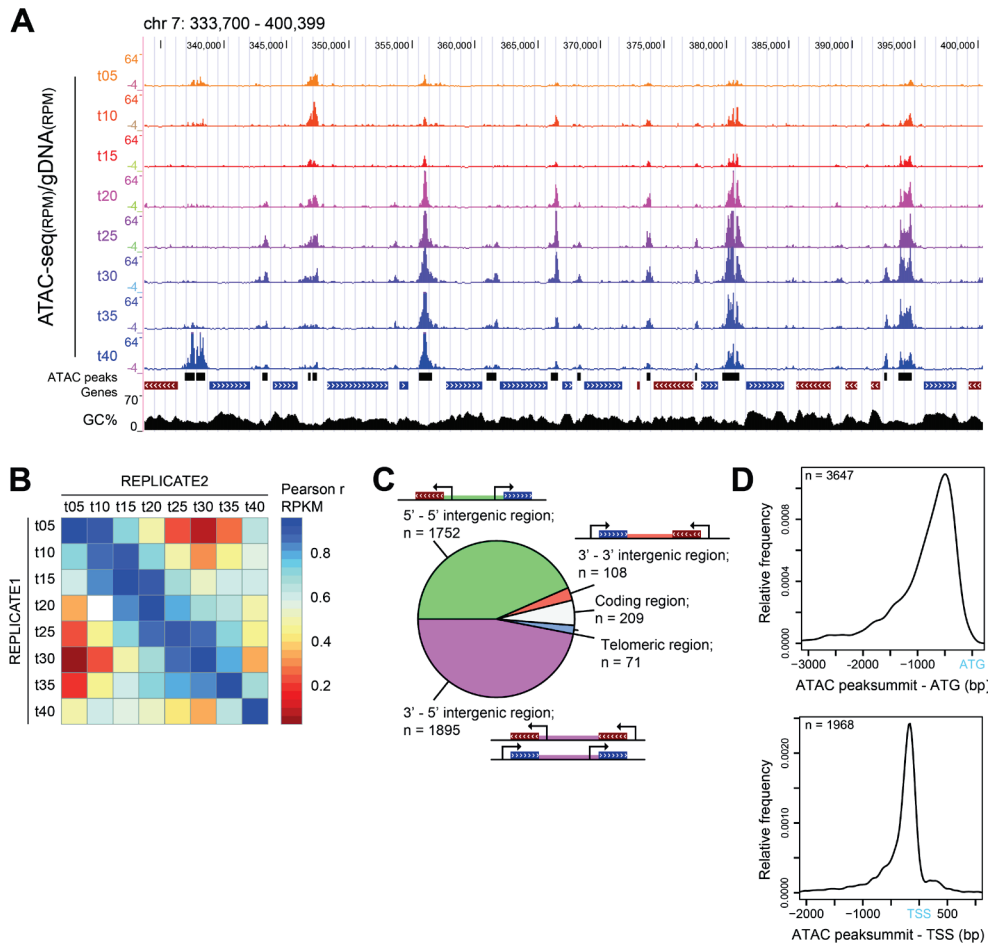


Figure 1: ATAC-seq detects chromatin accessibility during intraerythrocytic development of *P. falciparum*.

A) UCSC genome browser screenshot of a 66,700 bp region on chromosome 7 displaying chromatin accessibility profiles from eight time points (t05 - t40 hours post invasion [hpi]) as a ratio of normalized ATAC-seq tag count over background transposition in naked, genomic DNA (ATAC-seq/gDNA). Black bars below t40 track depict the peak regions as identified by MACS2 peak calling. Coding sequences are indicated as blue (positive strand) or red (minus strand) bars. GC%, the mean percentage of GC nucleotides, smoothed over 5bp windows. **B)** Heatmap depicting the Pearson correlation of RPKM values in peak regions from the above dataset (REPLICATE1) and an independent eight time point ATAC-seq dataset (REPLICATE2) demonstrating a high degree of reproducibility. **C)** The distribution of accessible regions among different genomic regions. Peaks were called on accessibility profiles of individual time points (t05-t40) and merged, yielding 4035 accessible regions. Intergenic regions were categorized based on the direction of transcription of the flanking coding regions (divergent, green; convergent, orange; tandem, pink). **D)** Peaks located in divergent and tandem intergenic regions were assigned to the closest downstream gene and the distance between the ATAC peak summit and the ATG or TSS was calculated. The line depicts the smoothed distribution of distances (kernel density estimate). See also Figure S1 and Table S1.

Furthermore, in the chromatin context, Tn5 transposition is known to give rise to (sub-) nucleosomal fragments (<150 bp) as well as fragments corresponding to mono-, di- and trinucleosomes as a result of transposition in the vicinity of TF binding sites and in linker regions between nucleosomes, respectively (Buenroostro et al., 2013). We therefore reasoned that selecting reads with a size between 50 and 150bp could increase the signal-to-noise ratio for the detection of TF binding sites. Indeed, compared to the other insert sizes, a higher proportion of 50-100 bp and 100-150 bp fragments mapped to intergenic, putative regulatory regions (Figures S1A and S1B) and to binding sites of an AP2 transcription factor (AP2-I [17], Figure S1B and next paragraph). Based on the above observations we decided to use only fragments with a size between 50 and 150bp and corrected the derived read counts with the read counts detected in the gDNA control library in all follow-up analysis (Figure 1A). Finally, the robustness of the data was assessed by preparing a replicate ATAC-seq dataset (replicate 2), which showed a high degree of correlation with the first dataset (Pearson correlation of 0.88 and higher; Figures 1B and S1C). Accordingly, our ATAC-seq approach enables robust and accurate identification of accessible chromatin regions despite AT-richness of the *P. falciparum* genome.

Dynamic chromatin accessibility in 5' intergenic regions highlights TF-binding events

Next, we identified local regions of increased accessibility for all eight time points of the IDC using the model-based analysis of ChIP-seq 2 (MACS2) algorithm for peak calling [34]. The number of identified accessible regions reflects the overall transcriptional output at the given stage of development [35-37], with ~500 regions in ring stages to ~3000 in late trophozoite/early schizonts (Figure S1D). After merging the peaks for all time points, a total 4035 regions were identified that show increased accessibility during one or more stages of the IDC (Table S1), 92% of which were also confirmed by the peaks called on the replicate ATAC-seq dataset (Data not shown). Ninety percent of the accessible regions locate to intergenic regions containing one or two putative promoter regions (Figure 1C). Within these regions, the majority of peaks locate up to 2 kb upstream of the ATG and, when a transcription start site (TSS) was known [30], within 1.5 kb upstream of the TSS (Figure 1D). In addition, these ATAC peaks captured 95% of the AP2-I binding sites detected by ChIP-seq (Figures 2A and 2B; [17]). Interestingly, two different clusters of AP2-I binding sites could be discriminated based on their accessibility profile during the IDC. A cluster of 64 regions (linked to 50 genes) becomes accessible in late trophozoites/early schizonts and a cluster of 105 regions (linked to 100 genes) that becomes accessible in mature schizonts and shows increased accessibility in t05 rings (Figures 2C and 2D). This subdivision is also evident at the molecular level with genes in cluster 1 being enriched for processes related to chromatin organization and cell cycle progression, while cluster 2 genes are clearly involved in host cell invasion (Table S2). Collectively, these results demonstrate that ATAC-seq detects dynamic chromatin accessibility in promoter regions of *P. falciparum* genes during the IDC and that

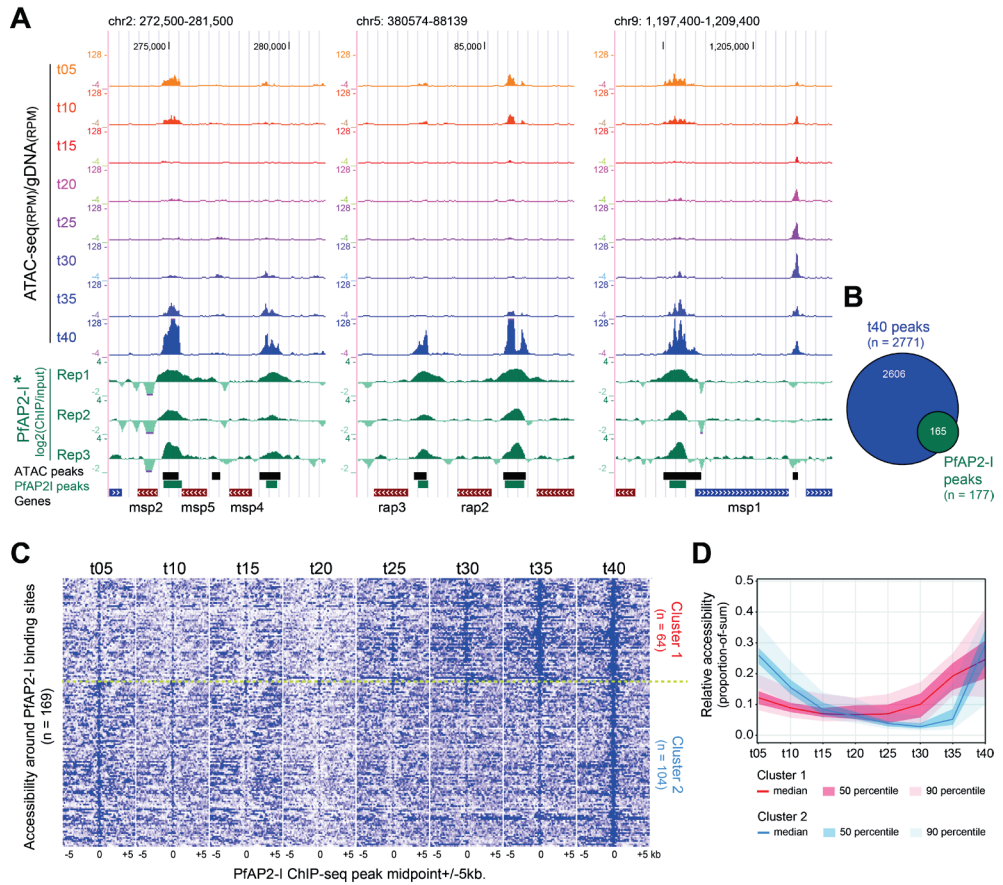


Figure 2: ATAC-seq predicts transcription factor binding events detected earlier by ChIP-sequencing. **A)** UCSC genome browser screenshots of three genomic regions with PfAP2-I transcription factors binding sites. Normalized ATAC-seq coverage is plotted as ratio over coverage in gDNA control. Data of the three PfAP2-I ChIP-seq replicates from [17], generated in the *P. falciparum* Dd2 line, were mapped against the *P. falciparum* 3D7 genome. Turquoise bars are PfAP2-I peaks from [17]. **B)** Overlap between PfAP2-I peaks and ATAC peaks from t40 time point (165 out of 2771 peaks overlap with 169 out of 177 PfAP2-I peaks) plotted as Venn diagram. **C)** Heatmap of non-gDNA corrected ATAC-seq accessibility profiles over PfAP2-I peaks (midpoint +/- 5 kb) that overlap with merged ATAC peaks (n = 169). Profiles were clustered using Pearson correlation calculated for the middle 100 bp window into 2 clusters by k-means clustering. **D)** Median accessibility profiles for the two clusters of ATAC peaks defined in (C) with the 50th and 90th percentile as a dark- and light-colored shades. Accessibility is calculated as proportion-of-sum of quantile normalized RPKM values over the time points. See also Table S2.

it can capture TF binding events. In addition, it demonstrates that data from ATAC-seq performed at multiple developmental stages can provide valuable temporal resolution to TF ChIP-seq data performed at a single time point.

Chromatin accessibility patterns are predictive for gene expression dynamics

To assess the relationship between chromatin accessibility and gene expression, we prepared directional RNA-seq libraries from the same parasite cultures as used for ATAC-seq. Overall, the chromatin accessibility pattern and the transcript abundance pattern of the downstream gene are positively correlated (see examples in Figure 3A). To quantify this correlation, accessible regions were assigned to the closest downstream located gene, yielding 3210 accessible region - gene pairs (accessible regions and/or genes with low signal and hence potentially noisy patterns were excluded; STAR Methods). Chromatin accessibility patterns during the IDC, which were highly reproducible between the replicates (Figure S1E, median correlation of $r=0.84$), were then used to group accessible regions by means of k-means clustering. Alignment with the assigned genes revealed a high degree of similarity between chromatin accessibility patterns and relative abundance of corresponding mRNAs (Figure 3B). In fact, the majority of the genes showed a clear positive correlation between chromatin accessibility and relative mRNA abundance (Pearson correlation > 0.6 , Figure 3C), demonstrating that chromatin accessibility is highly predictive of the gene expression pattern for the majority of genes. Moreover, this observation suggests that the gene regulatory events governing the IDC of *P. falciparum* are mainly activating events.

ATAC-seq regions are sufficient for regulating stage-specific gene expression

To study the potential of the identified accessible regions to drive stage-specific gene expression, parasite lines were generated with different accessible regions cloned upstream of the minimal *kahrp* promoter [38] and a *gfp-luciferase* (*gfp-luc*) reporter gene (Figures 4A, S2A and S2B). The region upstream of PF3D7_1372200 (*hrpIII*) has been characterized before and functioned as positive control [23] while the accessible regions upstream of PF3D7_0719000, PF3D7_1200700 and PF3D7_1222700 were selected based on their stage-specific accessibility and RNA abundance profiles (Figure S2C, blue framed rectangle). In addition, for PF3D7_0719000 and PF3D7_1200700 we created control parasite lines with a neighboring, not-accessible intergenic region cloned upstream of the minimal *kahrp* promoter (Figure S2C, red framed rectangle; for PF3D7_1222700 integration of the negative control construct could not be achieved). Remarkably, for all tested accessible regions, the temporal expression profile of the reporter gene matched the RNA expression profiles of the respective downstream located genes and was clearly above the background detected in the control lines (Figure 4B and S2C). This demonstrates that intergenic regions displaying dynamic chromatin accessibility are sufficient to induce stage-specific expression of the downstream located gene.

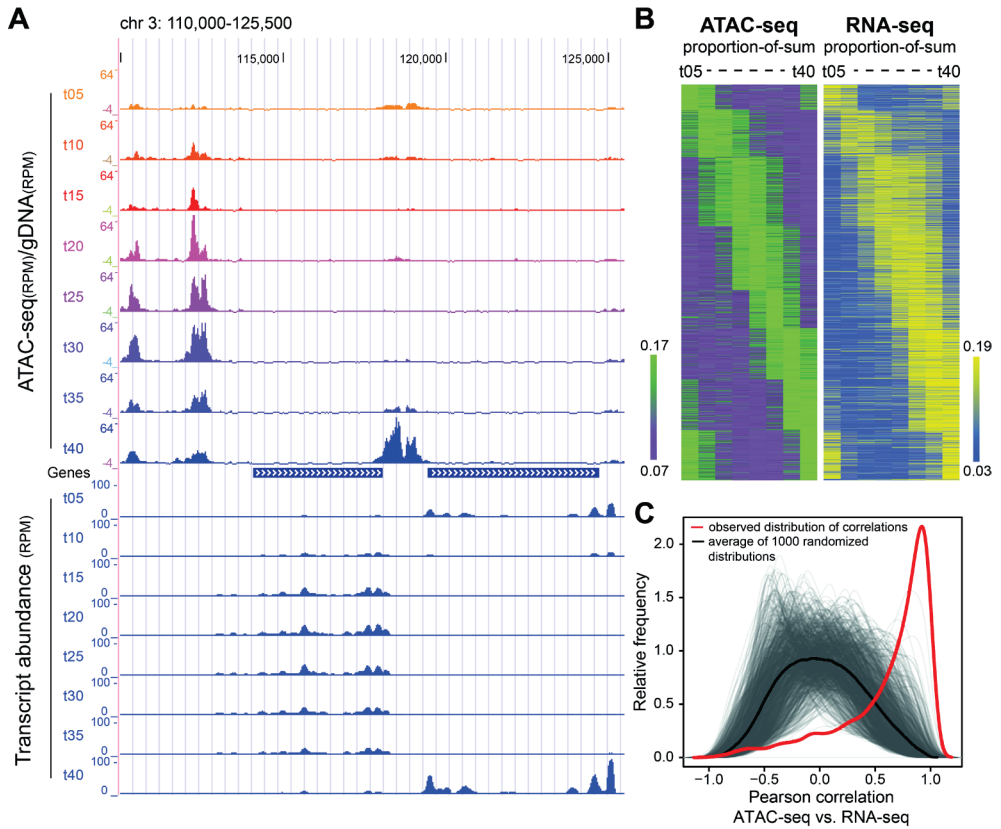


Figure 3: Chromatin accessibility positively correlates with mRNA abundance of the downstream gene.

A) UCSC genome browser screenshot (chr3: 110,000-125,500) of ATAC-seq chromatin accessibility profiles (top) and corresponding RNA-seq profiles (bottom; only reads mapping to the sense-strand are displayed). **B)** Heatmaps displaying relative chromatin accessibility and mRNA abundance profiles of the downstream gene through eight stages of intraerythrocytic development. Relative accessibility in ATAC-seq peaks located in divergent or tandem intergenic regions was calculated as a proportion-of-sum of quantile normalized RPKM values over the time points and clustered by k-means using the 1-Pearson correlation distance metric. The relative transcript abundances of the downstream located genes, expressed as proportion-of-sum of sense strand RPKM values over the time points, were plotted in the same order. Color scales range from the 20th to the 80th percentile for both datasets. **C)** Observed (red line) and 1000 randomized distributions (1000 gray lines with mean of all randomizations as black line) of Pearson correlation coefficients between relative chromatin accessibility and mRNA abundance of the downstream gene (as in (B)) plotted as smoothed kernel density distribution.

A



B

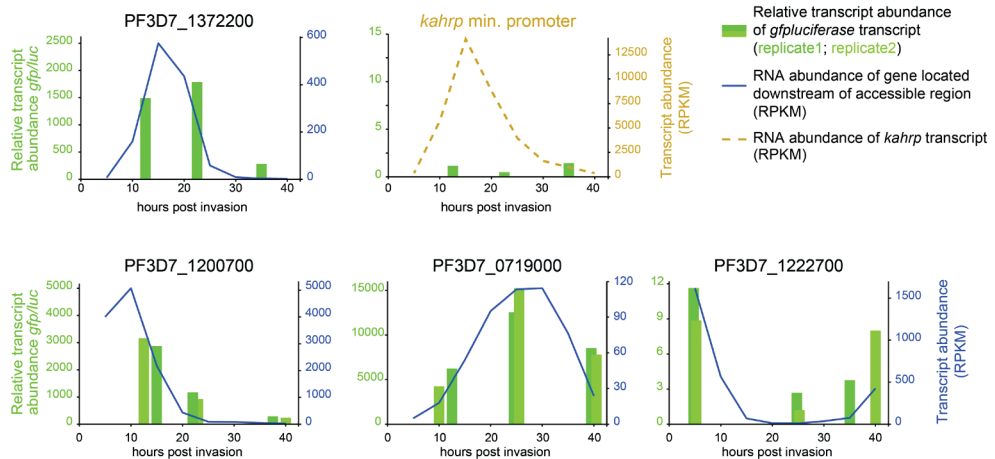


Figure 4: Accessible regions are sufficient in driving stage-specific reporter gene expression.

A) Schematic representation of the DNA constructs inserted into the $cg6$ locus of *attB*(+) 3D7 *P. falciparum* parasites using the site-specific integration system [58]. Accessible regions detected by ATAC-seq (light blue) were cloned upstream of the minimal *kahrp* promoter (yellow) followed by the *gfp/luciferase* fusion gene (green). **B)** Barplots showing the relative *gfp/luciferase* (*gfp/luc*) transcript abundance determined by RT-qPCR for ring, trophozoite and schizont stages from parasites carrying the reporter construct without accessible region (only the *kahrp* minimal promoter, yellow), with the accessible region of PF3D7_1372200 (*hrpIII*, positive control) or with selected accessible region located upstream of PF3D7_0719000, PF3D7_1200700 or PF3D7_1222700. For the latter three constructs with novel putative regulatory sequences, reporter gene expression was measured in biological duplicate. Relative *gfp/luc* transcript abundance was determined based on a standard, reference dilution series (see Methods). The relative abundances in replicate 2 were scaled to the average of replicate 1. The raw data are depicted in Figure S2. Transcript abundance (RPKM) of the respective gene as determined by RNA-seq is indicated as blue line for eight time points (t05-t40). See also Figure S2.

Specific sequence motifs are associated with dynamics of accessible regions

The ATAC-seq data revealed different patterns of accessibility over the IDC that showed an overall positive correlation with mRNA abundance. We reasoned that these were likely caused by the presence of different DNA motifs in promoter regions that are bound by specific transcription factors in a stage-specific manner. To identify DNA motifs that could perform this function, we first performed an exhaustive *de novo* motifs search using GimmeMotifs and seqGL (STAR Methods; [39, 40]). These *de novo* predicted motifs were combined with previously predicted *Plasmodium* motifs [41] and known vertebrate, invertebrate, and plant motifs from the CIS-BP database [42], yielding a comprehensive library of putative *cis*-regulatory sequences. Next, we identified gene sets with clear stage-

specific accessibility/expression profiles by selecting all accessible regions that positively correlated with transcript abundance (Pearson correlation > 0.6; n = 2118 regions; see Table S1) and clustered those considering both their accessibility as well as transcript abundance patterns over the IDC into eight clusters using k-means clustering (Figure 5A).

To identify motifs associated with specific accessibility/expression patterns, we used an ensemble of different regression and classification methods, as implemented in GimmeMotifs (van Heeringen and Veenstra, 2011), and searched for motifs from the above library that were consistently enriched in accessible regions of a specific cluster ($p < 0.01$, in at least two out of three runs; Table S3). After manually removing eight low-information content motifs (Figure S3A and Table S3), we clustered the remaining motifs yielding 41 non-redundant motifs (Figures 5B, Figure S3B and Table S3, for redundancy filtering see [39]). Interestingly, for all ATAC/RNA-seq co-clusters we observe enrichment of at least one predicted AP2 motif (in total 16 motifs predicted for 13 different AP2 proteins, blue font in Figure 5B), suggesting that the corresponding AP2 transcription factor is likely relevant in regulating these genes. Additionally, we detected motifs similar to the G-box element upstream of heat shock genes (motif vertebrate.C2H2_ZF_M6240 [22]). Importantly, in addition to these previously predicted motifs we identify 13 *de novo* motifs with potential regulatory capacity in *P. falciparum* (indicated with red font in Figure 5B).

DNA-pull down combined with quantitative proteomics reveals *cis-trans* regulatory interactions

We selected four motifs and identified their protein interactors by performing DNA pull-downs using short oligos representing actual accessible sequences containing the selected motifs and native nuclear extracts from non-synchronous, asexual *P. falciparum* 3D7 cultures. To identify proteins that specifically bind to the motif, but not to a control oligo with a scrambled motif, we analyzed pull-down and control samples by quantitative tandem mass-spectrometry. (See Table S4 for the complete list of motifs and identified proteins.)

First, we tested the CA-repeat motif predicted for protein PF3D7_0802100, which formed a “motif group” with similar motifs predicted for other AP2 proteins (PF3D7_0420300, PF3D7_1305200 and PF3D7_1456000, Figure S3B). The DNA pull-down confirmed the specific recruitment of PF3D7_0802100 and PF3D7_0420300 to the ACACACAT motif when compared to a scrambled control motif (ATCAAACC), but not the other two factors (Figure 6A). Next, we tested three *de novo* motifs (031, 028, and 050). All these motifs captured at least one ApiAP2 transcription factor (Figures 6B-D). The ApiAP2 factor PF3D7_0420300 was consistently identified among the interactors of the TTATTACAC motif (*de_novo_motif_031*, Figures 6B and S4A). Remarkably, this motif is more similar to the sequence preference of the second AP2 domain of this factor (TTATTACAC versus GIGTTACA [41]), potentially suggesting that this factor can bind to two different regulatory elements (CACACACA see above and TRTTACAC) using different AP2 domains.

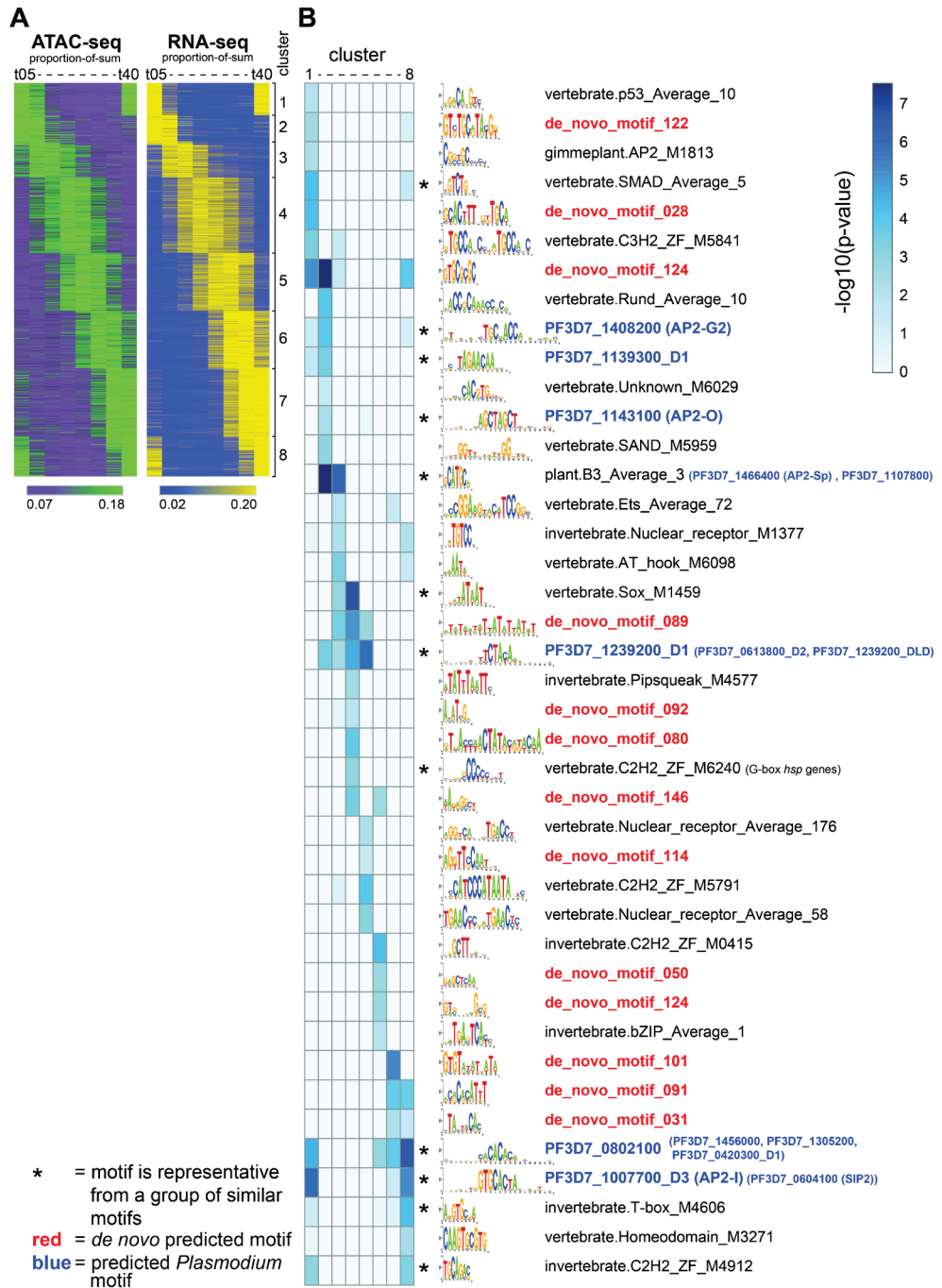
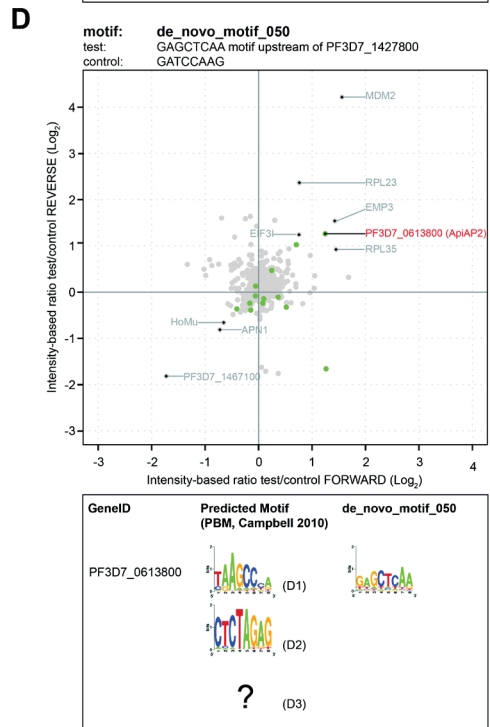
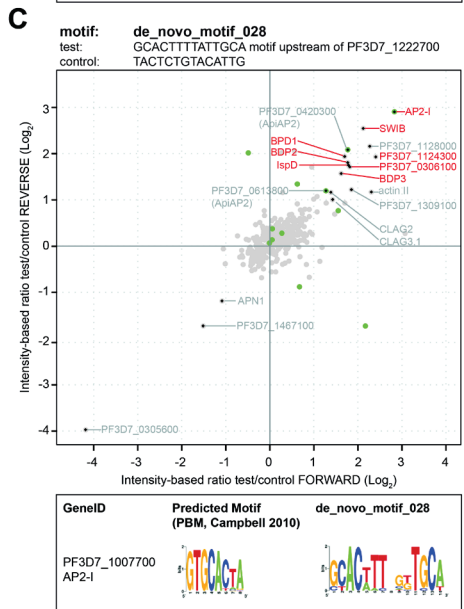
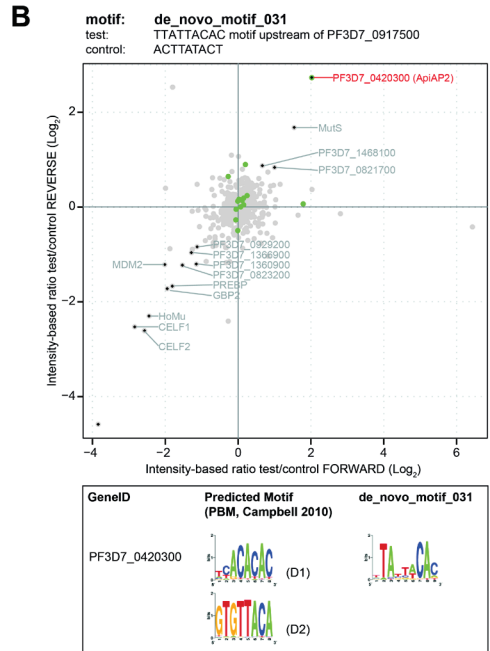
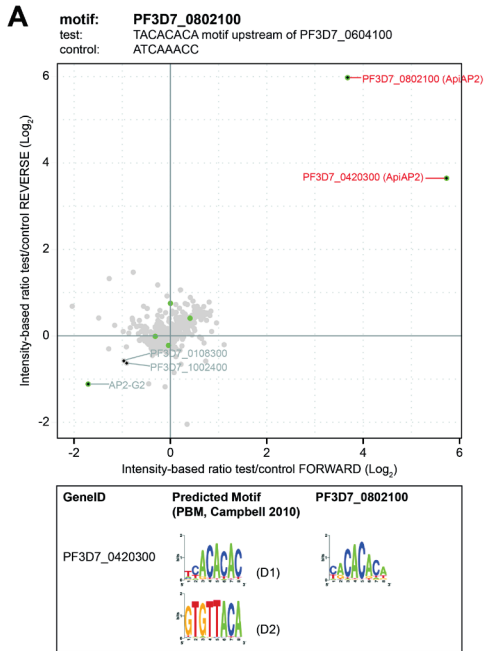


Figure 5: DNA motifs enriched in stage-specific ATAC regions.

A) Co-clustering of ATAC-seq and RNA-seq data, limited to peak-to-gene pairs with Pearson correlation > 0.6 (n = 2118; see Table S1). Accessibility and transcript abundance are expressed as proportion-of-sum of (qn)RPKM values over the time points and clustered by k-means using the 1-Pearson correlation distance metric into eight clusters. Color scales range from the 20th to 80th percentile per dataset. **B)** Heatmap of significance estimates for differential motif enrichment (expressed as $-\log_{10}(\text{P-value})$). Each column relates to a cluster generated by co-clustering the ATAC-seq and RNA-seq data by k-means cluster (see (A)). Each row refers to a motif or a 'motif group', with logo and name listed on the right. Asterisks indicate that the motif is a representative from a group of similar motifs (see Figure S3B and Table S3). Predicted binding sites for ApiAP2 transcription factors are reported in blue font with PlasmoDB geneID (name, if known in brackets). When a cluster contains one or more predicted *Plasmodium* motifs, these are reported in brackets behind the representative motif. See also Figure S3, Table S1 and Table S3.

For the second *de novo* motif (028), we tested three different probes and, interestingly, partially overlapping sets of chromatin-related factors were enriched together with the transcription factor AP2-I in each of them (Figured 6C and S4B and S4C). These included bromodomain proteins (BDP) 1, -2 and -3 (PF3D7_1475600); HMGB3; and a SWIB/MDM2 domain-containing protein (SWIB, PF3D7_0611400), but also two conserved unknowns (PF3D7_0306100 and PF3D7_1124300), NOP5 and IspD. This suggests that AP2-I is sufficient to recruit these chromatin factors to target gene promoters, in line with the current study of (Santos et al., 2017). Notably, in two of the three pull-downs another AP2-factor, SIP2, was enriched with the motif-containing oligos (Figures S4B and S4C). SIP2 has first been predicted to bind to a motif very similar to AP2-I (Figure S4B and S4C, [41]) but later shown to bind a longer sequence, named SPE2 (NVTGCA-4(5)-VGTGCR) upstream of subtelomeric *var* genes [12]. By chance sequences similar to a full SPE2 motif can be found in both of these oligos, including some flanking sequences (Table S5), explaining the binding of SIP2 to these two, but not the first oligo. Hence, SIP2 is likely not a genuine interactor of *de_novo_motif_028*. Lastly, we tested *de_novo_motif_050* (GAGCTCAA) using DNA probes from two different genomic regions. In this case, we observed a moderate, but consistent, interaction with the ApiAP2 factor PF3D7_0613800 (Figures 6D and Figure S4D). This motif is different from the predicted binding sites of the domains 1 and 2 of PF3D7_0613800 [41] and might be recognized by its third domain for which sequence preference has not been defined.

Collectively, these experiments indicate that the motifs predicted from the stage-specific regulatory elements indeed interact with specific transcription factors. In addition, they point to a delicate interplay between DNA elements, transcription and chromatin modifying factors in regulating intraerythrocytic development of malaria parasites.



- Protein with (putative) DNA-binding domain
- ◆ Significant, FDR=0.05
- red = Confirmed with motif from different genomic location

Figure 6: DNA pull-downs identify potential *cis-trans* regulatory interactors.

Scatter plots displaying the quantitative proteomic analysis of duplicate DNA pull-downs with label swap using (A) 58-60bp DNA probes with the CA-repetitive motif (ACACACAT) and three *de novo* predicted motifs; (B) *de_novo_motif_031*, TTATTACAC; (C) *de_novo_motif_028*, GCACWWT NNKTGCW; (D) *de_novo_motif_050*, GAGCTCAA). The same probes with a scrambled motif were used as controls. The statistically significant outliers (black diamond, intensity-based FDR < 5%) are the potential interactors to the motif. Red font indicates that the interaction was confirmed using a probe from a different genomic region but containing the same motif. Green dots are candidate DNA-binding factors derived from Table 4 of [10]. Earlier predicted binding sites for the identified AP2 factors are shown in relation to the motif used in the pull down below each plot. See also Figure S4, Table S4.

Discussion

Here we present high-resolution temporal chromatin accessibility data during intraerythrocytic development of *P. falciparum*. ATAC-seq, used in this study (Figure 1), clearly supersedes other chromatin-based technologies such as FAIRE-seq [29, 43], MNase-seq [30], or ChIP-seq [44] in identifying active regulatory sites on a global scale, both in terms of the number of sites identified as well as in terms of spatial resolution of the data. Furthermore, ATAC-seq shows a nearly complete overlap with ChIP-seq profiles of a TF, AP2-I (Figure 2, [17]). While due to the bias of the transposase against AT-rich sequences and stringent filtering we might not detect all regulatory events, we identified an accessible region for at least 60% of all *P. falciparum* genes and about 70% of the genes expressed during the IDC (comprising about 85% of all genes; Otto et al., 2010), providing a valuable resource for future studies that could range from targeted gene studies to predicting functional consequences for SNPs or other mutations in non-coding regions.

Notably, the coupling between accessible regions and genes in our analysis was based on the assumption that most genes in the *P. falciparum* genome are regulated by the nearest regulatory elements. While this assumption might not always be correct, it is remarkable that temporal accessibility and mRNA abundance profiles strongly correlate for the majority of the genes (Figure 3), suggesting functional interactions between these regulatory elements and the nearby genes. Furthermore, the DNA sequence of all four tested accessible regions was sufficient in dictating a stage-specific expression pattern to a reporter gene that is similar to that of the respective endogenous gene (Figure 4). Although our data does not exclude the existence of distant enhancers, at least the majority of identified regions in this dataset likely operates at close distance to their target gene. The scarcity of distant regulatory sequences in *P. falciparum* is also supported by the fact that long distance interactions identified so far in chromosome conformation studies in this parasite were restricted to centromeres, ribosomal DNA loci and subtelomeric regions [45, 46]. Collectively, our data together with earlier studies (e.g. [15, 17, 19, 23, 44, 47]), suggest that most *P. falciparum* genes have a compact regulatory unit like other unicellular eukaryotes (i.e. yeast,) with minimal promoter(s) and upstream regulatory element(s)

located within one or two kilobases of the target gene. However, our data, in combination with future high-resolution chromosome conformation studies, might reveal distant enhancers, if they exist.

The marked positive correlation between chromatin accessibility and mRNA abundance (Figure 3) also supports the notion that activating, rather than repressive, regulatory events drive gene expression in the IDC of *P. falciparum*. Alternatively, ATAC-seq might preferentially detect activator bound chromatin regions. Yet ATAC-seq has been shown to detect non-transcription-related DNA-binding events (e.g. CTCF binding to insulator elements, [31]) and bivalent promoters (co-occurrence of activating and repressive histone markings) in other organisms [48, 49]. Also, thus far, only two transcription factors, AP2-G2 and AP2-Sp, have been postulated to have repressive properties during blood stage development of *P. berghei* [16, 18, 19]. Regardless of the presence and specific functions of a few transcriptional repressors, our data suggest that a cascade of transcription activating events is mainly responsible for the stage-specific expression during blood stage development of *P. falciparum*.

Since their discovery, the AP2 gene family has been regarded as the major family of putative transcription factors in *Plasmodium*. However, even with these 27 putative transcription factors the proportion of transcription factors to the total number of genes remains to be low (~50-60 among ~5800 genes compared to, for example, 169 per ~6000 genes in yeast [50]). Therefore, to our surprise, besides AP2s, we did not consistently detect any other protein family in our DNA pull downs that could function as a sequence-specific DNA-binding factor. Hence, despite the existence of few other types of DNA-binding factors in *Plasmodium* (including C2H2-type [51], Myb-type [52], and HMGB-domain proteins [53]), so far all evidence suggests that the AP2 family can be regarded as the major TF family in *Plasmodium*, leaving researchers puzzled how such a small number of factors can govern such a delicate gene expression program. Combinatorial action of multiple transcription factors has been suggested to increase the regulatory potential of these factors in directing development of malaria parasites (e.g. [28, 54]). Such cooperative interaction between AP2-I and other stage-specific transcription factors could explain the different accessibility patterns observed for the AP2-I binding sites (Figures 2D). Yet we did not find any DNA motifs, other than the AP2-I binding site (GTGCA), strongly enriched in these clusters that could serve as a binding site for such factor (data not shown). Alternatively, post-translational modifications of DNA-binding domains from TFs [55] or protein-protein interactions between TFs and cofactors could affect TF sequence specificities and or recruitment of TFs to specific chromatin regions [56]. To this end, we (Figure 6C) and others [17, 57] detected a strong interaction between AP2-I and an epigenetic complex involving, among others, two acetylated histone binding proteins (BDP1 and BDP2). However, if and how these proteins contribute to stage- and/or sequence-specific binding patterns of AP2-I or enhanced binding of AP2-I to acetylated chromatin regions remains to be

determined. Furthermore, nearly half of the AP2 TFs has more than one AP2 domains. Our pull-down data suggests that in fact some of the AP2-factors could interact with different regulator elements using different domains (Figure 6). Eventually, it seems conceivable, that the limited number of sequence-specific transcription factors encoded by the *Plasmodium* genome use the combination of the above mechanisms to achieve the precision of regulation required to drive the gene expression program underlying blood stage development. Collectively, our work provides the in depth global view of the *in vivo* transcriptional regulatory events during intraerythrocytic development of *P. falciparum*. It also highlights some intricate details of the interplay between transcription factors and *cis*-regulatory elements that controls gene transcription, bringing us a big step closer to understand and fight this deadly parasite.

Author Contributions

C.G.T. and S.A-K.F. designed and performed experiments, analyzed data, prepared illustrations and wrote the manuscript. M.S.V-B., D.R.W. and S.J.v.H. provided expertise and help in the bioinformatics analyses and edited the manuscript; R.B. conceived and supervised the project, designed and performed experiments, and wrote the manuscript.

Acknowledgements

We are very grateful to Wieteke Anna Maria Hoeijmakers for her valuable suggestions on the project and the manuscript as well as for providing reagents, protocols and advice on mass spectrometry experiments. We thank Nina C. Hubner and Luan N. Nguyen for providing protocols on DNA pull-down experiments. We thank Marcus C.S. Lee and David Fidock for providing parasite strains and plasmids for rapid integration of our reporter constructs. We want to thank Luan N. Nguyen, Pascal Janssen and Eva Janssen-Megens for running mass spectrometry samples and sequencing libraries. This project also benefited greatly from the database provided by the GeneDB and PlasmoDB teams (www.genedb.org; www.plasmodb.org). This work was funded by the Netherlands Organization for Scientific Research [NWO-Vidi 864.11.007 to R.B.]. The authors declare no conflict of interest.

Star Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Requests for resources and reagents should be directed to the Lead Contact, Richárd Bártfai (r.bartfai@science.ru.nl).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Parasite culture conditions

Parasites were cultured in RPMI medium supplemented with 10% human serum, 0.2% NaHCO₃ and 2.5% or 5% human O+ red blood cells. Parasite lines were maintained in a shaking semi-automated 37 °C incubator in 10ml total volume and 5% hematocrit. For the ATAC-seq and RNA-seq parasite collections, the cultures were kept in T75 culture flasks with 20ml total volume and 2.5% or 1.25% hematocrit. For these collections the T75 flasks were placed in candle jars in a steady 37 °C incubator, as in [30]. For the collections of parasite RNA for RT-qPCR and parasite nuclei for the generation of nuclear protein extract, 20 or 50 ml parasite cultures with 2.5% hematocrit were kept in T75 or T175 flasks in a steady 37 °C incubator with gas composition of 3% O₂, 4% CO₂ and 93% N₂.

Parasite and bacterial strains

See Table S5 for details on parasite and bacterial strains used in this study.

METHOD DETAILS

Parasite culture synchronizations and collections

For combined ATAC-seq and RNA-seq collections, cultures were selected for *var2csa* expression, expanded and synchronized as follows. VAR2CSA panning was performed as in [60]. Petri dishes (150 × 15 mm, BD biosciences Falcon 351058) were coated overnight with Chondroitin sulfate A (0.05% CSA in PBS) and blocked with 1% Casein/PBS solution for at least one hour and rinsed twice with RPMI. Parasite cultures were centrifuged, resuspended in RPMI with 10% human serum, transferred to the CSA-coated petri dishes and incubated for 30 min at 37 °C in a candle jar. Afterwards, unbound parasites and non-infected erythrocytes were removed by gentle RPMI washes. Bound parasites were extensively resuspended in complete medium to detach them from CSA. Fresh blood was added to these parasites and they were put back in culture medium as described above in the shaking incubator. This selection was repeated four times before expansion. Before and during expansion of the culture, parasites were synchronized by sorbitol treatment and a Percoll gradient centrifugation. For the sorbitol treatment, parasites were spun down and the parasite pellet was gently resuspended in 6-7 pellet volumes of 5% D(-)-sorbitol (Merck, #107758) and incubated for 10 min at 37 °C while shaking. Parasites were spun down and new medium and fresh blood were added to 5% hematocrit. For percoll gradients, parasite cultures were spun down, resuspended in fresh medium to 10% hematocrit and an equal volume of 63% Percoll (GE Healthcare, #17-0891-01) in PBS was gently layered below the culture. The schizont interface was collected after spinning the gradient and fresh, Plasmodipur filtered RBCs (EuroProxima, the Netherlands) were added a 1.5 h later which was then set as time point zero (0 hours post invasion (hpi)) resulting in a synchronicity window of 7 h (i.e. 7h difference between the first and last invasion). Medium was changed every ten hours but not less than ten hours before collection. Cultures were mixed with every medium change and after 20 hpi kept at 1.25% hematocrit. Parasites were collected from 5 hpi onwards every 5 hours and ATAC-seq and RNA-seq collections were performed from the same synchronized culture. Giemsa stained blood

smears were made at each time point to monitor parasite growth and staging (See Figure S5 for representative microscope images and Table S6 with counts of parasite stages per time point).

For collections of parasites carrying the *attP(+)_minkahrp* expression constructs, site-specific integration was first confirmed and parasites were synchronized using sorbitol treatments and Percoll gradient centrifugations as described above. For each parasite line ring, trophozoite and schizont stages were collected (PF3D7_0719000 replicate 1 synchronized to a ~8 h window, collected 12 hpi, 25 hpi, 39 hpi; PF3D7_0719000 replicate 2 synchronized to a ~10 h window, collected 10.3 hpi, 25 hpi, 40.5 hpi; PF3D7_1200700 replicate 1 synchronized to a ~12 h window, collected 14 hpi, 23 hpi, 38 hpi; PF3D7_1200700 replicate 2 synchronized to a ~10 h window, collected 12.25 hpi, 23 hpi, 39.5 hpi; PF3D7_1222700 replicate 1 synchronized to a ~5 h window, collected 5 hpi, 25 hpi, 35 hpi; PF3D7_1222700 replicate 2 synchronized to a ~8 h window, collected 6.25 hpi, 24.5 hpi, 41 hpi; PF3D7_1372200 synchronized to a ~12 h window, collected 12 hpi, 23 hpi, 36 hpi; *kharp* minimal promoter only synchronized to a ~5 h window, collected 5 hpi, 25 hpi, 35 hpi; PF3D7_0719000 negative replicate 1 synchronized to a ~5 h window, collected 8 hpi, 25 hpi, 38.75 hpi; PF3D7_0719000 negative replicate 2 synchronized to a ~7 h window, collected 10 hpi, 25.5 hpi, 39 hpi; PF3D7_1200700 negative replicate 1 synchronized to a ~9 h window, collected 12 hpi, 24 hpi, 41 hpi; PF3D7_1200700 negative replicate 2 synchronized to a ~10 h window, collected 12.5 hpi, 25 hpi, 41 hpi).

ATAC-seq library preparation

Native parasite nuclei were isolated as in [36]. In short, after lysis of RBCs by 0.05% saponin treatment and separating nuclei from parasite debris using a cell lysis buffer (CLB: 10 mM Tris-HCl pH8.0, 10 mM NaCl, 3 mM MgCl₂, 0.2% NP-40) with 0.25 M sucrose cushion. A 10 µl sized nuclei pellet was resuspended with a cut-off pipet tip in 337.5 µl CLB and for ATAC-seq replicate 1 69 µl of nuclei was used for t05 to t20 and 23 µl of nuclei was used for t25 to t40 (these volumes were based on previous tests using a dilution series of nuclei). For ATAC-seq replicate 2 we had to optimize the amount of nuclei again due to the use of a kit from a different lot and this led us to use 466 µl of nuclei for t05 and t10 and 155 µl for t15 and t20. Nuclei were brought to 10.5 µl in CLB and used in a 25 µl ATAC reaction based on (Lara-Astiaso et al., 2014) with 2 µl Tn5 transposase and 12.5 µl TD buffer (Nextera DNA Library Prep Kit, #FC-121-1030, Illumina, USA). Reactions were incubated for 1 h in a 37 °C heat block. Nuclei were kept in suspension by gently tapping the tube every 10 minutes. The reaction was stopped by addition of 5 µl clean up buffer (900 mM NaCl, 300 mM EDTA), 2 µl 5% SDS and 2 µl proteinase K (Sigma-Aldrich #P6556) and incubated for 30 min at 40 °C. Tagmented DNA fragments were isolated using 2.4 sample volume of Agencourt AMPure XP beads (Beckman Coulter, #A63882, Inc. USA). Half of the isolated DNA was used for library preparation (the other half was stored as back-up) starting with a size selection using 0.85x volumes of AMPure XP beads to enrich for fragments of 500 bp and smaller. Size-selected fragments were amplified using the KAPA HiFi HotStart ready-mix (KAPA Biosystems, #KK2602, US) and Nextera index primers (Nextera DNA Sample Preparation Index Kit, #FC-121-1012) under the following conditions: 98 °C for 2 min; 16 cycles of 98 °C for 20 s, and 62 °C for 3 min; 62 °C for 5 min. Libraries were purified using 1x volumes Agencourt AMPure XP beads. The fragment size distribution of the libraries was evaluated in a High-Sensitivity Bioanalyzer run (Agilent, #5067-4626, US) and the size selection was repeated when there was a large proportion of fragments longer than 500 bp (replicate 1 t05, t15, t30, t35, t40). To control for sequence bias, the same ATAC protocol was applied to genomic DNA from synchronous wild type 3D7 *P. falciparum* ring stage parasites using 547.0 ng or 60.8 ng of input DNA. All ATAC-seq libraries were KAPA quantified (KAPA Library Quantification Kit, #KRO405).

RNA-seq library preparation

Parasite cultures were immediately placed on ice and washed once with ice-cold PBS. Pelleted cultures were resuspended in RLT buffer (Qiagen, #74106) supplemented with 1% β-mercaptoethanol and snap-frozen in liquid nitrogen. Total RNA was extracted using the RNeasy Mini Kit (Qiagen, #74106;

including RNA clean-up and two on-column DNase treatments) and RNA concentration was measured using the Qubit™ RNA HS Assay Kit (Invitrogen, #Q32852). RNA was then polyA-selected using the Oligotex mRNA Mini Kit (Qiagen, #70022) according to manufacturer's instructions. Subsequently, 2000 ng of polyA-selected total RNA equivalent were fragmented by alkaline hydrolysis (40 mM Tris acetate pH 8.2, 100 mM potassium acetate, 30 mM magnesium acetate) for 1 min 45 s at 85 °C in a 150 µl volume and precipitated as previously described in (Hoeijmakers et al., 2013). Next, polyA-selected RNA was cleaned from remaining genomic DNA (detected by qPCR) by two additional TURBO DNase treatments (Ambion, #AM2238). Strand-specific RNA-seq was performed as in [30]. Accordingly, first strand cDNA synthesis was performed with AT-corrected Random N9 primers (76% AT) in the presence of 0.2 µg Actinomycin D (Thermo Fisher Scientific #11805017). During second strand synthesis dTTPs were substituted with dUTPs to preserve strand-specific information. Next, 10 ng of each double stranded cDNA library was end repaired, extended with 3' A-overhangs, barcoded with NextFlex adapters (Bio Scientific, #514122) and treated with USER enzyme (NEB, #M5505L) to specifically degrade the dUTP-containing second strand. Libraries were amplified by PCR (98°C for 2 min; 4 cycles of 98 °C for 20 s, 62 °C for 3 min; 62 °C for 5 min) using KAPA HiFi HotStart ready mix (KAPA Biosystems, #KM2602) and NEXTflex primer mix (Bio Scientific, #514122) and subsequently gel size-selected for 300 - 400 bp using 2% E-Gel Size Select agarose gels (Invitrogen, #G6610-02). After an additional eight cycles of amplification (see above), libraries were purified and adapter dimers depleted using 1x volume of Agencourt AMPure XP beads (Beckman Coulter, #A63880).

Parasite transfection

Parasite transfections were performed as in [61] using a BTX electroporation system. Synchronized ring stage 3D7 *P. falciparum* attB(+) parasites were pelleted by centrifugation and 100% hematocrit packed cells were mixed with 75 µg of the pINT and 75 µg of the desired attP(+)_minkahrp plasmid in cytomix (120 mM KCl, 0.15 mM CaCl₂·2H₂O, 5mM MgCl₂·6H₂O, 25 mM HEPES, 2 mM EGTA, 10mM K₂HPO₄, 10mM KH₂PO₄) in 450 µl total volume in a 2 mm electroporation cuvette (BTX, #45-0125). After transfection, parasites were resuspended in warm culture medium and cultured at 2.5% hematocrit in the presence of 2.6 nM WR99210 (Jacobus Pharmaceutical Company, Inc.), 2.5 µg/ml Blasticidin S HCl (Gibco, #R210-01) and 250 µg/ml Geneticin (Gibco, #11811-031). After seven days, culturing was continued without Geneticin. When the parasite cultures became blood-smear, positive site-specific integration of the attP-containing plasmid into the parental line was confirmed by performing PCR using the primer combinations 'p1_for'/'p1_rev' and 'p2_for'/'p2_rev' (Table S7) on extracted genomic DNA (Qiagen, #51106). Genomic DNA of the non-transfected parental 3D7 line was used as negative control (Figure S2a and b). Afterwards parasites were cultured at a three week on/off schedule of 2.6 nM WR99210 and 2.5 µg/ml Blasticidin S HCl.

Plasmid DNA cloning

To examine the regulatory potential of the identified accessible several parasite lines were generated: four parasite lines with an integrated plasmid containing an accessible region detected by ATAC-seq upstream of a minimal *kahrp* promoter and a *gfp-luc* reporter gene (Figure 4a, Figure S2a), two parasite lines with a not-accessible, control region instead of the accessible region (a third line did not show successful integration) and one parasite with an integrated plasmid containing the minimal promoter followed by the reporter gene. To generate these parasites, we applied the Bxb1 integrase-mediated site-specific attP/attB integration system from [58] which results in directional integration of an attP-site containing plasmid into the *cg6* locus of a parental attB site-containing *P. falciparum* line. To generate the specific attP-plasmids, the pDC2 plasmid [58] was modified on several points. (All primers used for cloning, integration checking and RT-qPCR are listed in Table S7). The orientation of the 5' *cam-snf7-gfp-3'hsp86* cassette was reversed using the primers '5'Pfcam-F' and '3'hsp86-R', *PstI*/*Apal* digestion and ligation by the T4 ligase (Promega, #M1804) resulting in plasmid pOM1. The

snf7-gfp element was replaced by the *gfp-luc* sequence from the MV163 plasmid [62] using the primers 'GFPLuc-F' and 'GFPLuc-R', *AvrII/XhoI* digestion and ligation by the T4 ligase resulting in plasmid pOM2. Finally, the 5'*cam* was replaced by the *kahrp* minimal promoter [38] using the primers 'kahrp-F' and 'kahrp-R', digestion by *AvrII/Agel* and T4 ligation resulting in plasmid *attP(+)_minkahrp*. Accessible or control regions located upstream of the genes PF3D7_0719000, PF3D7_1200700 and PF3D7_1222700 or the accessible region upstream of PF3D7_1372200 were amplified and inserted upstream of the *kahrp* minimal promoter using their respective primers listed in Table S7 and *BglII/NotI* digestion and ligation by the T4 ligase. These accessible regions were selected because they showed clear, distinct, stage-specific accessibility patterns; the downstream gene showed a matching gene expression pattern; we favored accessible regions located in tandem intergenic regions (i.e. containing a single promoter) for clarity of the assignment between genes and accessible regions; and we excluded ATAC regions and parts of the peak that overlapped with a TSS.

RNA extraction, cDNA synthesis and qPCR

Total RNA was extracted as described in 'RNA-seq library preparation' and was checked for genomic DNA contamination by qPCR. If needed, the sample was additionally treated once or twice with TURBO DNase (Ambion, #AM2238). For each sample 500 to 1000 ng of total RNA was mixed with random hexamer primers (0.5 µg, Roche #11034731001), OligodT₁₂₋₁₈ (0.5 µg, Invitrogen #18418012) and dNTPs (0.5 mM in the final volume of 20 µl, Invitrogen 10297-018) and incubated for 5 min at 70 °C. First strand synthesis was performed for 1 h at 42 °C in First Strand Buffer (Invitrogen) supplemented with DTT (10 mM), Superscript III (200 units, Invitrogen, #18080044) and RNasin Plus RNase inhibitor (40 units, Promega, #N261B), after which superscript III was inactivated by incubation at 70 °C for 15 min. For all samples, a negative control reaction was performed in which Superscript III was replaced by water (RT minus control) under identical conditions. For each parasite line the same amount of RNA was used as template from the different time points.

To measure the relative *gfp-luciferase* (*gfp-luc*) transcript abundance, a qPCR was performed using SYBRgreen supermix (BioRad) and primers which were mixed according to the manufacturer's instructions. The qPCR was performed using the CFX96 Real Time Systems C1000 Touch Thermal Cycler (Bio-Rad) with the following program: 95 °C for 3 min, (94 °C for 10 s, 52 °C for 30 s, 68 °C for 30 s) 39 cycles, 95 °C for 1 min, 65 °C for 1 min and a gradient from 65 °C to 94.5 °C with a 0.5 °C increase every 10 s. Primers specific for *gfp-luc* served to assess the relative abundance of the reporter transcript ('GFP-1', 'GFPLuc', 'Luc-1', 'Luc-2') and primers for *blastidicin* and *actin* ('BSD-1', 'BSD-2', 'actin') controlled for successful cDNA synthesis (results not shown). All -RT controls reported 'not detectable' (NA) or in Cq values in the range of the H₂O control, which was included for all primer pairs (results not shown). The relative *gfp-luc* transcript abundance was measured against a standard dilution series prepared from *P. falciparum* 3D7 *attB(+)* genomic DNA and pOM2 plasmid DNA mixed in a close to 1 molar ratio (10-fold dilution series of genomic DNA ranging from 5 pg – 5000 pg, 10-fold dilution series of plasmid DNA ranging from 0.005pg - 5pg). As different standard series were used for the positive replicate 1 and replicate 2, the data of replicate 2 was scaled to the average of replicate 1 for Figure 4. The raw data for each replicate are depicted in Figure S2B.

Nuclear protein extract generation and DNA pull-down

For collections, asynchronous asexual *P. falciparum* 3D7 cultures were put on ice immediately and filtered over Plasmodipur filters (EuroProxima, Netherlands) to remove human white blood cells. Infected RBCs were washed once in PBS and resuspended in PBS with Protease Inhibitor Cocktail (PI at 1:100, Roche, #04693132001) and 0.05% saponin to a maximum of 6.25% hematocrit for a maximum of 15 minutes. Nuclei were isolated over a double sucrose gradient in CLB with PI (PI at 1:50, bottom layer of 0.25 M sucrose, top layer 0.1 M sucrose) and resuspended in CLB with 20% glycerol, pelleted by centrifugation, snap-frozen and stored at -80 °C until the generation of the

nuclear protein extract. Nuclear protein extract was generated as in [30] with two rounds of extraction in High Salt Extraction Buffer (50 mM HEPES pH7.5, 20% glycerol, 420 mM NaCl, 1.5 mM MgCl₂, 1 mM DTT, 0.4% NP-40, PI). Protein concentration was measured using a Qubit fluorometer (Qubit™ Protein Assay Kit, ThermoFisher Scientific, #Q33212). Nuclear protein extract was snap-frozen in aliquots and stored at -80 °C. Right before the pull-down, nuclear protein extracts were diluted to 0.909 mg/ml protein concentration in 50 mM HEPES pH7.5, 10% glycerol, 150 mM NaCl, 1.5 mM MgCl₂, 1 mM DTT, 0.125% NP-40, PI at 1:25, 9 ng/μl yeast tRNA (Sigma-Aldrich, #R5636), 9 ng/μl poly(dI:dC) (Sigma-Aldrich, #P4929) and 9 ng/μl poly(dA:dT) (Sigma-Aldrich, #P0883). Diluted extracts were spun once at 17000 x g for 25 minutes at 4°C to remove precipitates.

DNA pull downs were performed as in (Hubner et al., 2015; Kensche et al., 2016). Probes for DNA pull downs (ordered from Integrated DNA Technologies, US; Table S7) were dissolved in TE (10 mM Tris, 0.1 mM EDTA, pH 8.0) to 200 μM. 1000 pmoles of biotinylated forward probe was annealed to 1500 pmoles of reverse probe in annealing buffer (10 mM HEPES pH 8.0, 0.05 M NaCl, 1 mM EDTA, in DNase free water). For each pull-down, 50 pmoles of dsDNA probe was coupled to 10 μl of washed sepharose beads slurry (GE Healthcare, #17511301) in DNA Binding Buffer (DBB: 10 mM HEPES pH 8.0, 1 M NaCl, 10 mM EDTA, 0.05% NP-40 in DNase free water) in a total volume of 350 μl while rotating at RT for at least 1 h. Excess probes were removed by two washes with 500 μl DBB and two with 500 μl Protein Binding Buffer* (PBB*: 50 mM HEPES pH 8.0, 150 mM NaCl, 0.1% NP-40, 1 mM DTT, PI at 1:25). After the last wash, PBB* was removed almost completely for each reaction and 550 μl of diluted nuclear protein extract with 500 μg protein content was added and incubated for 1.5 h while rotating at 4 °C. Reactions were spun at 400 x g and supernatants were discarded. Beads (with probes and bound proteins) were then washed by once with 1 ml PBB*, twice with 1 ml PBB (PBB* without PI) and twice with 1 ml Wash Buffer (WB: 50 mM HEPES pH 8.0, 150 mM NaCl). After the last wash with PBB and the washes with WB, supernatants were removed as much as possible. Disulfide bonds were reduced by incubating the beads with 5 mM TCEP (Sigma-Aldrich, #C4706-2G dissolved to 100 mM in mass-spec grade Milli-Q and stored at -20 °C) in 100 mM TEAB (Sigma-Aldrich, #T7408-100 ml) for 1 h in a 37 °C shaking heat block. Beads were briefly spun down and incubated with 10mM (final concentration) of MMTS (Thermo Scientific, #23011, dissolved to 200 mM in isopropanol and stored at -20 °C) to alkylate disulfide bonds in a 37 °C shaking heat block for 10 min. Beads were briefly spun down and 0.4 μg Trypsin/LysC (dissolved to 0.4 μg/ul in Resuspension buffer (50 mM acetic acid (pH<3), Promega, #V5072) was added and incubated for 1 h in a 37 °C shaking heat block. Beads were spun for 1 min at 200 x g at RT and supernatants were collected in a new Eppendorf tube. 50 μl of 100 mM TEAB was added to the beads and these were incubated for another 5 min in a 37 °C shaking heat block and supernatants were added to the previously collected supernatants. Trypsin digestion in the supernatants was continued by overnight incubation in a 37 °C waterbath. Each probe was tested twice per experiment and peptides were labelled by dimethyl labelling (Boersema et al., 2009). NaBH₃CN (Merck, #818053) and CH₂O were used for 'light' labels and NaBD₃CN (Sigma-Aldrich, #190020-1G) and CD₂O for the 'heavy' labels. Labelling reactions were incubated for 1 h at RT while shaking and labelling was stopped by addition of 16 μl of 1% ammonia. Reactions of wild-type and mutated probes with different labels were then pooled and acidified by addition 10 μl of 100% trifluoroacetic acid (TFA, Biosolve BV, the Netherlands, #20234131). Samples were then cleaned and concentrated on C18 stage tips (Rappsilber et al., 2007) and stored at 4 °C until measurement.

QUANTIFICATION AND STATISTICAL ANALYSIS

ATAC-seq data Analysis

The ATAC-seq libraries were sequenced for 75 bp, paired-end on a NextSeq500 system (Illumina) using NextSeq500/550 HighOutput kit V2 (75 cycles) reagents (Illumina). Raw fastq reads were first evaluated using FastQC before continuing ([63] and reads obtained from the two gDNA control

libraries were combined after sequencing. Paired-end libraries were mapped with BWA-mem (version 0.7.10; [64]) against the *P. falciparum* 3D7 reference genome (PlasmoDB release 26; [65, 66]) and filtered for mapping quality ≥ 30 (samtools version 1.3.1; [67]). Duplicate reads were removed using Picard tools (version 1.139; (BroadInstitute)) and reads mapping to the apicoplast and mitochondrial DNA were removed as well as supplementary alignments (FLAG 2048). Finally, an *in silico* size selection was performed to select for read pairs with insert sizes between 50 and 150 bp (or different when indicated) and these libraries were used for further analysis (between 5.9 and 9.7 million reads for replicate 1; between 3.6 and 6.4 million per library for replicate 2; 36.9 million for the merged gDNA control library). For visualization, these libraries were converted to bedgraph files using bedtools genomecov (version 2.20.1; (Quinlan and Hall, 2010)) with the option '-pc' for paired end data and scaled per million reads (RPM). Alternatively, for genomic DNA-corrected tracks, the coverage in each of the t05 to t40 libraries (with an offset of +0.1) was divided by the coverage in the gDNA library (with an offset of +0.1). Bedgraph files were visualized in the UCSC genome browser [70].

Downstream analyses were performed using the data from ATAC-seq replicate 1. Peak calling we used the MACS2 subcommands 'macs2 pileup', 'macs2 bdgcmp' and finally 'macs2 bdgpeakcall' (MACS2 release 2.7; (Liu, 2016)). Because some MACS2 subcommands cannot handle paired end data, we first binned the libraries based on the insert size in steps of 5 bp. Next, the start site of reads aligning to the positive strand were shifted with +4 bp and those aligning to the minus strand with -5 bp to represent the center of the Tn5 transposon binding event as in [31]. Then a pileup track for each (binned) ATAC library was calculated by MACS2 pileup with --extsize set to half the mean insert size. The pileup tracks of the binned libraries were then summed, scaled per million of reads and a pseudocount of 0.1 was added to every position. Regions of local enrichment were identified with macs2 bdgcmp using the gDNA pileup track as background and scored in qvalues (-m qpois). Finally, macs2 bgpeakcall was used to identify regions with qvalue below 0.001 (-c 3.0). To prevent calling many small 'peaks' we allowed regions to be merged when they were within the maximum insert size of 150 bp (-g 150) and we set the minimum length of a peak to 100 bp (-l 100).

Peaks for all time points were merged and the highest scoring summit was selected as summit for the merged peak (4035 merged peaks in total). Peaks with a summit located in a coding region (209 merged peaks) or more than 3kb away from the first/last gene on each chromosome (71 merged peaks) were removed from further analyses. To assign the remaining 3755 intergenic peaks to genes we only selected peaks in intergenic regions flanking the 5' of a gene. For this purpose, intergenic regions (IGs) were categorized based on the flanking coding sequences: 'tandem IGs' are flanked by two genes both in 5'→3' or in 3'←5' orientation; 'divergent IGs' have a downstream gene with 3'←5' orientation and an upstream gene in 5'→3' orientation; 'convergent IGs' have a downstream gene in 5'←3' orientation and the upstream gene in 3'→5' orientation (see Figure 1c). Peaks with their summit located in 'tandem IGs' and 'divergent IGs' (3647 peaks) were assigned to the closest downstream gene using bedtools closest (version 2.20.1; (Quinlan and Hall, 2010)). To calculate accessibility per stage, tags were counted for each of the merged peaks in tandem or divergent IGs. Tag counts were offset by +1 and normalized to the number of reads per kb per million mapped reads (RPKM). For each peak the maximum RPKM value was determined across the stages and peaks with the lowest 10% of maximum values were removed. To correct for signal intensity differences among the time points we normalized the data on quantiles (using the normalized.quantiles command from the R package preprocessCore version 1.36.0; (Bolstad, 2017)). Then, for each peak we calculated the proportion of signal per time point compared to the summed signal over all time points. This proportion-of-sum value was used to calculate the accessibility pattern in each peak region over the time course (t05 – t40).

Directional RNA-sequencing

Strand-specific RNA-seq libraries were sequenced on the Illumina NextSeq 500 system to obtain 75 bp single-end reads (NextSeq500/550 HighOutput kit V2 (75 cycles) reagents (Illumina)). Reads were

evaluated using FastQC ([63] and mapped against the annotated *P. falciparum* 3D7 transcriptome from PlasmoDB release 26 [65, 66] using BWA samse (version 0.7.12-r1039; [72]). Single-end reads were filtered to mapping quality ≥ 15 (samtools version 1.2; [67]) and only uniquely mapped reads (between 9.2 and 11.6 million per library) were used for further analysis. To visualize RNA-seq data in the UCSC Genome browser, 75bp reads were additionally mapped against the annotated *P. falciparum* 3D7 genome from PlasmoDB version 26 [65, 66], filtered for uniquely mapped reads and mapping quality ≥ 15 . Reads were separated according to the strand they mapped to (sense strand FLAG16, antisense strand FLAG0) and normalized to the number of mapped reads per million (RPM). Bedgraph files were generated (version 2.20.1; [69]) and visualized in the UCSC genome browser [70].

To assess RNA abundance per gene, reads mapped against the transcriptome were separated based on alignment to the sense (FLAG 16) or antisense strand (FLAG 0) respectively. Only reads aligning to the sense strand of each transcript were used for further analysis. Tags were counted for all transcripts (excluding mitochondrial RNA and apicoplast RNA) and offset by +1. Transcript counts were normalized to the number of reads per kb per million mapped reads (RPKM) and the maximum RPKM value was determined per transcript. Low abundant transcripts with their maximum RPKM value across the stages were discarded (lowest 10 percentile). Relative transcript abundance to assess stage-specific expression patterns over the time course (t05 – t40) was calculated by dividing the RPKM value of each time point through the sum of RPKM values of all time points (proportion of sum).

Comparison of ATAC-seq and RNA-seq data

To compare accessibility and transcript abundance patterns, accessibility (proportion-of-sum) was clustered using the web-based Morpheus tool from the Broad Institute (BroadInstitute) into eight clusters by k-means clustering with the 1-pearson correlation coefficient as distance metric and 20,000 iterations. Relative transcript abundances (proportion-of-sum) of the downstream gene were then plotted in the same order. Accessibility and mRNA abundance profiles were visualized on a heatmap using color scale covering the 20th to 80th percentile of values. The Pearson correlation coefficient was calculated for each peak-to-transcript pair. Randomized correlations were calculated for 1000 shuffled peak-to-transcript matches.

For co-clustering of accessibility and transcript abundance patterns, peak-to-transcript matches were first filtered for a Pearson correlation coefficient above 0.6. The resulting matrix of accessibility and transcript abundance data (n = 2118 matches) was uploaded in Morpheus and again clustered into 8 k-means with the same settings as before.

Comparison with AP2-I ChIP-sequencing data

For visualization purposes, the AP2-I ChIP-sequencing data from Santos and co-workers was mapped against the *P. falciparum* 3D7 genome with settings as in [17]. In short, reads were trimmed with Trimmomatic (version 0.36; [74]), mapped with BWA-mem (version 0.7.10 [64]) against the *P. falciparum* 3D7 genome (PlasmoDB release 26 [65, 66]) and filtered for not being the primary alignment (FLAG 256), being a duplicate (FLAG 1024), being a supplementary alignment (FLAG 2048) and for mapping quality of 30 and higher. The MACS2 callpeak command was used to generate bedgraph files of the ChIP and input libraries (settings --m 5 50 --extsize 250 --call-summits -B -q 0.05 -g 2.2e7). These bedgraph files were used to make log₂ ChIP-over-input tracks that were uploaded in the UCSC genome browser. Bedtools intersect (version 2.20.1; [69]) was used to define the overlap between the ATAC-seq peaks with the trimmed AP2-I peaks (in 3D7 coordinates) reported in [17]. Fluff was used to generate the heatmap of accessibility over the AP2-I peaks that overlap with ATAC-seq peaks [75]. We used the build-in Gene Ontology tool of PlasmoDB with default settings to identify enriched GO terms [65].

Motif identification and enrichment analyses

For *de novo* motif identification, we used gimme motifs from the GimmeMotifs package (v0.11.0; [39]). Numerous *de novo* motif searches were performed on individual time points and using different number of clusters ranging from 4 to 16 in regions of 200 or 300 bp around the summit. The background consisted of either shuffled peak regions or the other clusters. Searches were performed for large (6 - 15 bp) or xl (6 - 20 bp) motifs. Motifs identified in these various searches were clustered using gimme cluster (-t 0.9) yielding a non-redundant list of *de novo* motifs. To identify motifs differentially enriched in one of the 8k means co-clusters of the ATAC-seq and RNA-seq data compared to other co-clusters, we ran an ensemble of different regression and classification methods, as implemented in GimmeMotifs [39]. As input motif library we used the clustered *de novo* motifs, predicted *Plasmodium* motifs [41, 76], and motifs from plants, vertebrates and invertebrates reported in CISBP [42], motifs from each subgroup were first clustered with gimme cluster at -t 0.9). Gimme maelstrom was ran three times and we selected motifs that had a P value= < 0.01 in at least two runs. From this list we manually removed eight low information content motifs. The remaining 70 motifs were grouped based on their similarities by gimme cluster (-t 0.9), resulting in 41 'motif groups'.

Mass spectrometry and MS data analysis

Loaded C18 stage tips were rehydrated with 25 μ l buffer A (0.1% formic acid) and peptides were eluted in PCR tubes using 30 μ l buffer B (80% acetonitrile, 0.1% formic acid). Acetonitrile was evaporated by a 15 min vacuum spin at room temperature and samples were reconstituted to 12 μ l with buffer A of which 5 μ l measured on a QExactive or Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). In both cases, the sample was separated over a 30cm C18-reverse phase column (1.8 μ m Reprosil-Pur C18-AQ, dr. Maisch 9852) and eluted using an Easy-nLC 1000 (Thermo Fisher Scientific). For the QExactive, elution was performed over a 94 min gradient (5.6% acetonitrile/0.1% formic acid - 25.6% acetonitrile/0.1% formic acid) and directly injected into the mass spectrometer. Data on the QExactive was acquired in TOP10 data-dependent acquisition mode with dynamic exclusion enabled for 20 s. Resolution for MS was set at 70.000 at m/z = 400 and for MS/MS at 17.5000. For the Fusion, elution was performed over a 114 min gradient (7.2% Acetonitrile/0.1% formic acid- 25.6% acetonitrile/0.1% formic acid) and directly injected into the mass spectrometer. Data on the Fusion was acquired in data-dependent top speed mode in a 3 s cycle with dynamic exclusion set at 60 s. Resolution was set at 120.000.

Raw MS spectra were analyzed as in [30] using MaxQuant (version 1.5.3.30, [77]). In short, standard settings were applied with the following modifications. Multiplicity was set at 2, adding a mass of 28.03 Da (light-labelled) or 36.08 Da (heavy-labelled) to the peptides N-terminus and lysine residues. Trypsin/P was set as the specific digestion mode with maximum 2 missed cleavages. Analyses were run with re-quantify set to 'match from and to'. MMTS (added mass of 45.99 Da) was specified as fixed modification of cysteines. The match-between-runs option was activated (with 0.7 min match time window and 20 min alignment time window) and calculation of iBAQ values was enabled. Peptide masses were searched against the *Plasmodium falciparum* 3D7 annotated proteome (PlasmoDB release 9.3 [65, 66]) with the entire human proteome included in the contaminants list using the integrated Andromeda search engine. Mass tolerance was set at 4.5 ppm for precursor ions and 20 ppm for fragment ions, and peptides and proteins were accepted with an 0.01 FDR cut-off. Protein quantification was set to minimally require a single peptide-ratio, but a more stringent downstream filtering on minimally 2 peptides (of which at least 1 unique) was applied for generation of scatterplots and determination of significance.

Downstream analyses were performed using the Perseus software package (version 1.4.0.20, [78]). Normalized H/L-ratios were log₂-transformed and intensity values were log₁₀-transformed. Significant outliers were determined using the intensity-based Significance B option (two-sided Benjamini-Hochberg test) with a FDR cut-off set to 0.05. The protein list was filtered for reverse hits, proteins that are only identified by site and potential contaminants. In addition, proteins

required a minimum of 2 peptides of which at least 1 unique in order to be considered as a hit in both the forward and reverse experiment. Data was plotted in R and significant outliers were labelled. Candidate TFs were retrieved from table 4 in [10] and highlighted as well.

DATA AND SOFTWARE AVAILABILITY

The accession number for the ATAC-seq and RNA-seq data reported in this paper is NCBI Sequence Read Archive, SRA: GSE104075.

References

1. Murray, C.J., et al., *Global malaria mortality between 1980 and 2010: a systematic analysis*. *Lancet*, 2012. **379**(9814): p. 413-31.
2. Cowman, A.F., et al., *Malaria: Biology and Disease*. *Cell*, 2016. **167**(3): p. 610-624.
3. Painter, H.J., T.L. Campbell, and M. Llinas, *The Apicomplexan AP2 family: Integral factors regulating Plasmodium development*. *Molecular and Biochemical Parasitology*, 2011. **176**(1): p. 1-7.
4. Bozdech, Z., et al., *The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum*. *Plos Biology*, 2003. **1**(1): p. 85-100.
5. Caro, F., et al., *Genome-wide regulatory dynamics of translation in the Plasmodium falciparum asexual blood stages*. *Elife*, 2014. **3**.
6. Foth, B.J., et al., *Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum*. *Mol Cell Proteomics*, 2011. **10**(8): p. M110.006411.
7. Voss, T.S., Z. Bozdech, and R. Bartfai, *Epigenetic memory takes center stage in the survival strategy of malaria parasites*. *Curr Opin Microbiol*, 2014. **20**: p. 88-95.
8. Flueck, C., et al., *Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors*. *PLoS Pathog*, 2009. **5**(9): p. e1000569.
9. Salcedo-Amaya, A.M., et al., *Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum*. *Proc Natl Acad Sci U S A*, 2009. **106**(24): p. 9655-60.
10. Bischoff, E. and C. Vaquero, *In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum*. *BMC Genomics*, 2010. **11**: p. 34.
11. Balaji, S., et al., *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*, 2005. **33**(13): p. 3994-4006.
12. Flueck, C., et al., *A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology*. *PLoS Pathog*, 2010. **6**(2): p. e1000784.
13. Iwanaga, S., et al., *Identification of an AP2-family protein that is critical for malaria liver stage development*. *PLoS One*, 2012. **7**(11): p. e47557.
14. Kafsack, B.F.C., et al., *A transcriptional switch underlies commitment to sexual development in malaria parasites*. *Nature*, 2014. **507**(7491): p. 248-+.
15. Kaneko, I., et al., *Genome-Wide Identification of the Target Genes of AP2-O, a Plasmodium AP2-Family Transcription Factor*. *PLoS Pathog*, 2015. **11**(5): p. e1004905.
16. Modrzynska, K., et al., *A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the Plasmodium Life Cycle*. *Cell Host Microbe*, 2017. **21**(1): p. 11-22.

17. Santos, J.M., et al., *Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I Transcription Factor*. *Cell Host & Microbe*, 2017. **21**(6): p. 731-+.
18. Sinha, A., et al., *A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium*. *Nature*, 2014. **507**(7491): p. 253-+.
19. Yuda, M., et al., *Global transcriptional repression: An initial and essential step for Plasmodium sexual development*. *Proc Natl Acad Sci U S A*, 2015. **112**(41): p. 12824-9.
20. Yuda, M., et al., *Identification of a transcription factor in the mosquito-invasive stage of malaria parasites*. *Mol Microbiol*, 2009. **71**(6): p. 1402-14.
21. Zhang, C., et al., *Systematic CRISPR-Cas9-Mediated Modifications of Plasmodium yoelii ApiAP2 Genes Reveal Functional Insights into Parasite Development*. *MBio*, 2017. **8**(6).
22. Militello, K.T., et al., *Identification of regulatory elements in the Plasmodium falciparum genome*. *Mol Biochem Parasitol*, 2004. **134**(1): p. 75-88.
23. Lopez-Estrano, C., et al., *An enhancer-like region regulates hrp3 promoter stage-specific gene expression in the human malaria parasite Plasmodium falciparum*. *Biochimica Et Biophysica Acta-Gene Structure and Expression*, 2007. **1769**(7-8): p. 506-513.
24. Sunil, S., V.S. Chauhan, and P. Malhotra, *Distinct and stage specific nuclear factors regulate the expression of falcipains, Plasmodium falciparum cysteine proteases*. *BMC Mol Biol*, 2008. **9**: p. 47.
25. Gunasekera, A.M., et al., *Regulatory motifs uncovered among gene expression clusters in Plasmodium falciparum*. *Mol Biochem Parasitol*, 2007. **153**(1): p. 19-30.
26. Young, J.A., et al., *In silico discovery of transcription regulatory elements in Plasmodium falciparum*. *BMC Genomics*, 2008. **9**: p. 70.
27. Elemento, O., N. Slonim, and S. Tavazoie, *A universal framework for regulatory element discovery across all genomes and data types*. *Mol Cell*, 2007. **28**(2): p. 337-50.
28. Russell, K., R. Emes, and P. Horrocks, *Triaging informative cis-regulatory elements for the combinatorial control of temporal gene expression during Plasmodium falciparum intraerythrocytic development*. *Parasit Vectors*, 2015. **8**: p. 81.
29. Ponts, N., et al., *Nucleosome landscape and control of transcription in the human malaria parasite*. *Genome Res*, 2010. **20**(2): p. 228-38.
30. Kensche, P.R., et al., *The nucleosome landscape of Plasmodium falciparum reveals chromatin architecture and dynamics of regulatory sequences*. *Nucleic Acids Res*, 2016. **44**(5): p. 2110-24.
31. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position*. *Nat Methods*, 2013. **10**(12): p. 1213-8.
32. Goryshin, I.Y., et al., *Tn5/IS50 target recognition*. *Proc Natl Acad Sci U S A*, 1998. **95**(18): p. 10716-21.
33. Gardner, M.J., et al., *Genome sequence of the human malaria parasite Plasmodium falciparum*. *Nature*, 2002. **419**(6906): p. 498-511.
34. Liu, T. *Advanced: Call peaks using MACS2 subcommands*. <https://github.com/taoliu/MACS.wiki.git> 2016 [cited 2016 19-08-2016]; Available from: <https://github.com/taoliu/MACS.wiki.git>.
35. Sims, J.S., et al., *Patterns of gene-specific and total transcriptional activity during the Plasmodium falciparum intraerythrocytic developmental cycle*. *Eukaryot Cell*, 2009. **8**(3): p. 327-38.
36. Bartfai, R., et al., *H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3*. *PLoS Pathog*, 2010. **6**(12): p. e1001223.

37. Lu, X.M., et al., *Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite Plasmodium falciparum*. *Nucleic Acids Res*, 2017. **45**(13): p. 7825-7840.
38. Brancucci, N.M.B., et al., *Identification of a cis-acting DNA-protein interaction implicated in singular var gene choice in Plasmodium falciparum*. *Cellular Microbiology*, 2012. **14**(12): p. 1836-1848.
39. van Heeringen, S.J. and G.J. Veenstra, *GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments*. *Bioinformatics*, 2011. **27**(2): p. 270-1.
40. Setty, M. and C.S. Leslie, *SeqGL Identifies Context-Dependent Binding Signals in Genome-Wide Regulatory Element Maps*. *PLoS Comput Biol*, 2015. **11**(5): p. e1004271.
41. Campbell, T.L., et al., *Identification and Genome-Wide Prediction of DNA Binding Specificities for the ApiAP2 Family of Regulators from the Malaria Parasite*. *Plos Pathogens*, 2010. **6**(10): p. 15.
42. Weirauch, M.T., et al., *Determination and inference of eukaryotic transcription factor sequence specificity*. *Cell*, 2014. **158**(6): p. 1431-1443.
43. Harris, E.Y., et al., *Chromatin-driven de novo discovery of DNA binding motifs in the human malaria parasite*. *BMC Genomics*, 2011. **12**: p. 601.
44. Ubhe, S., et al., *Genome-wide identification of novel intergenic enhancer-like elements: implications in the regulation of transcription in Plasmodium falciparum*. *BMC Genomics*, 2017. **18**(1): p. 656.
45. Ay, F., et al., *Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression*. *Genome Res*, 2014. **24**(6): p. 974-88.
46. Lemieux, J.E., et al., *Genome-wide profiling of chromosome interactions in Plasmodium falciparum characterizes nuclear architecture and reconfigurations associated with antigenic variation*. *Mol Microbiol*, 2013. **90**(3): p. 519-37.
47. Hasenkamp, S., et al., *Functional analysis of the 5' untranslated region of the phosphoglutamate 2 transcript in Plasmodium falciparum*. *Acta Trop*, 2013. **127**(1): p. 69-74.
48. Xu, J., et al., *Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells*. *Nat Genet*, 2017. **49**(3): p. 377-386.
49. Minoux, M., et al., *Gene bivalency at Polycomb domains regulates cranial neural crest positional identity*. *Science*, 2017. **355**(6332).
50. Hahn, S. and E.T. Young, *Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators*. *Genetics*, 2011. **189**(3): p. 705-36.
51. Bertschi, N.L., et al., *Malaria parasites possess a telomere repeat-binding protein that shares ancestry with transcription factor IIIA*. *Nat Microbiol*, 2017. **2**: p. 17033.
52. Gissot, M., et al., *PfMyb1, a Plasmodium falciparum transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation*. *J Mol Biol*, 2005. **346**(1): p. 29-42.
53. Briquet, S., et al., *High-mobility-group box nuclear factors of Plasmodium falciparum*. *Eukaryot Cell*, 2006. **5**(4): p. 672-82.
54. van Noort, V. and M.A. Huynen, *Combinatorial gene regulation in Plasmodium falciparum*. *Trends Genet*, 2006. **22**(2): p. 73-8.
55. Cobbold, S.A., et al., *Proteome-wide analysis reveals widespread lysine acetylation of major protein complexes in the malaria parasite*. *Scientific Reports*, 2016. **6**: p. 14.
56. Levo, M. and E. Segal, *In pursuit of design principles of regulatory sequences*. *Nat Rev Genet*, 2014. **15**(7): p. 453-68.

57. Josling, G.A., et al., *A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression*. Cell Host Microbe, 2015. **17**(6): p. 741-51.
58. Nkrumah, L.J., et al., *Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase*. Nat Methods, 2006. **3**(8): p. 615-21.
59. Walliker, D., et al., *Genetic analysis of the human malaria parasite Plasmodium falciparum*. Science, 1987. **236**(4809): p. 1661-6.
60. Fraschka, S.A., R.W. Henderson, and R. Bartfai, *H3.3 demarcates GC-rich coding and subtelomeric regions and serves as potential memory mark for virulence gene expression in Plasmodium falciparum*. Sci Rep, 2016. **6**: p. 31965.
61. Fidock, D.A. and T.E. Wellems, *Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil*. Proc Natl Acad Sci U S A, 1997. **94**(20): p. 10931-6.
62. Vos, M.W., et al., *A semi-automated luminescence based standard membrane feeding assay identifies novel small molecules that inhibit transmission of malaria parasites by mosquitoes*. Sci Rep, 2015. **5**: p. 18704.
63. Andrew, S., *FASTQC: a quality control tool for high throughput sequence data Institute*, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2010.
64. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. eprint arXiv:1303.3997v2 2013. **00**(00): p. 3.
65. Aurrecochea, C., et al., *PlasmoDB: a functional genomic database for malaria parasites*. Nucleic Acids Res, 2009. **37**(Database issue): p. D539-43.
66. Logan-Klumpler, F.J., et al., *GeneDB--an annotation database for pathogens*. Nucleic Acids Res, 2012. **40**(Database issue): p. D98-108.
67. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
68. BroadInstitute. *Picard*, <https://broadinstitute.github.io/picard/>. [cited 2015 22-9-2015]; Available from: <https://broadinstitute.github.io/picard/>.
69. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features*. Bioinformatics, 2010. **26**(6): p. 841-2.
70. Kent, W.J., et al., *The human genome browser at UCSC*. Genome Res, 2002. **12**(6): p. 996-1006.
71. Bolstad, B., *preprocessCore: A collection of pre-processing functions. R package version 1.40.0*, <https://github.com/bmbolstad/preprocessCore>. 2017.
72. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform*. Bioinformatics, 2010. **26**(5): p. 589-95.
73. BroadInstitute. *Morpheus*, <https://software.broadinstitute.org/morpheus/>. [cited 2016 28-06-2016]; Available from: <https://software.broadinstitute.org/morpheus/index.html>.
74. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data*. Bioinformatics, 2014. **30**(15): p. 2114-20.
75. Georgiou, G. and S.J. van Heeringen, *fluff: exploratory analysis and visualization of high-throughput sequencing data*. PeerJ, 2016. **4**: p. e2209.
76. De Silva, E.K., et al., *Specific DNA-binding by Apicomplexan AP2 transcription factors*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(24): p. 8393-8398.
77. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nat Biotechnol, 2008. **26**(12): p. 1367-72.
78. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data*. Nat Methods, 2016. **13**(9): p. 731-40.

Supplementary material

SUPPLEMENTARY FIGURES

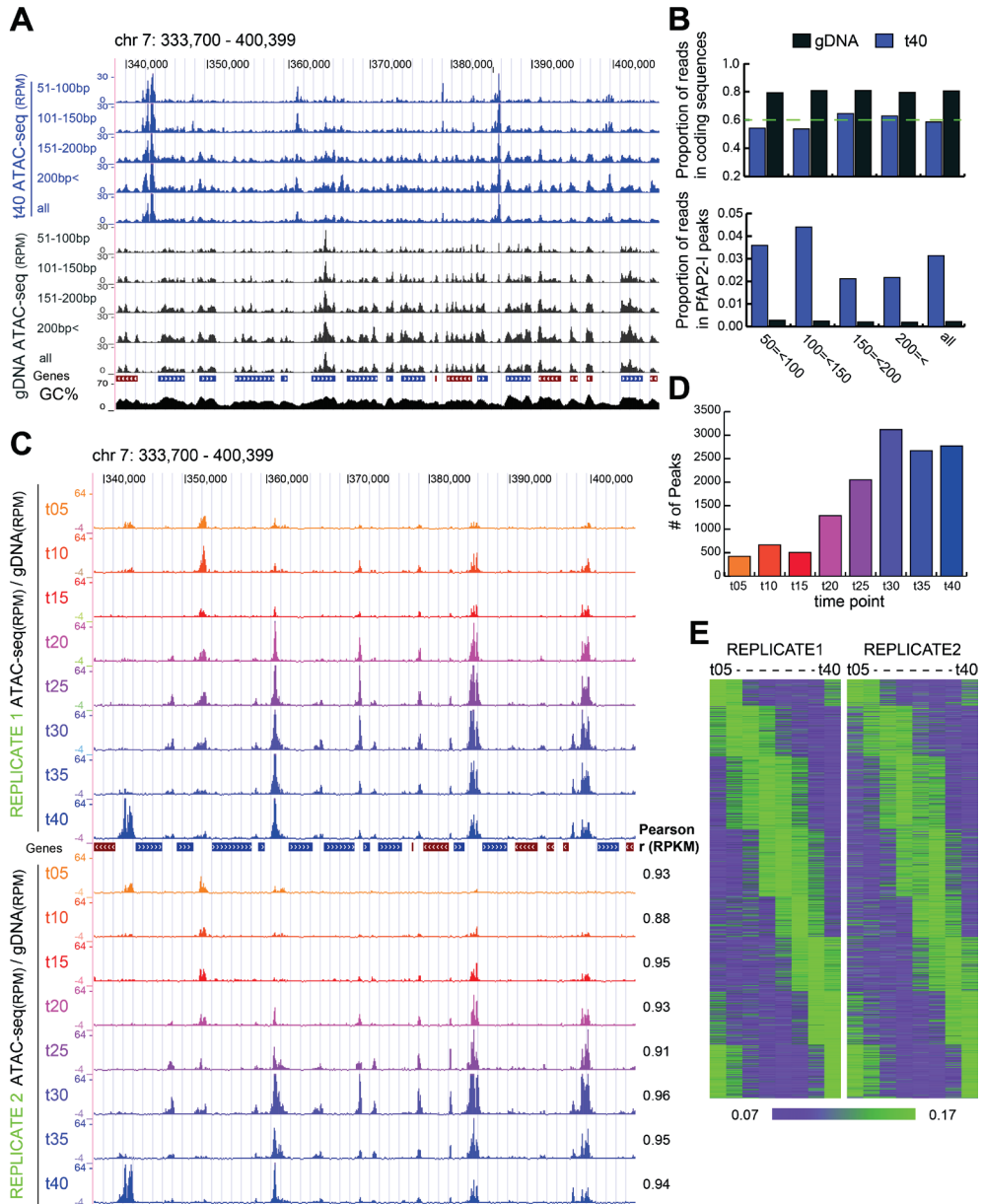


Figure S1 related to figure 1: ATAC-seq data processing and reproducibility.

A) UCSC genome browser screenshot of the region on chromosome 7 (333,700 - 400,399bp) showing libraries of t40 and gDNA without ('all') and after *in silico* selection for the indicated insert size ranges.

Coding sequences are indicated as blue (positive strand) or red (minus strand) bars. GC%, the mean percentage of GC nucleotides, smoothed over 5 bp. **B)** Barplots depict the proportion of reads located in coding sequences (top plot) or in AP2-I peaks (bottom plot) for t40 and gDNA ATAC-seq libraries without ('all') and after *in silico* selection for the indicated insert size ranges. **C)** UCSC genome browser screenshot as in Figure 1A showing the gDNA-corrected ATAC-seq profiles for the two replicates over eight timepoints of blood stage development. **D)** Barplot of the number of peaks as called by MACS2 at each point. **E)** Heatmaps of relative chromatin accessibility in the two ATAC-seq datasets for the regions and clustering depicted in Figure 3B. Relative accessibility was calculated as a proportion-of-sum of qnRPKM values over the time points and clustered by k-means using the 1-Pearson correlation distance metric. Color scales range from the 20th to the 80th percentile for both datasets.

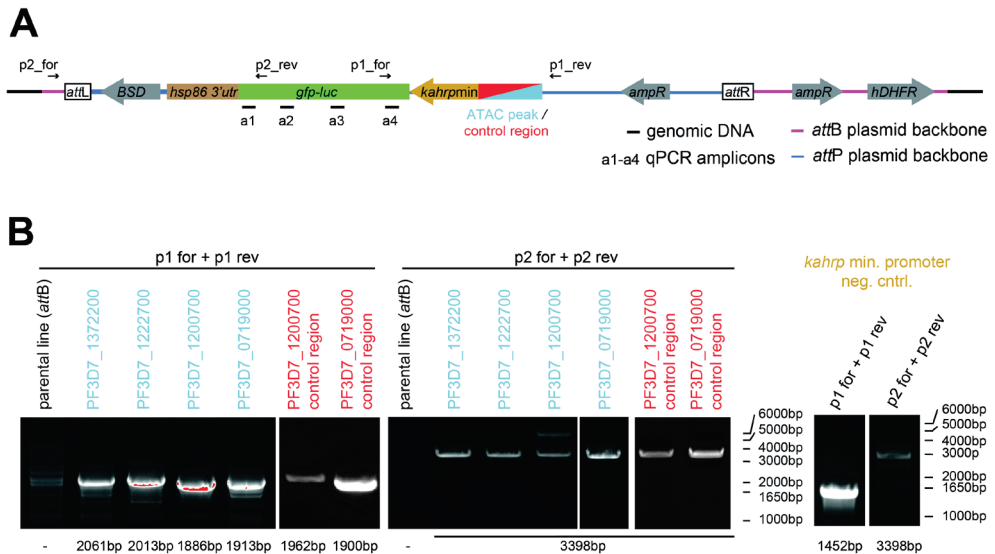
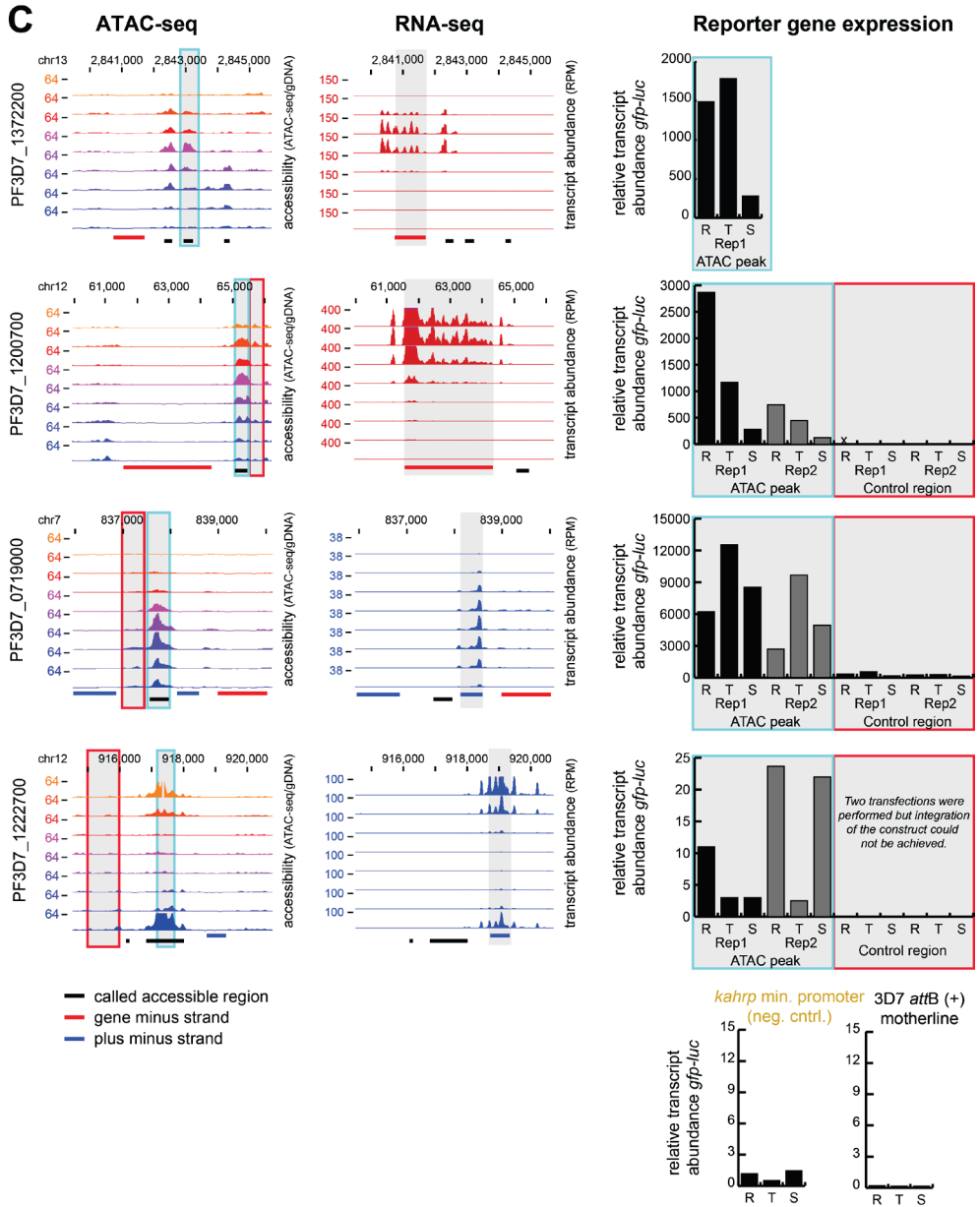


Figure S2 related to figure 4: Reporter constructs containing *gfp/luc* gene under the control of accessible or control regions.

A) Schematic representation of recombinant *cg6* locus (black) in the *attB*-containing parental *P. falciparum* 3D7 parasite line [58] after integration with an *attP*-containing plasmid carrying the *gfp/luc* gene (green) under the control of the *kahrp* minimal promoter (yellow) with or without accessible or control region (light blue/ orange). The *attB* plasmid backbone is indicated as a pink line, the *attP* integration plasmid backbone is blue. The position of primer pairs (p1 forward and reverse covering the accessible region and *kahrp* promoter or p2 forward and reverse covering the *gfp/luc* gene and plasmid backbone) used to confirm the generated parasite lines (see B) are indicated above the drawing. Location of amplicons (a1-4) measured by RT-qPCR (see C) are indicated below. The drawing is not to scale. **B)** Confirmation of insertion of the DNA constructs to the *cg6* locus by PCR. Genomic DNA from *attB*-site containing parental parasites served as negative control. **C)** (next page) UCSC genome browser screenshots showing gDNA-corrected ATAC-seq tracks and directional RNA-seq tracks (left and middle panel) for all eight time points. In the ATAC-seq screenshots, the cloned accessible regions upstream of PF3D7_0719000, PF3D7_1200700, PF3D7_1222700 and PF3D7_1372200 are shaded in grey with a blue rectangle. Selected control regions are highlighted with a dark orange rectangle. In the RNA-seq screenshots the relevant gene is shaded in grey. Black bars represent the called ATAC-seq peaks. Red and blue bars are genes encoded on minus and plus strand. Right panel: Relative *gfp/luc* transcript abundance measured by qRT-PCR (average of amplicon a1 to a4) for ring, trophozoite and schizont stages from parasites containing the accessible (or control) region assigned to the PF3D7_0719000, PF3D7_1200700, PF3D7_1222700 or PF3D7_1372200 gene upstream of the minimal *kahrp* promoter. Note that the replicate 2 schizont sample for PF3D7_1222700 has been collected at a later time point, explaining the higher reporter gene expression observed. The relative *gfp/luc* transcript abundance was assessed by taking the average of the four different primer pairs (see A, a1-4). 'x' depicts missing value.



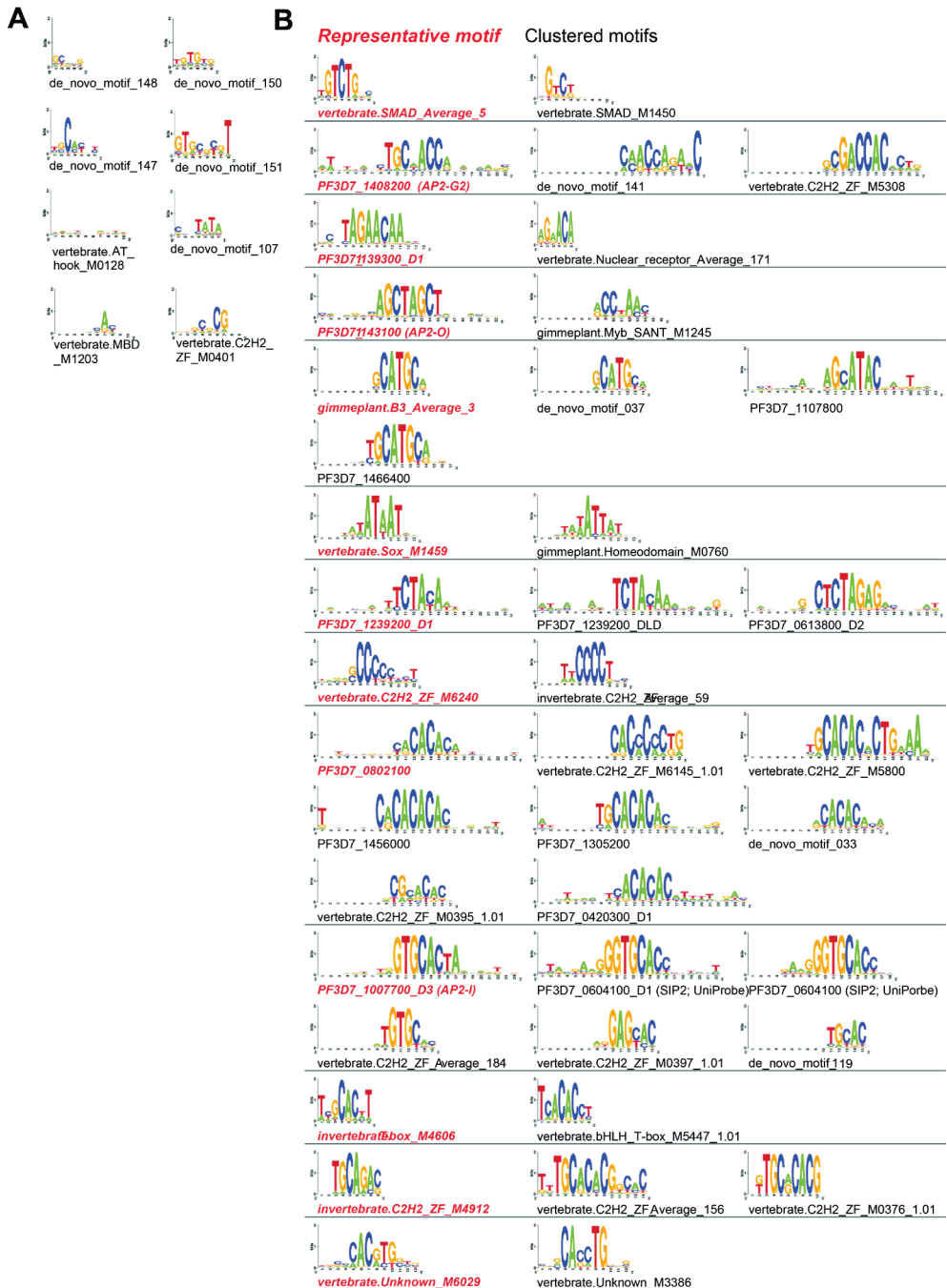


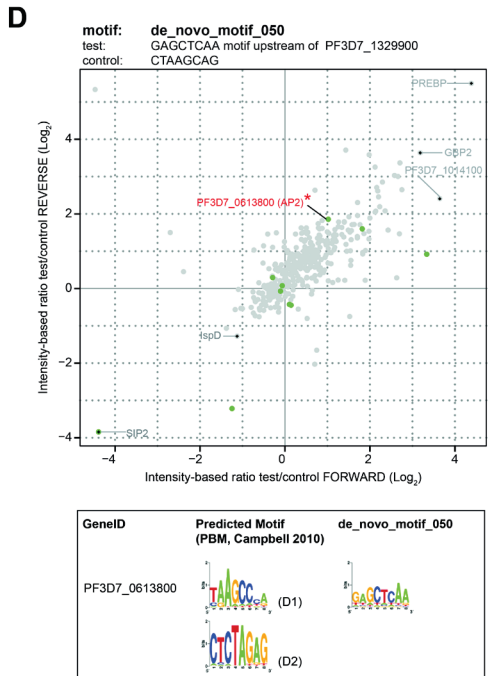
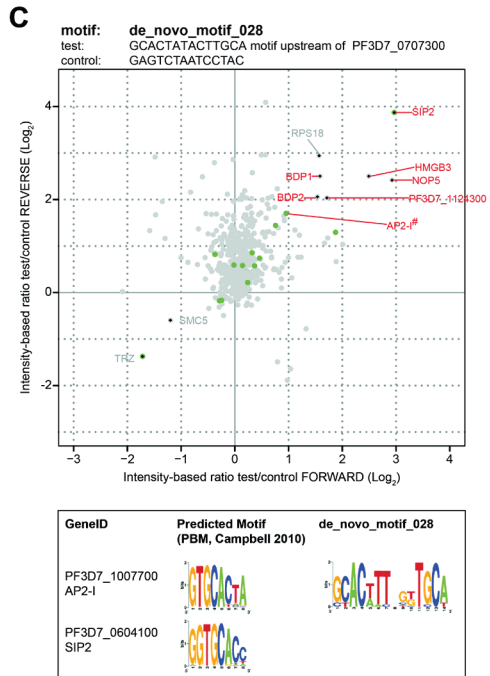
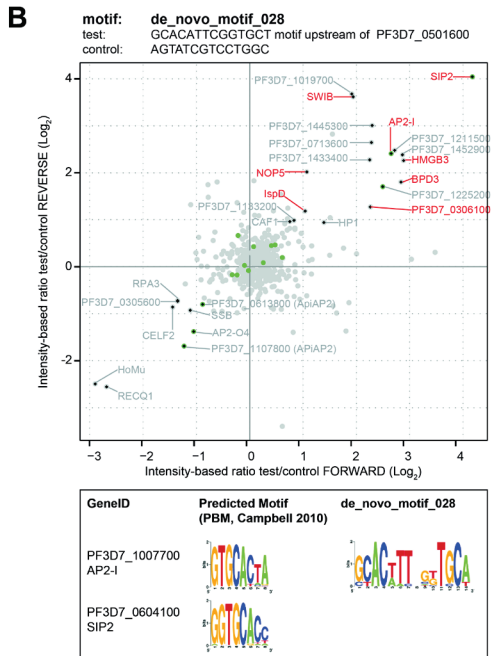
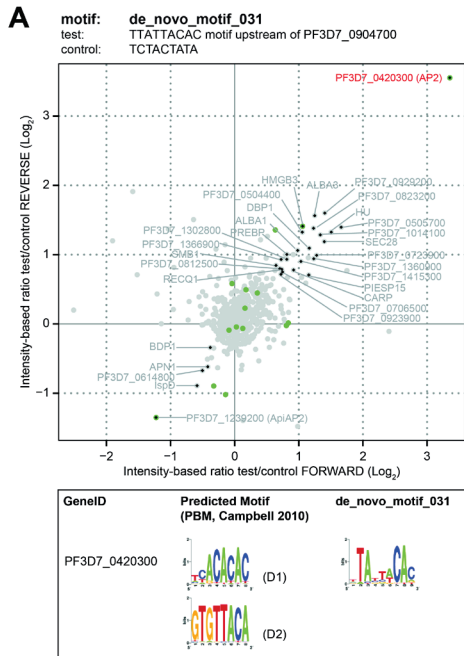
Figure S3 related to figure 5: Identification of putative DNA motifs.

A) List of eight repetitive and/or low-information content motifs excluded from further analysis. B) Motif groups generated by clustering similar motifs found to be differential enriched ($P < 0.01$ from rank aggregation test) for one or more ATAC/RNA-clusters. Representative motifs used in Figure 5 are indicated with red fonts.

Figure S4 related to figure 6: DNA pull downs for identification of potential protein interactors.

(See page 202)

Duplicate DNA pull downs with label swap performed using 60bp DNA probes with the (A) *de_novo_motif_031* (TTATTACAC); (B-C) *de_novo_motif_028* (GCACWWTNNKTGCW) or (D) *de_novo_motif_050* (GAGCTCAA). The same probes with a scrambled motif were used as controls. The statistically significant outliers (black diamond, intensity-based FDR <5%) are the proteins that interact with the motifs. Red font indicates that the interaction was confirmed using a probe from a different genomic region but containing the same motif. # and * indicate that the protein was significantly enriched in either the forward or reverse experiment at an FDR of 0.05 or 0.1, respectively. Green dots mark candidate DNA-binding factors derived from Table 4 of [10].



● Protein with (putative) DNA-binding domain
 ◆ Significant, FDR=0.05
 red = Confirmed with motif from different genomic location
 # or * = Significantly enriched in the forward or reverse experiment (FDR 0.05 or 0.1, respectively)

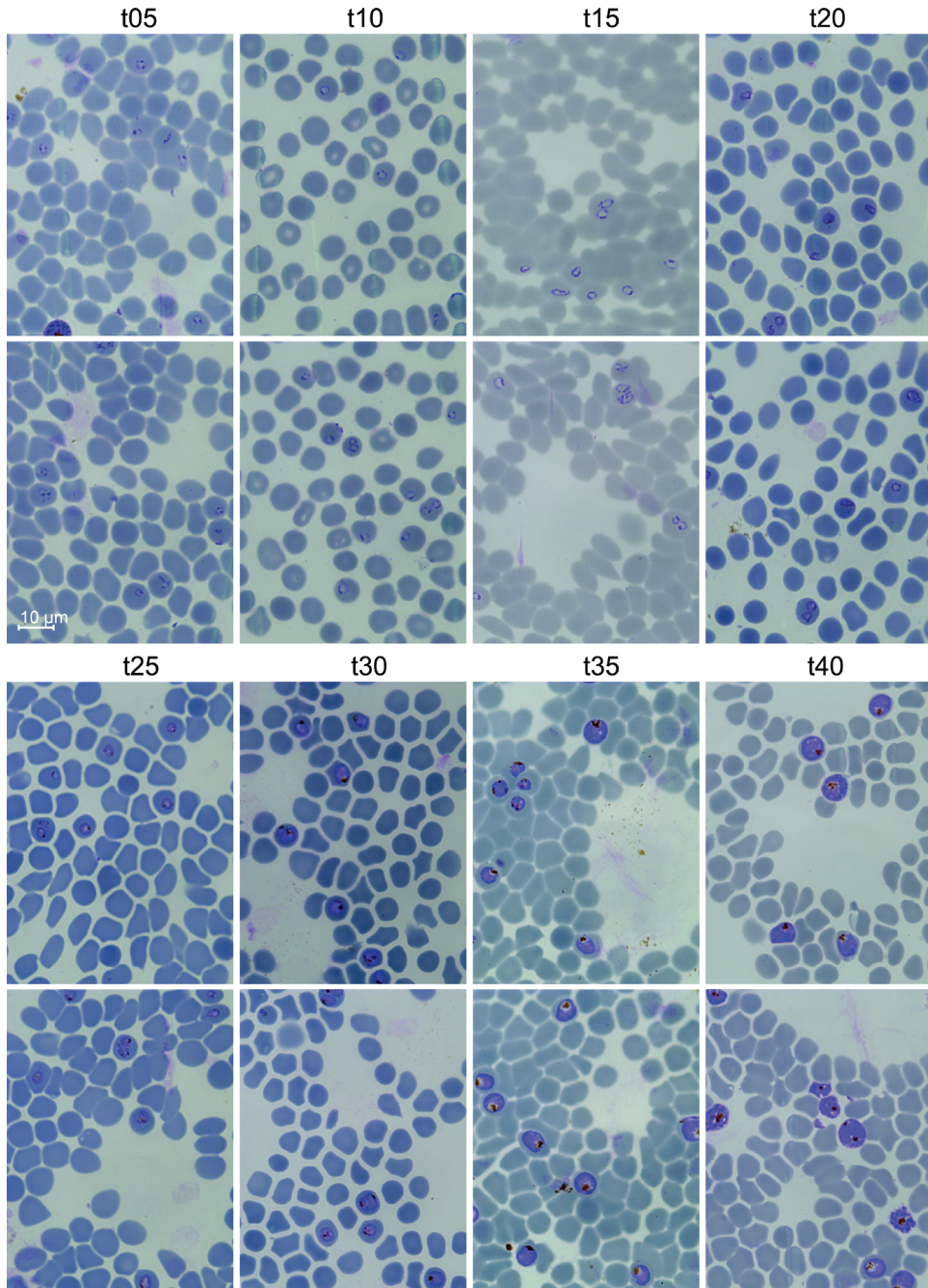


Figure S5 related to parasite culture synchronizations and collections: Representative images of Giemsa smears from time-course experiment (replicate1).

For each time point in the time-course experiment Giemsa stained blood smears were prepared. The two images show the representative parasite morphologies from each time point.

SUPPLEMENTARY TABLES

Table S1-S4, S6 and S7, please download online.

Table S1. Genome coordinates of ATAC peaks. Related to Figure 1, Figure S1 and Figure 5.

Table S2. Fluff heatmap clustering AP2-I peaks using ATAC-seq data. Related to Figure 2.

Table S3. Result of the motif enrichment analysis and associated position weight matrices. Related to Figure 5 and Figure S3.

Table S4. Proteins identified in DNA pull downs. Related to Figure 6 and Figure S4.

Table S5. Parasite and bacterial strains. Related to STAR Methods.

Parasite Strains

| Name | Description | Culture medium |
|--|--|--|
| <i>P. falciparum</i> 3D7 | wild-type <i>P. falciparum</i> strain (Walliker et al., 1987) | normal medium |
| <i>P. falciparum</i> 3D7 attB(+) | <i>P. falciparum</i> 3D7 strain with attB site in <i>cg6</i> locus (Nkrumah et al., 2006) | medium supplemented with 2.6nM WR99210 |
| <i>P. falciparum</i> 3D7 attB::attP_minkahrp | <i>P. falciparum</i> 3D7 attB(+) after transfection and confirmed genomic integration of plasmid attP_minkahrp | medium supplemented with 2.6nM WR99210, 2.5µg/ml Blasticidin S HCl (Gibco, R210-01) and 250µg/ml Geneticin for 7 days after transfection |
| <i>P. falciparum</i> 3D7 attB::attP_minkahrp_PF3D7_1372200 | <i>P. falciparum</i> 3D7 attB(+) after transfection and confirmed genomic integration of plasmid attP_minkahrp_PF3D7_1372200 | medium supplemented with 2.6nM WR99210, 2.5µg/ml Blasticidin S HCl (Gibco, R210-01) and 250µg/ml Geneticin for 7 days after transfection |
| <i>P. falciparum</i> 3D7 attB::attP_minkahrp_PF3D7_1200700 | <i>P. falciparum</i> 3D7 attB(+) after transfection and confirmed genomic integration of plasmid attP_minkahrp_PF3D7_1200700 | medium supplemented with 2.6nM WR99210, 2.5µg/ml Blasticidin S HCl (Gibco, R210-01) and 250µg/ml Geneticin for 7 days after transfection |
| <i>P. falciparum</i> 3D7 attB::attP_minkahrp_PF3D7_0719000 | <i>P. falciparum</i> 3D7 attB(+) after transfection and confirmed genomic integration of plasmid attP_minkahrp_PF3D7_0719000 | medium supplemented with 2.6nM WR99210, 2.5µg/ml Blasticidin S HCl (Gibco, R210-01) and 250µg/ml Geneticin for 7 days after transfection |
| <i>P. falciparum</i> 3D7 attB::attP_minkahrp_PF3D7_1200700negative | <i>P. falciparum</i> 3D7 attB(+) after transfection and confirmed genomic integration of plasmid attP_minkahrp_PF3D7_1200700negative | medium supplemented with 2.6nM WR99210, 2.5µg/ml Blasticidin S HCl (Gibco, R210-01) and 250µg/ml Geneticin for 7 days after transfection |

| | | |
|--|--|--|
| <i>P. falciparum</i> 3D7 <i>attB::attP_minkahrp_</i> PF3D7_0719000negative | <i>P. falciparum</i> 3D7 <i>attB</i> (+) after transfection and confirmed genomic integration of plasmid <i>attP_minkahrp_</i> PF3D7_0719000negative | medium supplemented with 2.6nM WR99210, 2.5µg/ml Blasticidin S HCl (Gibco, R210-01) and 250µg/ml Geneticin for 7 days after transfection |
|--|--|--|

Bacterial strains

| Name | Carried plasmid | Plasmid elements | Comment |
|--|------------------------------------|---|---|
| STBL3_pDC2 | pDC2 [56] | pDC2_3' <i>hsp86-gfp-snf7-5'pfcam-5'pcdt-bsd-3'hrp2-attP-ampR</i> | <i>attP</i> -containing plasmid |
| STBL3_pOM1 | pOM1 | pOM1_5' <i>pfcam-gfp-snf7-3'hsp86-5'pcdt-bsd-3'hrp2-attP-ampR</i> | expression cassette of pDC2 reverted and including <i>ApaI</i> site |
| STBL3_pOM2 | pOM2 | pOM2_5' <i>pfcam-gfp-luc-3'hsp86-5'pcdt-bsd-3'hrp2-attP-ampR</i> | <i>snf7-gfp</i> in pOM1 replaced by <i>gfp-luc</i> |
| XL10-Gold®_attP_minkahrp | <i>attP_min kahrp</i> | pOM2_5' <i>pfkahrp-gfp-luc-3'hsp86-5'pcdt-bsd-3'hrp2-attP-ampR</i> | pOM2 with 5' <i>pfcam</i> replaced by <i>kahrp</i> minimal promoter |
| XL10-Gold®_attP_minkahrp_PF3D7_1372200 | <i>attP_minkahrp_PF3D7_1372200</i> | pOM2_ATAc_1372200-5' <i>pfkahrp-gfp-luc-3'hsp86-5'pcdt-bsd-3'hrp2-attP-ampR</i> | <i>attP_minkahrp</i> with accessible region upstream of PF3D7_1372200 (or replaced by the accessible or control region located upstream of PF3D7_1372200, PF3D7_0719000 or PF3D7_1222700) |
| DH5a_pINT | pINT [56] | pINT-3' <i>hsp86-neoII-5'hsp86-5'PcDT-bxb1myc-3'hrp2-amp</i> | plasmid for expression of <i>mycobacteriophage bxb1 integrase</i> |

Table S6. Parasite staging for time-course experiments (replicate 1 and 2). Related to STAR Methods.

Table S7. Primers and oligos used for cloning, integration PCR, RT-qPCR and DNA pull down experiments. Related to Star Methods.

CHAPTER 7

General discussion

New insights into chromatin regulation of *Plasmodium* parasites

In recent years, significant progress has been made in identifying components of *P. falciparum*'s epigenetic machinery such as post-translational histone modifications (histone PTMs), numerous 'reader', 'writer' and 'eraser' enzymes and long non-coding RNAs (lncRNA) [1-10]. Components of the transcriptional machinery such as homologues of most general transcription factors (TFs) and RNA polymerase II (Pol II) subunits have been identified [11, 12], specific TFs discovered [13, 14] and regulatory sequences located or predicted (e.g. [15-17]). Several studies revealed crucial roles of epigenetic mechanisms in key biological processes such as antigenic variation and regulation of gametocyte conversion (**chapter 2**) as well as the roles of several specific TFs during parasite development [15, 18-25]. Yet, our knowledge about most epigenetic components and/or TFs is still very limited. Accordingly, we are far from understanding their precise function and how they interact to control chromatin-based processes such as gene expression throughout *Plasmodium*'s life cycle.

In the following paragraphs, I will discuss our findings from **chapter 3 to 6** and place them into the context of the state-of-the-art as well as highlight their contribution to our understanding of chromatin regulation in malaria parasites.

Histone variants and their role in chromatin biology of malaria parasites

Up until now, our knowledge about the function of individual components of *P. falciparum*'s epigenetic machinery is limited. The generation of genome-wide profiles can help to link epigenetic components such as histone variants, histone PTMs and their readers to genomic features. If profiled at consecutive time points, their dynamics can additionally reveal their relation to biological processes, such as gene expression or replication and hence help to predict their functionality in these processes.

Chapter 3 describes the genome-wide analysis of histone variant *PfH3.3* and completes our knowledge about the histone variants in the epigenetic landscape of *P. falciparum*. This study revealed that *PfH3.3*-variant nucleosomes primarily occupy euchromatic coding (~76% A+T) and subtelomeric repetitive sequences (~73% A+T), the GC-richest sequences in the extremely AT-biased *P. falciparum* genome (~81% A+T, [26]). Remarkably, intergenic regions, which exhibit a different base pair composition (~86% A+T) are marked by *PfH2A.Z/PfH2B.Z* double-variant nucleosomes and centromeres (~97% A+T) by *PfCenH3* probably as *PfCenH3/PfH2A.Z/PfH2B.Z* triple-variant nucleosomes [27-29]. Together the genome-wide profiles of all four *P. falciparum* histone variants support the idea that the individual histone variants and their combinations index functional distinct genomic domains in a base-composition dependent manner throughout the intraerythrocytic cycle (**Chapter 3**, [27-29]). Yet, *P. falciparum* has by far the most AT-biased genome and of the human-infecting species only closely related species from the same subgenus (*Laverania*)

such *P. reichenowi* (~81%) [30] and *P. gaboni* (~79%) [31] have a so extremely skewed genome composition. Accordingly, our observation that variant-nucleosomes index the genome in an AT-dependent manner might be a *P. falciparum* or *Laverania*-specific adaptation due to their evolving low-complexity genomes [32]. Thus, whether our observation is relevant to less AT-rich genomes of the human-infecting species *P. malariae* (~76%) [33], *P. ovale curtisi/wallikeri* (~71%) [33], *P. knowlesi* (~63%) [34] and *P. vivax* (~57.7%) [35] or to malaria model organisms such as the rodent-infecting species *P. yoelii*, *P. berghei* and *P. chabaudi* (~76-79%) [36] remains to be deciphered. Yet, GC-content differences in genomic regions such as intergenic regions versus coding sequences also occur in less extremely biased *Plasmodium* genomes just not as distinct as in *P. falciparum* [35].

As in other eukaryotes, also in *P. falciparum* variant-containing nucleosomes (likely) possess unique biophysical properties and provide distinct sites for PTMs, which serve as scaffold to recruit specific 'reader', 'writer' and 'eraser' proteins, and contribute to the regulation of chromatin-based processes [6, 37]. For instance, during asexual blood stage development, the histone tails of *PfH2A.Z* and *PfH2B.Z* are heavily acetylated compared to their canonical counterparts [4]. At intergenic regions, which contain regulatory sequences, *PfH2A.Z/PfH2B.Z* incorporation could potentially lead to reduced nucleosome stability/increased chromatin accessibility, which in turn promotes transcription more efficiently and/or be necessary for pre-initiation complex (PIC) recruitment. Interestingly, during blood stage development, these *PfH2A.Z/PfH2B.Z*-containing regions are dynamically marked by various histone modifications which correlate with transcriptional activity of the downstream located gene such as H3K9ac and H4K8ac or are incorporated in a stage-dependent manner such as H3K4me3 [27, 28, 38, 39]. In contrast to *PfH2A.Z* and *PfH2B.Z*, the histone tail of *PfH3.3* is almost identical to its counterpart *PfH3* (**chapter 3**). Yet, *PfH3.3* is mainly incorporated in coding sequences, depleted of the histone variants *PfH2A.Z/PfH2B.Z* and mostly devoid of the PTMs H3K4me3 and H3K9ac. Instead H2BK112ub co-occurs with *PfH3.3*-marked coding regions and correlates with gene transcription (my unpublished observation). Thus, the few differing amino acids between *PfH3.3* and *PfH3* (only eight in the entire sequence) seem to be sufficient to give each of them unique features and functions, potentially through different biophysical properties or distinct sites for PTMs. For instance, S10ph (as several other PTMs) has been solely detected for *PfH3* [40] and its deposition might be hampered in *PfH3.3* by the adjacent amino acid substitution [41]. Further PTM crosstalk and/or recruitment of specific reader complexes could hence promote unique functions of histone variants in processes such as transcription initiation and elongation or gene silencing (e.g. S10ph stimulates Gcn5 to acetylate H3K14 [42] whereas the double modification H3K9me3/S10ph associates with facultative heterochromatin at differentiated postmitotic cells [43])

Next to its incorporation into coding sequences, *PfH3.3* also occurs at the active and poised but not the inactive *var* gene promoter. Interestingly, the active/poised state of a certain *var* gene is maintained over multiple parasite generations. Thus, *PfH3.3* may serve as an epigenetic memory mark in the process of antigenic variation (**chapter 3**). However, how exactly *PfH3.3* incorporation contributes to the complex process of antigenic variation [44] has not been unraveled. Our efforts to tackle this question by generating conditional *PfH3.3* knock-down lines failed, potentially due to the interference of the knock-down tag with the proper function of *PfH3.3*. Yet, alternative approaches like conditional knock-down based on the glmS ribozyme [45] or the TetR-aptamer system [46] could circumvent this problem, and hence shed light on the unique function of *PfH3.3*. For instance, reduced *PfH3.3* levels may lead to an increased *var* gene switching rate and hence confirm its function as memory mark in *var* gene regulation.

Heritable gene-silencing in malaria parasites

In *Plasmodium* parasites, in contrast to their vertebrate hosts, HP1-mediated gene silencing seems to be the main mechanism of heritable silencing since the facultative heterochromatic mark H3K27me3 is absent in *P. falciparum* [4, 47] and the relevance of reported asymmetric DNA methylation [7] has not been established. Yet, other PTMs have been reported to (potentially) contribute to transcriptional repression but may not be involved in heritable gene silencing (e.g. H3K36me2 [47]), or just relevant for a subset of heterochromatic genes (e.g. H3K36me3 [48]).

Chapter 4 describes a very comprehensive ChIP-seq study we performed to identify the genome-wide binding sites of HP1 across different malaria species with the aim to expand our knowledge about conserved or species-specifically silenced genes across six different *Plasmodium* species. Additionally, we investigated variably silenced genes across four *P. falciparum* and two *P. knowlesi* strains as well as the role of HP1-mediated silencing in life cycle progression. Our results for *P. falciparum* were in line with previous findings, which showed that multi-gene family members (often involved in parasite pathogenicity) and few individual genes are targets of HP1-mediated gene silencing in *P. falciparum* schizont stages [49-51]. In our study, we also extended this knowledge to five additional species. Additionally, by exploring different parasite strains of the same species, we revealed that many heterochromatic genes are silenced in a strain-specific manner. Most of these genes were also variably expressed in *P. falciparum* isogenic parasite strains [52] and field isolates [53]. Hence, variant HP1-mediated gene silencing may serve as a mechanism to increase the adaptive capacity, for instance, to host immune responses but also to erythrocytes with different surface receptor repertoires. Previously, it has been shown that *P. falciparum* parasites can use several alternative pathways to invade human erythrocytes, for instance, when cultured with blood of human donors lacking certain surface receptors or with erythrocytes with a restricted surface receptor repertoire due to enzyme

treatment [54]. Thus, genome-wide HP1 profiles generated from parasites cultured in different blood types or differently treated erythrocytes may reveal the epigenetic changes underlying alternative invasion pathways. Indeed, genome-wide HP1 profiles of two *P. knowlesi* clones adapted to different host cells (Macaque versus human blood cells) revealed a dozen of genes which were clearly differently marked by HP1. Subsequent culturing of the human blood-adapted parasite strain in Macaque blood and *vice versa* as performed previously [55] could reveal whether this host cell adaptation has been purely due to epigenetics. In that case, genes important for the adaption to human erythrocytes but unbeneficial for Macaque erythrocytes might be silenced again whereas genes important for the adaption to Macaque erythrocytes would be derepressed and *vice versa*. In addition, genome-wide HP1 profiles from independent *P. knowlesi* strains or isolates could reveal other variantly silenced genes, which could help to uncover epigenetically regulated processes in *P. knowlesi*.

Most strikingly, we discovered that heterochromatin is strictly maintained during asexual proliferation but is extensively remodeled during sexual differentiation of *P. falciparum* parasites. This appears to be crucial for the activation of key gametocyte-specific genes and the adaptation of the erythrocyte remodeling machinery. Interestingly, heterochromatic domains mainly expand leading to an increased number of silenced genes in gametocytes compared to asexual parasites. More recently generated H3K9me3 ChIP-seq data [56] showed that these heterochromatic domains expand even further in *P. falciparum* oocysts and salivary gland sporozoites, implying that an additional subset of genes gets silenced during development in the mosquito vector. Thus, our observation suggests that asexually proliferating blood stages have the most accessible chromatin structure with almost 80% of the genome being transcriptionally active during blood stage development [26, 57] and that the percentage of active genes decreases in succeeding life cycle stages. A reset of the heterochromatic state could potentially occur during liver stage schizogony, which would also be in line with a previous study, suggesting that the mosquito passage may be a prerequisite for the reset of the epigenetic state of virulence genes in *P. chabaudi* [58]. However, further analysis will be needed to investigate when partial or complete resetting of heterochromatin occurs and expansion of our study and the study of Gomez-Diaz and co-workers by HP1 and/or H3K9me3 ChIP-seq on other life cycle stages such as mature liver schizonts and first generation merozoites could clearly contribute to clarify this open question. ChIP-seq analysis of stages which are very difficult to collect in large numbers, such as *P. falciparum* liver stages, could be a particular challenge. However, further processes in the ChIP-seq technology to reduce the starting material and/or improving the signal-to-noise ratio could overcome these challenges. Drop-ChIP [59], for instance, uses a droplet-based microfluidics system [60] to barcode digested chromatin from single cells before immunoprecipitation, which decreases the background noise caused by unspecific

binders. However, further improvements regarding the low sequencing depth/resolution due to the limited number of valid sequencing reads from individual cells, would be needed. The above described gradual expansion of heterochromatic domains results in an increased number of silenced genes throughout life cycle progression, which are likely unbeneficial or even incompatible with the stage-specific requirements. At the same time, few genes, which have been silenced in asexual stages, get de-repressed in these succeeding life cycle stages, likely in an active manner. For instance, *ap2-g*, the gene encoding the master regulator of gametocyte commitment [21, 23, 61]. Thus, independent mechanisms could be responsible for activating and/or silencing these different genes.

Epigenetic regulation of commitment to gametocytogenesis

The *ap2-g* locus is usually silenced in the vast majority of asexual parasites and only transcribed in a small parasite sub-population, which committed to develop to gametocytes [21, 23]. Interestingly, the proportion of these committed parasites/the gametocyte conversion rate can be modulated through environmental conditions such as host lysophosphatidylcholine (LysoPC) levels as recently shown for *P. falciparum* [62]. In **chapter 5** we aimed to decipher the mechanism upstream of AP2-G expression, which resets *ap2-g* heterochromatinization in *P. falciparum* [61]. Interestingly, we revealed GDV1 [63] as upstream activator of sexual differentiation. The expression of GDV1 itself is regulated by a long non-coding *gdv1* antisense RNA and absence of this lncRNA increased GDV1 expression levels. Increased GDV1 levels, in turn, lead to reduced HP1 occupancy at a particular set of genes including *ap2-g*, *dblmsp2* and early gametocyte markers [21, 61, 63, 64] and consequently also to increased transcript levels of these genes. Thus, increased GDV1 levels induce AP2-G expression resulting in increased gametocyte conversion rates in a parasite population. Interestingly, the second highly induced gene *dblmsp2* encodes a merozoite surface protein, which has previously been shown to be only expressed in a small percentage of schizonts [65, 66], which are likely committed, and the expressed protein DBLMSP2 may solely be anchored to the surface of committed merozoites [67]. Interestingly, DBLMSP2 has been shown to be bound by unspecific IgM, which seems to mask parasites from the adaptive host immune system (IgG) [67, 68]. Thus, its expression in committed parasites may contribute to successful transmission and survival of malaria parasites. Notably, orthologues of GDV1 have not been identified in all malaria species, for instance, not in the rodent-infecting species examined in **chapter 4**, which raises the question how HP1 removal of the *ap2-g* locus could occur in these species. Do other proteins undertake the function of GDV1 in rodent species, and if yes which one(s)? The identification of *PbHP1* interacting proteins may reveal interesting candidate proteins. Also, two key questions remained open in regard to GDV1-induced gene activation in *P. falciparum*: How does GDV1 remove/destabilize HP1 at few selected loci? And which upstream events modulate *gdv1* antisense RNA transcription, GDV1 stability or activity?

LysoPC found in human serum represses sexual differentiation [62]. Accordingly, reduced levels of LysoPC and hence of its metabolite PC leads to increased gametocyte commitment rates [62]. But how is this linked to reduced *gdv1* antisense RNA and/or increased GDV1 transcript abundance (the latter has been detected after LysoPC depletion, [62]) and thus in varying gametocyte commitment rates? Unveiling this missing link will not only be a further milestone in understanding the mechanisms of gametocyte commitment in *P. falciparum* and likely in other species with GDV1 orthologues but will also extend our knowledge about the connection of external signals and/or parasite metabolism and epigenetic gene regulation.

Regulation of euchromatic genes in malaria parasites

From previous studies we learned that in *P. falciparum* heterochromatin mediated gene silencing is of major importance ([69], **chapters 2, 4 and 5**) and controls, depending on the species, ~4%-16% of all genes during blood stage development (**chapter 4**). Yet, during the asexual blood stage cycle, the largest part of the parasite genome is in a transcriptionally permissive, euchromatic state, and encompasses roughly 92% of all genes in *P. falciparum* [49, 51]. In *P. falciparum*, typical characteristics of euchromatin are *PfH2A.Z/PfH2B.Z*-double variant nucleosomes at intergenic regions, ‘active’ PTMs such as H3K9ac, H3K4me3, H4K8ac and H2BK112ub which correlate with gene expression and/or promoter strength and general high acetylation levels of the tails of the histones H3 and H4 (unpublished data, [38, 51, 70, 71]). Yet, how exactly epigenetic marks and their ‘reader’, ‘eraser’ and ‘writer’ proteins interact to contribute to transcription regulation in malaria parasites has not been deciphered. Additionally, also in *P. falciparum*, binding of specific *in trans*-acting transcription factors (TFs) [13, 14] to regulatory DNA sequences (e.g. [16, 17, 72]) likely stimulates or inhibits the assembly and/or activity of the preinitiation complex (PIC) at the core (or minimal) promoter [11, 12] as shown for other eukaryotes [73]. Yet, an accurate, genome-wide overview of *cis*-regulatory DNA elements is still lacking for *P. falciparum*. Up until now, we do not know how and which specific *in trans*-acting TFs and *cis*-regulatory elements interact to produce the typical ‘just-in-time’ transcription pattern during blood stage development in *P. falciparum* [74] and to which extent and/or how epigenetic mechanisms contribute to ‘just-in-time’ transcription.

In **chapter 6** we aimed to decipher the gene regulatory networks in *P. falciparum* during blood stage development. By applying ATAC-seq [75], we generated a genome-wide overview of accessible chromatin regions throughout the asexual blood stage cycle. The vast majority of these regions were found in euchromatic intergenic regions located within one or two kilobases upstream of the next downstream located gene. Based on previous observations, which suggested a close proximity of regulatory elements and their target genes (e.g. [15, 16, 20]), we assigned each accessible region to the nearest downstream located gene. Indeed, for the vast majority of genes, chromatin accessibility dynamics and

transcript abundance pattern strongly positively correlated. Additional *in vivo* studies showed on the basis of examples that these identified accessible regions were sufficient to dictate stage-specific gene expression to a reporter gene similar to the expression pattern of the respective endogenous gene. Motif screening within accessible regions clustered according to their stage-specific accessibility/expression profile of the assigned gene uncovered 'stage-specific' *Plasmodium* [17, 76] and *de novo* motifs. Subsequently, a set of identified motifs was used for DNA-pull-downs combined with quantitative proteomics. Specific TFs were found among the motif interacting proteins, suggesting that the discovered *cis-trans* regulatory interactions are likely involved in the 'just-in-time' transcription regulation during blood stage development. Expanding this analysis to the remaining motifs will definitively reveal further *cis-trans* regulatory interactions during asexual proliferation. In addition, applying the same methodology to other life cycle stages such as gametocytes, could highly contribute to unravel yet unknown stage-specific gene transcription regulatory mechanisms.

Furthermore, our data suggest that i) gene expression in *P. falciparum* is predominantly regulated by *cis*-regulatory elements, which are in close proximity to the target gene and ii) gene expression in *P. falciparum* is primarily regulated by transcription activating events. Distant enhancers as key regulators of gene expression as observed in many metazoan [77] are hence rather unlikely. In line with this, long distant interactions were restricted to centromeres, telomeres, ribosomal rDNA genes and virulence gene clusters in previous chromosome conformation studies [78, 79]. The largely activating nature of transcription regulating events is supported by previous TF data. So far, TFs have been shown to primarily act as activators (e.g. AP2-G [21, 23], AP2-O [15, 25] AP2-I [24]) except for AP2-Sp, which may act as both by activating sporozoite genes in *P. berghei* oocysts [80] and potentially repressing female gametocytes-specific genes in *P. berghei* gametocytes or asexual blood stages [22] and AP2-G2, which has been identified as transcriptional repressor in *P. berghei* gametocytes [20, 22].

However, how these stage-specific *cis-trans* regulatory interactions between *cis*-regulatory elements and TFs are related and/or coupled to epigenetic mechanisms has not been deciphered yet. What is the role of histone PTMs and histone variants at and/or next to accessible regions/TF binding sites? In many eukaryotes H3.3/H2A.Z double variant nucleosomes are particularly unstable and additional acetylation of H3.3 at K122, which occurs near the histone-DNA interphase [81], is assumed to contribute to their destabilization at promoter and enhancer elements, thereby facilitating TF binding [82]. PfH2A.Z/PfH2B.Z-double variant nucleosomes found at promoter regions in *P. falciparum* [28] are also rather labile and their heavy acetylation or PTMs such as the recently detected H3K112ac [40] might function in a similar way. In addition, certain PTMs or ncRNAs next to TF binding sites could modulate TF binding affinity or recruit TFs as shown for other eukaryotes [73, 83]. Thus, further investigations of the transcription regulatory network and

the establishment of the link between TF binding events and epigenetic mechanisms will remain one of the major encounters of future malaria research.

Conclusion and Outlook

By combining our novel findings, and integrating them into the latest state-of-the-art regarding gene transcription regulation in *P. falciparum*, the following, yet incomplete, picture emerges:

During blood stage development, the vast majority of genes is expressed in a ‘just-in-time’ manner with peak mRNA abundances correlating with the need for the products they encode for [74]. This ‘just-in-time’ gene expression is regulated by stage-specific TF binding to *cis*-regulatory elements, which are in close proximity to the respective target genes, and lead to largely transcription activating *cis-trans* regulatory interactions (**chapter 6**). The role of the chromatin composition at these *cis*-regulatory elements is still elusive but may influence TF binding by nucleosome stability (e.g. H2A.Z/H2B.Z double-variant nucleosomes) and/or the presence of specific PTMs. In contrast, members of multi-gene families implicated in parasite virulence and immune evasion as well as few individual genes including *ap2-g* are heritably silenced via HP1 during asexual proliferation (**chapter 2 and 4**). Nonetheless, the chromatin state of some of these genes can change during asexual proliferation, for instance, in case of the *var* gene family members. Time after time, a heritably silenced *var* gene is activated, while the previously activated one is silenced, and remains active for multiple generations (**chapter 2**). The incorporation of the histone variant H3.3 into the *var* gene promoter likely contributes to the propagation of this active state to the next generations (**chapter 3**). Frequently, few parasites escape the asexual proliferation cycle and commit to gametocytogenesis (**chapter 5**, [84]). This cell fate decision is likely stochastic, but also responsive to environmental factors [62, 84] and linked to *gdv1* levels or activity by a yet unknown mechanism. Decreased *gdv1* antisense RNA levels induce GDV1 expression and derepress *ap2-g* (**chapter 5**). This leads to transcription of the TF AP2-G, which in turn induces a cascade of downstream regulatory events leading to sexual differentiation [21, 23]. These downstream regulatory events are accompanied by extensive chromatin remodeling. Early gametocyte markers, which have been heritably silenced in asexual parasites, get activated and expressed in gametocytes (**chapter 4**). At the same time, heterochromatic domains expand and lead to silencing of previously active genes, which are potentially needed for proper erythrocyte remodeling during asexual proliferation (**chapter 4**). Yet, the link between heterochromatin formation and AP2-G induced gametocytogenesis remains to be deciphered. Moreover, heterochromatic domains are even further extended in oocysts and salivary gland sporozoites (**chapter 4**, [56]), giving the impression that heterochromatin continuously spreads from asexual blood stages to liver stages. But why and how, remains to be unveiled.

Taken together, these new insights gained in **chapter 3 to 6** contribute to our understanding of the process of gene regulation in *Plasmodium* parasites. Still, many questions remain unanswered, for instance, what is the precise function of other epigenetic components, regulatory sequences and specific TFs? What role do epigenetic mechanisms play in malaria parasites during transcription regulating and other chromatin-based processes such as DNA replication? How do individual TFs interact to form a regulatory network and which regulatory events occur during, for instance, asexual proliferation, stage transition or response to external stimuli? Thus, understanding these individual mechanisms, linking them and integrating the gained knowledge into a comprehensive model of chromatin regulation/transcription regulation in malaria parasites is one of the major challenges for the upcoming years.

References

1. Lasonder, E., et al., Insights into the *Plasmodium falciparum* schizont phospho-proteome. *Microbes Infect*, 2012. 14(10): p. 811-9.
2. Miao, J., et al., The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene*, 2006. 369: p. 53-65.
3. Treck, M., et al., The phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe*, 2011. 10(4): p. 410-9.
4. Trelle, M.B., et al., Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*. *J Proteome Res*, 2009. 8(7): p. 3439-50.
5. Ponts, N., et al., Unraveling the ubiquitome of the human malaria parasite. *J Biol Chem*, 2011. 286(46): p. 40320-30.
6. Cui, L. and J. Miao, Chromatin-mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*. *Eukaryot Cell*, 2010. 9(8): p. 1138-49.
7. Ponts, N., et al., Genome-wide mapping of DNA methylation in the human malaria parasite *Plasmodium falciparum*. *Cell Host Microbe*, 2013. 14(6): p. 696-706.
8. Broadbent, K.M., et al., A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol*, 2011. 12(6): p. R56.
9. Broadbent, K.M., et al., Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC Genomics*, 2015. 16: p. 454.
10. Amit-Avraham, I., et al., Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. *Proc Natl Acad Sci U S A*, 2015. 112(9): p. E982-91.
11. Callebaut, I., et al., Prediction of the general transcription factors associated with RNA polymerase II in *Plasmodium falciparum*: conserved features and differences relative to other eukaryotes. *BMC Genomics*, 2005. 6: p. 100-100.
12. Coulson, R.M.R., N. Hall, and C.A. Ouzounis, Comparative Genomics of Transcriptional Control in the Human Malaria Parasite *Plasmodium falciparum*. *Genome Research*, 2004. 14(8): p. 1548-1554.
13. Balaji, S., et al., 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*, 2005. 33(13): p. 3994-4006.

14. Bischoff, E. and C. Vaquero, In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics*, 2010. 11: p. 34-34.
15. Kaneko, I., et al., Genome-Wide Identification of the Target Genes of AP2-O, a *Plasmodium* AP2-Family Transcription Factor. *PLoS Pathogens*, 2015. 11(5): p. e1004905.
16. López-Estraño, C., et al., An enhancer-like region regulates hrp3 promoter stage-specific gene expression in the human malaria parasite *Plasmodium falciparum*. *Biochim Biophys Acta*, 2007. 1769(7-8): p. 506-13.
17. Campbell, T.L., et al., Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog*, 2010. 6(10): p. e1001165.
18. Flueck, C., et al., A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog*, 2010. 6(2): p. e1000784.
19. Iwanaga, S., et al., Centromere plasmid: a new genetic tool for the study of *Plasmodium falciparum*. *PLoS One*, 2012. 7(3): p. e33326.
20. Yuda, M., et al., Global transcriptional repression: An initial and essential step for *Plasmodium* sexual development. *Proceedings of the National Academy of Sciences of the United States of America*, 2015. 112(41): p. 12824-12829.
21. Kafsack, B.F., et al., A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, 2014. 507(7491): p. 248-52.
22. Modrzynska, K., et al., A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the *Plasmodium* Life Cycle. *Cell Host & Microbe*, 2017. 21(1): p. 11-22.
23. Sinha, A., et al., A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature*, 2014. 507(7491): p. 253-7.
24. Santos, J.M., et al., Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I Transcription Factor. *Cell Host & Microbe*, 2017. 21(6): p. 731-741.e10.
25. Yuda, M., et al., Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol*, 2009. 71(6): p. 1402-14.
26. Gardner, M.J., et al., Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 2002. 419(6906): p. 498-511.
27. Bártfai, R., et al., H2A.Z demarcates intergenic regions of the *Plasmodium falciparum* epigenome that are dynamically marked by H3K9ac and H3K4me3. *PLoS Pathog*, 2010. 6(12): p. e1001223.
28. Hoeijmakers, W.A., et al., H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the *Plasmodium falciparum* genome. *Mol Microbiol*, 2013. 87(5): p. 1061-73.
29. Hoeijmakers, W.A., et al., *Plasmodium falciparum* centromeres display a unique epigenetic makeup and cluster prior to and during schizogony. *Cell Microbiol*, 2012. 14(9): p. 1391-401.
30. Otto, T.D., et al., Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat Commun*, 2014. 5: p. 4754.
31. Sundararaman, S.A., et al., Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun*, 2016. 7: p. 11078.
32. Hamilton, W.L., et al., Extreme mutation bias and high AT content in *Plasmodium falciparum*. *Nucleic Acids Research*, 2017. 45(4): p. 1889-1901.
33. Rutledge, G.G., et al., *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature*, 2017. 542(7639): p. 101-104.
34. Pain, A., et al., The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*, 2008. 455(7214): p. 799-803.

35. Carlton, J.M., et al., Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*, 2008. 455(7214): p. 757-763.
36. Otto, T.D., et al., A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol*, 2014. 12: p. 86.
37. Doerig, C., et al., Post-translational protein modifications in malaria parasites. *Nat Rev Microbiol*, 2015. 13(3): p. 160-72.
38. Gupta, A.P., et al., Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite *Plasmodium falciparum*. *PLoS Pathog*, 2013. 9(2): p. e1003170.
39. Petter, M., et al., H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite *Plasmodium falciparum*. *Mol Microbiol*, 2013. 87(6): p. 1167-82.
40. Coetzee, N., et al., Quantitative chromatin proteomics reveals a dynamic histone post-translational modification landscape that defines asexual and sexual *Plasmodium falciparum* parasites. *Scientific Reports*, 2017. 7: p. 607.
41. Sullivan, W.J., Histone H3 and H3.3 variants in the protozoan pathogens *Plasmodium falciparum* and *Toxoplasma gondii*. *DNA Seq*, 2003. 14(3): p. 227-31.
42. Lo, W.-S., et al., Phosphorylation of Serine 10 in Histone H3 Is Functionally Linked In Vitro and In Vivo to Gcn5-Mediated Acetylation at Lysine 14. *Molecular Cell*, 2000. 5(6): p. 917-926.
43. Sabbattini, P., et al., A novel role for the Aurora B kinase in epigenetic marking of silent chromatin in differentiated postmitotic cells. *The EMBO Journal*, 2007. 26(22): p. 4657-4669.
44. Deitsch, K.W. and R. Dzikowski, Variant Gene Expression and Antigenic Variation by Malaria Parasites. *Annual Review of Microbiology*, 2017. 71(1): p. 625-641.
45. Prommana, P., et al., Inducible knockdown of *Plasmodium* gene expression using the glmS ribozyme. *PLoS One*, 2013. 8(8): p. e73783.
46. Goldfless, S.J., J.C. Wagner, and J.C. Niles, Versatile control of *Plasmodium falciparum* gene expression with an inducible protein-RNA interaction. *Nature communications*, 2014. 5: p. 5329-5329.
47. Karmodiya, K., et al., A comprehensive epigenome map of *Plasmodium falciparum* reveals unique mechanisms of transcriptional regulation and identifies H3K36me2 as a global mark of gene suppression. *Epigenetics & Chromatin*, 2015. 8: p. 32.
48. Jiang, L., et al., PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature*, 2013. 499(7457): p. 223-7.
49. Flueck, C., et al., *Plasmodium falciparum* Heterochromatin Protein 1 Marks Genomic Loci Linked to Phenotypic Variation of Exported Virulence Factors. *PLoS Pathog*, 2009. 5(9): p. e1000569.
50. Lopez-Rubio, J.J., et al., 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol*, 2007. 66(6): p. 1296-305.
51. Salcedo-Amaya, A.M., et al., Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A*, 2009. 106(24): p. 9655-60.
52. Rovira-Graells, N., et al., Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res*, 2012. 22(5): p. 925-38.
53. Mok, S., et al., Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science (New York, N.Y.)*, 2015. 347(6220): p. 431-435.
54. Wright, G.J. and J.C. Rayner, *Plasmodium falciparum* erythrocyte invasion: combining function with immune evasion. *PLoS Pathog*, 2014. 10(3): p. e1003943.

-
55. Moon, R.W., et al., Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc Natl Acad Sci U S A*, 2013. 110(2): p. 531-6.
 56. Gómez-Díaz, E., et al., Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression during development in *Anopheles gambiae*. *Sci Rep*, 2017. 7: p. 40655.
 57. Siegel, T.N., et al., Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. *BMC Genomics*, 2014. 15: p. 150.
 58. Spence, P.J., et al., Vector transmission regulates immune control of *Plasmodium* virulence. *Nature*, 2013. 498(7453): p. 228-231.
 59. Rotem, A., et al., Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol*, 2015. 33(11): p. 1165-72.
 60. Mazutis, L., et al., Single-cell analysis and sorting using droplet-based microfluidics. *Nature protocols*, 2013. 8(5): p. 870-891.
 61. Brancucci, N.M., et al., Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe*, 2014. 16(2): p. 165-76.
 62. Brancucci, N.M.B., et al., Lysophosphatidylcholine Regulates Sexual Stage Differentiation in the Human Malaria Parasite *Plasmodium falciparum*. *Cell*, 2017. 171(7): p. 1532-1544.e15.
 63. Eksi, S., et al., *Plasmodium falciparum* gametocyte development 1 (*Pfgdv1*) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS Pathog*, 2012. 8(10): p. e1002964.
 64. Eksi, S., et al., Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum*. *Mol Biochem Parasitol*, 2005. 143(1): p. 90-9.
 65. Amambua-Ngwa, A., et al., Population Genomic Scan for Candidate Signatures of Balancing Selection to Guide Antigen Characterization in Malaria Parasites. *PLoS Genetics*, 2012. 8(11): p. e1002992.
 66. Tetteh, K.K.A., et al., Analysis of Antibodies to Newly Described *Plasmodium falciparum* Merozoite Antigens Supports MSPDBL2 as a Predicted Target of Naturally Acquired Immunity. *Infection and Immunity*, 2013. 81(10): p. 3835-3842.
 67. Crosnier, C., et al., Binding of *Plasmodium falciparum* Merozoite Surface Proteins DBLMSP and DBLMSP2 to Human Immunoglobulin M Is Conserved among Broadly Diverged Sequence Variants. *The Journal of Biological Chemistry*, 2016. 291(27): p. 14285-14299.
 68. Barfod, L., et al., Evasion of immunity to *Plasmodium falciparum* malaria by IgM masking of protective IgG epitopes in infected erythrocyte surface-exposed PfEMP1. *Proc Natl Acad Sci U S A*, 2011. 108(30): p. 12485-90.
 69. Voss, T.S., Z. Bozdech, and R. Bártfai, Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr Opin Microbiol*, 2014. 20: p. 88-95.
 70. Bártfai, R., et al., H2A.Z Demarcates Intergenic Regions of the *Plasmodium falciparum* Epigenome That Are Dynamically Marked by H3K9ac and H3K4me3. *PLoS Pathog*, 2010. 6(12): p. e1001223.
 71. Westenberger, S.J., et al., Genome-wide nucleosome mapping of *Plasmodium falciparum* reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. *BMC Genomics*, 2009. 10: p. 610.
 72. Russell, K., R. Emes, and P. Horrocks, Triaging informative cis-regulatory elements for the combinatorial control of temporal gene expression during *Plasmodium falciparum* intraerythrocytic development. *Parasites & Vectors*, 2015. 8(1): p. 81.
 73. Lelli, K.M., M. Slattery, and R.S. Mann, Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. *Annual review of genetics*, 2012. 46: p. 43-68.

74. Bozdech, Z., et al., The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*. *PLoS Biology*, 2003. 1(1): p. e5.
75. Buenrostro, J.D., et al., Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. *Nature methods*, 2013. 10(12): p. 1213-1218.
76. Weirauch, Matthew T., et al., Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell*, 2014. 158(6): p. 1431-1443.
77. Kolovos, P., et al., Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin*, 2012. 5: p. 1-1.
78. Ay, F., et al., Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res*, 2014. 24(6): p. 974-88.
79. Lemieux, J.E., et al., Genome-wide profiling of chromosome interactions in *Plasmodium falciparum* characterizes nuclear architecture and reconfigurations associated with antigenic variation. *Molecular Microbiology*, 2013. 90(3): p. 519-537.
80. Yuda, M., et al., Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol Microbiol*, 2010. 75(4): p. 854-63.
81. Rothbart, S.B. and B.D. Strahl, Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta*, 2014. 1839(8): p. 627-43.
82. Pradeepa, M.M., Causal role of histone acetylations in enhancer function. *Transcription*, 2017. 8(1): p. 40-47.
83. Takemata, N. and K. Ohta, Role of non-coding RNA transcription around gene regulatory elements in transcription factor recruitment. *RNA Biology*, 2017. 14(1): p. 1-5.
84. Josling, G.A. and M. Llinás, Sexual development in *Plasmodium* parasites: knowing when it's time to commit. *Nat Rev Microbiol*, 2015. 13(9): p. 573-87.

SUMMARY

Malaria is one of the deadliest infectious diseases claiming close to half a million human lives annually despite extensive elimination efforts. The causal agent are single-celled parasites of the genus *Plasmodium*. The life cycle of these parasites involves a mosquito vector and a vertebrate host. In both hosts the parasite develops into various morphologically and functionally distinct forms and infects and grows within very different cell types and tissues to complete the life cycle. A prerequisite for these complex environmental adaptations and phenotypic changes are molecular mechanisms acting on the chromatin, the highly organized structure that compacts the DNA through its interaction with various proteins and RNA. These so-called epigenetic mechanisms mediate the composition and compaction of the chromatin and hence influence essential biological processes including gene transcription, the process, in which DNA is transcribed into RNA, the template for subsequent protein synthesis.

In this thesis, I aim to contribute to a better understanding of epigenetic regulatory mechanisms as well as transcriptional gene regulation in malaria parasites. In **chapter 1**, I introduce malaria parasites and the disease they cause as well as chromatin-regulatory mechanisms in eukaryotic organisms. An overview about the different components of the epigenetic machinery in malaria parasites as well as key *Plasmodium*-specific examples of epigenetic regulation are provided in **chapter 2**. One of these epigenetic components, the histone variant *PfH3.3*, is characterized in **chapter 3**, where we uncovered its genome-wide localization to GC-rich regions as well as its potential role in the process of antigenic variation in the human-infecting species *P. falciparum*. **Chapter 4** focuses on the heterochromatic mark HP1, which we profiled in different *Plasmodium* species, strains and stages to provide a detailed catalog of heritably silenced genes. In addition, we revealed conserved and specialized features of HP1-mediated gene silencing across the genus *Plasmodium*, explored variably silenced genes across different strains and unveiled a major role for HP1-mediated gene silencing in the process of sexual differentiation. In **chapter 5** we unveiled the initial regulatory events that lead to sexual commitment in *P. falciparum*, namely the activation of the master transcription factor *PfAP2-G*. We showed that binding of the nuclear protein *PfGDV1* to the heterochromatic mark *PfHP1* at the *pfap2-g* locus counteracts HP1-mediated silencing and induces gametocyte conversion. The expression of *PfGDV1* is regulated by its own antisense RNA. **Chapter 6** addresses the transcriptional regulatory network in *P. falciparum* during blood stage development. Here, we revealed *cis*-regulatory elements in a genome-wide scale and investigated their role in relation to transcription. We identified 'stage-specific' novel and previously identified (*Plasmodium*) motifs and uncovered yet unknown *cis-trans* regulatory interactions. In **Chapter 7**, I discuss our findings and how they contribute to a better understanding of the biology of malaria parasites.

SAMENVATTING

Malaria is één van de meest dodelijke infectieziekten en is verantwoordelijk voor bijna een half miljoen mensenlevens per jaar, ondanks verwoede preventie- en eliminatiepogingen. De ziekte wordt veroorzaakt door eencellige parasieten van het geslacht *Plasmodium*. De levenscyclus van deze parasieten vindt deels plaats in een muggenvector en deels in een gewervelde gastheer. In beide gastheren ontwikkelt de parasiet zich in morfologisch en functioneel verschillende vormen, en infecteert en groeit hij in verschillende celtypen en weefsels. Om zich aan te passen aan deze veranderende omgeving maakt de parasiet onder andere gebruik van moleculaire mechanismen die werken op het chromatine, een zeer georganiseerde structuur die DNA verpakt door zijn interactie met verschillende eiwitten en RNA. Deze zogenoemde epigenetische mechanismen mediëren de samenstelling en verpakking van het chromatine en beïnvloeden zo essentiële biologische processen waaronder transcriptie, het proces waarin DNA wordt overgeschreven tot RNA, een blauwdruk voor de navolgende eiwitsynthese.

In deze scriptie is het mijn doel bij te dragen aan een beter begrip van de epigenetische regulatie mechanismen alsmede van transcriptionele genregulatie in malariaparasieten. In **hoofdstuk 1** introduceer ik de malariaparasieten en de ziekte die zij veroorzaken, evenals de chromatine-gereguleerde mechanismen in eukaryoten. Een overzicht van de verschillende componenten in de epigenetische machinerie van malariaparasieten als wel enkele belangrijke *Plasmodium* specifieke voorbeelden van epigenetische regulatie worden besproken in **hoofdstuk 2**. Één van deze componenten, histon variant *PfH3.3*, wordt gekarakteriseerd in **hoofdstuk 3**, waar we zijn genoomwijde lokalisatie in GC-rijke gebieden als wel zijn potentiële rol in het proces van antigene variatie in de mens-infecterende soort *P. falciparum* aan het licht brengen. **Hoofdstuk 4** focust op de heterochromatine markering HP1, die wij hebben geprofileerd in verschillende *P. falciparum* soorten, -stammen en -stadia, om een gedetailleerd overzicht te maken van genen waarvan de transcriptierepressie erfelijk wordt doorgegeven. Daarnaast hebben wij geconserveerde en gespecialiseerde kenmerken van HP1-gemedieerde repressie over het gehele geslacht *Plasmodium* aangetoond, genen onderzocht die tussen verschillende stammen verschillen in hun repressie of activatie staan en een belangrijke rol voor HP-1 gemedieerde transcriptie repressie in het proces van seksuele differentiatie aangetoond. In **hoofdstuk 5** onthullen we de initiële regulatoire gebeurtenissen die ervoor zorgen dat aseksuele *P. falciparum* parasieten gaan differentiëren naar seksuele stadia, namelijk de activatie van de transcriptiefactor *PfAP2-G*. We tonen aan dat de binding van het nucleaire eiwit *PfGDV1* aan de heterochromatine marker *PfHP1* op de *pfap2-g* locus HP-1 gemedieerde repressie tegenwerkt en gametocyte conversie induceert. De expressie van *PfGDV1* is gereguleerd door zijn eigen antisense RNA. In **hoofdstuk 6** bespreken we het transcriptionele regulatoire netwerk in *P. falciparum* tijdens de bloedstadium ontwikkeling. Hier onthullen we *cis-*

regulatoire elementen op een genomewijde schaal en onderzoeken we hun rol in relatie tot transcriptie. We hebben stadium specifieke nieuwe en eerder aangetoonde (*Plasmodium*) motieven geïdentificeerd en nog onbekende *cis-trans* regulatoire interacties ontdekt. In **hoofdstuk 7** bediscussieer ik onze bevindingen en hoe zij bijdragen aan een beter begrip van de biologie van de malariaparasieten.

ACKNOWLEDGEMENTS

I am very grateful to all those who have encouraged and supported me throughout the last five and a half years. Without you this thesis would not have been possible!

First and foremost, I would like to thank my supervisor Dr. Richárd Bártfai for giving me the wonderful opportunity to join his lab, for his excellent supervision and guidance. Richárd, thank you so much for everything you taught me, your critical advice, your encouraging words after minor and major setbacks, your confidence in me and your continuous support. It has been a challenging journey, but also exciting and rewarding, and filled with many humorous and memorable moments. I wish you all the best for the future!

In addition, I am very grateful to Prof. Henk Stunnenberg for being my promotor and hence giving me the opportunity to join the department of Molecular Biology at the RIMLS with its fruitful research environment.

I would like to express my gratitude to the manuscript committee members Prof. Hans van Bokhoven, Prof. Andy Waters and Dr. Blandine Franke-Fayard for their time and efforts to assess and approve my thesis.

I am very grateful to all my colleagues and collaboration partners within and outside the institute for their direct or indirect contribution to this thesis.

First of all, I would like to say a big thank you to all members and former members of the malaria epigenome group – Wieteke, Ben, Maaïke, Philip, Rob and Christa. Thanks a lot for your help, advice and feedback, for sharing protocols, buffers and tasks, for helping out when needed and so on, and of course, for all the good chats and laughs we had. Wieteke, I am very thankful for your warm welcome and highly appreciated all your explanations, advice and valuable suggestions, which have been very conducive to the outcome of this thesis. Philip, thank you for your patience and explanations. Rob, thank you very much for your contribution to the first research chapter of this thesis, your positive attitude and funny anecdotes. Christa, thanks a lot for your input, our coffee breaks, and the very successful collaboration, which led to a great chapter of this thesis. Additionally, I would like to thank Rob V., with whom I teamed up during the first weeks of my PhD project, and my students Maud, Laurie and Marjolein for their hard work, motivation and enthusiasm. Thanks a lot!

I am very grateful for the great collaboration with the Voss lab in Basel, Switzerland, which resulted in two very nice research articles that are included in this thesis. Till, Michael, it has been a great pleasure working with you! Thank you so much! Also, many thanks to Nicole and Igor for the very nice time I had in Basel while culturing and harvesting parasites. Vielen lieben Dank euch allen!

Importantly, I would like to express my gratitude to all members of the Marie Curie Innovative Training Network Paramet. It has been a great pleasure getting to know you and I highly enjoyed our meetings and workshops – a fantastic and inspiring environment to

learn and discuss science and a good opportunity to travel throughout Europe. Especially, I would like to thank our coordinator Prof. Sylke Müller as well as Prof. David Westhead, Prof. Christine Clayton and Prof. Andy Waters for suggestions and valuable advice, and of course, all my peers. Maja, Julie, Maria, Daniela, Marion, Ewelina, Diana, Tina, Francis, Marco, Fernando, Sanu, Freddy, it has been such a nice time with plenty of good talks and many fun moments. Thanks a lot for that!

At this point, I would also like to thank Dr. Ann-Kristin Müller, Heidelberg, for introducing me to the field of malaria research, for her guidance and enduring support. Ann-Kristin, vielen herzlichen Dank!

I would like to thank my colleagues and former colleagues from the departments of Molecular Biology and Molecular Developmental Biology. Especially Eva for her help and advice and the former sequencing team members Bowon, Kim, Martijn and Nilofar for their great work and technical support. I am grateful to Kees-Jan, Martin, Simon, Arjen, George, Wout and Rita for patiently answering bioinformatic or server-related questions, support and valuable advice. Anita and Josephine, thank you very much for managing the lab and for your help. Siebe, thanks a lot for your advice and being so helpful whenever something was broken or missing. I am also very grateful to Maria for handling the paperwork and helping me in many different ways. Georgina, Ann, Jieqiong, Eduardo, Kriti and Laura, thank you so much for all the nice chats, good talks and fun evenings. Laura, my almost U-body, I am very lucky to have you as a friend. Many thanks for proofreading different sections of my thesis and translating my summary into Dutch.

Furthermore, I remember many helpful suggestions and comments during and after seminars, in the lab, hallway or offices; had many interesting work-related and unrelated discussions, nice chats and good laughs during lunch breaks or in the lab; or got a helping hand here and there. Nina, Luan, Ning-Qing, Jani, Menno, Roberta, Jessi, Klaas, Sabine, Roderick, Mark, Jan, Bilge, Naomi, Leonie, Marco, Julien, Karolina, Jo, Evelyn, Jacob, Matteo, Ila, Saartje, Sarita, Gert Jan, Boris, Roya, Ehsan, Giovanni, Ali, Farid, Nader, Yaser, Shuang-Yin, Onkar, Koen, Esther, Guoqiang, Joost, Tianran, Cheng, Tanya, Colin, Hendrik, Rene, Guido, Simone, Kees, Michiel, Rik, Lisa, Susan, Matthew, Nelleke, Ino, Christina, Pascal, Arne, Marijke, Lidwien, Marion, thank you everyone!

In addition, I would like to thank all malaria people who directly or indirectly contributed to my PhD research, despite being scattered throughout different floors and buildings. Especially, I would like to thank Marga, Wouter and Rianne for the very smooth communication regarding media, serum and blood orders. Also, I would like to thank Koen, Karin, Judith and Martijn as well as Sanna, Maarten and Geert-Jan for advice, help, nice chats and the laughs we had when we still shared the old malaria unit.

Aga, many thanks for being so nice and helpful during my first and very exiting week in Glasgow.

In particular, I would like to thank Laura and Marije for being my paranymphs. I could not have imagined the last years without you! Susan, dir auch vielen lieben Dank für all die guten Gespräche!

There are also many other friends to thank: Maïke, Ili, Jennifer, you immediately made me feel welcome in the Netherlands; Melisa, Nicole, Maarten, Yasemin, Secil, Yorick, I am so glad we met during my time in Nijmegen. Thanks to you I have many great memories. Vielen lieben Dank euch allen! Thank you all!

Steffi, du hast mir mehr als einmal den passenden Schups zur rechten Zeit gegeben. Danke, dass ich auf dich zählen kann! Franzi, Sandra, Britta, ich freue mich immer auf unsere Treffen in der Heimat. Herzlichen Dank! Anke, Karla, Mirjam, Yasmin, Julia, schön, dass wir in Kontakt bleiben und vielen Dank für eure Unterstützung aus der Ferne! Ronny, Irena, it is really nice to have you here in the Netherlands and it is very much appreciated that we are always welcome! Thanks to you, I also met someone very special. Dank euch habe ich auch einen ganz besonderen Menschen kennengelernt.

Uwe, ich bin so froh, dass ich dich habe. Du warst in dieser turbulenten und nicht immer einfachen Zeit für mich da. Ich weiß nicht, was ich ohne dich getan hätte. Danke!

Last but not least, I would like to thank my family for always being there when needed. Oma, ich weiß, dass du wahnsinnig stolz auf mich gewesen wärest. Danke!

Words are not sufficient to express my gratitude towards my parents. Mutti, Papa, ich kann euch gar nicht genug danken für alles was ihr für mich getan habt. Ihr habt mich in allem unterstützt und euer Glaube an mich hat mich auf allen Wegen begleitet. Worte reichen nicht, von Herzen Danke!

At the very end, I would like to apologize to anyone I may have forgotten.

I wish you all the best!

Sabine

April 24, 2018

CURRICULUM VITAE

Sabine Anne-Kristin Fraschka was born on April 19, 1986 in Mosbach, Germany. She studied Biomedicine at the Julius-Maximilians-University Würzburg, Germany, and obtained her Bachelor of Science and Master of Science in 2008 and 2010, respectively. She performed her master's internship in the laboratory of Dr. A.K. Müller at the Department of Infectious Diseases, Parasitology, Heidelberg University Medical School, Germany, focusing on liver stages of the rodent-infecting malaria parasite *Plasmodium berghei*. After graduation, she worked for two years in the field of malignant melanoma at the Department of Dermatology and Allergology, Ludwig-Maximilian University, Munich, Germany. In 2013, Sabine joined the laboratory of her co-supervisor Dr. R. Bártfai (supervisor Prof. Dr. ir. H.G. Stunnenberg) at the Department of Molecular Biology, RIMLS, Radboud University, Nijmegen, The Netherlands, for her PhD research. From 2013 until 2016, she was a member of the Marie Curie Innovative Training Network ParaMet, an interdisciplinary PhD program. During her PhD project, Sabine explored chromatin-based processes in the human-infecting malaria parasite *Plasmodium falciparum*. The results of her research are described in this thesis.

PUBLICATIONS

Toenhake, C.G.*, [Fraschka, S.A.*](#), Vijayabaskar, M.S., Westhead, D.R., van Heeringen, S.J., and Bártfai, R.# (2018). Chromatin accessibility-based characterization of the gene regulatory network underlying Plasmodium falciparum blood stage development. *Cell Host & Microbe* 23, 557-569.e559.

Filarsky, M., [Fraschka, S.A.](#), Niederwieser, I., Brancucci, N.M.B., Carrington, E., Carrió, E., Moes, S., Jenoe, P., Bártfai, R., and Voss, T.S.# (2018). GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science* 359, 1259.

[Fraschka, S.A.*](#), Filarsky, M.*, Hoo, R., Niederwieser, I., Yam, X.Y., Brancucci, N.M.B., Mohring, F., Mushunje, A.T., Huang, X., Christensen, P.R., Nosten, F., Bozdech, Z., Russell, B., Moon, R. W., Marti, M., Preiser, P. R., Bártfai, R.# and Voss, T. S.# (2018). Comparative heterochromatin profiling reveals conserved and unique epigenome signatures linked to adaptation and development of malaria parasites. *Cell Host & Microbe* 23, 407-420.e408.

[Fraschka, S.A.](#), Henderson, R.W., and Bártfai, R.# (2016). H3.3 demarcates GC-rich coding and subtelomeric regions and serves as potential memory mark for virulence gene expression in Plasmodium falciparum. *Sci Rep* 6, 31965.

[Fraschka, S.A.](#), and Bártfai, R.# (2016). Epigenetic gene regulation: Key to development and survival of malaria parasites, In: *A comprehensive analysis of parasite biology: From metabolism to drug discovery* pp. 399-420.

Hohenauer, T., Berking, C., Schmidt, A., Haferkamp, S., Senft, D., Kammerbauer, C., [Fraschka, S.](#), Graf, S.A., Irmeler, M., Beckers, J., Flaig, M., Aigner, A., Höbel, S., Hoffmann, F., Hermeking, H., Rothenfusser, S., Endres, S., Ruzicka, T., and Besch, R.# (2013). The neural crest transcription factor Brn3a is expressed in melanoma and required for cell cycle progression and survival. *EMBO Mol Med* 5, 919-934.

* These authors contributed equally

Corresponding author