Deciphering the role of an RNA Pol III-transcribed non-coding RNA in Plasmodium falciparum

Gretchen Diffendall

Thèse de doctorat en parasitologie moléculaire de

Sorbonne Université / Universite Paris Cité

Ecole Doctorale Complexité du Vivant (ED515)

Unité de Biologie des Interactions Hôte-Parasite

présentée par

Gretchen Diffendall

pour obtenir le grade de Docteur de Sorbonne Université sur le sujet:

Deciphering the role of an RNA Pol III-transcribed non-coding RNA in *Plasmodium falciparum*

Membres du jury :

Dr. Olivier Silvie	Président du jury
Dr. Benoit Gamain	Rapporteur
Dr. Julien Guizetti	Rapporteur
Dr. Antoine Claessens	Examinateur
Dr. Lucy Glover	Examinateur
Pr. Artur Scherf	Directeur de thèse

Présentée et soutenue publiquement à Paris, le 07-07-2022

Summary

The protozoan parasite *Plasmodium falciparum* is the causative agent of the deadliest form of human malaria. This pathogen uses monoallelic expression of variant surface adhesion molecules, encoded by the *var* gene family, to evade the host immune system and cause pathogenesis. It remains unclear how monoallelic expression of *var* gene activation works at the molecular level and if environmental factors can modulate *var* gene expression. Our laboratory showed a Pol III transcribed GC-rich non-coding RNA gene family, termed RUF6, acts as a trans-activator of *var* genes. A physical association between the transcribed RUF6 ncRNA and the active *var* gene locus was observed through FISH. Transcriptional repression of all RUF6 by a specific CRISPR interference strategy resulted in transcriptional down regulation of the entire *var* gene family, suggesting a potential enhancer-like function to *var* gene expression. An understanding of how RUF6 ncRNA mediates *var* gene activation is lacking. Here we developed a robust RNA-directed proteomic discovery (ChIRP-MS) protocol to identify *in vivo* RUF6 ncRNA protein interactions. Biotinylated antisense oligonucleotides were used to purify the RUF6 ncRNA interactome. Mass spectrometry identified several uniquely enriched proteins that are linked to gene transcription such as RNA Pol II subunits, nucleosome assembly proteins, and a homologue of the Dead-Box Helicase 5 (DDX5). Affinity purification of PfDDX5 identified several proteins originally found by our RUF6-ChIRP protocol, validating the robustness of the technique for the identification of ncRNA interactomes in *P. falciparum*. Inducible displacement of nuclear Pf-DDX5 resulted in the significant down-regulation of the active *var* gene. Our work identifies a RUF6 ncRNA protein complex that interacts with RNA Pol II to sustain *var* gene expression. We postulate that DDX5 helicase may resolve G-quadruplex secondary structures highly enriched in *var* genes to facilitate transcriptional activation and progression.

Furthermore, we discovered environmental factors that trigger downregulation of *var* gene transcription. We observe that isoleucine starvation and high $MgCl_2$ concentrations in the medium inhibit RNA Polymerase III transcribed genes. Importantly, this includes a *P. falciparum*-specific regulatory ncRNA gene family (encoded by the RUF6 gene family) that is a key regulator in *var* gene activation. We identified a homologous gene to the highly conserved eukaryotic Maf1, as a negative effector of RUF6 ncRNA transcription. Elevated $MgCl_2$ concentrations led to a shift of cytoplasmic PfMaf1 to the nuclear compartment. We used an inducible protein degradation system to show that external stimuli depend on PfMaf1 to trigger lower expression of RUF6 genes. Our results point to a TOR independent pathway that responds to changes in the environment and represses Pol III transcription. This work provides new and important conceptual insights into PfMaf1-dependent repression of parasite virulence that may be highly relevant for establishing subclinical parasite persistence in the dry season.

Taken together, these results help to better understand the function and regulation of a ncRNA involved in regulating the antigenic variation and pathogenesis in *P. falciparum*. Our validation of the ChIRP-MS technique allows for future studies in identifying RNA-binding proteins for ncRNAs whose function remains to be fully characterized.

Résumé

Le parasite protozoaire *Plasmodium falciparum* est l'agent causal de la forme la plus mortelle de paludisme humain. Ce pathogène utilise l'expression monoallélique de molécules d'adhésion de surface variantes, codées par la famille de gènes *var*, pour échapper au système immunitaire de l'hôte et provoquer une pathogenèse. On ne sait toujours pas comment l'activation du gène *var* fonctionne au niveau moléculaire et si des facteurs environnementaux peuvent moduler l'expression du gène *var*. Notre laboratoire a montré qu'une famille de gènes d'ARN non codants transcrits par Pol III, appelée RUF6, agit comme un trans-activateur des gènes *var*. Une association physique entre l'ARNnc RUF6 transcrit et le locus du gène *var* actif a été observée par FISH. La répression transcriptionnelle de tous les RUF6 par une stratégie d'interférence CRISPR spécifique a entraîné une régulation négative de la transcription de toute la famille des gènes *var*, suggérant une fonction potentielle de type amplificateur pour l'expression des gènes *var*. Une compréhension de la façon dont l'ARNnc RUF6 médie l'activation du gène *var* fait défaut. Ici, nous avons développé un protocole robuste de découverte protéomique dirigée par l'ARN (ChIRP-MS) pour identifier les interactions in vivo des protéines ARNnc RUF6. Des oligonucléotides antisens biotinylés ont été utilisés pour purifier l'interactome d'ARNnc RUF6. La spectrométrie de masse a identifié plusieurs protéines enrichies de manière unique qui sont liées à la transcription génique, telles que les sous-unités d'ARN Pol II, les protéines d'assemblage des nucléosomes et un homologue de la Dead-Box Helicase 5 (DDX5). La purification par affinité de PfDDX5 a identifié plusieurs protéines trouvées à l'origine par notre protocole RUF6-ChIRP, validant la robustesse de la technique pour l'identification des interactomes d'ARNnc chez *P. falciparum*. Le déplacement inductible de PfDDX5 nucléaire a entraîné une importante régulation à la baisse du gène *var* actif. Notre travail identifie un complexe protéique RUF6 ARNnc qui interagit avec l'ARN Pol II pour soutenir l'expression du gène *var*. Nous postulons que l'hélicase DDX5 peut résoudre les structures secondaires G-quadruplex hautement enrichies en gènes *var* pour faciliter l'activation et la progression de la transcription.

De plus, nous découvrons des facteurs environnementaux qui déclenchent une régulation négative de la transcription du gène *var*. Nous observons que la privation d'isoleucine et les concentrations élevées de $MgCl_2$ dans le milieu inhibent les gènes transcrits par l'ARN polymérase III. Il est important de noter que cela inclut une famille de gènes ARNnc régulateurs spécifiques de *P. falciparum* (codée par la famille de gènes RUF6) qui est un régulateur clé de l'activation du gène *var*. Nous avons identifié un gène homologue à l'eucaryote Maf1 hautement conservé, en tant qu'effecteur négatif de la transcription de l'ARNnc RUF6. Des concentrations élevées de $MgCl_2$ ont entraîné un déplacement de PfMaf1 cytoplasmique vers le compartiment nucléaire. Nous avons utilisé un système de dégradation inductible des protéines pour montrer que les stimuli externes dépendent de PfMaf1 pour déclencher une expression plus faible des gènes RUF6. Nos résultats indiquent une voie indépendante de TOR qui répond aux changements de l'environnement et réprime la transcription Pol III. Ce travail fournit des informations conceptuelles nouvelles et

importantes sur la répression de la virulence du parasite dépendante de PfMaf1 qui peuvent être très pertinentes pour établir la persistance subclinique du parasite pendant la saison sèche.

Pris ensemble, ces résultats aident à mieux comprendre la fonction et la régulation d'un ARNnc impliqué dans la régulation de la variation antigénique et de la pathogenèse chez P. falciparum. Notre validation de la technique ChIRP-MS permet de futures études dans l'identification des protéines de liaison à l'ARN pour les ARNnc dont la fonction reste à caractériser entièrement.

Acknowledgments:

My thesis could not have been completed without the help, support, and mentorship that I have received throughout my 3 years with BIHP at Institut Pasteur. I first want to thank my supervisor, Professor Artur Scherf, for not only initially accepting me into his lab but for helping me in every step of this project. I am forever grateful for the independence that was given to me, by him, which has allowed me to grow as a scientist more than I ever could have hoped for. Looking back, I appreciate the weekly meetings and constant feedback. I have learned to pay a greater attention to detail while still seeing the big picture of the work done. I am additionally appreciative for Artur giving me the responsibility to supervise two master students and present my work at an international conference.

I want to specifically thank my reviewers, Julien Guizetti and Benoit Gamain, for accepting the role and taking the time to read through the entirety of my thesis. I value their opinion and comments as accomplished scientists in their fields. The other members of my jury, Dr. Antoine Claessens, Dr. Lucy Glover, and Dr. Olivier Silvie, I thank for accepting to be present during my defense. I also value and respect all of their input related to this thesis project.

My project would not have been possible without the work done by Anna Barcons-Simon. I cannot thank her enough for training me during my first 6 months as a PhD student. Even while she was pregnant with Roc she was always available for my constant questions and assisted in the project above and beyond. I have her to also thank for being available even after she left BIHP, and France itself, for questions and comments especially while writing the manuscript.

I would not even be here in France without the amazing Anne Cozanet. She truly has helped in every administrative aspect and made what would have been a challenging and complicated process so easy for me. Since the beginning I also must thank Jessica Bryant. For being the most supportive (first and last) office-mate. Her help on everything from lab-related problems to navigating all the obstacles that come with moving to Paris means so much to me. I look up to her and all that she has done as a scientist and a woman in the field. She has continued to be a great friend and I am ecstatic for her and Sebastian to become parents to a very lucky child.

One positive consequence of COVID-19 was the rearrangement in offices. My second office-mate, Elie Hammam, helped me grow as a person in the lab. Our constant supply of snacks, never ending

laughing, and his patience for all of my stupid questions helped make each day better. He is truly a great person and I will forever value his friendship. Another person whose friendship means a great deal to me is Catariba Maria Da Silva Rosa. Our picnic venting sessions, daily lab discussions, and Paris outings are some of my top memories from the past 3 years. Lab wouldn't be the same without her, and her singing.

In fact, everyone in BIHP throughout the entire 3 years has made such a positive impact on me. Starting with Carlos and his optimism, Patty and her constant support and conversation, Flore for always going out of her way to help, Camilla and her positive attitude, and Irina for always making me laugh. I thank Aurélie too for her invaluable and incredibly hard work on the Maf1 project, Lilianna for her input during lab meetings, and Parul and Lexi for being cheerful additions to the lab. Lastly, I thank Sebastian Baumgarten and his lab for all of their help and positive attitude.

Outside of Institut Pastuer, I thank my family and friends. Their constant love and support have kept my moral and spirits high during challenging times, like when every experiment seemed to go wrong. My sister especially, even from far away. I thank Seb for being the reason I even considered moving to Paris as well his encouragement when I first started. Jen, and all the people from the FEU at the CiuP, I must thank for allowing me to still have a social life outside of the P2. Finally, I thank Barry. For being there for me in so many ways. Your love, support, and praise helped so much in me being able to pull through and finish my thesis as well as grow as a person.

Again, thank you all. I cannot express in words how truly grateful I am for all of the experiences, good and bad, rewarding and challenging. I wouldn't take it back for anything.

Contents:

List of Figures:

List of Tables:

Abbreviations:

AT adenine-thymine
ChIP-seq ChIP sequencing
ChIRP chromatin isolation by RNA purification
EMSA electrophoretic mobility shift assay
FISH fluorescence in situ hybridization
GC guanine-cytosine
HAT histone acetyltransferase
HDAC histone deacetylase
HDM histone demethylase
Hi-C genome-wide 3C
HMT histone methyltransferase
HP1 heterochromatin protein 1
hpi hours post infection
iRBC infected red blood cell
kb kilobase
lncRNA long non-coding RNA
Mb megabase
miRNA micro RNA
mRNA messenger RNA
NAT natural antisense transcript
ncRNA non-coding RNA
ORF open reading frame
PfEMP1 P. *falciparum* erythrocyte membrane protein 1
Pfmc-2TM P. *falciparum* Maurer's clefts two transmembrane protein
PIC preinitiation complex
Pol polymerase
PTM post-translational modification
RBC red blood cell
RIFIN repetitive interspersed family of polypeptides
RNA-seq RNA sequencing
RNAi RNA-interference
STEVOR subtelomeric *var*iable open reading frame TAD topologically associating domains
TAS telomere-associated sequences
TF transcription factor
TSS transcription start site
UTR untranslated region

Part 1

Introduction

In Part 1 I introduce background information about both the pathogen studied, the human malaria parasite *Plasmodium falciparum*, and the particular topics of this PhD work on antigenic variation and non-coding RNA. This introductory part is not meant to be an all-encompassing description of the subject, rather an overview aimed for an easier understanding of the body of work. It concludes with a section highlighting the aim of the thesis.

1 Introduction

1.1 Introduction to malaria

Malaria is a deadly, vector borne disease caused by eukaryotic single-celled parasites of the phylum Apicomplexa and genus *Plasmodium*. *Plasmodium* parasites can infect birds, reptiles, and mammals, including humans. *Plasmodium* parasites are transmitted to their vertebrate host during a blood meal from their invertebrate vector, female mosquitoes from the genus *Anopheles*. Human malaria is an ancient disease and detailed description of its symptoms appear as early as 2700 BC in the Chinese Canon of Medicine, clay tablets from Mesopotamia from c. 2000 BC and Egyptian papyri from c. 1500 BC [Cox, 2010]. However, the disease has most likely affected humans much longer, perhaps throughout the entire history of our species. Currently, the oldest proof of its causative agent, *Plasmodium*, is parasitic DNA isolated from a mosquito fossil dated 30 million years ago [Poinar, 2005]. It wasn't until 1718 when Francesco Torti gave the name malaria, believing that the disease was air borne and came from the "bad air," *mal-aria* in Italian, of swamps. In 1880 Charles Louis Alphonse Laveran discovered the cause of the disease by identifying the protozoan parasites in a blood smear from a malaria patient. Finally, in 1897, a British medical officer, Dr. Ronald Ross, discovered that the vector that transmits the disease was mosquitoes. Not long after, an Italian professor, Giovanni Battista Grassi and colleagues, demonstrated that human malaria could only be transmitted by *Anopheles* mosquitoes [Cox, 2010]. From approximately 400 species of *Anopheles*, about 60 are malaria vectors, and 30 are of major importance.

1.1.1 Malaria epidemiology

While everyone is susceptible to malaria, pregnant women and children under five bear the greatest burden of malaria. Different factors influence the transmission of the disease including those related to the: vector, parasite, human host, and conditions within the environment. For example, transmission is highly dependent upon conditions related to the climate like rainfall patterns and levels in an area, temperature, and humidity. Usually, transmission is seasonal and divided into a rainy and dry season. Most often there is a peak of malaria transmission both during

and following the rainy season. Transmission can *var*y from year to year within a location, making tracking transmission very complex.

Malaria affects a large number of countries home to over two billion people, with more than 40% of the world's population at risk of contracting the disease [WH0 2021]. In fact, in the 20th century alone, between 150 million and 300 million people died from the disease, accounting for 2 to 5 percent of all deaths [Carter and Mendis, 2002]. The World Health Organization, WHO, launched the Global Malaria Eradication Campaign in 1955 that successfully eliminated malaria from a number of countries. Unfortunately, the ultimate goal of global eradication was not met. It was later recognized that malaria control, rather than an eradication program, was more practical for the remaining endemic countries and a global malaria control strategy was adopted in 1992. This strategy, The Roll Back Malaria Partnership, helps implement coordinated action against malaria.

Countries where malaria is still prevalent include India, Brazil, Afghanistan, Sri Lanka, Thailand, Indonesia, Vietnam, Cambodia, China; the majority of all malaria cases occurs in countries of sub-Saharan Africa [Snow *et al.*, 1999; Breman *et al.*, 2001]. In Africa, Malaria is estimated to cost more than $12 billion annually and account for nearly 25% of all deaths in children under the age of five [Rowe *et al.*, 2006]. According to the latest WHO World malaria report, in 2020 there were 241 million estimated malaria cases in the 85 endemic countries. The disease caused 627,000 deaths in 2020, 12% more than the previous year, 2019. It is estimated that 47,000 of those deaths were the result of service disruptions due to the COVID-19 pandemic [WHO, 2021]. The malaria endemic areas comprise almost exclusively tropical and subtropical regions of the world (Fig. 1.1). Africa carries the highest share of global malaria burden, being home to 95% of cases and 96% of deaths, with just six countries accounting for over half of all malaria deaths globally. Young children under 5 years old are at the highest risk of malaria and accounted for 77% of the global malaria deaths in 2020 [WHO 2021]. These at-risk children are susceptible to the disease because their protective immunity has not yet been developed against the most severe forms of the disease. Pregnant women and their newborns are also vulnerable due to the increased risk of miscarriage or newborn death.

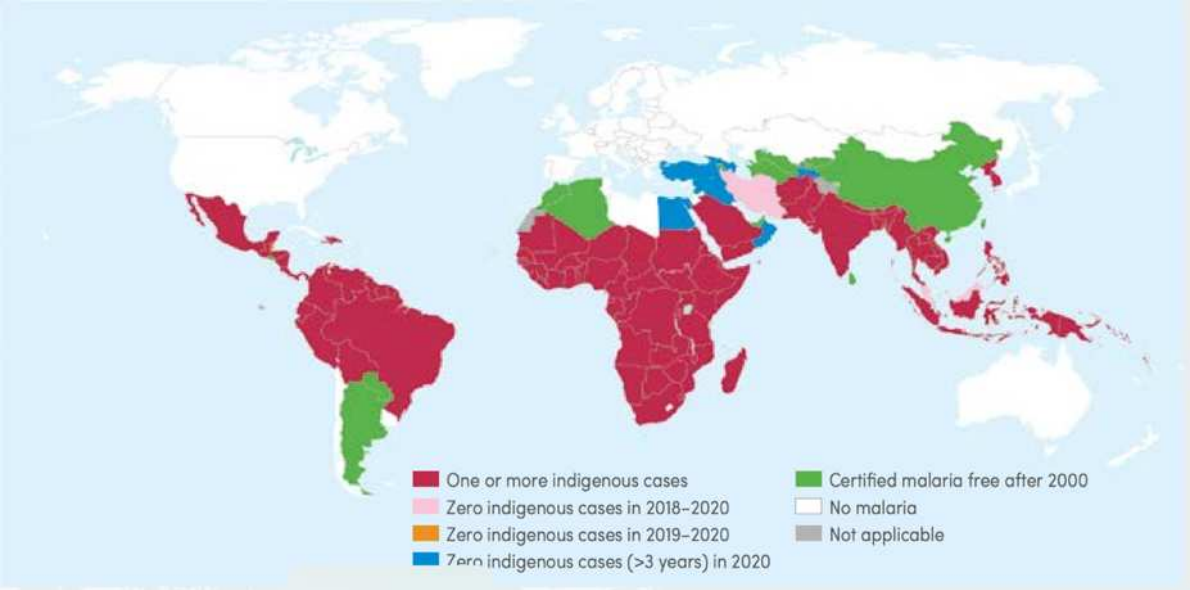


Figure 1.1: Worldwide distribution of countries with malaria cases in 2000 and their status by 2020. Countries with zero indigenous malaria cases for at least three consecutive years are considered malaria free. Taken from World malaria report 2020 [WHO, 2021].

While over 100 *Plasmodium* species can infect animals, the only five that infect humans are: *Plasmodium falciparum*, *Plasmodium ovale, Plasmodium vivax* and *Plasmodium malariae and P. knowlesi.* The latter one is primarily a zoonosis. Importantly, two species (*P. vivax and P. ovale*) form dormant forms in the liver called hypnozoites. These quiescent stages are resistant to most common drug treatments. *P. vivax* is the dominant malaria parasite outside of the African continent while *P. falciparum* is the deadliest and most prevalent on the African continent. This parasite is responsible for the vast majority of severe malaria cases which can become fatal if untreated.

1.1.1.1 Current drugs and control strategies for malaria

Currently, the dominant malaria prevention and control efforts focus on minimizing human contact with mosquito vectors through the use of mosquito nets, insecticides, and prophylactic and treatment drugs. However, the lack of an accessible and effective vaccine and the alarming emergence of resistance to current antimalarial drugs and insecticides pose a great challenge to the fight against malaria and emphasize the necessity for novel therapies and transmission blocking

strategies. Emergence of drug resistant parasites has been reported since the introduction of the first malaria drugs (Fig. 1.2) [Calderón *et al*., 2013]. For most recent drugs, resistance has appeared soon after drug introduction [Calderón *et al*., 2013, Cottrell *et al*., 2014]. Amongst all currently available drugs, artemisinin and its derivates (ARTs) are the most effective against *P. falciparum*, the species responsible for the deadliest form of the disease. Artemisinin is produced by Artemisia annua, a plant used in traditional Chinese medicine as a fever treatment and was isolated by Youyou Tu in 1972 who was awarded The Medicine Nobel Prize in Physiology or Medicine in 2015 for its discovery [Tu, 2011, Miller and Su, 2011, Paddon *et al*., 2013]. ARTs compounds are often used in combination with other antimalarials such as Artemisinin Combination Therapies (ACT). The combination of drugs with different targets and mechanisms of action considerably reduces the likelihood of parasite resistance [WHO, 2020].

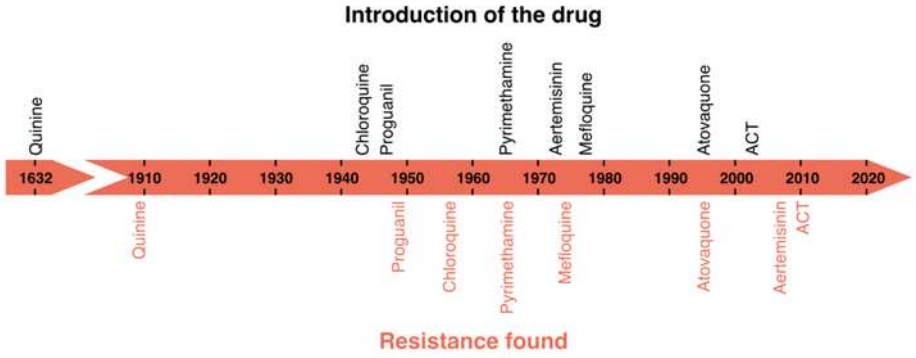


Figure 1.2: Timeline of antimalarial drugs resistance. Top panel indicates the year of the introduction of each drug, and lower panel, the year of appearance of resistance. Modified from [Calderón *et al*., 2013].

ACTs are still recommended by WHO as the first line treatment for P. *falciparum* malaria. However, the recent appearance of *P. falciparum* parasites resistant to artemisinin chemotherapy plays a key role in the decline of success to control malaria observed in the recent years and represents one of the major challenges for malaria elimination [WHO, 2021; Zhu *et al*., 2018; Dondorp *et al*., 2009; Ashley *et al*., 2014]. *P. falciparum* resistance to artemisinin was first reported in Cambodia in 2009 [Dondorp *et al*., 2009] and appeared in other Southeast Asia regions shortly after [Ashley *et al*., 2014, Tun *et al*., 2015]. Artemisinin resistance is associated with nonsynonymous single nucleotide polymorphisms (SNPs) in the *P. falciparum* gene Kelch13

(PfK13) [Straimer *et al*., 2015]. The identification of this molecular marker [Ariey *et al*., 2014] allows the tracking and monitoring of emergent resistant parasite populations. As a matter of fact, rise in artemisinin resistance in Africa is a concerning public health threat and would have a devastating effect in malaria burden, but also in the economic situation of the countries affected [Dondorp *et al*., 2010; Lubell *et al*., 2014; Ariey and Ménard, 2019].

Altogether, there is a need for an effective malaria vaccine, key to eradication of the disease. The current vaccine, RTS,S, is a recombinant protein which includes epitopes from *P. falciparum* circumsporozoite protein (CSP), together with a viral envelope from hepatitis B virus (HBsAg) and a chemical adjuvant (AS01). WHO now recommends that the RTS,S malaria vaccine be used for the prevention of *P. falciparum* malaria in children living in regions with moderate to high transmission as defined by WHO. This is the first vaccine against a human parasite to receive a WHO recommendation. However, downsides to this vaccine are the suggested four doses and its modest protection, 28% in children and 18% in young infants, which decreases over time [RTSS Clinical Trials Partnership, 2015]. Thus, development of new drugs, insecticides and vaccines, together with epidemiological surveillance are crucial to face the current challenges to malaria control. Hence, both basic research to better understand the molecular basis of parasite biology and interactions with the host and further application studies are of pivotal importance for the development of novel therapeutic targets.

1.2 *Plasmodium* biology

Plasmodium species belong to the phylum *Apicomplexa*, along with *Toxoplasma*, *Cryptosporidium* and *Theileria* parasites. This large phylum of alveolates comprises exclusively intracellular parasites [White and Suvorova, 2018]. Most apicomplexan parasites have an apicoplast, a non-photosynthetic organelle, which evolved from a red algal endosymbiont [van Dooren and Striepen, 2013]. The life cycle of these parasites is particularly complex due to the multiple stages they undergo and the wide range of environments they live in. Each stage of the life cycle has a distinct morphology and biochemistry [White and Suvorova, 2018; Frénal *et al*., 2017, Francia and Striepen, 2014]. Specifically, alternation of generation accompanied with the changing between vertebrate and invertebrate hosts, and asexual and sexual cycles.

The name *Apicomplexa* originates from an apical complex structure, characteristic of the extracellular invasive stages of these parasites, zoites. These stages are polarized and one pole contains the apical complex which comprises two types of vesicles, rhoptries and micronemes, that facilitate parasite entrance into the host cell [Bargieri *et al*., 2014]. Zoites have gliding motility and invade the host cell by forming a junction with the host cell membrane, further invagination of the membrane forms the parasitophorous vacuole in which the parasites grow and multiply [Bargieri *et al*., 2014].

1.2.1 *P. falciparum* life cycle

All malaria parasites have a complex, yet similar, life cycle involving both sexual and asexual replication in the mosquito vector and host, respectively. The particularities in pathogenesis for each species are due to differences in structural and biochemical aspects, the tropism for particular types of red blood cells (reticulocytes or erythrocytes) as well as in the length of each parasite cycle, which corresponds to the periodicity of the symptoms [Fujioka and Aikawa, 2002].

The life cycle can be classified into three primary stages depending on the location of the parasites: liver, the bloodstream or the midgut of the mosquito. *P. falciparum* is transmitted through the bite of an infected female anopheline mosquito, primarily *Anopheles gambiae*. During a blood meal, the mosquito vector injects sporozoites into the skin that travel through the bloodstream and within 30 minutes, reach the liver. Sporozoites multiply in liver hepatocytes by a type of asexual reproduction, exoerythrocytic schizogony, yielding thousands of merozoites that are released into the bloodstream after 7-14 days. Released merozoites go on to invade red blood cells, RBCs, and develope through ring, trophozoite, and schizont stages within ~48 hours. Upon rupturing of the schizonts, newly formed merozoites reinvade other erythrocytes, thus perpetuating the infection. A small subset of parasites do not progress and divide and instead differentiate, by gametocytogenesis, into the sexual forms, female and male gametocytes. Gametocytes are taken up by the *Anopheles* mosquito during a blood meal and undergo sexual reproduction in the mosquito midgut, transforming into oocysts. Thousands of sporozoites develop within the oocyst and are released into the mosquito body cavity to travel to the salivary glands where they are ready

to be injected into the human in the saliva of the mosquito during a blood meal [Miller *et al.*, 2002, Josling and Llinás, 2015]. An overview of the *P. falciparum* life cycle is shown in Fig.1.3.

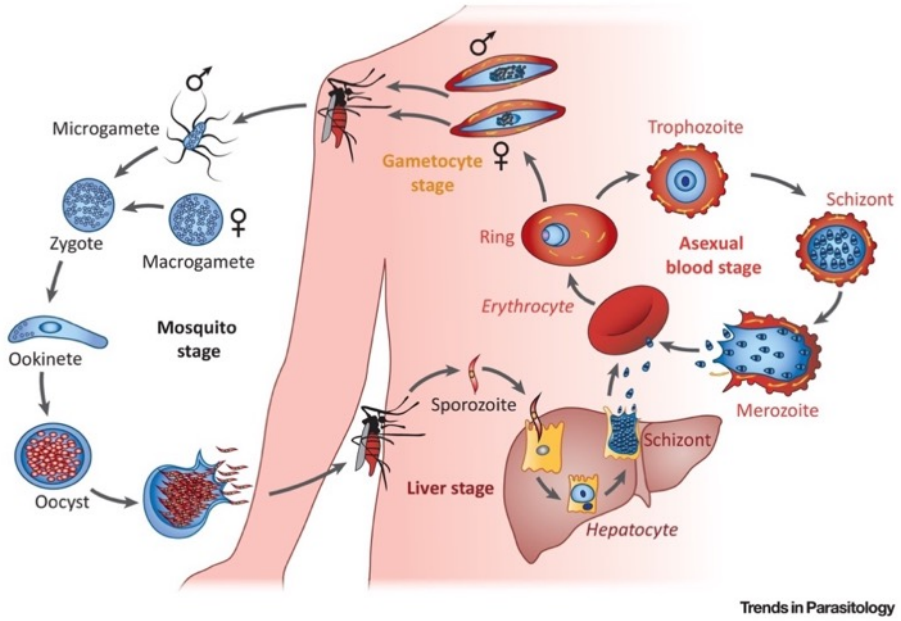


Figure 1.3: Illustration of *P. falciparum* life cycle. After transmission into the skin of the human host during a mosquito bite, sporozoites reach the liver and invade hepatocytes. Within the hepatocyte one sporozoite generates tens of thousands of hepatic merozoites, which enter the bloodstream and invade erythrocytes. The parasite asexual blood cycle causes the clinical symptoms of malaria and is initiated by erythrocyte invasion by a merozoite followed by its further development through ring, trophozoite and schizont stages, generating 16-32 daughter merozoites. These merozoites are in turn released during egress and invade new erythrocytes completing the cycle. One asexual cycle typically lasts 48 hours and <5% of intraerythrocytic parasites develop into male or female gametocytes, which are taken up by a mosquito during a blood meal to undergo sexual reproduction in the mosquito midgut lumen. An ookinete, the motile zygote, crosses the gut epithelium and transforms into an oocyst. Thousands of sporozoites develop within the oocyst and are released into the mosquito body cavity to travel to the salivary glands and allow the process to begin again. Taken from [Alexander G. Maier 2018].

1.2.1.1 The intraerythrocytic cycle

The intraerythrocytic developmental cycle, IDC, is a tightly controlled development program beginning with the invasion of merozoites into RBCs. Here, the parasite forms a parasitophorous vacuole where it can grow and develop in the host erythrocyte cytoplasm [Zuccala

and Baum, 2011, Cowman *et al*., 2017]. During the ~48-hour intraerythrocytic asexual cycle the haploid parasite develops through different morphological stages while enlarging its cytoplasm within the parasitophorous vacuole. Invaded merozoites first develop into low metabolic ring stage (~0-24 hours post invasion, hpi) parasites, that then rapidly grow into trophozoites (24-36hp) where the parasite takes up more of the infected RBC (iRBC). In the final stage, named schizont stage, ~36-48 hpi, mitotic divisions are first followed by nuclear membrane division and later cytokinesis, that results in the formation of 16-32 daughter merozoites. Schizont bursting and erythrocyte rupture, release the merozoites in the bloodstream, enabling the cycle to restart by invading new erythrocytes.

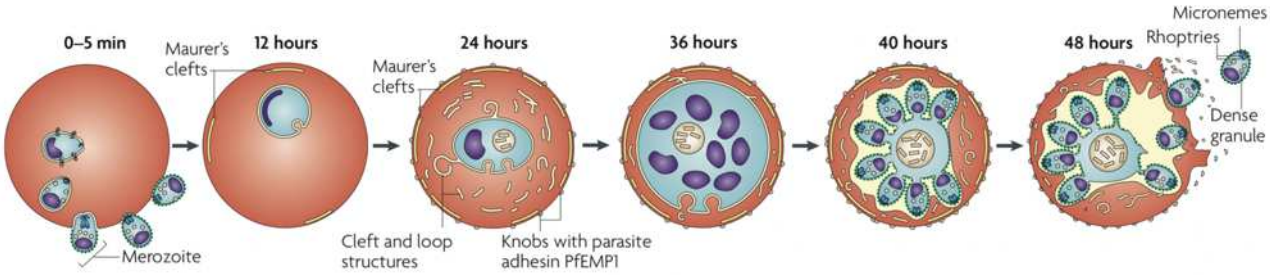


Figure 1.4: Stages of the intraerythrocytic asexual cycle. After erythrocyte invasion by a merozoite, the parasite develops through the ring (~0-24 hpi), trophozoite (~24-36 hpi) and schizont (~36-48 hpi) stages. After 24 hpi, membranous structures can be found in the erythrocyte cytoplasm and knobs are formed in the erythrocyte membrane where exported parasite antigens like PfEMP1 are displayed. Hemoglobin degradation, to obtain amino acids for protein synthesis, causes the accumulation of hemozoin crystals in the parasite digestive vacuole. At ~48 hpi the infected erythrocyte bursts, releasing 16–32 daughter merozoites. Taken from [Maier et al., 2009].

Specialized protein expression during the IDC is needed for the parasite's survival and evasion of host immune responses. Once inside the host RBC, the parasite remodels the host erythrocyte by trafficking and displaying proteins on the surface of the infected erythrocyte that help the parasite acquire nutrients and evade the host immune system [Maier *et al*., 2009]. This protein trafficking occurs mainly via parasite-derived membranous structures, termed Maurer's clefts, found in the erythrocyte cytosol [Mundwiler-Pachlatko and Beck, 2013]. The surface of the iRBC becomes distorted from a particular type of protrusions called knobs [Leech *et al*., 1984a]. These structures are primarily composed of knob-associated histidine-rich protein (KAHRP) where exported surface antigens, involved in cytoadhesion, like *P. falciparum* erythrocyte membrane protein 1

(PfEMP1) are anchored (Fig. 1.4). iRBCs containing ring stage parasites appear in peripheral blood circulation. Cytoadhesion occurs in mature stage parasites (18-24 hours post-invasion) where the iRBCs are sequestered in blood vessels in different organs, thereby preventing them from circulating in capillaries of critical target organs [David *et al*., 1983; Miller *et al.*, 2013]. As a consequence, this adhesion phenotype prevents phagocytic clearance by the spleen and is also involved in lethal complications of the disease [Maier *et al.*, 2009]. The parasite uses antigenic *var*iation of these surface antigens as a mechanism to escape the host immune system and establish chronic infection, explained later in section 1.4 [Scherf *et a*l., 1998]. It is during this time, the IDC, that is responsible for the pathogenesis of malaria infection. Because of this, symptoms, including the acute fever, usually take a week to appear.

1.2.2 *P. falciparum* malaria pathology and pathogenesis

Disease symptoms caused by *P. falciparum* infection can be very diverse, varying from asymptomatic to lethal and include a wide range of clinical symptoms like fever, chills, headache, vomiting, diarrhea, muscular and abdominal pain [Bartoloni and Zammarchi, 2012]. The characteristic waves of fever, associated with malaria, are the result of periodic, and exponentially increasing, merozoite egress and invasion during the asexual blood stage. In <0.5% of infections [Sachs and Malaney, 2002; WHO, 2021] mild malaria can progress to severe with associated symptoms such as: anemia, acute renal failure, pulmonary edema, generalized seizures, circulatory collapse, and prostration, which can be followed by coma and death [Bartoloni and Zammarchi, 2012]. Each of these complications are associated with *P. falciparum* iRBC sequestration in different tissues and organs. Particularly, this adhesion is mediated by parasite variant surface antigens (VSA) displayed on the iRBC [Wahlgren *et al.*, 2017].

Cerebral malaria is the most severe pathology associated with this parasite and is caused by iRBC sequestration in the brain microvasculature which impairs the blood brain barrier and leads to an unarousable coma. Young children, pregnant women and travelers are at highest risk of developing severe malaria [WHO, 2021]. Pregnant women are at risk of a particular pathology termed pregnancy-associated malaria (PAM), caused by iRBC sequestration in the placenta resulting in low birthweight that can cause infant death [Sharma and Shukla, 2017].

Individuals can acquire natural immunity to malaria through ongoing exposure to infections allowing for the production of anti-malarial antibodies including those against

VSAs [Chan *et al.*, 2014]. This is especially important to those living in areas endemic to the severe form of the disease.

1.3 Genome organization and gene regulation in *P. falciparum*

P. falciparum development and survival within its host is largely dependent on gene regulation and expression. In eukaryotes, this control occurs at multiple levels, including transcriptional, post-transcriptional and translational stages. Transcriptional regulation, specifically, is known to play a key role where it is primarily controlled through the binding of transcription factors to regulatory DNA elements. Typically, eukaryotic genes are composed of a promoter and *cis*-acting elements, binding sites for proteins, as well as distal elements such as enhancers, silencers and insulators (Fig. 1.5) [Maston *et al*., 2006]. Additionally, similar to most eukaryotes, *P. falciparum* transcription occurs predominantly in a monocistronic fashion [Lanzer *et al*., 1992a]. Transcriptional control is evident when comparing the expression of genes in both sexual [Young *et al*., 2005] and asexual stages [Bozdech *et al*., 2003; Le Roch *et al*., 2003] of the parasite. It has been shown that functionally related genes are specifically expressed in the different stages.

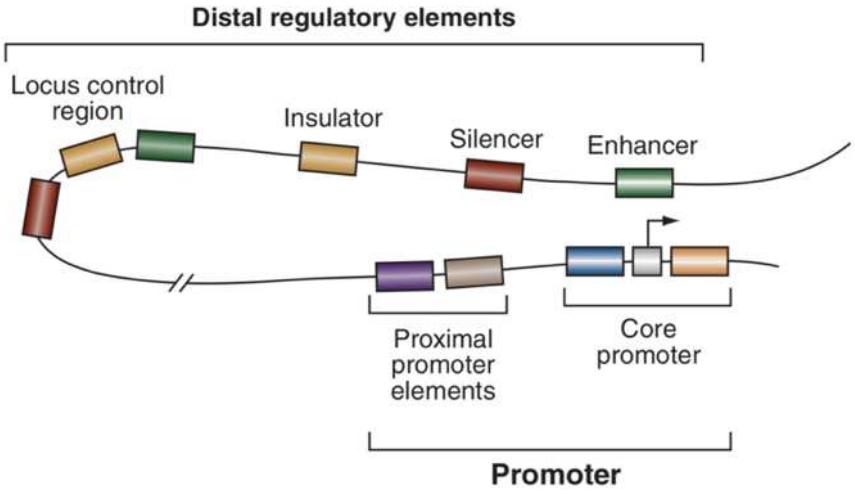


Figure 1.5: Typical eukaryotic gene regulatory region. The core promoter and proximal promoter elements make up the promoter. Upstream distal regulatory elements include: enhancers, silencers insulators, and other control regions. These are located up to 1 Mb from the promoter and interact via DNA looping. Taken from [Maston *et al*., 2006].

A second layer of transcriptional control is seen when DNA is packaged into chromatin, thereby regulating the accessibility of transcription factors and polymerases. Epigenetic factors like histone modifications and nucleosome repositioning can modify different chromatin condensation states [Kornberg and Lorch, 1999]. Additionally, the nuclear architecture and genome organization contribute to transcriptional regulation [Branco and Pombo, 2007].

Genome organization of the parasite and its different layers of gene regulation: epigenetic regulation with specific transcription factors, alternative splicing, and post-transcriptional regulation are further described below.

1.3.1 *P. falciparum* genome and features

The *P. falciparum* nuclear genome is composed of 23.3 megabase (Mb) with an additional 35 kilobase (kb) circular apicoplastic and a 6 kb mitochondrial genome [Gardner *et al.*, 2002]. The haploid nuclear genome is distributed over 14 chromosomes, ranging in size from 0.64 to 3.3 Mb. *P. falciparum* has the most AT-rich genome sequenced so far, with a striking ~79.6% AT-content genome-wide and rising up to 93% in intronic and intergenic regions [Gardner *et al*., 2002]. There are currently ~5700 open reading frames (ORF) annotated, with over 33% of predicted genes remaining functionally unannotated in the parasite genome. Two distinct regions make up *P. falciparum* chromosomes: an internal region, where house-keeping genes are located, and chromosomal ends, comprising the telomeric DNA and the telomere-associated sequences (TAS), which have coding and non-coding regions (Figure 1.6) [Scherf *et al*., 2001]. Noncoding regions in the TAS are located adjacent to G-rich tandem repeats in the telomeric DNA and composed of six different telomere-associated repetitive elements (TAREs 1-6) [Figueiredo *et al*., 2000]. Coding regions in the chromosomal ends contain most members of clonally variant multigene families associated with antigenic variation and cytoadhesion [Gardner *et al*., 1998, Bowman *et al*., 1999, Gardner *et al*., 2002].

In internal regions, tandem repeat-rich centromeric regions occupy a 2.3 to 2.5 kb region on each of the 14 chromosomes [Kelly *et al*., 2006]. These areas are enriched in the centromeric histone *var*iant, PfCENH3, and histone variant, PfH2A.Z [Hoeijmakers *et al.*, 2012]. Non-coding

RNAs (ncRNAs) from centromeric regions have been reported [Li *et al*., 2008] and have been suggested to facilitate the loading of PfCENH3 to centromeres [Vembar *et al*., 2014].

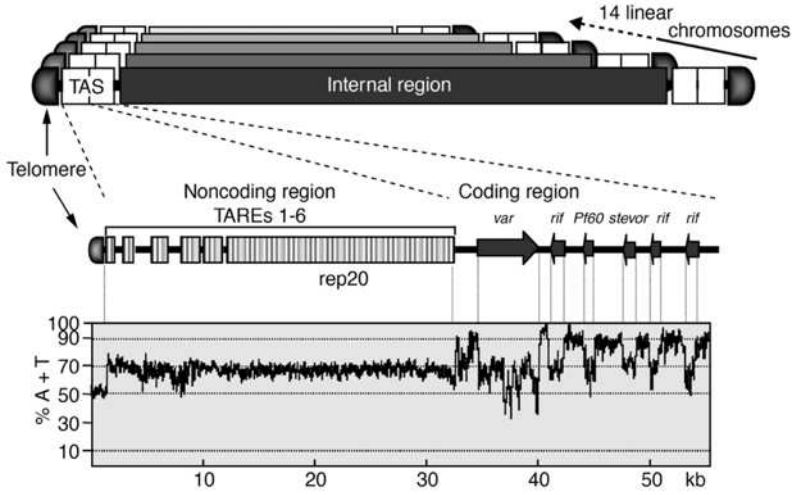


Figure 1.6: Chromosome organization in P. *falciparum*. Top: The 14 lineal chromosomes are composed of an internal region and the chromosome ends. Middle: End of chromosome 3 is represented, as an example, to show the telomere and TAS. TAS is comprised of both a non-coding region, with the 6 TAREs, and a coding region, where the locus of gene families associated to virulence like *var*, rif and stevor are found. Lower panel: AT-content plotted for the different regions of the TAS. Taken from [Scherf *et al*., 2001].

1.3.2 Transcriptional regulation by epigenetics and chromatin structure

Throughout the complex life cycle in its two different hosts, *P. falciparum* relies heavily on a combination of gene regulatory mechanisms, largely dependent on epigenetic processes. Chromatin remodeling, DNA methylation and nuclear architecture are a few examples of how the parasite is able to undergo such controlled gene regulation. (Long ncRNAs, which also play an important role, are reviewed further on in section 1.5).

1.3.2.1 Chromatin structure

Eukaryotic chromatin is tightly packed into nucleosomes, composed of 147 bp of DNA wound around histone octamers [Olins and Olins, 1974; Kornberg, 1974]. *P. falciparum* contains both the canonical histones, H2A, H2B, H3 and H4, and the four histone variants, H2A.Z, H2Bv, H3.3 and CenH3, but lacks the linker histone, H1 [Miao *et al*., 2006].

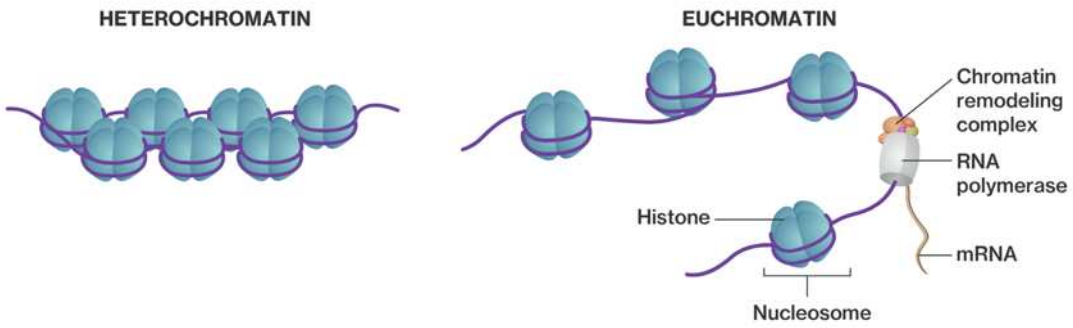


Figure 1.7: Chromatin states. Chromatin is classified as heterochromatin or euchromatin depending of its compaction level. Heterochromatin is compact and comprises a nucleosome array condensed into a 30 nm fiber. The condensation into heterochromatin reduces the access of the transcriptional machinery and is associated with repressed genes. Euchromatin is less compact and is described as a 11 nm fiber which resemble 'beads on a string', where the beads are the nucleosomes and the string is the DNA. The euchromatic state is open and permissive for chromatin remodeling complexes and RNA polymerases to access the DNA, therefore associated with active gene transcription.

Chromatin exists in one of two states (shown in Figure 1.7): euchromatin, open and accessible to transcriptional machinery, or heterochromatin, condensed and transcriptionally repressive and is able to switch between the two. To do so, chromatin structure is altered, by nucleosome occupancy, thus affecting levels of transcription. In eukaryotes, including *P. falciparum,* nucleosome-depleted regions (NDRs) have been found in promoter regions, to allow for a more open chromatin structure, thus leading to higher levels of transcription [Kensche *et al.*, 2016; Lee *et al*., 2004, Struhl and Segal, 2013]. This is evident in the parasite at transcription start sites (TSSs) and core promoters. Such regions are nucleosome-free during the asexual stages, to allow for a general transcriptionally active state, whereas the opposite is seen in the sexual stage [Bunnik *et al*., 2014, Kensche *et al*., 2016; Ponts *et al*., 2011]. In fact, a study showed that transcription at strong TSSs is coupled by eviction of nucleosomes and nucleosome positioning and dynamics can indicate the function of regulatory DNA elements [Kensche *et al*., 2016].

Developmentally regulated histone H3 protease clipping at the N-terminal region (aa 21) was observed in asexual blood stages of *P. falciparum* [Herrera-Solorio *et al.*, 2019]. This truncated form primarily integrates into chromatin regions upstream of six DNA replication gene loci hinting at the existence of a highly specific cellular targeting machinery for truncated histones.

This report identified a novel epigenetic mechanism employed by *P. falciparum* that is linked to DNA metabolism.

1.3.2.2 Histone post-translational modifications

Post-translational modifications (PTMs) of histones play a vital role in the epigenetic regulation of genes, with functions like regulation of gene expression, chromatin condensation, DNA repair and replication (Table 1.1) These modifications alter the total charge of the histones which in turn affect the interaction with DNA, ultimately disrupting nucleosome occupancy. PTMs occur at the most accessible region of the core histones, in the N-terminal tail region. Among the many specific types, the most common are: acetylation, generally associated with gene activation, and methylation of lysine (K) residues, associated with both activation or silencing, depending on the affected residues [Kouzarides, 2007]. Such modifications either allow or eliminate sites where factors like readers or writers and erasers (which can generate or remove PTMs, respectively) can interact to regulate gene expression.

Table 1.1: Histone PTMs and their functions. Modified from [Kouzarides, 2007].

Modifications	Residues Modified	Functions Regulated
Acetylation	K-ac	Transcription, DNA repair, Replication, Condensation
Phosphorylation	S-ph, T-ph	Transcription, DNA repair, Condensation
Methylation (lysines)	K-me1/me2/me3	Transcription, DNA repair
Ubiquitylation	K-ub	Transcription, DNA repair
Methylation (arginines)	R-me1/2a/me2s	Transcription
Sumoylation	K-su	
ADP ribosylation	E-ar	
Deimination	R > Cit	
Proline Isomerization	P-cis > P-trans	

Histone writers add, while erasers remove, PTMs. For example, histone acetyltransferases (HATs) and histone methyltransferases (HMTs) are writers while histone deacetylases (HDACs) and demethylases (HDMs) are erasers. Histone readers are proteins that recognize specific PTMs

and orchestrate the different functions in DNA transcription. Reader proteins bind to PTMs at histone tails and recruit proteins associated with transcription or chromatin remodeling.

Evidence for the role of gene regulation by histone PTMs in *P. falciparum* was first observed by studying antigenic variation in the Scherf laboratory [Freitas-Junior *et al*., 2005]. Throughout the years, mass spectrometry and antibody-based analyses have demonstrated that *P. falciparum* has over 232 histone PTMs [Miao *et al*., 2006; Cui *et al*., 2007; Issar *et al*., 2008; Trelle *et al*., 2009; Treeck *et al*., 2011; Dastidar *et al*., 2012], 88 unique to *Plasmodium* [Saraf *et al*., 2016]. The most studied are histone H3 lysine 9 acetylation (H3K9ac) and lysine 4 methylation (H3K4me3) which are markers for euchromatin, and lysine 9 mono-, di, and tri-methylation (H3K9me1, H3K9me2, H3K9me3) mark heterochromatin (Fig. 1.8) [Lopez-Rubio *et al*., 2007; Chookajorn *et al*., 2007; Cui *et al*., 2007; Cui *et al*., 2008b].

The most characterized histone reader in *P. falciparum* is the heterochromatin protein 1 (PfHP1), which binds H3K9me3 in heterochromatic domains of subtelomeric and central chromosomal clusters of virulence genes [Flueck *et al*., 2009; Pérez-Toledo *et al*., 2009].

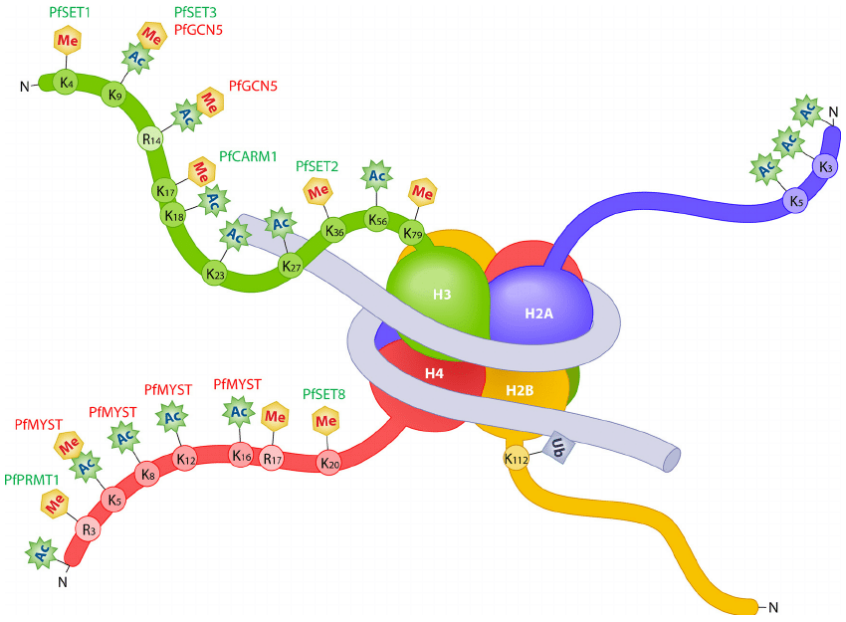


Figure 1.8: Schematic drawing of a nucleosome with the four canonical histones (H3, H4, H2A, and H2B) in *P. falciparum*. The covalent PTMs of the histone tails (methylation [Me], acetylation [Ac], and ubiquitination [Ub]) and enzymes catalyzing the addition of the PTMs (PfGCN5, PfSET1, PfSET2, PfCARM1, PfPRMT1, PfMYST, and PfSET8) are colored in red or green above the modification. [Taken from Cui *et al*., 2010]

1.3.2.2.1 Histone variants

Histone variants, which substitute for the core canonical histones, often have specific structural and functional features that play a role in establishing and maintaining epigenetic states, chromosome segregation, transcriptional regulation and DNA repair. The process of replacing canonical histones with histone variants results in the generation of functionally specialized chromatin domains [Henikoff and Smith, 2015]. During the IDC in *P. falciparum*, PfH2A.Z variant is found in euchromatic intergenic regions with H3K4me3 and H3K9ac [Bártfai *et al*., 2010] and has been shown to dimerize with PfH2B.Z (PfH2Bv) at the most highly transcribed promoters [Hoeijmakers *et al*., 2013; Bártfai *et al*., 2010]. Additionally, *P. falciparum* PfH3.3 variant is found in euchromatic GC-rich coding regions and subtelomeric repetitive sequences [Fraschka *et al*., 2016]. Lastly, centromeric histone variant PfCenH3 along with PfH2A.Z variant are located in centromeric regions [Hoeijmakers *et al*., 2012; Lopez-Rubio *et al*., 2009; Grewal and Jia, 2007].

1.3.2.3 DNA methylation

DNA methylation is involved in regulating gene expression through the recruitment of proteins that lead to gene repression or by inhibiting transcription factor binding to DNA. While the most widely accepted role for DNA methylation is gene silencing [Jones, 2012], methylation of DNA in the 5th carbon of cytosines (5mC) in CpG islands is also associated with genomic imprinting, regulation of gene expression, and chromosome stability. DNA methyltransferases (DNMTs) enzymes catalyze the transfer of a methyl group to DNA. Different DNA base modifications include the most abundant, 5mC, which is further oxidized by ten-eleven-translocatin (TET) enzymes into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [Rasmussen and Helin, 2016]. In other organisms, 5hmC, specifically, has been implicated in the regulation of many cellular and developmental processes, however its presence in *P. falciparum* remains unclear. DNA 5mC and its oxidized form have also been reported in the extremely AT-rich genome (>80%) of the protozoan human malaria parasite *P. falciparum* [Hammam *et al.*, 2020]. Malaria parasites belong to a small group of eukaryotes denoted as “DNMT2-only organisms” (absence of DNMT1 and DNMT3A/B genes) that include

a wide range of phylogenetically diverse group of organisms such as *D. melanogaster, S. mansoni, Schizosaccharomyces pombe* and *Entameoba histolytica* [Jeltsch *et al.*, 2017]. Similarly, to what is observed in other eukaryotes, plasmodial DNMT2 methylates $tRNA^{Asp}$ at position C38 [Hammam *et al.*, 2020]. Recent experimental evidence from the Scherf laboratory hints to Pf-DNMT2 as having a dual function in DNA and RNA cytosine methylation [Hammam *et al.*, 2021]. The plasmodial TET enzymes involved in 5hmC synthesis remain elusive.

1.3.3 Nuclear genome organization

The spatial organization and distribution of eukaryotic chromatin is related to gene expression. Mechanisms and molecular players involved in genome organization are therefore essential to fully comprehend nuclear organization and genome function. Evidence in higher eukaryotes show genome architecture plays a role in regulating transcription. Compartmentalization of the nucleus, specifically topologically associating domains (TADs), chromatin loops, involving the CCCTC-binding factor (CTCF) and cohesion, and long-range interactions, such as bringing together a distant enhancer to its target promoter, all contribute to genome organization [Rowley and Corces, 2018; Merkenschlager and Nora, 2016].

In *P. falciparum*, while it appears that genome organization and its effect on gene expression is crucial, much is still left to be discovered. It was shown that the transcription site of ribosomal DNA (rDNA) cluster and colocalize in the nucleolus, the most prominent domain and the site of transcription of rDNA by RNA polymerase I (Pol I) [Mancio-Silva *et al*. 2010; Lemieux *et al*. 2013]. Similarly, RNA Pol II and Pol III transcription sites cluster in distinct loci domains during early and late stages (Fig. 1.9) [Mancio-Silva *et al*., 2010]. Telomeres have also been shown to form clusters that are tethered to the nuclear periphery [Freitas-Junior *et al*., 2000].

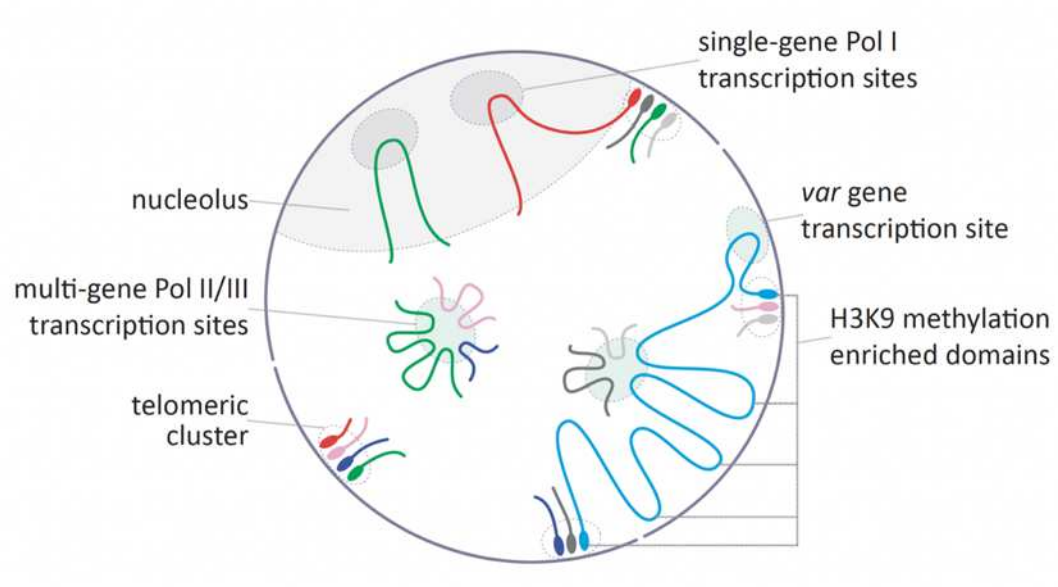


Figure 1.9: Nuclear organization in *P. falciparum*. Schematics of the genome organization within the nucleus for ring stage parasites, displaying the nucleolus, the location of rDNA transcription by Pol I, the transcription sites for the RNA Pol II and Pol III machinery, and the telomeric and virulence gene families clusters in the nuclear periphery. Taken from [Scherf *et al*., 2017].

Studies have shown that genes involved in the control of parasite virulence (*var* genes, discussed in section 1.4.1) are arranged in repressive centers along the nuclear periphery [Duraisingh *et al.*, 2005; Lopez-Rubio *et al.*, 2009; Freitas-Junior *et al.*, 2005]. These genes are located in subtelomeric regions, predominately, and internal chromosomal clusters [Gardner *et al*., 2002]. Subtelomeric and central virulence genes are localized to the nuclear periphery where they form 4-7 repressive clusters [Lopez-Rubio *et al*., 2009, Lemieux *et al*., 2013; Ay *et al*., 2014]. This perinuclear localization is associated with heterochromatin and H3K9me3 enrichment in repressed virulence genes [Lopez-Rubio *et al*., 2009]. Specifically, for the virulence *var* gene family, the single active member is spatially segregated away from the repressed clusters thereby allowing its transcription [Ralph *et al*., 2005b; Lopez-Rubio *et al*., 2009].

Lastly, throughout the IDC, the nuclear architecture of the parasite is remodeled, correlating with transcriptional changes. Namely, the transition between ring and trophozoite stages, where transcriptional activity increases, is correlated with an increase in nuclear pores, to facilitate the export of messenger RNA (mRNA) to the cytoplasm, and nucleosome eviction [Weiner *et al*., 2011]. These changes are later reversed once the parasite enters schizogony.

1.3.3.1 Transcription factors and DNA elements

Eukaryotic transcription is carried out by three RNA polymerases that transcribe different types of genes. RNA Pol I, responsible for rRNA synthesis, RNA Pol II for transcribing protein-coding genes into mRNA as well as micro RNAs (miRNAs) and other ncRNA, and RNA Pol III that transcribed tRNA, 5S rRNA, and other small ncRNAs. These polymerases rely on specific DNA element-binding transcription factors (TFs) to be recruited onto gene promoters. Canonical eukaryotic promoters contain a binding site for the preinitiation complex (PIC), containing general TFs, proximal to the TSS, and upstream regulatory regions where specialized TFs bind, in order to enhance or repress transcription [Vannini and Cramer, 2012].

While many putative DNA regulatory elements have been identified bioinformatically [Gunasekera *et al*., 2007; Wu *et al*., 2008; Young *et al*., 2008], only a few have been validated experimentally (summarized in Table 1.2) [Lanzer *et al.*, 1992b; Voss *et al.*, 2003; López-Estraño *et al.*, 2007; Olivieri *et al.*, 2008; Osta *et al.*, 2002; Santos *et al.*, 2017]. DNA elements with enhancer and repressive activities have been described, however, insulator elements and a CTCF homolog have not yet been identified. 202 genes, making up 3.5% of the *P. falciparum* genome, code for either TFs or proteins with DNA binding domains [Bischoff and Vaquero, 2010]. Among them, 73 are chromatin-associated proteins and components of the transcriptional machinery; furthermore, 27 of them belong to a highly conserved Apicomplexa-specific family of transcription factors [Balaji *et al*., 2005]. This family of TFs, termed ApiAP2, contains a modified version of the Apetala 2 (AP2)-integrase DNA binding domain found in a family of plant TFs [Riechmann and Meyerowitz, 1998]. TFs from the ApiAP2 family are a central component of *P. falciparum* transcription.

Table 1.2: Validated DNA binding proteins regulating transcription or chromatin remodeling in *P. falciparum*. Modified from [Scherf *et al*., 2017; Toenhake *et al.*, 2019].

DNA binding protein	Gene ID	Class / Subclass	Biological process	Reference
PfMyb1	PF3D7_1315800	Specific TF / Helix-turn-helix	Transcriptional regulation of genes involved in cell cycle regulation and progression	[Gissot *et al*., 2005]
PfMyB2	PF3D7_1033600	Specific TF/ Helix-turn-helix	Pre-mRNA splicing factor CEF1	[Sorber *et al*., 2011]
PfAP2-Sp	PF3D7_1466400	Specific TF / ApiAP2	Transcriptional regulation of genes involved in sporozoite development	[Campbell *et al*., 2010]
PfSIP2	PF3D7_0604100	Specific TF / ApiAP2	Binding subtelomeric non-coding regions	[Flueck *et al*., 2010]
PfAP2-variant	PF3D7_1107800	Specific TF / ApiAP2	Binding to intron element of *var* genes in a complex with nuclear actin	[Zhang *et al*., 2011]
PfAP2G	PF3D7_1222600	Specific TF/ ApiAp2	Principal regulator of gametocytogenesis switching	[Josling and Llinás, 2015]
PfTRZ	PF3D7_1209300	Specific TF / Zinc finger	Binding to telomeric TT(T/C)AGGG repeats and regulates expression of 5S rDNA loci	[Bertschi *et al.*, 2017]
PfEGXP	Pf3D7_1466200	Homodomain-like TF	Early gametocyte enriched phosphoprotein	[Toenhake *et al.*, 2019]
PfPREP	PF3D7_1011800	Specific TF / K homology	Transcriptional regulation of IDC stages	[Komaki-Yasuda *et al*., 2013]
PfAlbas	PF3D7_0814200 PF3D7_1346300 PF3D7_1006200 PF3D7_1347500	Alba	Non-specific DNA binding enriched in subtelomeres	[Chêne *et al*., 2012, Goyal *et al*., 2012]
PfHMGBs	PF3D7_1202900 PF3D7_0817900	Chromatin remodeling factor / high mobility group box 3 protein	Non-specific DNA binding; regulates transcription of gametocyte genes and ookinete formation	[Briquet *et al*., 2006, Kumar *et al*., 2008]

Accessibility of chromatin correlates with sites of active transcription where DNA elements such as TSS or enhancers are also present [Ruiz *et al.*, 2018]. The identification of genome-wide chromatin accessibility in the *P. falciparum* IDC genome allowed for the discovery of novel regulatory regions [Toenhake *et al.*, 2018; Ruiz *et al.*, 2018]. Transposase accessible chromatin sequencing (ATAC-seq) identified ~4000 novel regulatory regions, the majority found within 2 kb upstream of a transcribed gene [Toenhake *et al.*, 2018]. Motif analyses of these sites predicted ApiAp2 transcription factor binding sites and functions [Toenhake *et al.*, 2018; Ruiz *et al.*, 2018].

1.3.4 Post-transcriptional and translational gene regulation

Following control at the transcriptional level, gene expression can further be regulated at the translational and post-translational level. Such regulation is related to mRNA splicing and mRNA stability, which relies on RNA-protein interactions to either target mRNA for degradation or prevent access of the ribosome to its translation start codon [Day and Tuite, 1998]. For example, translation initiation can be repressed through eukaryotic translation initiation factor 4E (eIF4E)-binding proteins. In *P. falciparum*, a large number of proteins code for mRNA decay and translational rate control, however, detailed post-translational regulation has not been fully characterized. Additionally, antisense ncRNAs can participate in this layer of regulation, discussed in further detail in section 1.5.

The best-known form of RNA processing is splicing. In order for transcribed mRNA to translated, it needs to be further processed. Specifically, alternative splicing, where introns, non-protein-coding regions, are removed by a large complex of RNAs and proteins, termed the spliceosome. This process is essential to *P. falciparum* with approximately half of its genes containing more than one intron [Yeoh *et al.*, 2019].

Once mRNA exits the nucleus, the mature mRNA has one of three outcomes: translation, degradation, or sequestration followed by repression. Cytoplasmic mRNA transcripts are susceptible to degradation by the RNA exosome [Chen and Shyu, 2011]. Particularly in the case of the *P. falciparum*, its RNA exosome complex, has been found to be is associated with degradation of diverse RNAs, as well as play a particularly important role in degradation of cryptic

and antisense RNA [Droll *et al*., 2018]. Recently, an orthologue of a eukaryotic RNA exosome-associated RNAse, PfRrp6, has been found in *P. falciparum* [Fan *et al.*, 2020]. Knockdown of this gene was shown to impact regulation of virulence gene expression [Fan *et al.*, 2020]. Repression of cytoplasmic mRNA can also occur in complexes called mRNA nuclear ribonucleoprotein (mRNP) granules [Baum *et al.*, 2009; Balu *et al.*, 2011]. An additional layer of post-transcriptional regulation is mediated by a chromatin-associated exoribonuclease (PfRNase II) that silences a subset of the *var* gene family, independently of the RNA exosome, through degradation of nascent cryptic RNA [Zhang *et al*., 2014].

Translational repression is seen in some invasion-related genes during the trophozoite stage. In fact, translational delay occurs for over 30% of transcribed genes in *P. falciparum* [Vembar *et al.*, 2016]. A DNA/RNA-binding protein, PfAlba1, was found to bind to mRNAs to maintain their stability and removal of the protein, in late stage, correlates with protein synthesis [Vembar *et al.*, 2015].

1.3.4.1 Epitranscriptomics

Similar to epigenetic modifications of histones and DNA during transcriptional regulation, RNA can also be reversibly modified. Identifying the biochemical modifications to RNA, epitranscriptomics, is an emerging field of study. Chemical modifications of tRNA have been reported in *P. falciparum* to be highly coordinated throughout the life cycle [Ng *et al*., 2018]. There are a higher number of modifications seen during trophozoite stage, consistent with increased translation, that have been suggested to be related to protein abundancy through enhanced translational efficiency [Ng *et al*., 2018]. The dominate modification on mRNA for the parasite is the methylation of N6-methyladenosine (m6A), associated with mRNA stability or translational efficiency [Baumgarten *et al.*, 2019].

Epitranscriptomics are additionally important to the parasite during its switch between its host and vector. Various forms of stress can influence this change, requiring the parasite to maintain stable protein synthesis through coordinated gene regulation. One such example, recently identified, is the cytosine methylation of $tRNA^{Asp\ (GTC)}$ that allows the parasite to maintain blood-stage homeostasis during cellular stress [Hammam *et al.*, 2020].

1.4 Antigenic variation in *P. falciparum*

In order for infectious agents, including viruses, bacteria, fungi, and protozoa, to survive inside their host, they must avoid clearance by the host immune system. Similar to other microorganisms, *P. falciparum* uses antigenic variation as an evasion mechanism to evade the host immune response. This process consists of coordinated changes in the expression of its VSAs, which are displayed on the surface of RBCs and recognized by the host immune system [Deitsch *et al*., 2009].

As a result of the displayed VSAs, the host produces antibodies against them, leading to a reduction in iRBCs and thus decreases the overall parasitic load in an infected individual. However, before complete clearance, sub-populations arise, expressing a different VSA form, not yet recognized by the host antibody response. Surface antigen switching forces the host to continually attempt to detect and clear the pathogen, ultimately resulting in waves of chronic infection (Fig 1.10) [Miller *et al*., 1994],

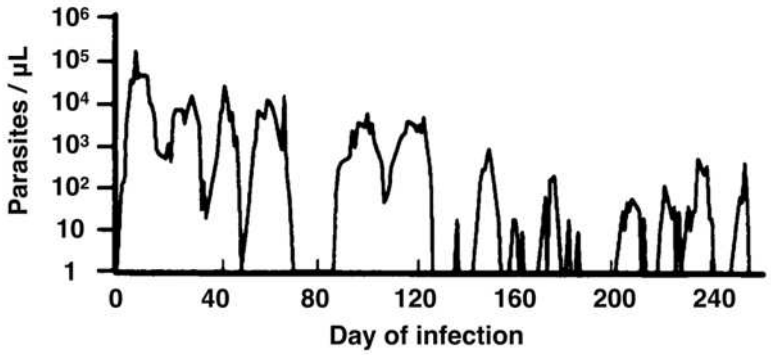


Figure 1.10: Waves of *P. falciparum* parasitemia. *P. falciparum* chronic infection in RBCs after a single mosquito bite. Levels of parasitemia are shown for a period of 260 days. Each wave represents a distinct sub-population that expresses a different VSA. Taken from [Miller *et al*., 1994].

There are two main mechanisms that coordinate the switching of the VSAs, genetic and epigenetic. Genetic mechanisms refer to mutation and recombination in DNA of the antigen or its regulatory elements, whereas epigenetic mechanisms do not alter the primary nucleotide sequence.

1.4 Antigenic variation in P. falciparum

Genetic mechanisms are used by African Trypanosomes in a kind of 'programmed variation' leading to a diversified antigen repertoire [Borst, 2003]. Epigenetic mechanisms allow for variation by switching between the expression of members of a gene family in a non-random order in cultured *P. falciparum* parasites [Noble *et al.*, 2013]. At this stage of our knowledge, epigenetically-mediated switching remains puzzling [Noble *et al.*, 2013]. Despite an underlying mechanism regulating switching, programmed variation may still result in random changes over time [Borst, 2003]. An overview of protozoan parasites using programmed antigenic variation and the involved gene families is shown below in Table 1.3.

Table 1.3: Multigene families encoding variant surface antigens in different protozoan parasites. Modified from [Deitsch *et al*., 2009].

Organism	Gene	Copy number	Variant surface antigen
Trypanosoma brucei	*vsg*	>1000	Variant surface glycoprotein
Babesia bovis	*ves and vesβ*	~130-160	Variant expressed surface antigen
Giardia lamblia	*vsp*	~150	Variant surface protein
Plasmodium falciparum	*var*	~60	Erythrocyte membrane protein 1

The most important VSA involved in antigenic variation and pathogenesis in *P. falciparum* are PfEMP1 proteins (*P. falciparum* erythrocyte membrane protein 1), encoded by the *var* multigene family [Leech *et al*., 1984b]. Unlike VSGs of trypanosomes, whose primary function is to delay the production of anti-trypanosome antibodies to avoid clearance by the humoral immune system, PfEMP1 is the main VSA responsible for the strict cytoadherence of mature iRBCs [Baruch *et al*., 1996]. This process is only seen in *P. falciparum* because the other human malaria species do not encode *var* genes. The sequestration of iRBCs obstructs blood flow and is responsible for the phenotypes of severe disease, cerebral and placental malaria (Fig 1.11) [*Kyes et al.*, 2001]. Cytoadherence, and its causative sequestration, of iRBCs prevents peripheral circulation of the parasitized cells through the spleen where they would be recognized and destroyed [Barnwell *et al*., 1983; Miller *et al*., 1994]. In addition to endothelial cells, late stage-iRBCs can also adhere to uninfected RBCs (uRBCs), forming rosettes [Udomsangpetch *et al*.,

1989]. This process is believed to help shield the VSAs on the iRBCs from immune-recognition by the host.

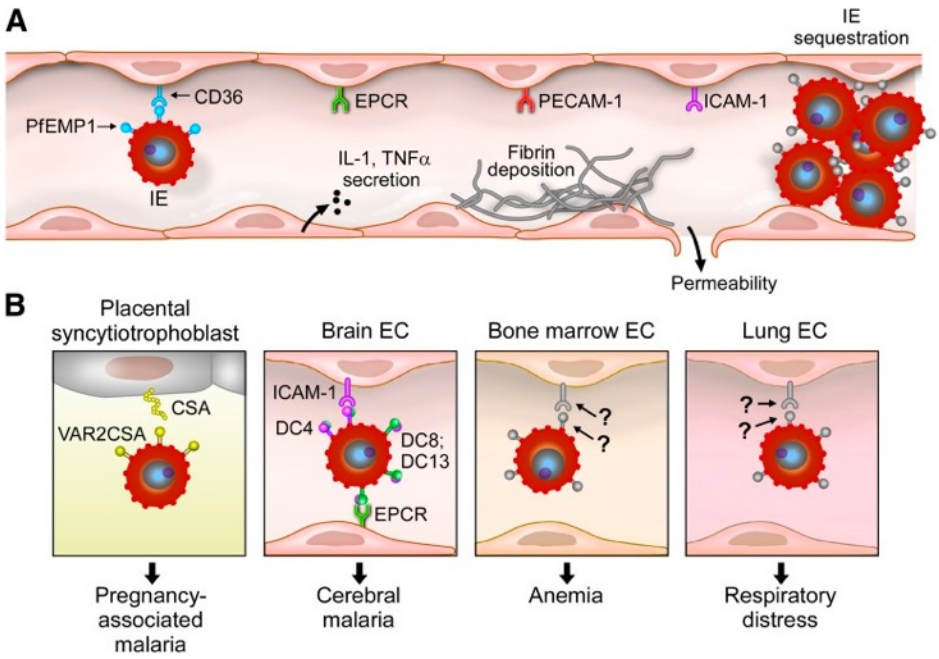


Figure 1.11: PfEMP1-endothelial receptor interactions mediate sequestration in infected erythrocytes (IEs). (A) Typical microvessel found in a variety of organs and tissues in patients with malaria. The parasite shown on the left is expressing a PfEMP1 variant that binds CD36 on endothelial cells (ECs). In addition to CD36, ECs may also express EPCR, PECAM-1, and ICAM-1. (B) IE sequestration in four different organs. Pregnancy-associated malaria is an organ-specific syndrome initiated by the expression of the PfEMP1 variant *var*2csa, which mediates IE binding to placental CSA. In brain ECs, cerebral malaria (CM) is initiated by the expression of a PfEMP1 *var*iant containing DC8 or DC13 (green circles), which mediates IE binding to EPCR. Respiratory distress and anemia are organ-specific malaria syndromes that may occur alone or in combination with CM. Taken from [Aird *et al.*, 2014].

In addition to *var* genes, several other clonally variant multigene families exist in *P. falciparum* that code for VSAs and undergo antigenic variation. However, only the *var* gene family shows strict singular gene expression, whereas the other families express a subset of family members at a given time. A description of these gene families is summarized below in Table 1.4.

Table 1.4: Multigene families encoding variant surface antigens in *P. falciparum*. Abbreviations: PV (parasitophorous vacuole), MC (Maurer's clefts).

	var	***rif***	***stevor***	***Pfmc-2TM***
Copy number	~60	~150	~30	~13
Gene location	Subtelomeric and central	Mainly subtelomeric	Mainly subtelomeric	Subtelomeric
mRNA	3-18 hpi	12-27 hpi	22-32 hpi	18-30 hpi
Function	Cytoadherence, Immune evasion	Cytoadherence, Immune evasion	Rosetting	Unknown
Immunogenicity	Yes	Yes	Yes	Unknown
Location	iRBC surface, MC	iRBC surface, MC, PV, merozoite	iRBC surface, MC, PV, merozoite	iRBC surface, MC, PV
Molecular weight	200-360 kDa	30-45 kDa	30-40 kDa	20-25 kDa

1.4.1 Mutually exclusive expression of the *var* gene family

P. falciparum additionally uses mutually exclusive gene expression for antigenic variation. This mechanism allows for the expression of one member of a large, multicopy gene family like *var* genes, while the remaining members are silenced. PfEMP1-coding *var* genes are a major player in immune evasion and pathogenesis for *P. falciparum* parasites [Scherf *et al*., 1998; Smith *et al*., 1995]. This gene family is conserved among species from the *Laverania* subgenus of *Plasmodium*, which includes *P. falciparum* and several ape species [Larremore *et al.*, 2015; Otto *et al*., 2018].

A recent study showed that var gene expression is not limited to blood stage parasites. A particular *var* gene (called var^{sporo}) was identified on the surface of salivary gland sporozoites (NF54 strain). Antibodies raised against the N-terminal var^{sporo} domains significantly reduced sporozoite invasion into hepatocytes [Zhangi *et al.*, 2018]. The biological role of var^{sporo} awaits further studies.

Despite advances in the understanding of regulation of this multigene family expression, major mechanistic questions concerning how a single member is activated, while the others remain silenced, are still unanswered. Epigenetics have been shown to play a major role in this highly coordinated regulation and several layers of control are required to achieve activation of one

member while simultaneously silencing the rest. The following sections discuss the organization, structure, and regulation of *var* genes.

1.4.1.1 *var* gene organization and structure within the genome

The different isolated *P. falciparum* strains share the same general organization and types of *var* genes. The genome of *P. falciparum* 3D7 strain, the strain used in culture throughout the course of this thesis, contains 60 *var* genes. The 60 *var* genes are classified into different groups depending on the 5' upstream sequence, position, and orientation within the chromosome, summarized below in Table 1.5 and Figure 1.12 [Gardner *et al*., 2002; Lavstsen *et al*., 2003; Kraemer *et al*., 2007]. Subtelomeric *var* genes are found in 13 of the 14 chromosomes whereas central *var* genes are restricted to chromosomes 4, 7, 8, and 12. The *var*2csa, coding for the PfEMP1 involved in pregnancy associated malaria, is found in the subtelomeric region of chromosome 12 and belongs to its own group, upsE [Lavstsen *et al*., 2003].

Table 1.5: Characteristics of *P. falciparum var* gene groups. Taken from [Kyriacou *et al*., 2006].

var gene group	Upstream sequence	Position	Orientation (direction of transcription)	No. of genes in 3D7
A	UpsA	Subtelomeric	Telomeric	10
B/A	UpsB	Subtelomeric	Centromeric	4
B	UpsB	Subtelomeric	Centromeric	21
B/C	UpsB	Central	Telomeric	10
C	UpsC	Central	Telomeric	13

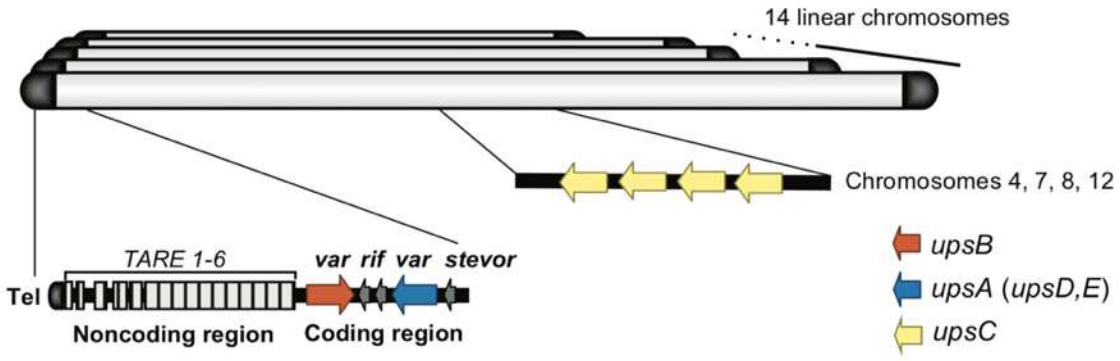


Figure 1.12: Genomic location and organization of *var* genes. Subtelomeric *var* genes are located adjacent to TAREs. Commonly, the most proximal to the TAREs is a member from the *var* upsB type followed by a *var* upsA type member transcribed in the opposite direction, towards the telomere. Four chromosomes present *var* clusters of the upsC type in central chromosomal regions and arrayed in tandems. *rif* genes can often be found interspersed within *var* genes. Taken from [Scherf *et al*., 2008].

Subtelomeric *var* genes tend to undergo more frequent recombination events, due to their location, contributing to a greater diversity in *var* repertoire [Freitas-Junior *et al*., 2000]. Also found in subtelomeric regions are other clonally variant multigene families encoding for VSA like *rif, stevor* and *Pfmc-2TM* [Gardner *et al*., 2002].

All *var* genes are composed of a 5′ upstream promoter region, followed by exon I, a relatively conserved intron, and exon II (Fig 1.13). Exon I codes for the extracellular domain of PfEMP1 while exon II codes for the intracellular parts of the PfEMP1 protein [Su *et al*., 1995].

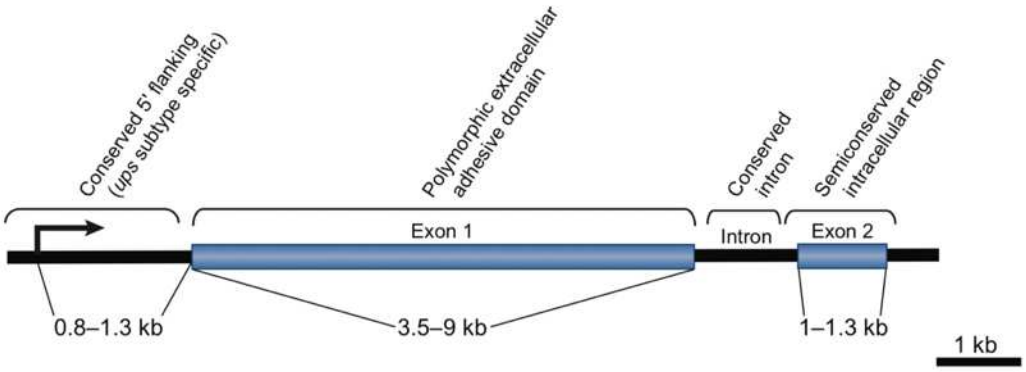


Figure 1.13: Structure and features of *var* genes. All *var* genes are composed of two exons with one intron between. Exon 1 codes for the polymorphic extracellular domain which contains variable numbers of Duffy-binding-like (DBL) adhesion domains and cysteine-rich interdomain regions (CIDR). Exon 2 encodes the intracellular domain, containing the transmembrane domain and a highly conserved acidic terminal segment (ATS). Taken from [Scherf *et al*., 2008].

1.4.1.1.1 *var* gene regulation

Throughout the IDC, *var* gene expression is tightly regulated [Kyes *et al*., 2007, Schieck *et al*., 2007]. The active *var*, transcribed by RNA Pol II, reaches its peak expression during ring stage, around 12hpi. In late stages, when the expressed PfEMP1 proteins are on the surface of the iRBC, the active *var* is repressed, while remaining poised in order to allow for transcription to be initiated in the following cycle (Fig 1.14) [Kyes *et al*., 2000; Gardner *et al*., 1996; Kriek *et al*., 2003; Lopez-Rubio *et al*., 2007].

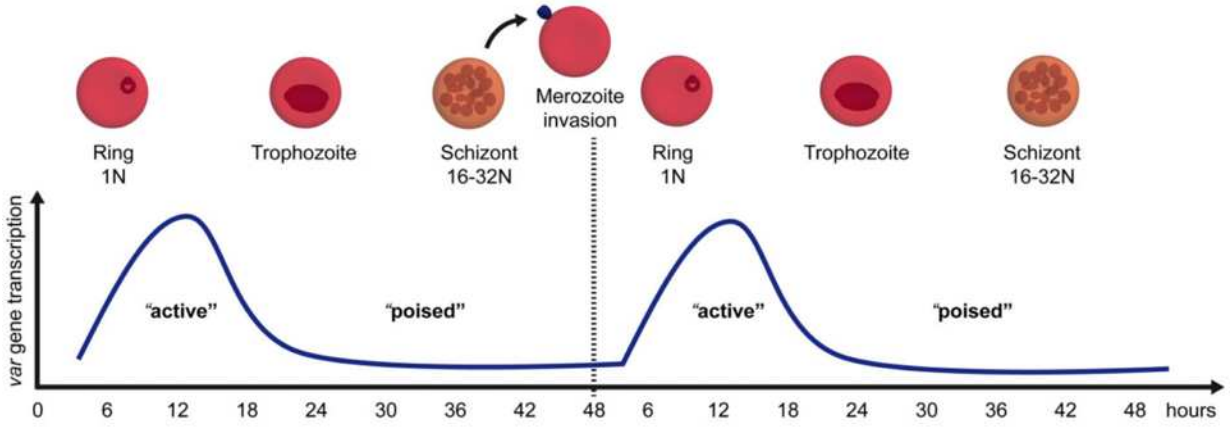


Figure 1.14: *var* gene transcription throughout the intraerythrocytic cycle. A single *var* gene is transcribed in ring stages parasites, and stops transcribing before DNA replication. The active *var* remains in a poised state during trophozoite and schizont stages and is reactivated during the next cycle in the majority of the population, while some parasites switch to another member. Taken from [Guizetti and Scherf, 2013].

As mentioned in section 1.3.3, *var* genes have distinct characteristics related to histone PTMs, chromatin state, nuclear localization, and binding-proteins. The single active *var* gene is enriched in euchromatic modifications, H3K9ac and H3K4me2/3, near TSSs [Lopez *et al*, 2007]. In addition, H4Kac and its potential writer PfMYST are also enriched in the promoter of the active *var* gene [Freitas-Junior *et al*., 2005; Miao *et al*., 2010]. The double histone variant PfH2A.Z/PfH2B.Z is only found during transcription at the active *var* gene [Bártfai *et al*., 2010; Petter *et al*., 2011; Petter *et al*., 2013]. PfH3.3 is enriched in the promoter region of the active and later poised *var* gene [Fraschka *et al*., 2016]. The poised *var* gene also maintains the activating mark H3K4me2 throughout the cycle [Lopez-Rubio *et al*., 2007] as well as the protein PfSET10 [Volz *et al*., 2012].

Silent *var* genes are commonly enriched in H3K9me3 which facilitates heterochromatin formation by recruiting PfHP1 [Lopez *et al*, 2007; Freitas-Junior *et al*., 2005; Chookajorn *et al*., 2007; Flueck *et al*., 2009; Pérez-Toledo *et al*., 2009]. Additionally, the marker H3K36me3 is characteristic of silent *var* genes and is controlled by histone methyltransferase (HMT) PfSETvs [Jiang *et al*., 2013]. Studies have also shown that the protein PfSir2, sirtuin (silent information

regulator) HDAC, is involved in *var* gene silencing [Tonkin *et al*., 2009]. Features of silent, active, and poised *var* genes are summarized below in Figure 1.15.

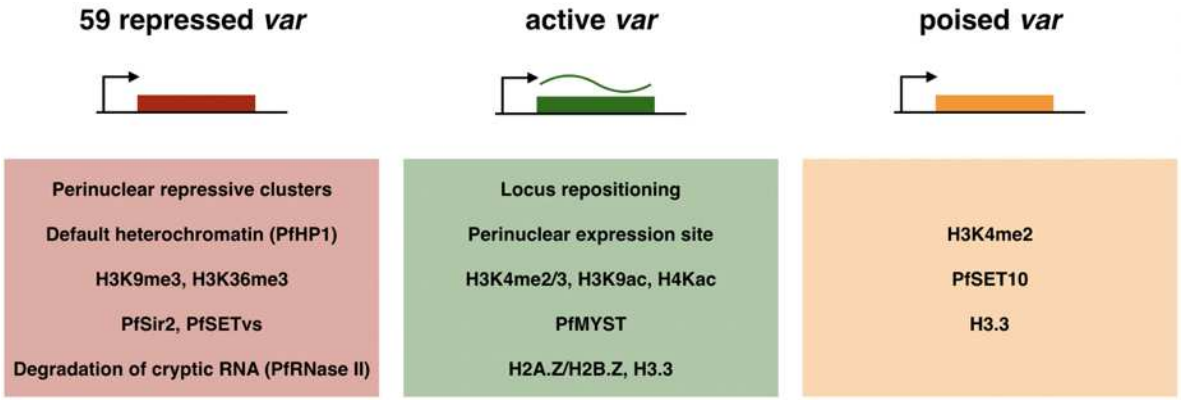


Figure 1.15: Characteristics of silent, active and poised *var* genes. Mechanisms, modifications and factors associated with the different states of *var* genes are summarized.

FISH studies revealed that nuclear organization of *var* genes is comprised of silent *var* genes clustered and tethered to the nuclear periphery, with central and telomeric *var* genes clustered independently of one another [Ralph *et al*., 2005b; Lopez-Rubio *et al*., 2009], and the sole active *var* gene localized to a perinuclear expression site apart from the repressive clusters (Fig 1.9) [Ralph *et al*., 2005b; Mok *et al*., 2008]. Upon switching, the newly active *var* gene repositions to a transcriptionally permissive site along the nuclear periphery, as shown by DNA and RNA fluorescence in situ hybridization (FISH) [Ralph *et al*., 2005b; Freitas-Junior *et al*., 2005; Voss *et al*., 2006; Lopez-Rubio *et al*., 2009].

Switching to a new active *var* gene occurs at low frequency with the rate dependent on the location [Horrocks *et al*., 2004]. For example, central *var* genes appear to be more stable, with switch rates of ~0-0.3% per cycle, while subtelomeric members have higher switch rates of at least ~1-2% per cycle [Frank *et al*., 2007]. It remains to be determined what are the chromatin factors that determine *var* switch rates.

More factors have been proposed to be associated or related to different *var* gene states. An ApiAP2 protein, PfSIP2, seems to promote heterochromatin formation and thus *var* silencing through its binding to specific promoter sequences in *var* genes [Flueck *et al*., 2010]. Antisense

ncRNA originating from the bidirectional *var* intron promoter has been reported to regulate activation of *var* genes [Amit-Avraham *et al*., 2015; Jing *et al*., 2018] although these studies relied on artificial systems and episomal expression. While deletion of the intron, in its genomic location, has shown it is not essential for either activation nor silencing of *var*2csa, it might potentially play a role in switching [Bryant *et al*., 2017]. Additionally, a family of ncRNA has recently been shown to be essential to *var* gene monoallelic expression, discussion further in section 1.5.2.3.

1.4.2 Other clonally variant gene families associated with virulence

In addition to *var* genes, the clonally variant *rif*, *stevor, surf*, and *Pfmc2TM* gene families are encoded within the *P. falciparum* genome. These gene families are expressed during the IDC and their protein products, RIFINS, STEVOR, SURFINSm and Pfmc2TM, are similarly displayed on the iRBC surface. It was when *rifin*, *stevor* and *Pfmc-2TM* families were shown to undergo switching [Kyes *et al*., 1999; Lavazec *et al*., 2007; Niang *et al*., 2014], that their involvement in antigenic variation was first suggested. However, while studies have shown these gene families are associated with virulence, their exact function is still largely unknown.

The multigene families *rif* (repetitive interspersed families of polypeptides), *stevor* (subtelomeric *var*iant open reading frame) and *Pfmc-2TM* (Maurer's cleft two transmembrane protein) are large families that transcribe 2 transmembrane (2TM) proteins. For these three families the 2TM domains are found flanking a hypervariable region that codes for the exposed part of the antigen (Fig. 1.16) [Lavazec *et al*., 2006]. They additionally contain a PEXEL motif required for the export of the proteins [Hiller *et al*., 2004; Marti *et al*., 2004].

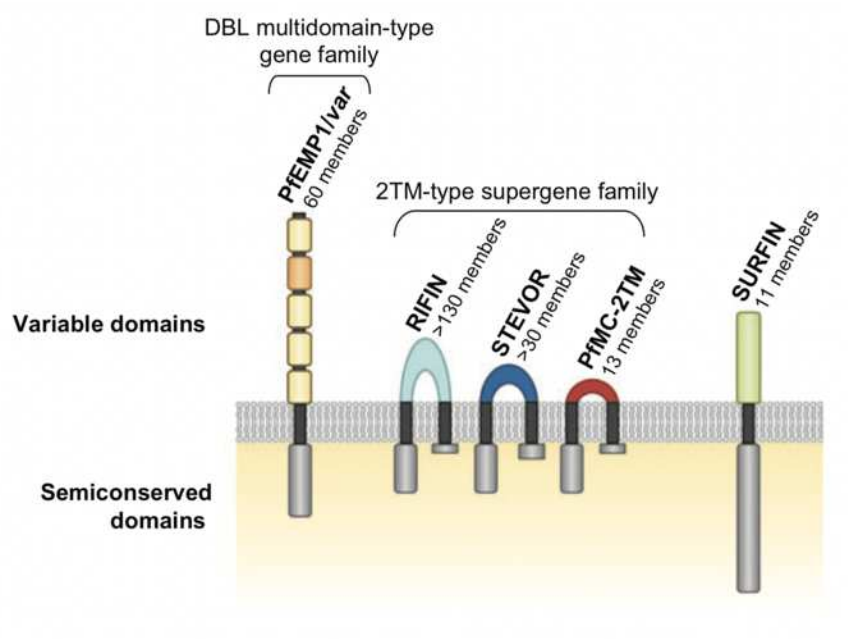


Figure 1.16: VSAs displayed on the surface of a *P. falciparum* iRBC. All molecules represented are trafficked via Maurer's clefts to the erythrocyte membrane. PfEMP1 are first exposed, at 18 hpi, with the others following later in the IDC. These proteins are coded by clonally variant gene families. Two predicted trans-membrane regions for RIFIN, STEVOR and PfMC-2TM are shown, suggesting a loop-formation at the iRBC surface. Taken from [Scherf *et al*., 2008].

These three gene families are conserved in the same species as *var* genes, in *Laverania* species [Otto *et al*., 2018b] with the majority located in subtelomeric regions, and the remaining within central chromosomal *var* gene clusters [Gardner *et al*., 2002]. RIFINs have been shown to participate in immune evasion by targeting two immune inhibitory receptors. Specifically, by binding leucocyte immunoglobulin-like receptor B1 (LILRB1) to inhibit the activation of B cells and natural killer (NK) cells, associated with severe malaria [Saito *et al*., 2017].

Resembling RIFINs protein structure, STEVOR proteins have been shown to mediate rosetting [Goel *et al*., 2015]. Unlike PfEMP1, RIFIN, STEVOR, and SURFIN proteins are additionally found on the surface of merozoites, to possibly mediate invasion [Petter *et al*., 2007; Khattab *et al*., 2008; Khattab and Meri, 2011]. An exact role for *Pfmc-2TM* remains unclear.

1.5 Regulatory non-coding RNA

Gene expression at the transcriptional and post-transcriptional levels, can additionally be regulated with the help of RNAs not transcribed into proteins, called ncRNAs. In recent years, ncRNAs emerged as key modulators of gene expression and have been shown to be involved in a broad variety of mechanisms such as mRNA transcription, mRNA splicing, and translational efficiency [Pearson & Jones, 2016; Wright & Bruford, 2011].

ncRNAs are mainly classified into two categories, according to their length: small RNA (<200 nucleotides (nt)) and long non-coding RNA (lncRNA) (>200 nt), although this is an arbitrary cutoff as the 180 nt B2 RNA is referred to as an 'honorary' lncRNA [reviewed by Long *et al*, 2017]. Indeed, mid-size ncRNAs have now recently begun to be characterized, which now refers to ncRNA between 50 and 400 nt (Fig 1.17) [Boivin *et al*, 2019].

Canonically, functional ncRNAs are small RNAs and include traditional RNA molecules, such as transfer RNA (tRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), and telomerase RNA component (TERC). Functional tRNA and rRNA are involved in mRNA translation, snoRNA is involved in rRNA modification, snRNA is involved in RNA splicing, and finally, TERC plays a role in telomere elongation. Regulatory ncRNAs, often referred simply as ncRNAs, include all the other ncRNAs and can have very diverse functions. Regulatory ncRNAs include microRNAs (miRNAs), mid-size ncRNAs, and long ncRNAs (lncRNAs) [Ponting *et al*., 2009; Ma *et al*., 2013]. Finally, natural antisense transcripts (NATs) are RNAs transcribed from the opposite strands of DNA that overlap completely or partially with the sense RNA [Li *et al*., 2008; Pelechano *et al*., 2013].

Mid and long ncRNAs can play roles in diverse functions, including transcription, RNA processing, RNA degradation, and translation, while being predominantly nuclear [Boivin *et al*, 2019; Kopp *et al*, 2018]. They are commonly transcribed by RNA Polymerase II and can interact with DNA, RNA, and proteins to modulate chromatin structure and function, and regulate transcription of genes in *cis* or *trans*. RNA-binding proteins (RBPs) can stabilize these ncRNA and assist in their transport to promoter regions of genes.

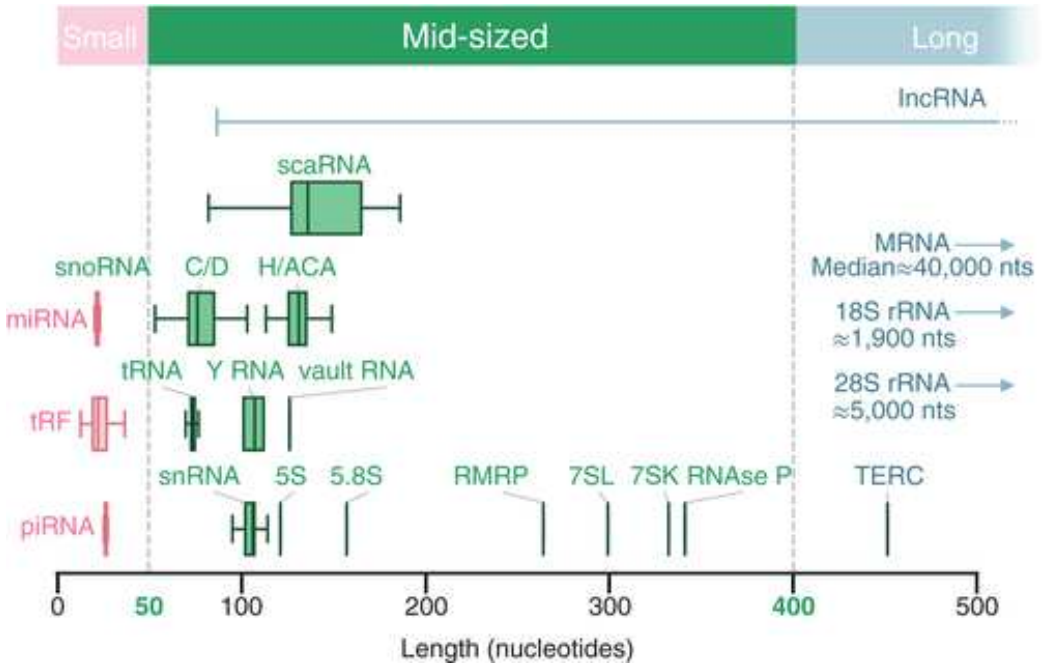


Figure 1.17: Types and size distribution of different RNA. The different types of RNAs are illustrated as a function of their length. The average size of small, mid-size and long RNA is indicated in red, green and blue boxes, respectively. Dotted lines indicate the size delimitation between these RNA families. The median size of mRNAs and the size of rRNAs that are too large to be visible on the figure are indicated on the right. Examples are for human RNA, taken from [Boivin *et al*., 2019].

1.5.1 Functions and mechanisms of action of regulatory ncRNA

As previously mentioned, regulatory ncRNA refers to short, mid-size, and long ncRNAs. Short ncRNAs, including small interfering RNA (siRNA) and miRNA have well defined functions related to mRNA degradation and stability (Table 1.6) including the RNA-interference (RNAi) pathway that mediates mRNA silencing [Fire *et al*., 1998]. Mid and long ncRNAs, on the other hand, are still being characterized and have been shown to play an important role in a broad range of functions with regulation of transcription by chromatin modulation as the top studied. For example, lncRNAs have been shown to facilitate the formation of chromatin loops between distant enhancers and promoters [Arnold *et al*, 2020; Postepska-Igielska *et al*, 2015; Zhu *et al*, 2020; Cajigas *et al*, 2018; Tan *et al*, 2019]. Because ncRNAs have only recently been further classified into mid-size ncRNAs, most studies refer to the ncRNA studied as lncRNA. For this reason, in the following sections, mid-size ncRNA will be referred to as long ncRNA.

Table 1.6: Examples of regulatory ncRNAs. Modified from [Gomes *et al.*, 2013; Li *et al.*, 2020].

	ncRNA full name	Class, size	Function
siRNA	small interfering RNA	short ncRNA	mRNA degradation
miRNA	micro RNA	short ncRNA, 22-23 nt	mRNA stability
piRNA	piwi-interacting RNA	short ncRNA, 26-31 nt	Transposon silencing
paRNA	promoter-associated RNA	medium ncRNA	Transcription repression
eRNA	enhancer RNA	lncRNA	Promoter activation
NAT	natural antisense transcript	lncRNA	mRNA stability
lincRNA	long intergenic ncRNA	lncRNA	Transcription control
iRNA	intronic RNA	lncRNA	Transcription control
satellite	satellite ncRNA	lncRNA	Centromere assembling
repeat ncRNA	repeat-associated ncRNA	lncRNA	Repeat silencing
pseudogene RNA	pseudogene ncRNA	lncRNA	mRNA stability

1.5.1.1 Chromatin remodeling and transcriptional regulation by ncRNA

lncRNAs constitute an important class of epigenetic regulators to control chromatin structure. Unique to lncRNAs is their biochemical ability to form RNA-RNA, RNA-DNA, and RNA-protein complexes through the interaction with a wide range of molecules to specific RNA functional domains [Quinn *et al*., 2014]. Additionally, the folding of RNA into its 3D structure allows for protein recognition and binding.

Recently, lncRNAs have been shown to control chromatin structure thereby regulating gene expression specifically by directly interacting with histone- and DNA-modifying enzymes to modify histones or DNA (Fig 1.18) [Rinn *et al*., 2014; Rinn *et al*., 2012; Devaux *et al*., 2015]. For example, they can directly bind to chromatin-modifying enzymes to serve as a guide for chromatin modifiers to target genomic regions. Long non-coding RNAs (lncRNAs) can interact with chromatin modifiers and recruit them to target-gene promoters in order to activate or suppress their transcription in *cis* [Yap *et al*., 2010; Rosa *et al*., 2016; Csorba *et al*., 2014], or in *trans* at distant loci [Holdt *et al*., 2013]. An example is shown in Figure 1.18.

LncRNAs can regulate transcription by acting as scaffolds for chromatin- and histone-modifying complexes and orchestrate their organization spatially and temporally [Pandey *et al*.,

2008; Tsai *et al*., 2010; Khalil *et al*., 2009; Spitale *et al*., 2011; Yap *et al*., 2010]. Additionally, chromatin can switch between heterochromatin to its open conformation, euchromatin, from RNA binding to histones. The negative charge on RNA can neutralize positively charged histone tails leading to their de-compaction [Dueva *et al*., 2019]. Direct binding of lncRNAs to DNA in *cis* or *trans,* can facilitate or inhibit proteins from binding and thus, their activity. Furthermore, lncRNAs transcription can influence the binding of regulatory factors [Hirota *et al*., 2008; Lefevre *et al*., 2008; Baumgarten *et al*., 2012].

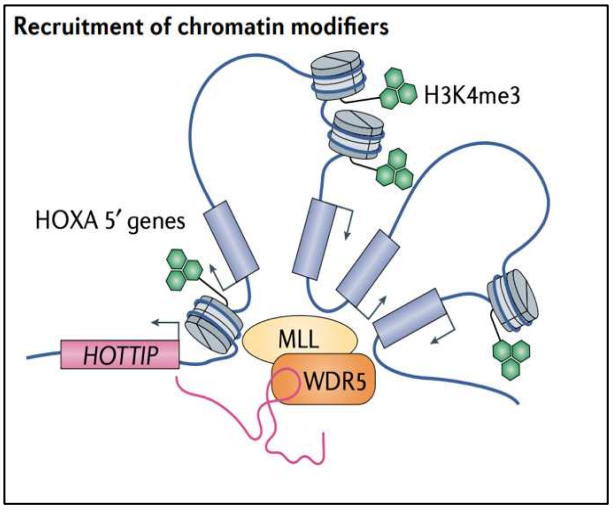


Figure 1.18: Chromatin regulation mediated by long non-coding RNAs. Schematic of a lncRNA interacting with chromatin modifiers and recruiting them to target-gene promoters, to activate or suppress transcription, in cis or trans. HOXA transcript at the distal tip (HOTTIP) [Wang et al., 2011; Luo et al., 2019] acts in cis at the 5′ genes of the HOXA gene cluster, with which it interacts through chromatin looping. HOTTIP interacts with WD repeat-containing protein 5 (WDR5), thereby targeting the complex to the promoters and promoting histone H3 Lys4 trimethylation (H3K4me3). Taken from [Statello et al., 2021].

Transcription can also be regulated by nuclear lncRNAs, through the modification of histone or DNA modifications, primarily methylation and acetylation. lncRNAs can regulate histone methylation or acetylation alone [Zuckerman *et al*., 2020], or can act as a scaffold that interacts with enzymes or methylation and acetylation complexes, thereby regulating histone methylation and acetylation simultaneously [Carlevaro-Fita *et al*., 2016].

Another mechanism by which lncRNAs regulate transcription is by directly binding to RNA Pol II. For example, the 331nt 7SK lncRNA binds RNA Pol II in *trans* and prevents the PTEFβ transcription factor binding, thus affecting transcriptional output of the cell [Peterlin *et al*, 2012]. Another example are ncRNAs from Alu SINEs (short interspersed nuclear elements) that also bind to RNA Pol II and repress transcription, of specific genes, upon heat shock [Mariner *et al*., 2008]. The assembly of the transcription protein initiation complex (PIC) can be hindered as a

result of lncRNAs forming RNA:DNA triplex structures [Martianov *et al*., 2007] or interact directly with TFs to either activate or repress transcription [Shamovsky *et al*., 2006; Kino *et al*., 2010]. Schematics showing some mechanisms of transcriptional regulation by ncRNA are illustrated below in Fig. 1.19.

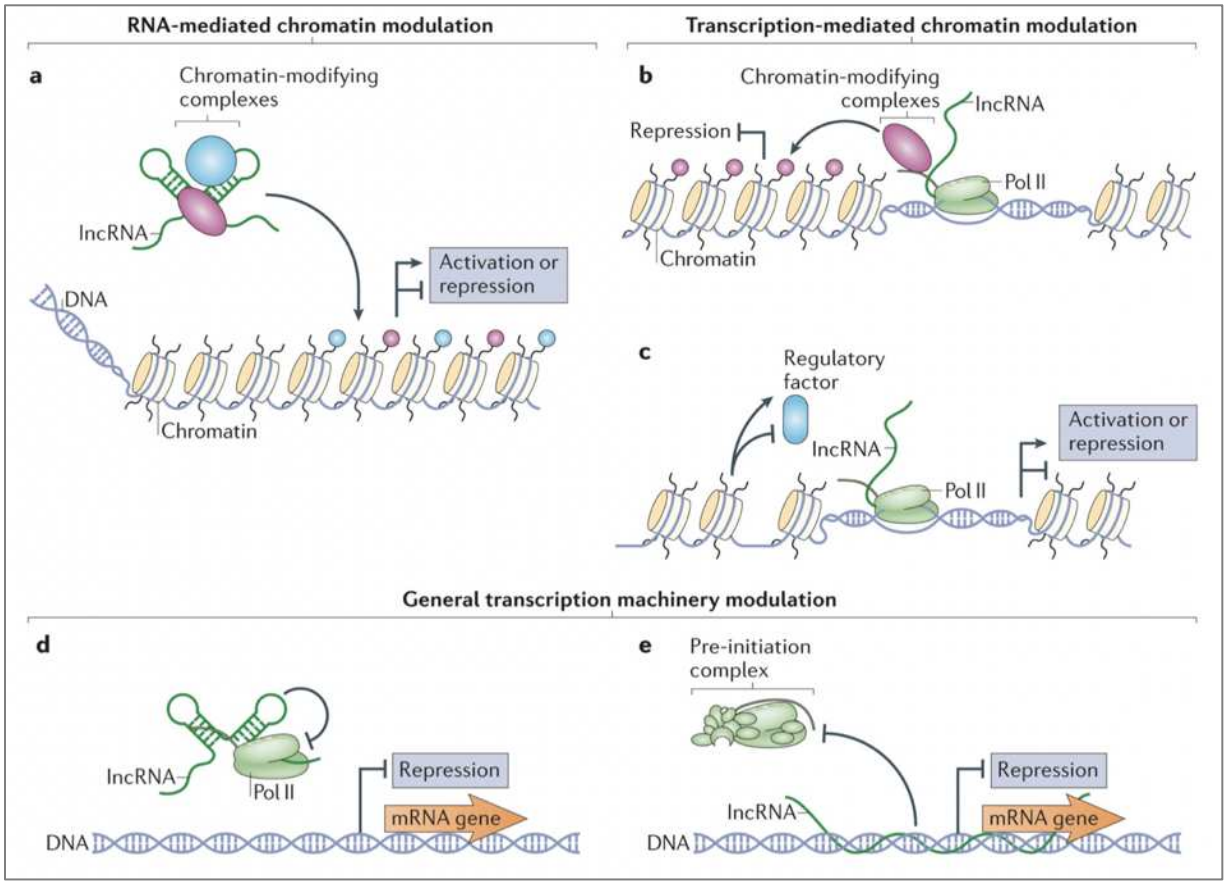


Figure 1.19: Examples of ncRNA-mediated transcriptional regulation. (a) LncRNAs can target chromatin-modifying complexes, or TFs, to specific loci. (b) Chromatin-modifying complexes can be bound to RNA Pol II and act while lncRNAs are being transcribed. (c) Transcription of lncRNAs can also cause either activating or repressing chromatin remodeling. (d) LncRNAs binding to RNA Pol II represses transcription of target genes. (e) LncRNA-DNA triplex structures inhibit assembly of the PIC and repress transcription. Taken from [Geisler and Coller, 2013].

Surprisingly, enhancer regions, short DNA sequences where proteins bind to increase transcription potential, are actively transcribed [De Santa *et al*., 2010; Kim *et al*., 2010] giving way to an additional class of ncRNAs, enhancer RNAs (eRNAs). While their size correlates to

lncRNAs, their specific histone methylation markers make them a unique category. Studies have shown that these eRNAs have many diverse roles and in some cases the eRNA transcript is required for the activity of its enhancer (Fig. 1.20) [Ørom *et al*., 2010; Wang *et al*., 2011].

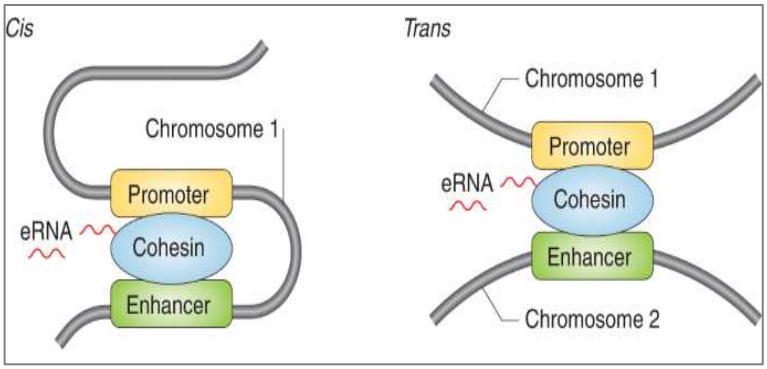


Figure 1.20: Stabilization of enhancer-promoter looping. Schematic diagram depicting the roles of eRNAs acting in *cis* (left) to regulate specific enhancer-promoter looping by supporting cohesin binding at enhancers and influencing chromatin accessibility by regulating recruitment of cohesin complex in *trans* (right) to facilitate gene expression. Taken from [Sartorelli *et al*., 2020].

1.5.1.2 Post-transcriptional regulation by ncRNA

Post-transcriptional regulation, specifically mRNA processing, by lncRNAs occurs by forming lncRNA-protein complexes (lncRNPs) that can affect mRNA splicing and output. This happens when splicing factors bind to the same motif on lncRNAs rather than on pre-mRNAs [Yin *et al*., 2012; Wu *et al*., 2016; Yap *et al*., 2018]. Furthermore, while the mechanism is still unknown, NATs can regulate splicing of their associated mRNAs mRNAs [Beltran *et al*., 2008; Hastings *et al*., 1997; Munroe and Lazar, 1991; Faghihi and Wahlestedt, 2009].

After mRNA processing, lncRNAs can be involved in regulation of translational efficiency [Carrieri *et al*., 2012] by positively or negatively regulating mRNA stability [Faghihi *et al*., 2008; Gong and Maquat, 2011]. lncRNAs can also stabilize target mRNA by binding to regions where miRNAs, involved in mRNA repression, would bind [Faghihi *et al*., 2008]. More so, miRNAs can additionally bind to lncRNAs rather than mRNA, ultimately preventing the interaction of miRNA with mRNA targets [Salmena *et al*., 2011; Cesana *et al*., 2011; Wang *et al*., 2010].

1.5.2 ncRNA in P. *falciparum*

Interest in the involvement of ncRNAs in *P. falciparum* has been growing, notably in the regulation of genes associated with virulence. This interest began largely after 189 RNA-binding proteins were discovered, upon sequencing the *P. falciparum* genome, which highlighted the potential of RNA-mediated regulation in the parasite [Gardner *et al.*, 2002]. Furthermore, for a large number of genes, widespread and developmentally regulated long antisense transcripts has been reported [Siegel *et al.*, 2014; Vembar *et al.*, 2014]. Overall, ncRNA studies are still an emerging field in malaria parasites with only a few examples that have a biological function determined.

1.5.2.1 Human miRNA able to modulate *P. falciparum* RNA

RNA interference (RNAi) pathway uses double-stranded RNA to regulate the expression of protein-coding genes by silencing of genes through a sequence-specific fashion. miRNAs are the sequence-specific component of the RNA-induced silencing complex (RISC). While it appears that *P. falciparum* parasites lack the components of the RNAi pathway [Gardner *et al*., 2002; Baum *et al*., 2009] and no associated miRNAs have been detected [Xue *et al*., 2008], evidence suggests that human-derived miRNAs are imported into *P. falciparum* where the parasite uses them to modulate its own gene expression. It is suggested that these imported RNAs interact with mRNAs by a process of *trans*-splicing and prevent ribosomes from loading thus repressing translation [LaMonte *et al*., 2012]. The biological impact of this work needs further validation.

1.5.2.2 Subtelomeric lncRNA

Using a high-resolution DNA tiling array, Broadbent *et al*. [Broadbent *et al.*, 2011] investigated the role of lncRNAs and identified a family of 22 telomere-associated lncRNAs, termed lncRNA-TARE (telomere-associated repeat elements). These lncRNA-TAREs were shown to exhibit coordinated expression during DNA replication and contain the binding sites, SPE2, for PfSip2 suggesting that PfSip2 regulates the lncRNA-TARE locus [Flueck *et al.*, 2010; Broadbent *et al.*, 2011]. Sierra-Miranda *et al*., additionally discovered that TARE-3 and TARE-6

lncRNA that colocalizes to a perinuclear compartment, linked with heterochromatin, during ring stage. Later, in schizont stage, the full length ncRNAs form nuclear foci [Sierra-Miranda *et al*. 2012].

1.5.2.3 Natural antisense transcripts

NATs have been identified in *P. falciparum* and are thought to have a role in gene regulation. Transcription of a particular NAT has been shown to be highly correlated to that of its related mRNA, coded by the gene PfGDV1 (P. falciparum gametocyte development 1), an activator of sexual commitment [Broadbent *et al*., 2015]. Sexual commitment to gametocytogenesis, triggered by environmental factors, has been shown to be activated by GDV1 expression under the control of its NAT [Filarsky *et al.*, 2018]. Specifically, PfGDV1 NAT has been shown to target heterochromatin to evict the *ap2-g* repressor, HP1 [Filarsky *et al.*, 2018].

1.5.2.4 Pf-*var*-aslncRNAs

A new candidate in *var* gene regulation are *var* gene antisense long non-coding RNAs (aslncRNAs). Transcription of a particular one of these nuclear ncRNAs, originating from the *var* intron until exon I (Fig 1.22) [Epp *et al*., 2009], has been proposed to activate its corresponding *var* gene [Jiang *et al*., 2013; Avraham *et al*., 2015; Jing *et al*, 2018]. These studies are based on episomal constructs. Contrarily, other studies state that these ncRNAs are either not overall correlated with *var* gene activation or silencing, or essential [Ralph *et al*., 2005a; Bryant *et al*., 2017]. Bryant *et al*. demonstrated that upon deletion of the *var2csa* intron, *var* gene activation was not inhibited implying that aslncRNA is not essential for *var* gene activation [Bryant *et al*., 2017]. It still debated what is the biological role of intron derived ncRNA in *var* transcriptional regulation.

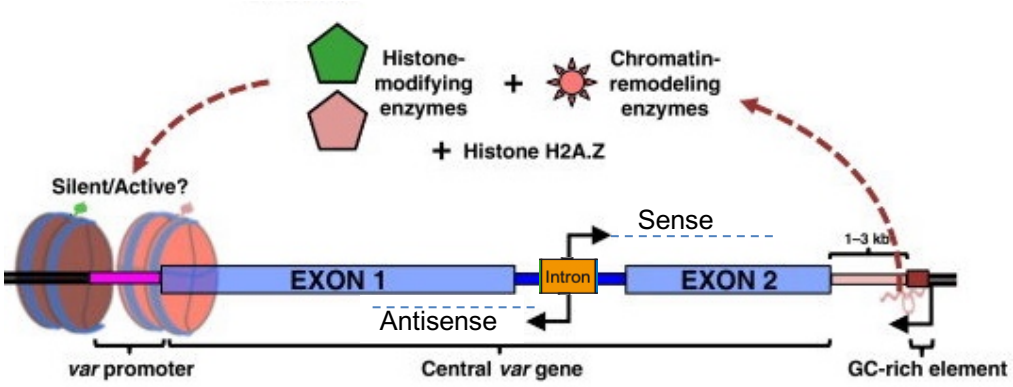


Figure 1.22: Illustration of *var* gene and regulatory elements. A central *var* gene, composed of a promoter, exon I, intron, and exon 2. Sense and antisense lncRNAs are transcribed from the intron. Upstream at the 5' end of some central *var* genes are GC-rich elements, transcribed into ncRNA. Modified from [Vembar, *et al.*, 2014].

1.5.2.5 RUF6 ncRNA

In addition to the *var* gene intron-derived lncRNAs, another group of ncRNAs have been found to be related to virulence in the pathogen. A family of mid-sized ncRNAs, 136 nt, was discovered after identifying conserved sequences with high GC-content, more than 50%, in intergenic regions of central chromosomal *var* clusters (Fig 1.23) [Hall *et al*., 2002; Hyman *et al*., 2002; Gardner *et al.*, 2002]. This GC-content is high for *P. falciparum*, considering its genome has ~20% GC content genome-wide with the highest AT-content than any other organism known to date and the transcriptional peak for these ncRNAs occurs during late ring stages, as shown by transcriptomic studies [Mourier *et al.*, 2008; Otto *et al*., 2010; Broadbent *et al*., 2011; Siegel *et al*., 2014; Wei *et al.,* 2014]. Interestingly, this ncRNA gene family is conserved only in *Plasmodium* species from the Laverania subgenus that also contain *var*, *rif* and *stevor* virulence gene families [Otto *et al*, 2018], suggesting a regulatory role in virulence gene expression.

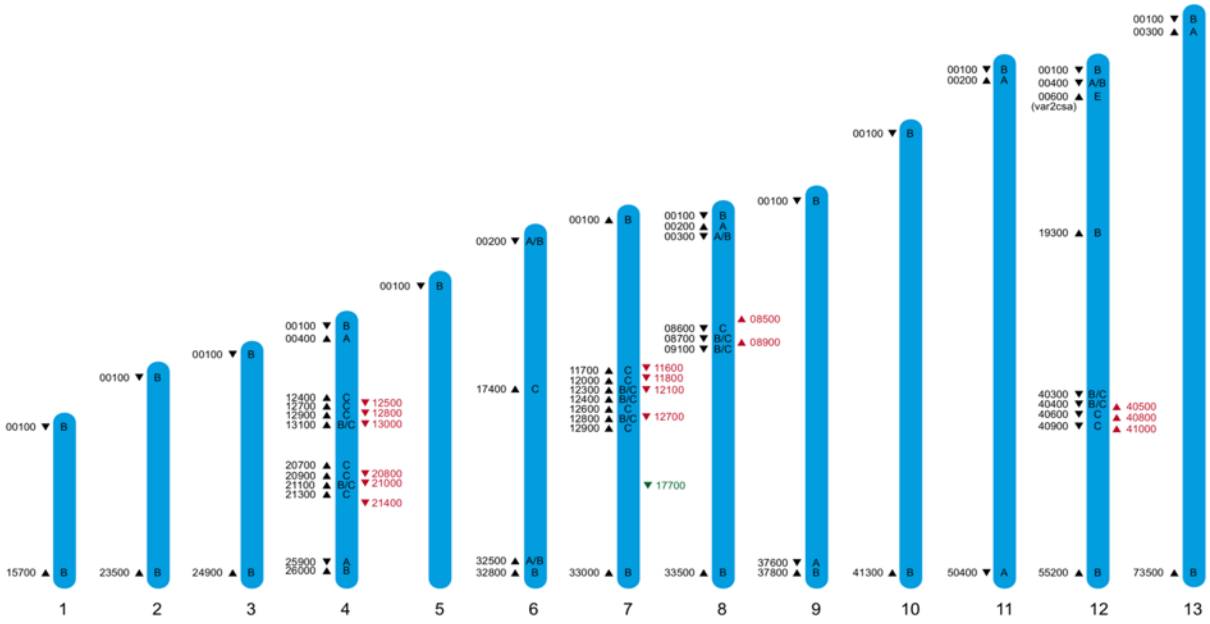


Figure 1.23: Genomic organization and sequence alignment of RUF6 ncRNA gene family. Chromosomal location of *var* genes (black arrowheads), GC-rich ncRNA elements (red arrowheads) and T-serine ligase gene (green arrowhead) within the 3D7 genome (modified from Fastman *et al*., 2012). Only five last digits of gene numbers are displayed (PF3D7_chr#xxxxx). Direction of the arrowhead indicates orientation of gene. Drawing is not to scale and chromosome 14 is omitted since it does not contain any *var* genes or GC-rich elements. Promoter subtype for each *var* genes is shown (ups A, A/B, B, B/C, C, E). Taken from [Guizetti *et al.*, 2016].

The GC-rich sequences code for a novel family of 15 highly homologous ncRNAs (89.9-100% conserved, Fig 1.24), annotated as RNA of unknown function 6 (RUF6), with loci interspersed between central *var* and *rif* genes, [Chakrabarti *et al*., 2007; Mourier *et al*., 2008; Panneerselvam *et al*., 2011]. Their positioning relative to *var* genes and their high sequence conservation amongst the different members first suggested they play a role in the regulation of *var* expression (Fig. 1.28). Unlike other regulatory mid and long ncRNAs, RUF6 members contain the A- and B-box that is only found at RNA Pol III-transcribed tRNA genes, and short interspersed nuclear elements (SINE) in other organisms [Guizetti *et al*., 2016].

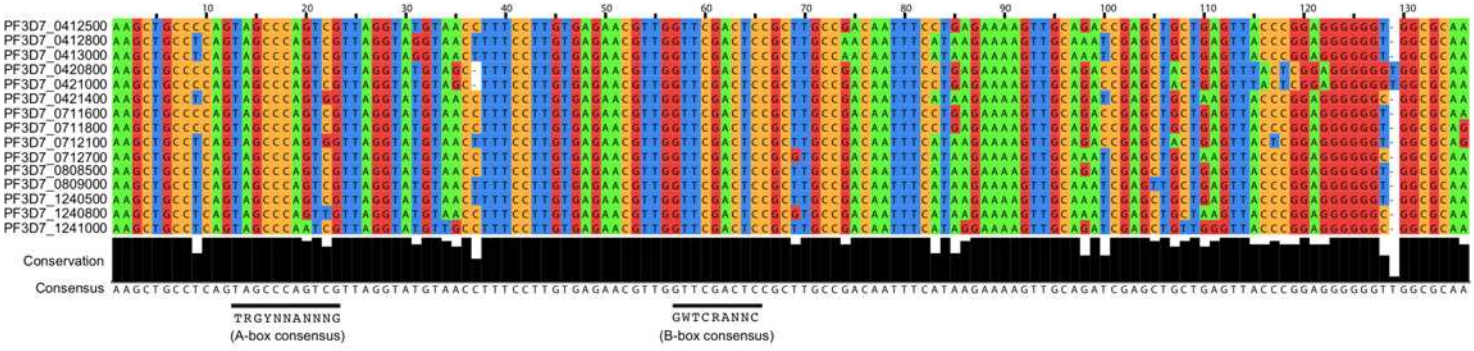


Figure 1.24: Genomic organization and sequence alignment of RUF6 ncRNA gene family. Multiple sequence alignement of highly conserved GC-rich non-coding RUF6 RNA elements. 15 members of RUF-6 gene family aligned by Clustal Omega (http://www.ebi.ac.uk) and presented in Jalview (http://www.jalview.org). Degree of conservation per base and consensus sequence are displayed below. Black lines show position of potential A- and B-box consensus motifs (as assessed in [Dieci *et al*., 2013]). Taken from [Guizetti *et al.*, 2016].

Recently, two studies have shown that RUF6 ncRNA is linked to the regulation of mutually exclusive transcription of the *var* multigene family [Guizetti *et al,* 2016; Barcons-Simon *et al,* 2020]. First, through fluorescence *in situ* hybridization (FISH), these ncRNAs were shown to colocalize to the *var* gene expression site. Second, episomal overexpression of a distinct RUF6 member resulted in the deregulation of mutually exclusive *var* gene expression [Guizetti *et al,* 2016]. A small subset of *var* genes was upregulated in RUF6 transfected parasites. In addition, downregulating the entire RUF6 gene family, provided further evidence for the necessity of RUF6 ncRNA in mutually exclusive *var* gene activation [Barcons-Simon *et al,* 2020]. Barcons-Simon *et al*. developed a CRISPRi approach, targeting all 15 GC-rich members simultaneously, and confirmed their implication in *var* activation, as ncRNA down-regulation led to *var* gene family repression.

Steady state transcriptional analysis, of the 15 RUF6 ncRNA members, indicated that when a single RUF6 gene is predominantly transcribed, it is always found adjacent to the 5' region of an active central *var* gene [Barcons-Simon *et al*., 2020]. Conversely, when an active subtelomeric or central *var* lacking a RUF6 gene, with this configuration, RUF6 transcription occurs from multiple loci at lower levels (Fig 1.25) [Barcons-Simon *et al*., 2020]. It has previously been reported that switch rates of *var* genes for central members are less prone to switch than

subtelomeric [Frank *et al*., 2007]. It is therefore speculated that adjacent RUF6 transcription could aid in opening chromatin for transcriptional machinery to be accessible at the *var* promoter or stabilize the open conformation. In fact, a recent study, using the Assay for Transposase-Accessible Chromatin by sequencing (ATAC-seq), showed increased accessibility at GC-rich genes adjacent to actively transcribed *var* and *rif* genes [Ruiz *et al.,* 2018].

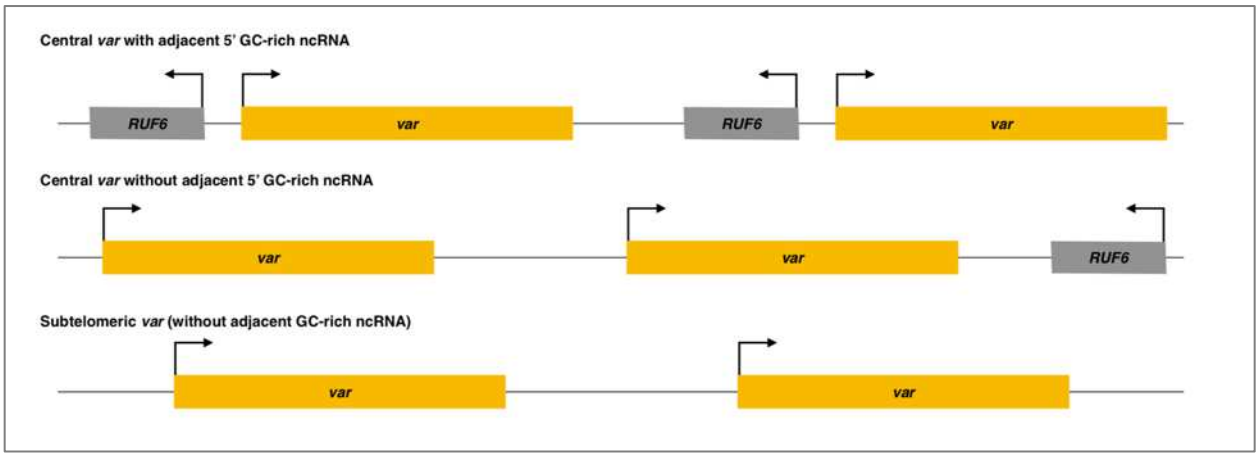


Figure 1.25: Illustration showing GC-rich RUF6 ncRNA and *var* genes relative chromosomal location. Three different possibilities for the chromosomal location of *var* genes relative to that of GC-rich RUF6 genes are shown. Only a subset of central *var* genes harbor RUF6 genes adjacent to the 5' upstream sequence of a *var*. These *var* genes, once activated, are less prone to switch then subtelomeric var genes or central *var* genes with no adjacent RUF6.

1.6 Aim of thesis:

Because the vast functions of mid and long-ncRNAs are primarily facilitated by RNA-binding proteins (RBPs), it is likely that interacting proteins play a role in the mechanism of action for RUF6 ncRNA and the activation of a *var* gene locus (Fig 1.26) [Mohibi *et al*, 2019; Gerstberger *et al*, 2014]. Out of the 189 RNA-binding proteins, discovered in *P. falciparum*, functions for the majority are still undetermined [Gardner *et al,* 2002; Vembar *et al,* 2014; Reddy *et al*, 2015]. These data all led to formation of the aim of my PhD project, to identify the RUF6 ncRNA-binding protein interactome to gain insight into key questions: how RUF6 ncRNA is targeted to the *var* expression site and how this ncRNA promotes singular *var* gene activation. At the beginning of my thesis in April 2019, the schematic model, shown below, of the proposed action of RUF6 lacked any interacting protein partners.

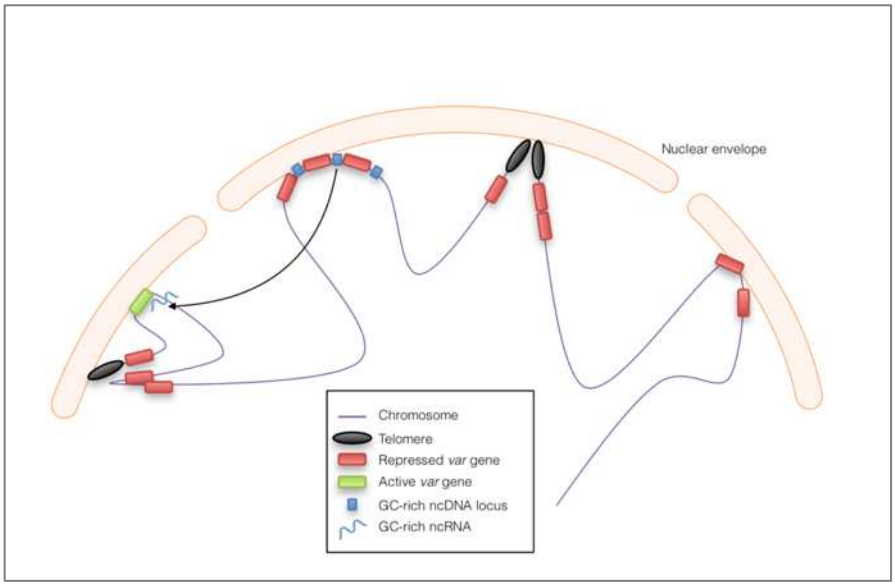


Figure 1.26: Schematic model of the role of RUF6 ncRNA in *var* activation in the perinuclear expression site. According to the model, RUF6 ncRNA acts in *trans* to activate a singular member of the *var* gene family.

1.6 Aim of thesis:

1.6 Aim of thesis:

Part 2

2 Results

Contents:

This second part contains the main results of my thesis, linked to the study of RUF6 ncRNA. Each section is presented in the format of an independent manuscript including the summarized methods used and a brief discussion. Section 2.1 corresponds to results currently submitted for review. Section 2.2 describe results in preparation for submission. Additional preliminary results can be found in Appendices A-C.

2.1 Discovery of RUF6 ncRNA interacting proteins

The highlights of this section are:

- Optimizing ChIRP-MS in *P. falciparum*
- Validating RUF6-binding candidate proteins
- Functional studies of RUF6-binding DDX5 protein

Results from this manuscript are currently submitted for review. My contribution to the work was conceptualization, experimental design, constructs and strains generation, performing all of the experiments (with the exception of the EMSAs), analyzing the data, and writing the manuscript.

Discovery of RUF6 ncRNA interacting proteins involved in *P. falciparum* immune evasion

Diffendall, G[1,2]., Barcons-Simon, A[1,2,5]., Baumgarten, S.[3], Dingli, F[4]., Loew, D[4]., Scherf, A[1,*]

[1]*Universite Paris Cité, Institut Pasteur, Biology of Host-Parasite Interactions Unit, INSERM U1201, CNRS EMR9195, Paris, France.*

[2]*Sorbonne Université Ecole doctorale Complexité du Vivant ED515, Paris, France.* [3]*Plasmodium* RNA Biology, Institut Pasteur, Paris, France

[4]*Institut Curie, PSL Research University, Centre de Recherche, CurieCoreTech Mass Spectrometry Proteomics, Paris, France*

[5]*Biomedical Center, Division of Physiological Chemistry, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich 82152, Germany*

**Corresponding author: artur.scherf@pasteur.fr*

Abstract:

Non-coding RNAs (ncRNAs) are emerging regulators of immune evasion and transmission of *Plasmodium falciparum.* RUF6 is a ncRNA gene family that is transcribed by RNA Polymerase III but actively regulates the Pol II-transcribed *var* virulence gene family. Understanding how RUF6 ncRNA connect to downstream effectors is lacking. We developed an RNA-directed proteomic discovery (ChIRP-MS) protocol to identify *in vivo* RUF6 ncRNA protein interactions. The RUF6 ncRNA interactome was purified with biotinylated antisense oligonucleotides. Quantitative label-free mass spectrometry identified several unique proteins linked to gene transcription including: RNA Pol II subunits, nucleosome assembly proteins, and a homologue of Dead-Box Helicase 5 (DDX5). Affinity purification of Pf-DDX5 identified proteins originally found by our RUF6-ChIRP protocol, validating the technique's robustness for identifying ncRNA interactomes in *P. falciparum*. Inducible displacement of nuclear Pf-DDX5 resulted in significant down-regulation of the active *var* gene. Our work identifies a RUF6 ncRNA protein complex that interacts with RNA Pol II to sustain *var* gene expression, including a helicase that may resolve G-

quadruplex secondary structures in *var* genes to facilitate transcriptional activation and progression.

Introduction:

Non-coding RNAs (ncRNAs) are increasingly recognized as key players in all aspects of the cellular life cycle (Cech & Steitz, 2014) ncRNAs are mainly classified into two categories, according to their length: small RNA (<200 nucleotides (nt)) and long non-coding RNA (lncRNA) (>200 nt), although this is an arbitrary cutoff as mid-size ncRNAs ranging from 40 to 400 nt emerge as regulatory RNAs with diverse functions (Boivin et al., 2019). Mid and long ncRNAs can play roles in diverse functions, including transcription, RNA processing, RNA degradation, and translation, while being predominantly located in the nucleus (Boivin et al., 2019; Kopp & Mendell, 2018). They are characterized by the absence of protein-coding capabilities, and commonly transcribed by RNA Polymerase II (Erdmann et al., 2001; Fernandes et al., 2019). Of note, the B2 RNA is transcribed by RNA Pol III yet it can directly bind to RNA Pol II and regulate its activity (Allen et al., 2004). ncRNAs can directly interact with DNA or RNA, while RNA-binding proteins commonly contribute in regulatory functions. RNA-binding proteins (RBPs) can stabilize ncRNA and assist in their transport to promoter regions of genes. RBPs have also been shown to facilitate the formation of chromatin loops between distant enhancers and promoters (Arnold et al., 2020; Cajigas et al., 2018; Postepska-Igielska et al., 2015; Tan-Wong et al., 2019; Zhu et al., 2020).

The human malaria protozoan parasite *P. falciparum* remains a global health burden that claims hundreds of thousands of lives each year (W.H.O., 2021). This pathogen lacks the canonical RNAi machinery, including the key enzyme Dicer that processes transcribed shRNAs into siRNAs (Baum et al., 2009)**.** Developmentally regulated long ncRNAs have been first described, originating from the promotor located in the intron of a virulence gene encoded by the *var* gene family (Calderwood et al., 2003; Su et al., 1995). Transcription of lncRNA originating from the *var* gene intron was observed to correlate with *var* gene activation (Amit-Avraham et al., 2015; Jiang et al., 2013; Jing et al., 2018). However, Bryant *et al* showed that deletion of the endogenous *var* gene intron using genome editing did not block transcriptional silencing or activation of the targeted *var* gene but did lead to higher rates of *var* gene switching (Bryant et al., 2017).

2.1 Discovery of RUF6 ncRNA interacting proteins

Subtelomeric non-coding regions, producing long ncRNAs from repetitive DNA regions, were also reported (Broadbent et al., 2011, 2015; Raabe et al., 2012; Siegel et al., 2014; Sierra-Miranda et al., 2012). Although the biological role for most ncRNA is missing, only a few functional studies connect ncRNA mechanistically to sexual commitment and antigenic variation of malaria parasites. The activation of the Ap2G Master regulator of sexual commitment, is controlled by an antisense RNA of the GDV1 gene (Filarsky et al., 2018). *P. falciparum* relies on the mutually exclusive expression of virulence gene families to survive within its host (reviewed in (Guizetti & Scherf, 2013; Wahlgren et al., 2017). In particular, the 60-member *var* multigene family codes for PfEMP1, an important variant surface adhesion molecule central to the development of the pathogenesis of the disease (Miller et al., 2002). All *var* genes are composed of a 5′ upstream promoter followed by exon I, a relatively conserved intron, and exon II at subtelomeric and central locations in each chromosome. Transcription of this gene family is tightly regulated by multiple epigenetic layers to ensure mutually exclusive expression at a specific expression site. Remarkably, G4 quadruplex motifs are enriched in *var* gene promotor and coding regions (Gage & Merrick, 2020; Gazanion et al., 2020; Smargiasso et al., 2009) **S**tabilized coding-strand G4 can inhibit transcription of reporter genes in transfected *P. falciparum* (Harris et al., 2018), raising the possibility that specific helicases may be involved in *var* gene activation (Smargiasso et al., 2009).

The active *var* gene is enriched in histone marks H3K4me3 and H3K9ac as well as the histone variant H2A.Z, and localizes to a distinct perinuclear expression site (Lopez-Rubio et al., 2007, 2009; Petter et al., 2011; Ralph et al., 2005). All other *var* genes remain transcriptionally silenced, tethered in repressive clusters enriched in histone modifications H3K9me3, H3K36me3 and heterochromatin protein 1 (Pf-HP1) and at the nuclear periphery (Flueck et al., 2009; Jiang et al., 2013; Lopez-Rubio et al., 2009; Pérez-Toledo et al., 2009). While the histone mark signatures for active, poised and silent *var* genes have been described (Guizetti & Scherf, 2013), it is unclear how the RNA Pol II transcriptional machinery is recruited to the *var* gene expression site.

The RUF6 ncRNA gene family (15 members) has recently been proposed to be linked to mutually exclusive transcription of the *var* multigene family (Barcons-Simon et al., 2020; Guizetti et al., 2016; Wu et al., 2019). RUF6 ncRNA was shown to colocalize in *trans* to the *var* gene expression site by fluorescence *in situ* hybridization (FISH), and the episomal overexpression of RUF6

resulted in the upregulation of previously silenced *var* genes, thus disrupting monoallelic expression (Guizetti et al., 2016). Downregulating the entire RUF6 gene family, using CRISPR interference (CRISPRi), led to a concurrent down-regulation in all *var* genes, (Barcons-Simon et al., 2020). RUF6 genes are 135 nucleotides long and have a GC-content higher than 50%, which is striking when compared to ~20% GC content genome-wide (Chakrabarti et al., 2007; Gardner et al., 2002). Interestingly, this ncRNA gene family is conserved only in *Plasmodium* species from the *Laverania* subgenus that encode a *var* gene family (Otto et al., 2018). Unlike other ncRNAs, RUF6 members contain canonical A- and B-box required for RNA polymerase III (Pol III) transcription. Although RUF6-mediated *var* gene activation has been genetically validated, an understanding of how nuclear RUF6 ncRNA interacting proteins mechanistically connect to their downstream effectors are missing.

In this study, we shed light on the mode of action of the RUF6 ncRNA by exploring its interaction partners. To identify specific RUF6 ncRNA-binding proteins in their native cellular context, we used a recently reported method termed Chromatin Isolation by RNA Purification followed by liquid chromatography-mass spectrometry (ChIRP-MS) (Chu et al., 2015; Chu & Chang, 2018). Chu and colleagues first used the technique to newly identify novel protein partners of the lncRNA Xist, essential for X chromosome inactivation in mammals. We adapted ChIRP-MS to *P. falciparum* for the first time, which allowed for the systematic discovery of RUF6 ncRNA binding proteins. The identified interactome indicates that RUF6 ncRNA interacts directly with RNA Pol II transcription at the *var* expression site. Using RNA immunoprecipitation, proteomics, and an inducible knock sideways approach, we identify the Deadbox RNA helicase, Pf-DDX5 as being a regulator of transcription of virulence genes.

Results:

Identification of RUF6 ncRNA-interacting proteins by ChIRP-MS

Given the lack of sequence homology between *var* gene loci and RUF6 ncRNA we hypothesize that RUF6 ncRNA associates with the active *var* loci via RNA-binding proteins (Fig. 1A). To test this hypothesis, we conducted electrophoretic mobility shift assays (EMSAs) using a biotinylated probe of a ncRNA RUF6 member (PF3D7_1241000). We observed a specific gel shift when incubating the probe with a nuclear extract of asynchronous parasites that was reproducible among

four independent experiments. An excess of an identical unlabeled probe competed for the binding factor, resulting in lack of a shift in the labeled probe. The same result was not obtained if a nonspecific competitor probe was used (Fig. 1B). These data suggest the presence of a specific RUF6 ncRNA binding factor.

To identify potential RUF6 ncRNA binding proteins *in vivo*, we adapted the Chromatin Isolation by RNA Purification (ChIRP) method developed by Chu et al (Chu et al., 2012, 2015; Chu & Chang, 2018) to *P. falciparum* (Fig. 1C). This technique cross-links cells in their native context using formaldehyde before hybridization of the target RNA with biotinylated oligonucleotides. Biotin-associated proteins are digested on-beads and identified after peptides desalting by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Negative controls like non-interacting probes and RNase A treatment of lysate prior to ChIRP are included.

To investigate the interactome of RUF6, we preformed ChIRP-MS on a 3D7 clone which expresses the RUF6 gene PF3D7_0412800 and the adjacent *var* gene PF3D7_0412700 (Barcons-Simon et al., 2020). We designed two sets of antisense biotinylated (3'-TEG Biotin) DNA oligo probes (odd and even) that hybridize to the active ncRNA and target the regions less prone to be located within a loop according to the predicted secondary structure (Fig. 1D, Table S1). As a control, we used a pool of two different scrambled probes that had similar GC-content to the target probes but were not predicted to bind anywhere in the genome. To eliminate non-RNA-specific protein interaction partners, one additional experimental group was performed in which cells were first treated with RNAase A, then incubated with RUF6-targeted probes. Quantitative reverse transcription PCR (RT-qPCR) on RNA isolated from each ChIRP replicate showed that RUF6 ncRNA was specifically enriched with both odd and even targeted probes, but not with the scrambled probes or the samples treated with RNase A (Fig. 1E).

Having demonstrated the specificity of the ChIRP technique with RUF6 ncRNA, we performed ChIRP and on-bead digestion followed by LC-MS/MS at 18 hpi, when both the active *var* and RUF6 ncRNA are transcribed, to identify proteins that interact with RUF6 ncRNA *in vivo*. We used four biological replicates, grown and harvested at separate times, and performed a label-free mass spectrometry quantification to identify enriched proteins in the targeted even probe samples

compared to the control samples. This led to the quantification of 571 total *P. falciparum* proteins (false discovery rate [FDR] of 1%, number of proteotypic peptides used ≥ 3, Fig. 2A, Table EV1). Target probes were compared to each of the two controls separately to determine a list of candidate proteins. We selected proteins that were significantly enriched or unique in the target sample compared to one or both controls (ratio ≥ 2.0, adjusted p-value ≤ 0.05 or unique proteins). The targeted RUF6 probes for all four replicates yielded a total of 386 significantly enriched or unique proteins (Fig. 2A, Tables EV2-3). Unique proteins in the target samples included three ApiAP2 transcription factors, two subunits of RNA Pol II, chromodomain-helicase-DNA-binding protein (Pf3D7_1023900.1), DNA topoisomerase 2 (Pf3D7_1433500.1), transformer-2 (Pf3D7_1002400.1), ATP-dependent RNA helicase DDX5 (Pf3D7_1445900.1), and multiple RNA-binding proteins and proteins with unknown functions. An RNA-binding protein (Pf3D7_1330800) was significantly enriched in target samples compared to both control groups (ratio = 4.92 adj. p-value = 9.44E-03 vs scrambled probes and ratio = 3.73 adj p-value = 1.49E-05 vs RNase treated comparisons). Significantly enriched proteins in target samples compared to control scrambled probe samples included: a DNA/RNA-binding protein Alba2 (Pf3D7_1346300; ratio = 12.13, p = 0.05), proteins of unknown function (Pf3D7_0813300; ratio= 5.71, p= 6.81E-04 and Pf3D7_0821400; ratio = 3.74, p = 0.02). Proteins significantly enriched or uniquely found with targeted probes gave gene ontology terms related to mRNA- (p = 9.96E-20), RNA- (p = of 1.14E-18), and nucleic acid- (p = 1.87E-15) binding (Fig. 2C, Table EV4). Such terms are consistent with transcription and gene regulation.

Analysis of ChIRP-MS interacting proteins

For downstream analysis, we selected candidate proteins if they met one or more of the following selection criteria: 1) have predicted RNA-binding potential, 2) have a function in gene activation, and/or 3) are conserved specifically in the *var* gene-containing *Laverania* species but not in malaria species that do not encode *var* genes (Plasmodium species). This selection left a list of 15 candidates: 11 unique proteins in the target sample compared to both controls and 4 unique proteins in the target sample compared to the scrambled probe control (Fig. 2B, D, Tables EV5-6). The top six candidates were selected to begin with further validation due to their predicted RNA-binding properties and involvement in gene regulation (Fig 2D). We were successful in C-terminal tagging two of the six candidates for further validation. We used the Selection Linked Integration (SLI)

strategy (Birnbaum et al., 2017) to generate strains in which the candidate proteins Pf3D7_1445900 Pf-DDX5 and Pf3D7_1423700 unknown function was tagged with a GFP epitope. To confirm that the candidate proteins bind to RUF6 ncRNA, we performed RNA immunoprecipitation (RIP) followed by RT-qPCR with three biological replicates, grown and harvested at separate times, for wild-type 3D7 parasites and the GFP-tagged proteins of interest in nuclear extracts of ring stage parasites. A significant enrichment ($p = 0.0078$) was observed for RUF6 in Pf-DDX5-GFP parasites, but not in Pf3D7_1423700-GFP parasites (Fig. 3A). No enrichment was observed for controls mRNA fructose-bisphosphate aldolase (FBA, Pf3D7_1444800) or tRNA Alanine (Pf3D7_0411500) (Fig 3C). These data suggest that the Pf-DDX5, but not the other candidate protein (Pf3D7_1423700), binds to RUF6 ncRNA. Western blot (Fig. 3B) and Immunofluorescence (Fig. 3C) analyses showed that Pf-DDX5 was present in the nucleus in ring and trophozoite stages, during which *var* gene and *RUF6* transcription peaks. Having demonstrated the association of Pf-DDX5 and RUF6 ncRNA, we set out to gain insight into the protein interactome of Pf-DDX5. We performed immunoprecipitation followed by quantitative mass spectrometry of Pf-DDX5-GFP and GFP-tagged Pf3D7_1423700, which served as a control since it was found to not bind RUF6 by RIP (Fig. 3A). A total of five biological replicates, with the parasites grown and harvested at separate times, was done. Analysis of the quantitative mass spectrometry data revealed significant enrichment of the control protein (ratio = 3.00, $p = 1.13E-14$) and Pf-DDX5 (ratio = 4.88, $p = 3.92E-28$) in their respective samples (Fig. 4A, Table E7).

A comparison of the proteins that were significantly enriched in the Pf-DDX5 IP-MS with those in the RUF6 ChIRP-MS revealed key similarities and differences. 110 proteins were significantly enriched only in the Pf-DDX5 IP-MS, but not in the RUF6 ChIRP-MS, suggesting that DDX5 could be part of multiple protein complexes with various cellular functions not limited to RUF6 mediated *var* gene regulation. In fact, while GO analysis for all significantly enriched proteins in the Pf-DDX5 sample showed the top results were translation ($p = 1.90$) and structural component of the ribosome ($p = 1.36$, Tables EV8-9), these changed greatly when only looking at shared proteins in the Pf-DDX5 sample that were also found in the targeted RUF6 ChIRP-MS. Twenty three proteins that were unique or significantly enriched in the RUF6 ChIRP-MS were also significantly enriched in the DDX5 IP-MS (Fig. 4B, Table EV10).

These shared proteins have functions mainly related to regulation of DNA-templated transcription initiation by RNA Pol II (p = 1.27E-04, Fig. 4C Table EV11). GO analysis of these shared proteins revealed that three of the topmost significantly enriched molecular functions were protein binding (p = 4.49E-06), RNA binding (p = 0.006), and single-stranded DNA exodeoxyribonuclease activity (p = 0.009, Table EV12). Biological process GO analysis showed chromatin assembly (p = 2.16E-03) and two out of the four proteins having to do with 'regulation of transcription initiation from RNA polymerase II promoter', 'positive regulation of RNA polymerase II transcription preinitiation complex assembly' were found in the shared proteins between RUF6 and Pf-DDX5 samples (p = 1.27E-04, Table EV11).

Importantly, seven proteins from our original shortlisted candidate proteins from the RUF6 ChIRP-MS (Fig. 2D) were found to be significantly enriched in the Pf-DDX5 IP-MS : proteins of unknown function (Pf3D7_1455300.1, ratio = 2.34, p = 1.99E-04, Pf3D7_0721100.1, ratio = 1.99, p = 2.51E-06, and Pf3D7_0819600.1, ratio = 2.05, p = 6.44E-04), polyadenylate-binding protein (Pf3D7_1107300.1, ratio = 3.14 , p = 3.95E-05), chromatin remodeling protein (Pf3D7_1104200.1 ratio = 6.39, p = 2.59E-05), pre-mRNA-processing factor (Pf3D7_1110200.1, ratio = 2.39 , p = 3.39E-02), and nucleosome assembly protein (Pf3D7_0919000, ratio = 2.03, p = 1.72E-03). Finding shared proteins between the two techniques further strengthens the evidence that Pf-DDX5 is part of a larger protein complex associated with RUF6 ncRNA. Shared proteins have GO analysis results for nucleosome positioning, chromatin organization, and protein-DNA complex subunit organization (p = 1.23E-03, 1.84E-02, 0.04).

Nuclear DDX5 is involved in *var* gene regulation

We hypothesized that RUF6 ncRNA-associated Pf-DDX5 may have a direct role in the immune evasion mechanism. We used a knock sideways approach (described in Birnbaum, J., et al. 2017) to study the function of Pf-DDX5. Knock sideways is based on the ligand-induced dimerization of FK506-binding protein (FKBP), which is fused to the endogenous Pf-DDX5, and FKBP-rapamycin binding protein (FRB), that is separately expressed and fused to a mislocalizer, anchored to the parasite plasma membrane (PPM). Upon addition of rapamycin, Pf-DDX5 levels decreased substantially in the nucleus (60-hour treatment) (Fig. 5B), and colocalization with the

FRB mislocalizer increased in the PPM (Fig. 5A). Parasite growth appeared to be unaffected after two growth cycles, but lower parasitemia was observed after 5 days of treatment (p = 0.031) (Fig. 5C), indicating a fitness cost for Pf-DDX5 displacement. This observation supports the phenotype observed for piggyBac transposon mutagenesis for this gene (M. Zhang et al., 2018). To determine the effects of Pf-DDX5 knock sideways on transcription, we treated synchronized parasites with rapamycin for 60 hours, which is within a window of time when growth should not yet be affected, and harvested ring stage parasites when *var* gene transcription reaches its peak. Giemsa staining of the harvested parasites confirm that both control and rapamycin-treated parasites were at ring stage (Fig. S2). Additionally, the age of the parasites was estimated to be ring stage (Fig. S3), shown in a heatmap of Pearson *r* correlation coefficients. Three biological replicates were used with the parasites grown and harvested at separate times. Rapamycin induced mislocalization of Pf-DDX5 resulted in significantly down-regulated (568) and up-regulated genes (301); Benjamini–Hochberg-adjusted *p*-value (i.e., *q*) ≤ 0.05) (Table EV13).

RUF6 transcript levels were not affected indicating that Pf-DDX5 does not interfere with the stability of this ncRNA (Fig. 6A). Importantly, the single active *var* gene in this clone was one of the topmost significantly down-regulated genes (*q* = 2.59E-06) (Fig. 6B, C, D).

GO enrichment analysis revealed that down-regulated genes were enriched in the biological processes of translation (p = 9.47E-34) and RNA processing (p = 9.69E-22, Table EV14). Genes included in these categories encode splicing factors, a subunit of RNA Pol III, and proteins with unknown functions). GO enrichment analysis revealed that up-regulated genes were enriched in the biological processes: movement in host environment (p = 2.41E-21) and entry into host (p = 4.75E-21, Table EV15). Together, these data indicated that the recruitment of Pf-DDX5 by RUF6 ncRNA to the active *var* gene promotes transcription, possibly by resolving secondary structures that are highly enriched in this gene family (Gazanion et al., 2020).

Discussion

Despite the association of ncRNAs as regulators of malaria parasite pathogenesis and transmission, the underlying molecular mechanisms remain elusive. To overcome technical challenges and explore key open-ended questions, we developed a new tool for malaria parasites aimed to identify

ncRNA-protein interactions in their native cellular context. We established a robust ChIRP-MS protocol using biotinylated tiling oligo probes for RUF6 (see Fig. 1D). After validating our probes in pull down assays, this new tool was used to identify RUF6 associated proteins. Importantly, ChIRP-MS method allowed us to isolate and identify specific RUF6 RNA-associated proteins that were not detected using nuclear extracts in RUF6 ncRNA affinity purification assays (unpublished data). This result highlights that the "native" ChIRP-MS method detects interactions that cannot be reconstituted once the nucleus is disrupted. This method may find broad applications in the malaria field to identify specific RNA-binding proteins.

The rationale of this study was to gain biological insight into the RUF6 ncRNA-mediated activation of the *var* gene family, which encodes a major virulence factor and contributes to immune evasion from the host immune response. The RUF6 gene family is strictly linked to a small subset of malaria species that encode the *var* virulence gene family (Otto et al., 2018.). This evidence has motivated previous studies to explore its role in the monoallelic expression of the *var* gene family (Barcons-Simon et al., 2020; Guizetti et al., 2016; Wei et al., 2015). It remains unknown how the ncRNA RUF6 is targeted in *trans* to an active *var* gene expression site and which proteins RUF6 recruits and interacts with to promote efficient singular *var* gene transcription. Here we established a comprehensive RUF6 ncRNA protein interactome and validated the helicase DDX5 as a regulator of *var* gene transcription. Two controls were used (scrambled probes and RNase A-treated samples) to raise the prospect of identifying protein candidates that either bind directly or are associated with RUF6 ncRNA. We pulled out binding proteins that were detected with label free quantitative proteomics (LC-MS/MS) (Fig. 2A). One challenging aspect of proteomic analysis is the amount of input material needed to overcome non-specific background hits. To overcome this obstacle, we used 4.0E10 parasites in ChIRP-MS (n=4) (and later 1.5E09 parasites in Co-IP MS, n=5). Quantitative analysis showed more proteins that were uniquely found in our RUF6 target samples rather than proteins more statistically enriched in our target samples compared to controls ($p< 0.05$). Therefore, our candidate proteins were uniquely identified in our target samples and absent in one or both control groups. Among those candidates were genes that encode proteins implicated in gene transcription and candidates with unknown function. We hypothesized that some proteins bind specifically to the ncRNA and others form complexes via protein-protein interactions.

Candidates involved in transcription are homologous to two subunits of RNA Pol II, to DNA topoisomerase 2, DDX5 RNA helicase, to nucleosome assembly proteins and CHD1, a chromatin remodeling protein. These data point to an interaction of the RUF6 complex with the *var* gene RNA Pol II transcription machinery. It is noteworthy in this context that RNA Pol II transcribes *var* genes but RUF6 is transcribed by RNA Pol III. In fact, a ncRNA transcribed by RNA Pol III (called B2 RNA) has been shown to act in *trans* to bind directly to RNA Pol II to regulate transcription in mice (Espinoza et al., 2004). Notably, two candidate proteins have been previously reported to bind to or near *var* gene promoters, Alba DNA/RNA-binding proteins (*Archeal* chromatin protein family) and a member of the ApiAp2 transcription factor family (Chêne et al., 2012; Goyal et al., 2012; Martins et al., 2017). Additionally, candidate protein transformer-2, not yet characterized in *Plasmodium*, is known to act in insects as an upstream regulatory element in sexual regulation (Nguyen et al., 2021). Furthermore, polyadenylate-binding protein was among the candidate proteins and has been found in *Drosophila* to interact with RNA Pol II at the promoter during transcription (Kachaev et al., 2019). These findings corroborate our previously proposed mode of function, namely the co-localization of RUF6 ncRNA with actively transcribed *var* genes (Guizetti *et al*, 2016) and that RUF6 transcription is linked to *var* gene activation (Barcons-Simon et al., 2020; Guizetti et al., 2016). We performed functional validation of one of the identified interactome proteins, which is predicted to bind to RNA. It is a homolog of human DEAD box protein DDX5, an RNA helicase, that can also unwind secondary structures of DNA (Wu et al., 2019). Pf-DDX5 was selected for further validation not only because of its canonical role in unwinding RNA/DNA (Wu et al., 2019), but also for its role in transcriptional regulation and elongation and its interaction directly with RNA Pol II (Clark et al., 2013). Typically, helicases function in the separation of double-stranded RNA, DNA and RNA/DNA hybrid structures and contribute to the process of gene regulation (Bourgeois et al., 2016). The *P. falciparum* genome has a predicted set of 63 helicases (Reddy et al., 2015), but only very few have been functionally characterized in malaria parasites (Tuteja, 2017). The RNA helicase DOZI has an important role in translational repression of mRNA during sexual development (Mair et al., 2006) and the DNA RecQ helicase is an important chromatin factor associated with multiple roles including DNA replication, genome stability and heterochromatin organization of clonally variant genes (Claessens et al., 2018; Z. Li et al., 2019).

We obtained genetically modified Pf-DDX5 GFP-tagged transfectants for functional analysis using the knock sideways system (Birnbaum et al., 2017). IFA localization studies and Western blot analysis confirmed the nuclear presence of Pf-DDX5 during ring stage and RIP RT-qPCR confirmed its association with RUF6 ncRNA (Fig. 3). We used GFP-tagged Pf-DDX5 parasites to perform Co-IP assays to explore the interactome of Pf-DDX5. Mass spectrometry (LC-MS/MS) revealed many proteins found in the Pf-DDX5 sample that were also found in the original RUF6 ChIRP-MS. This result validates the ChIRP-MS technique as a method to identify ncRNA interactomes in *P. falciparum* (Fig. 4D). Shared proteins had functions related to RNA binding and involved in processes related to initiation and regulation of transcription from RNA Pol II, strengthening the notion that Pf-DDX5 is associated with RUF6 ncRNA at the *var* gene RNA Pol II promoter. Of note is that histone variant H2A.Z was significant in the Pf-DDX5 IP with a *p* value of 1.10E-03 and a ratio of 2.73. In the RUF6 ChIRP-MS however the *p* value was not significant (p= 0.89). H2A.Z was previously shown to be enriched at active *var* gene promoters (Petter et al., 2011, 2013). The combination of two methods CHIRP-MS and Pf-DDX5 Co-IP-MS provides the first comprehensive catalogue of proteins that interacts with RUF6 ncRNA. It contains numerous proteins of unknown function.

Since the functional role in gene regulation of Pf-DDX5 in malaria parasites has not been reported, we used the Pf-DDX5 knock sideway strategy to displace the majority of nuclear Pf-DDX5 (Fig. 5A). The nuclear depletion results in a significant decrease in the active *var* gene transcription without disrupting monoallelic expression (Fig. 6). This result identifies Pf-DDX5 as a novel regulator of transcriptional activation of *var* genes.

What is the possible mechanism of Pf-DDX5 at the *var* promotor? Human DDX5 has recently been reported to be involved in transcriptional regulation resolving secondary structures near RNA Pol II promoters (Wu et al., 2019). One of these structures are G-quadruplexes (G4s). G4s have been found to play a role in various cellular processes including gene transcription (Siddiqui-Jain et al., 2002). A recent report demonstrated that DDX5 can bind not only to RNA G-quadruplexes, but also to DNA G-quadruplexes (G. Wu et al., 2019). In this context it is important to note that a recent study predicted DNA G4s in *P. falciparum*, using the tool G4Hunter algorithm (Gazanion

et al., 2020). The authors showed that the *var* multi-gene family, specifically, showed a significant enrichment in DNA G4s found in the promoter and first exon (Gazanion et al., 2020). We therefore predict that plasmodial DDX5, helps unwind these structures to allow for proper transcription by RNA Pol II throughout the *var* locus. Notably, an RNA G4 quadruplex is predicted in RUF6. This raises the possibility that this structure recruits Pf-DDX5 to this ncRNA.

Two ribonucleases have been reported to regulate clonally variant RUF6 ncRNA in steady state RNA levels by degrading nascent RNA; an exosome-independent PfRNase II (Zhang et al., 2014) and an RNA exosome-associated Rrp6 (Fan et al., 2020). Inactivation of those ribonucleases disrupt heterochromatin-dependent silencing of clonally variant gene families. General activation of RUF6 gene family in Rrp6 mutant parasites indicated that this ncRNA family is enriched in heterochromatin regions. This observation does support our findings, that RUF6 is involved in the activation process of silent var genes in heterochromatin islands.

A recent report developed a catalytically inactive Cas9 system to investigate the chromatin associated with *var* gene loci (promoter and intron) (Bryant et al., 2020). We did find thirty-six shared proteins between the two studies including CHD1, chromatin remodeling protein, AP2 domain transcription factor, and ISWI, which was validated in their study to be associated with the active *var* promoter. Three proteins were shared between our RUF6 ChIRP-MS and Pf-DDX5 IP with the proteins found significantly in the *var* promoter: protein of unknown function Pf3D7_0704300, Phax domain-containing protein Pf3D7_1021900, and chromatin remodeling protein Pf3D7_1104200. Pf-DDX5 was not significantly found in the study by Bryant and colleagues, and this may be explained by the fact that the isolated chromatin primarily is from silent *var* gene promotors.

In conclusion, here we developed a robust ChIRP-MS technique to identify the first ncRNA-protein interactome that sustains the *P. falciparum* immune evasion strategy. This work opens up new avenues to identify unprecedented regulatory chromatin factors that will shed light on the mechanisms of RUF6 ncRNA recruitment to the *var* gene expression site and singular virulence gene expression in this deadly pathogen. This insight may provide new strategies to target pathogenesis of malaria parasites.

Materials and Methods:

Parasite culture and synchronization

Asexual blood stage 3D7 *P. falciparum* parasites were cultured as previously described in (Lopez-Rubio et al., 2009; W.H.O., 2021). Parasites were cultured in human RBCs (obtained from the Etablissement Français du Sang with approval number HS 2016-24803) in RPMI-1640 culture medium (Thermo Fisher 11875) supplemented with 10% v/v Albumax I (Thermo Fisher 11020), hypoxanthine (0.1 mM final concentration, C.C.Pro Z-41-M) and 10 mg gentamicin (Sigma G1397) at 4% hematocrit and under 5% O_2, 3% CO_2 at 37 °C. Parasite development was monitored by Giemsa staining. Parasites were synchronized by sorbitol (5%, Sigma S6021) lysis at ring stage, plasmagel (Plasmion, Fresenius Kabi) enrichment of late stages 24 h later, and an additional sorbitol lysis 6 h after plasmagel enrichment. Parasites were cultured under static conditions with the exception of shaking during late schizont until early ring stage. Parasites were harvested at 1–5% parasitemia.

RNA electrophoretic mobility shift assay

Probe sequences for RUF6 (PF3D7_1241000) and its antisense were amplified with primers containing the SP6 promoter sequence. In vitro transcription was performed on 0.2 ug of the PCR products with the MAXIscript T7/SP6 Kit (Ambion) using the SP6 enzyme mix. RNAs were checked for size on a denaturing urea polyacrylamide gel and biotin was incorporated on the 3' end of the RNA fragments using the Pierce RNA 3' End Biotinylation Kit (Thermo Fisher Scientific). Alternatively, synthesized 5' biotinylated RNA probes were also used. RNA electrophoretic mobility shift assays (RNA EMSA) were performed based on LightShift Chemiluminiscent RNA EMSA Instructions (Thermo Fisher Scientific). RNAs were relaxed and refolded by incubation to 64°C and gradual cooling to 4°C prior to binding. 20 uL of binding reaction containing 5 g of nuclear extract, 2 g of tRNA, 10, 20 or 40 fmol of biotinylated probe and when indicated 8 pmol of unlabeled probe were incubated in REMSA buffer (10mM HEPES, 20mM KCl, 1mM MgCl2, 1 mM DTT) with RNAsin Ribonuclease Inhibitor (Promega) at room temperature for 25min. After electrophoresis on a native polyacrylamide TBE gel, transferred RNA was cross-linked to the nylon membrane at 120mJ/cm2. Detection of biotin-labelled RNA was performed using the Chemiluminiscent Nucleic Acid Detection Module (Thermo Fisher Scientific) and imaged with the ChemiDoc XRS+ system (Bio-Rad).

Chromatin isolation by RNA purification and proteomics mass spectrometry analysis (ChIRP-MS)

ChIRP was performed as previously described (Chu & Chang, 2018) with the following modifications. $4x10^{10}$ parasites per sample were harvested and lysed with saponin prior to 3% formaldehyde crosslinking. After final wash buffer washes, five additional washes with 500 mM NaCl followed by another five washes with 25 mM NH_4HCO_3 buffer (ABC) were performed. Bound proteins were analyzed by proteomics mass spectrometry. A total of 4 biological replicates, cultured separately and grown at different times, for each of the four samples, scrambled probes, RNase A treated, even probes and odd probes, were performed.

For proteomics mass spectrometry, bound proteins were first on-beads digested for one hour with 0.6 µg of trypsine/LysC (Promega), then loaded onto a homemade C18 StageTips for desalting. Peptides were eluted using 40/60 MeCN/H2O + 0.1% formic acid and vacuum concentrated to dryness. Online liquid chromatography (LC) was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled to a Q Exactive HF-X with a Nanospay Flex ion source (Thermo Scientific). Peptides were first trapped on a C18 column (75 µm inner diameter × 2 cm; nanoViper Acclaim PepMap™ 100, Thermo Scientific) with buffer A (2/98 MeCN/H2O in 0.1% formic acid) at a flow rate of 2.5 µL/min over 4 min. Separation was then performed on a 50 cm x 75 µm C18 column (nanoViper Acclaim PepMap™ RSLC, 2 µm, 100Å, Thermo Scientific) regulated to a temperature of 50°C with a linear gradient of 2% to 30% buffer B (100% MeCN in 0.1% formic acid) at a flow rate of 300 nL/min over 91 min.

MS full scans were performed in the ultrahigh-field Orbitrap mass analyzer in ranges *m*/*z* 375–1500 with a resolution of 120 000 at *m*/*z* 200. The top 20 intense ions were subjected to Orbitrap for further fragmentation via high energy collision dissociation (HCD) activation and a resolution of 15 000 with the intensity threshold kept at 1.3×10^5. We selected ions with charge state from 2+ to 6+ for screening. Normalized collision energy (NCE) was set at 27 and the dynamic exclusion of 40s.

For identification, the data were searched against the Plasmodium falciparum FASTA database *(PlasmoDB-36 Pfaciparum3D7 AnnotatedProtein containing cas9 and the common contaminants)*

using Sequest^{HT} through proteome discoverer (version 2.2 or 2.4). Enzyme specificity was set to trypsin and a maximum of two-missed cleavage sites were allowed. Oxidized methionine, carbamidomethyl cysteine and N-terminal acetylation were set as variable modifications. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.02 Da for MS/MS peaks. The resulting files were further processed using myProMS (Poullet et al., 2007) v3.9.3 (https://github.com/bioinfo-pf-curie/myproms). False-discovery rate (FDR) calculation used Percolator [The et al, 2016] and was set to 1% at the peptide level for the whole study. The label free quantification was performed by peptide Extracted Ion Chromatograms (XICs) computed with MassChroQ version 2.2.1 (Valot et al., 2011). For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN matching) with two-missed cleavages were used. Median and scale normalization was applied on the total signal to correct the XICs for each biological replicate (n=4). To estimate the significance of the change in protein abundance, a linear model (adjusted on peptides and biological replicates) was performed and *p*-values were adjusted with a Benjamini–Hochberg FDR procedure. Proteins with at least 2-fold increase between probe and control condition, an adjusted *p*-value under 0.05, and at least three total peptides in all replicates were considered significantly enriched in sample comparisons. Candidate proteins that display at least 3 total peptides in all replicates were chosen if they were unique in the target sample, even probes compared to at least one of the controls.

RNA isolation and reverse transcription-quantitative PCR (RT-qPCR)

RNA was harvested from synchronized parasite cultures after saponin lysis in 0.075% saponin in PBS, followed by one wash in PBS and resuspension in the QIAzol reagent. Total RNA was extracted using an miRNeasy minikit and performing on-column DNase treatment (Qiagen 217004). Reverse transcription from ChIRP eluted RNA was achieved using SuperScript VILO (Thermo Fisher Scientific) and random hexamer primers. cDNA levels were quantified by quantitative PCR in the CFX384 real time PCR detection system (BioRad) using Power SYBR Green PCR Master Mix (Applied Biosystems) and primers from a previous study (Guizetti et al., 2016). Starting quantity means of three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA. RUF6 and housekeeping fructose-bisphosphate aldolase (PF3D7_1444800) transcript levels were compared in ChIRP and input samples. The starting

quantity means from three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA.

Generation of genetically modified *P. falciparum* strains

All cloning of episomal constructs was performed using KAPA HiFi DNA Polymerase (Roche 07958846001), In-Fusion HD Cloning Kit (Clontech 639649), and XL10-Gold Ultracompetent *E. coli* (Agilent Technologies 200315). Transgenic GFP-tagged parasites were generated as previously described in (Birnbaum et al., 2017). For localization and knock sideways, the last 500-1000 bp of each candidate protein target gene was cloned into pSLI-2xFKBP-GFP. Each sequence started with an in-frame stop codon but the stop codon at the end of the gene was removed. 50 µg of plasmid DNA was transfected into ring stage 3D7 *P. falciparum* parasites using the protocol described elsewhere (Hasenkamp et al., 2013). Transfected parasites were selected with constant drug selection pressure of 4 nM WR99210 (Jacobus Pharmaceuticals) to obtain a cell line containing the episomal plasmid. A second drug selection using 400 µg/ml of G418 was done to select for integrants. Once parasites emerged, gDNA of each integration cell line was collected using a commercial kit (DNeasy Blood & Tissue Kit) and checked by PCR to show that integration occurred at the correct locus. Both genome and vector specific primers for the 5' and 3' region were used so that the PCR product would cover the plasmid/genome junction. Vector primers used were the same as described (Birnbaum et al., 2017). Once proper size gel bands from PCR were seen, localization of the GFP tagged protein was checked with a fluorescence microscope (Delta Vision Elite microscope (GE Healthcare)). Image overlays were generated using Fiji (Schindelin et al., 2012).

Procedure for knock-sideways

pSLI-2xFKBP-GFP tagged parasites were further transfected with 50 µg of mislocalizer plasmid DNA. The PPM (parasite plasma membrane) mislocalizer (pLyn-FRB-mCherry) for nuclear targets was used. Parasites were selected with 2 µg/ml Blasticidin S. Once transfectants were obtained, mislocalizer mCherry expression was assessed using a fluorescence microscope. For knock sideways experiments, 20 µL rapalog working solution (resulting in 250nM final) was added to cause mislocalization of pSLI-2xFKBP-GFP tagged parasites to the PPM. Parasites were cloned

by limiting dilution, and the targeted genomic locus was sequenced to confirm tag and FKBP integration as well as epitope mislocalizer BSD.

Western blot analysis

iRBCs were washed once with Dulbecco's phosphate-buffered saline (DPBS, Thermo Fisher 14190) at 37°C and lysed with 0.075% saponin (Sigma S7900) in DPBS at 37°C. Parasites were washed once with DPBS, resuspended in 1 ml cytoplasmic lysis buffer (25 mM Tris–HCl pH 7.5, 10 mM NaCl, 1.5 mM $MgCl_2$, 1% IGEPAL CA-630, and 1× protease inhibitor cocktail ["PI", Roche 11836170001]) at 4°C, and incubated on ice for 30 min. Cells were further homogenized with a chilled glass douncer, and the cytoplasmic lysate was cleared with centrifugation (13,500 *g*, 10 min, 4°C). The pellet (containing the nuclei) was resuspended in 100 µl nuclear extraction buffer (25 mM Tris–HCl pH 7.5, 600 mM NaCl, 1.5 mM $MgCl_2$, 1% IGEPAL CA-630, PI) at 4°C and sonicated for 10 cycles with 30 s (on/off) intervals (5 min total sonication time) in a Diagenode Pico Bioruptor at 4°C. This nuclear lysate was cleared with centrifugation (13,500 *g*, 10 min, 4°C). Protein samples were supplemented with NuPage Sample Buffer (Thermo Fisher NP0008) and NuPage Reducing Agent (Thermo Fisher NP0004) and denatured for 5 min at 95°C. Proteins were separated on a 4-15% TGX (Tris-Glycine eXtended) (Bio-Rad) and transferred to a PVDF membrane. The membrane was blocked for 1 h with 5% milk in PBST (PBS, 0.1% Tween 20) at 25°C. GFP-tagged proteins and histone H3 were detected with anti-GFP (Chromotek, 1:1,000 in 5% milk-PBST) and anti-H3 (Abcam ab1791, 1:1,000 in 5% milk-PBST) primary antibodies, respectively, followed by donkey anti-rabbit secondary antibody conjugated to horseradish peroxidase ("HRP", Sigma GENA934, 1:5,000 in 5% milk-PBST). Aldolase was detected with anti-aldolase-HRP (Abcam ab38905, 1:5,000 in 5% milk-PBST). HRP signal was developed with SuperSignal West Pico chemiluminescent substrate (Thermo Fisher 34080) and imaged with a ChemiDoc XRS+ (Bio-Rad).

Immunofluorescence assay

For live cell fluorescent microscopy, 500 µl of parasite cultures (3-5% parasitemia, 3-4% hematocrit) was spun down for 45 sec at 2.4 rpm. iRBC pellet was resuspended in PBS-DAPI for 20 min before spinning down for 45 sec at 2.4 and removing supernatant. iRBC pellet was then washed with PBS before mounting. Unattached cells were washed out with PBS and finally

covered with culture medium prepared with phenol red free RPMI 1640. Images were captured using a Delta Vision Elite microscope (GE Healthcare). Image overlays were generated using Fiji (Schindelin et al., 2012).

SLI transfected cultures were used with GFP-booster antibodies, anti-GFP V_HH/ Nanobody conjugated to a fluorescent dye (Chromotek) and Chromotek RFP antibody (6G6) to visualize mCherry mislocalizer. 10 µl of iRBCs were washed with PBS and fixed with for 30 min in 0.0075% Glutaraldehyde/4% PFA/PBS. After PBS washing, parasites were permeabilized with 0.1% TritonX100/PBS for 10-15 min before quenching free aldehyde groups with $NaBH_4$ solution for 10 min. Next, parasites were blocked with 3% BSA–PBS for 30 min. Primary antibody incubation lasted for 1 h before three washes with PBS, and secondary antibody incubation for 30-60 min, Alexa Fluor 488-conjugated anti-mouse IgG (Invitrogen) diluted 1:2,000 in 4% BSA–PBS. After three final washes in PBS, cells were mounted in Vectashield containing DAPI for nuclear staining. Images were captured using as described before.

RNA Immunoprecipitation

Wild type 3D7, pSLI-2xFKBP-GFP-DDX5 (Pf3D7_1445900) tagged parasites and pSLI-2xFKBP-GFP-protein of unknown function (Pf3D7_1423700) tagged parasites were synchronized and harvested at 18hpi (n=3 biological replicates). Each culture (10^9 parasites) was centrifuged, and RBCs were lysed with six volumes of 0.15% saponin in DPBS for 5 min at 4°C. Parasites were centrifuged at 4,000 *g* for 5 min at 4°C, and the pellet was washed twice with DPBS at 4°C. Parasites were then cross-linked with 3% formaldehyde for 15 min at room temperature and quenched with 125mM glycine for 5 min on ice. Each sample was resuspended in 2 mL lysis buffer (10 mM HEPES pH 8, 10 mM KCl, 0.1 mM EDTA pH 8, 0.25% final concentration of IGEPAL CA-630) + CI (Roche, EDTA free) at 4°C and incubated with gentle agitation at 4°C for 30 min. Extracts were centrifuged in microcentrifuge tubes for 10 min at 13,500 g at 4°C. Once the supernatant was removed, the pellet was resuspended in 300 µL SDS lysis buffer (50 mM Tris–HCl pH 8, 10 mM EDTA pH 8, 1% SDS + CI) at 4°C. Next, 300uL were transferred to Diagenode 1.5 mL sonication tubes. Samples were sonicated with the Diagenode Pico sonicator for 22 cycles (30 sec on/30 sec off), vortexing and spinning occasionally. Lysates were centrifuged at 16000 × *rcf* for 10min at 4°C and supernatants were transferred to fresh tubes. Beads were prepared by

washing 25 µL per sample with 500 µL of ice-cold dilution buffer. Samples then had 2x volume of dilution buffer (10mM TRIS pH 7.5, 150mM NaCl, 0.5mM EDTA) + CI (Roche, EDTA free) added and washed beads (Chromotek gtma-10 GFP-Trap® Magnetic Agarose beads) and were incubated overnight at 4°C. The next morning, beads were collected and washed with 500 µL of wash buffer (2xSSC, 0.5% SDS, 1mM PMSF) for 5 min with rotation at 4°C. After the final wash, beads were resuspended in 100 uL pK buffer. Input sample volume was adjusted to 100 µL with pK buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.0, 1 mM EDTA, 0.5% SDS, 5% pK (Proteinase K (New England Biolabs P8107S)). Samples were incubated at 50°C for 45 min with constant mixing. Next, samples were boiled at 95°C for 10 min before being resuspended in 700 µL of QIAzol reagent (Qiagen 79306). RNA was extracted using an miRNeasy minikit and performing on-column DNase treatment (Qiagen 217004). Reverse transcription from RNA-IP eluted RNA was achieved using SuperScript VILO (Thermo Fisher Scientific) and random hexamer primers. cDNA levels were quantified by quantitative PCR in the CFX384 real time PCR detection system (BioRad) using Power SYBR Green PCR Master Mix (Applied Biosystems) and primers from a previous study (Guizetti et al., 2016). Starting quantity means of three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA. RUF6, housekeeping fructose-bisphosphate aldolase (PF3D7_1444800), and Alanine tRNA (PF3D7_0411500) transcript levels were compared in RNA IP and input samples.

Co-Immunoprecipitation followed by Mass spectrometry (Co-IP-MS)

pSLI-2xFKBP-GFP-DDX5 tagged parasites (n = 5 biological replicates) and pSLI-2xFKBP-GFP-unknown function tagged parasites (n = 5 biological replicates) were synchronized. At 18 hpi, each culture (1.5×10^9 parasites) was centrifuged and RBCs were lysed with six volumes of 0.15% saponin in DPBS for 5 min at 4°C. Parasites were centrifuged at 4,000 *g* for 5 min at 4°C, and the pellet was washed twice with DPBS at 4°C. Parasites were then cross-linked with 3% formaldehyde for 15 min at room temperature and quenched with 125 mM glycine for 5 min on ice. Cross-linked parasites were washed twice with DPBS and then lysed with 5 ml of lysis buffer (10 mM Tris–HCl pH 7.5, 1 mM EDTA, 0.5% IGEPAL CA-630) supplemented with protease inhibitors (Thermo Fisher 78440) and 1 U/µl of Benzonase (Merck 71206) at 4°C and incubated with rotation for 30 min at 4°C. Extracts were centrifuged for 8 min at 380g at 4°C and the cytoplasmic supernatant was removed. The nuclear pellet was resuspended in 900 µL nuclear lysis

buffer (10 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1%sodium deoxycholate, 0.1% SDS, PI) at 4°C and transferred to 1.5 ml sonication tubes (300 µL per tube, DiagenodeC30010016). Samples were sonicated for 10 min (30 s on/off) in a Diagenode Pico Bioruptor at 4°C. Lysates were then centrifuged for 10 min at 13,500g at 4°C and supernatant was transferred to a fresh tube. 900 µL of the supernatant was mixed with 1.35 mL (2:3) dilution buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mmM EDTA). Nuclear fraction supernatants were incubated with 25 µl of α-GFP trap beads (Chromotek), first washed twice in dilution buffer, overnight with rotation at 4°C. The next day, the beads were collected on a magnet and the supernatant was removed. While on the magnetic stand, beads were washed twice with 500 µL washing buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.05% NP40), once with 25 mM NH_4HCO_3 (Sigma 09830) buffer, and then transferred to new tube. Finally, the beads were resuspended in 100 µl of 25 mM NH_4HCO_3 (Sigma 09830) and digested by adding 0.2 µg of trypsin-LysC (Promega) for 1 h at 37 °C. Samples were then loaded into custom-made C18 StageTips packed by stacking one AttractSPE® disk (#SPE-Disks-Bio-C18-100.47.20 Affinisep) and 2mg beads (#186004521 SepPak C18 Cartridge Waters) into a 200 µL micropipette tip for desalting. Peptides were eluted using a ratio of 40:60 CH3CN:H2O + 0.1% formic acid and vacuum concentrated to dryness with a SpeedVac apparatus. Peptides were reconstituted in 10 of injection buffer in 0.3% trifluoroacetic acid (TFA) before liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Online LC was performed as described previously in the ChIRP-MS protocol and coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) by modifying the peptide trapping flow to 3.0 µL/min over 4 min and the separation temperature to 40 °C with a linear gradient of 3% to 29% buffer B.

Parasite growth assay

Parasite growth was measured as described previously (Vembar et al., 2015). A clone of pSLI-DDX5-GFP-FKBP with mCherry-FRB was tightly synchronized and diluted to 0.2% parasitemia (5% hematocrit) at ring stage using the blood of two different donors separately. Each culture was split, and 20 µL rapalog working solution was added (250nM final concentration of rapamycin) to one half. The growth curve was performed with three technical replicates per condition per blood. Parasitemia was measured every 24 h by counting 10 randomly selected different fields on Giemsa stained slides each day for a total of 5 days.

Stranded RNA sequencing and analysis

Parasites were synchronized by sorbitol (5%, Sigma S6021) lysis at ring stage, plasmagel (Plasmion, Fresenius Kabi) enrichment of late stages 24 h later, and an additional sorbitol lysis 3 h after plasmagel enrichment before the parasites were separated into the +/- Rapamycin groups. After a full cycle, another sorbitol was done (at 3hpi) prior to harvesting at 12hpi (9 hours later and with 60 hours of rapamycin exposure to the + Rapamycin group). +/- Rapamycin infected RBCs containing synchronized (12 hpi ± 3 h) parasites were lysed in 0.075% saponin (Sigma S7900) in DPBS at 37°C. The parasite cell pellet was washed once with DPBS and then resuspended in 700 µL QIAzol reagent (Qiagen 79306). Total RNA was subjected to rRNA depletion to ensure ncRNA and mRNA capture using the RiboCop rRNA Depletion Kit (Lexogen) prior to strand-specific RNA-seq library preparation using the TruSeq Stranded RNA LT Kit (Illumina) with the KAPA HiFi polymerase (Kapa Biosystems) for the PCR amplification. Multiplexed libraries were subjected to 150 bp paired-end sequencing on a NextSeq 500 platform (Illumina). Sequenced reads (150 bp paired end) were mapped to the *P. falciparum* genome (Gardner et al., 2002) (plasmoDB.org, version 3, release 57) using "bwa mem" (H. Li & Durbin, 2009) allowing a read to align only once to the reference genome (option "–c 1"). Alignments were subsequently filtered for duplicates and a mapping quality ≥ 20 using samtools (H. Li & Durbin, 2009). Three biological replicates for -Rap and +Rap samples were analyzed for the knockdown experiment.

Estimation of cell cycle progression

To estimate the parasite age (i.e. hours post infection), we correlated transcripts levels (average FPKM >= 10) from all three replicates of the +rapamycin and -rapamycin samples with the transcript levels of a reference microarray dataset (Bozdech et al., 2003). Pearson r correlation coefficients were calculated and visualized in R () with options cor (R Core Team (2020)) and heatmap2(), respectively.

Differential gene expression analysis

A clone of integrated pSLI-DDX5-GFP-FKBP with episomal mCherry-FRB was synchronized and split into two cultures. Rapamycin (250nM final concentration) was added to one culture at 0

hpi, and parasites were harvested 60h later during ring stage (12 hpi). RNA sequencing reads for three technical replicates of the rapamycin-treated and three technical replicates of the untreated pSLI-DDX5-GFP-FKBP with mCherry-FRB clone were mapped to the *P. falciparum* genome and quality filtered as described above for RNA-seq. Strand-specific gene counts were calculated using htseq-count (Anders et al., 2015). Differential gene expression analysis was performed using DESeq2 (Love et al., 2014) with significantly differentially expressed genes featuring a Benjamini–Hochberg-adjusted *p*-value (i.e., *q*) ≤ 0.05. MA plots were generated using the "baseMean" (mean normalized read count over all replicates and conditions) and "$\log_2$FoldChange" values (rapamycin-treated over control) as determined by DESeq2. RPKM values were calculated in R using the command rpkm() from the package edgeR (Robinson et al., 2010). Gene Ontology enrichments were calculated using the built-in tool at PlasmoDB.org (Aurrecoechea et al., 2017).

Statistical analysis

All statistical analyses were performed using GraphPad Prism version 9.1.0 (216) for Mac. To test for a normal distribution of the data, the Shapiro-Wilk normality test was used. To test for significance between two groups, a two-sided independent-samples t test was used. Gene ontology enrichments were calculated using the build-in tool at https://plasmoDB.org.

Data availability

The data generated in this study are available in the following databases:

Acknowledgements

We would like to thank Jessica Bryant for helpful comments and critically reading the manuscript. This work was supported by ERC grants to Artur Scherf (ERC AdG PlasmoSilencing) and Sebastian Baumgarten (ERC StG PlasmoEpiRNA) and the Agence Nationale de Recherche (grant ANR-11 LabEx -0024-01 ParaFrap to A. Sc).

Author Contributions

GD and AS conceptualized the project and conceived experiments. AB performed the EMSA experiment and designed the primers for the ChIRP-MS experiments. FD carried out the mass

spectrometry experimental work and DL supervised mass spectrometry and data analysis. SB performed the bioinformatic data analysis. GD and AS wrote the manuscript. All authors discussed and approved the manuscript.

Conflict of Interest

The funders had no role in the study design, data collection and interpretation or decision to submit the work for publication. The authors declare that they have no conflict of interest.

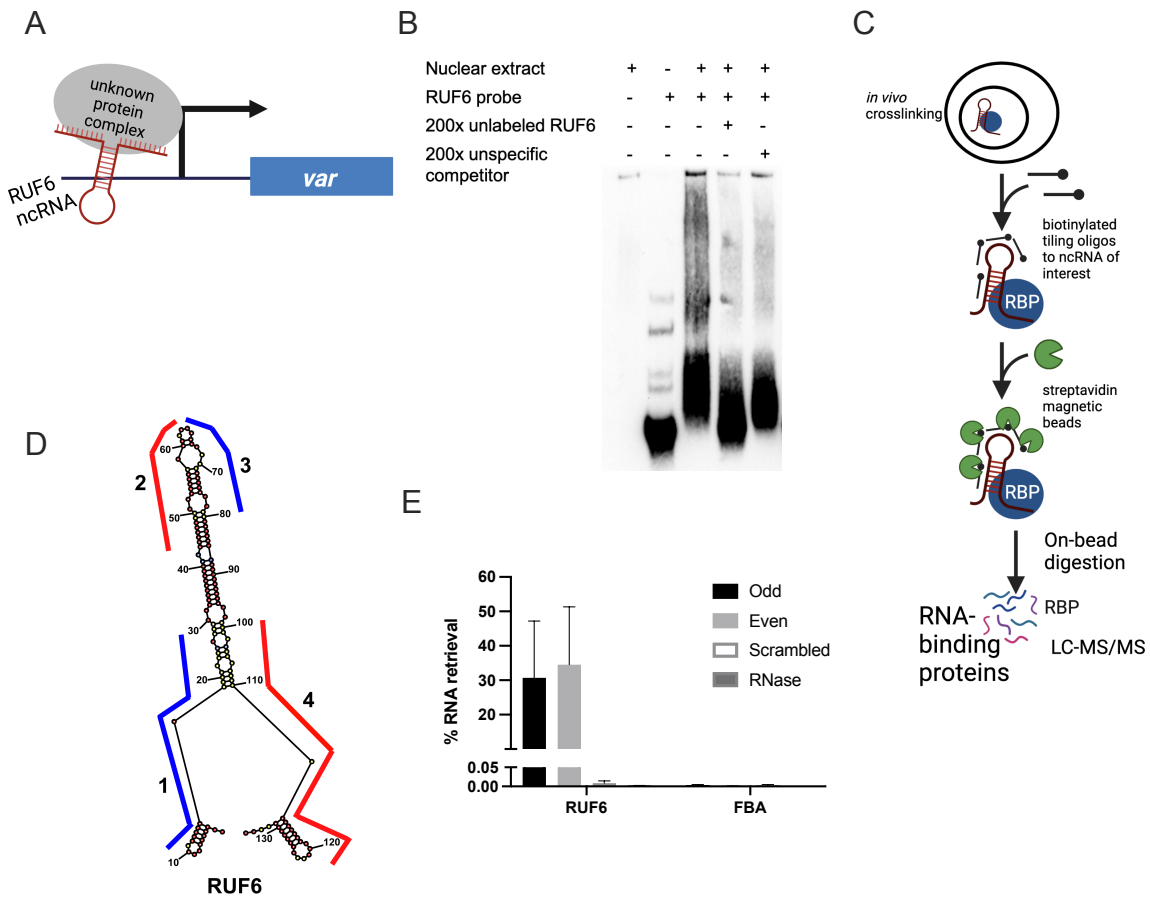
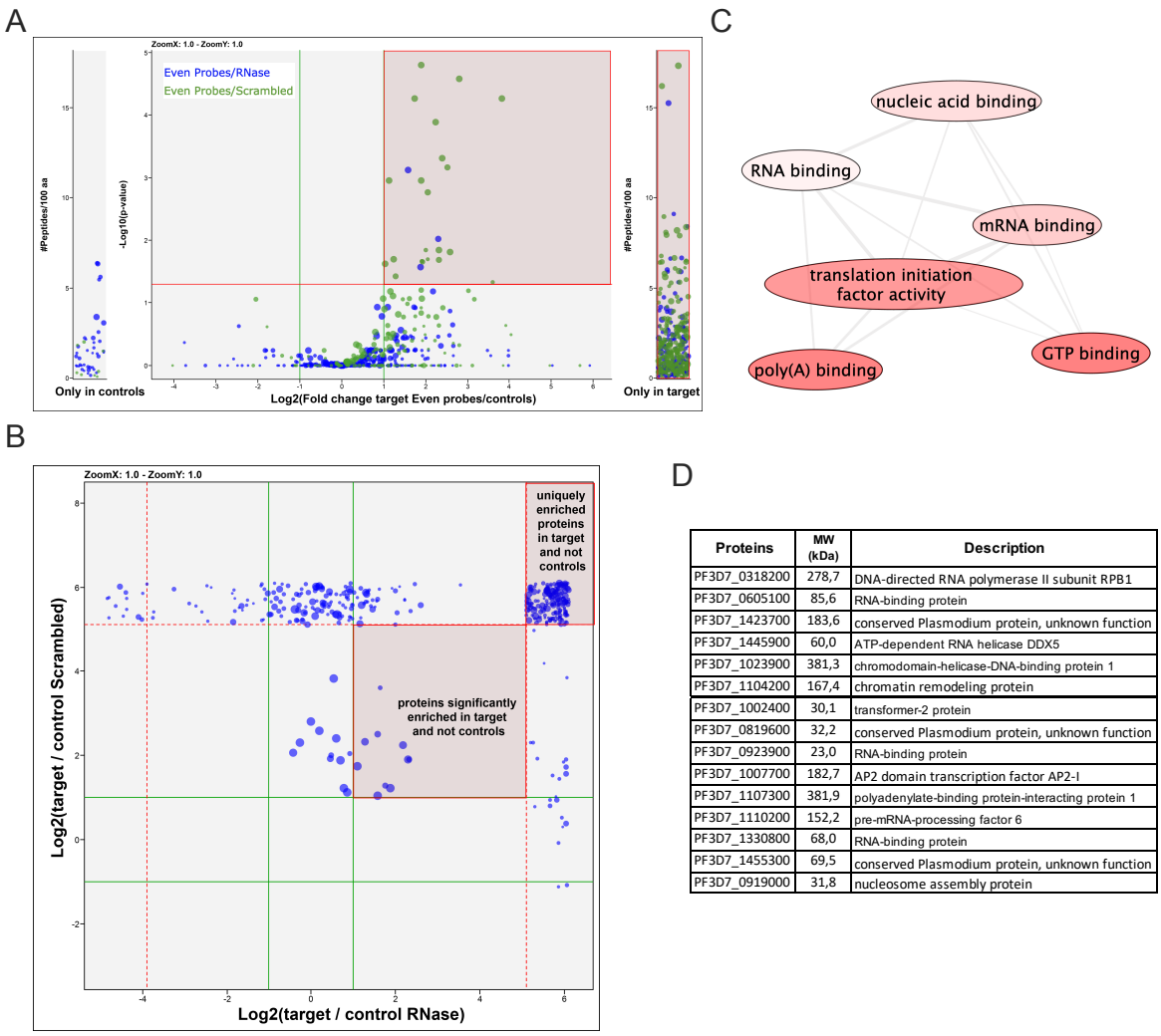


Figure 1. Developing ChIRP-MS to identify RUF6 ncRNA-binding proteins (A) Schematic showing hypothetical model of RUF6 ncRNA associating with proteins at active *var* gene promoter. (B) Electrophoretic mobility shift assay (EMSA) using biotinylated RUF6 ncRNA probe (Pf3D7_1241000) gives a specific shift in presence of nuclear extract that is competed using an unlabeled probe in excess (200x). (C) Outline of ChIRP-MS protocol. RNP complexes are crosslinked *in vivo* before addition of biotinylated anti-sense oligos. Target ncRNA are pulled out with beads and RNA binding proteins are identified through LC-MS/MS. Created with BioRender.com (D) Lowest free energy secondary structure of the RUF6 coded by PF3D7_0412800, as predicted by RNA structure bioinformatics web server. The four oligonucleotide probes used for ChIRP-MS are shown along their binding regions of the ncRNA structure. Odd probes 1 & 3 are colored in blue and even probes 2 & 4 are colored in red. (E) Percentage of RNA retrieval in ChIRP-MS compared to input samples using odd, even, and scrambled sets of probes and RNase treated samples. Transcript levels were assessed by RT-qPCR and fructose-bisphosphate aldolase (FBA, PF3D7_1444800) levels were used as a negative control.



D

Proteins	MW (kDa)	Description
PF3D7_0318200	278,7	DNA-directed RNA polymerase II subunit RPB1
PF3D7_0605100	85,6	RNA-binding protein
PF3D7_1423700	183,6	conserved Plasmodium protein, unknown function
PF3D7_1445900	60,0	ATP-dependent RNA helicase DDX5
PF3D7_1023900	381,3	chromodomain-helicase-DNA-binding protein 1
PF3D7_1104200	167,4	chromatin remodeling protein
PF3D7_1002400	30,1	transformer-2 protein
PF3D7_0819600	32,2	conserved Plasmodium protein, unknown function
PF3D7_0923900	23,0	RNA-binding protein
PF3D7_1007700	182,7	AP2 domain transcription factor AP2-I
PF3D7_1107300	381,9	polyadenylate-binding protein-interacting protein 1
PF3D7_1110200	152,2	pre-mRNA-processing factor 6
PF3D7_1330800	68,0	RNA-binding protein
PF3D7_1455300	69,5	conserved Plasmodium protein, unknown function
PF3D7_0919000	31,8	nucleosome assembly protein

Figure 2. ChIRP-MS identification of RUF6 ncRNA interacting proteins (A) RUF6 ChIRP-MS Volcano plot of label-free quantitative proteomic analysis of 571 *P. falciparum* proteins present for all four replicates in target samples (even probes) compared to the control groups: scrambled probes and RNase treated samples. Blue dot color represents target (even probes) versus control (RNase treated), green dot color represents target (even probes) vs control (scrambled probes) quantifications. Each dot represents a protein, and its size corresponds to the sum of peptides from both conditions used to quantify the ratio of enrichment. *x*-axis = $\log_2$(fold-change), *y*-axis = $-\log_{10}$(p-value), horizontal red line indicates adjusted p-value = 0.05, and vertical green lines indicate absolute fold-change = 2.0. Side panels indicate proteins uniquely identified in either sample (*y*-axis = number of peptides per 100 amino acids). All individual comparisons can be found in supplementary figures. Red highlighted boxes show 386 proteins significantly enriched or unique in target samples compared to each control. (B) Correlation plot comparing target (even probes) vs control (RNase treated) and target (even probes) vs control (scrambled probes) of the 386 proteins from (A). Red highlighted boxes show proteins significantly enriched or unique in target samples compared to both controls. (C) Molecular function groupings displayed represent gene ontology terms related to molecular function of genes enriched only in RUF6 target samples but not in control samples, prepared for visualization using PlasmoDB (plasmodb.org) and the REViGO tool (http://revigo.irb.hr/). Proteins had a p-value cut-off at 0.05. (D) Protein IDs of selected candidate proteins.

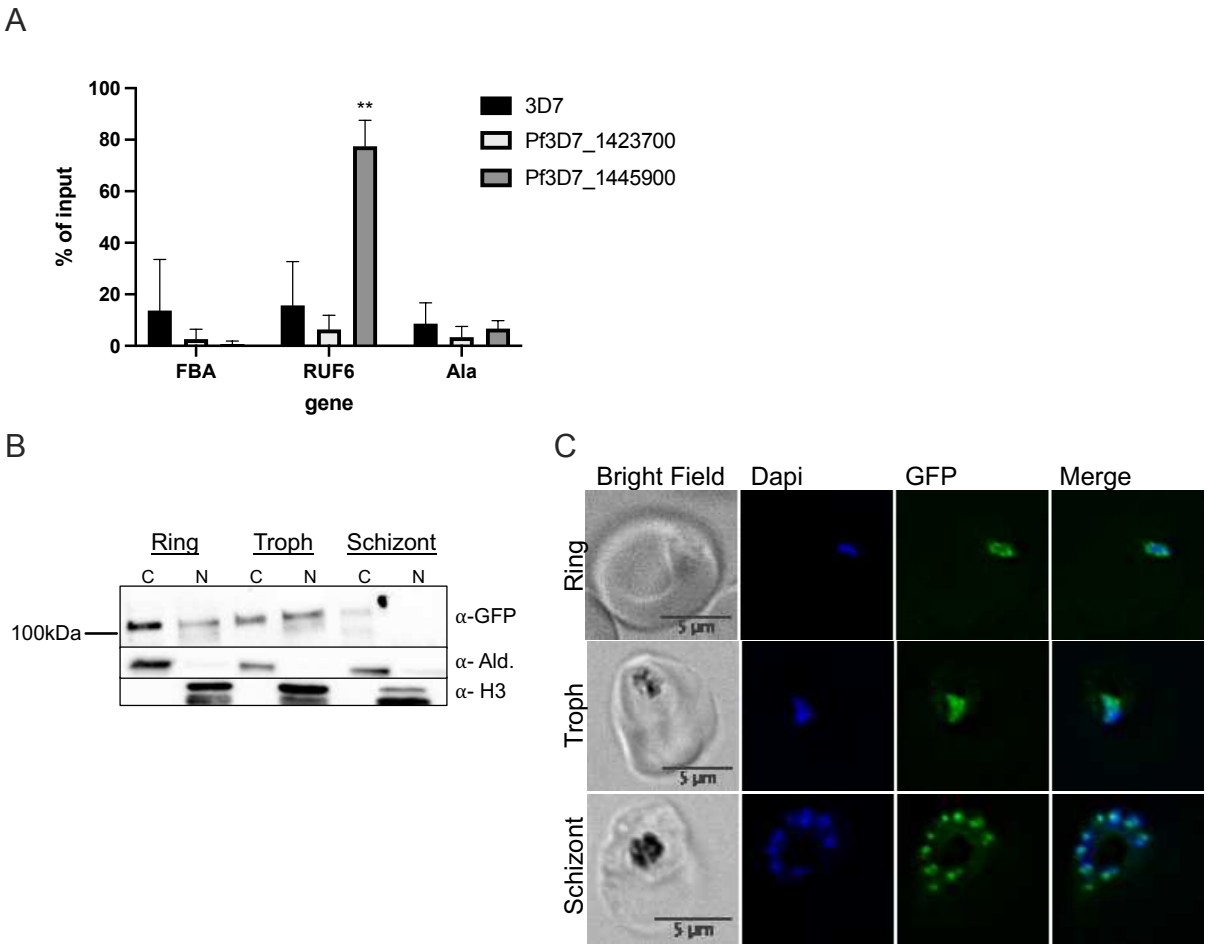
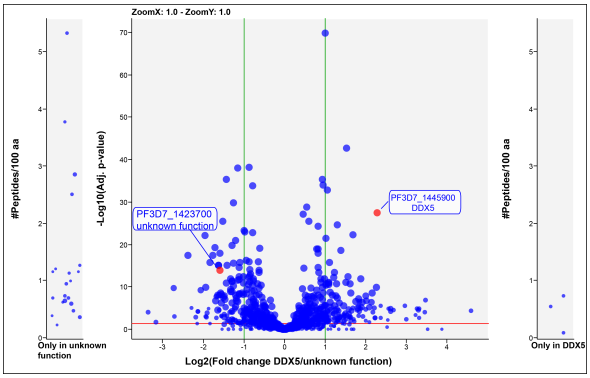
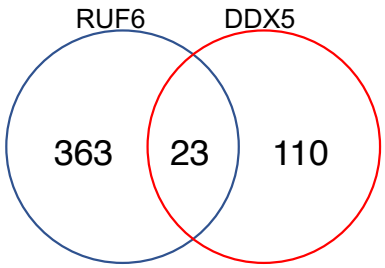


Figure 3. Validation of ChIRP-MS candidate proteins (A) RT-qPCR results from RNA Immunoprecipitation analysis on wild type 3D7 nuclear extracts: protein of unknown function (Pf3D7_1423700) and Pf-DDX5 RNA helicase (Pf3D7_1445900). Primers used for RT-qPCR were FBA, Fructose-bisphosphate aldolase Pf3D7_1444800, RUF6 for entire gene family, and tRNA Alanine, Pf3D7_0411500. Results are displayed as % of input. Error bars are displayed from 3 biological replicates. (B) Western blot of Pf-DDX5 in cytoplasmic and nuclear extracts throughout the IDC (rings, trophs, and schizonts). (C) Representative immunofluorescence images show brightfield, Dapi, GFP, and Dapi-GFP merge for candidate protein Pf-DDX5.

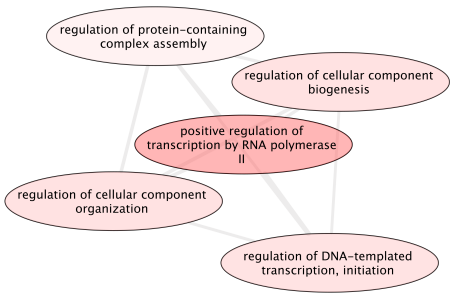
A



B



C



D

Shared Pf-DDX5 proteins with targeted RUF6 ChIRP-MS

Proteins	MW (kDa)	Description
PF3D7_1445900	60,0	ATP-dependent RNA helicase DDX5, putative
PF3D7_0919000	31,8	nucleosome assembly protein
PF3D7_1455300	69,5	conserved Plasmodium protein, unknown function
PF3D7_1107300	381,9	polyadenylate-binding protein-interacting protein 1, putative
PF3D7_0704300	217,5	conserved Plasmodium membrane protein, unknown function
PF3D7_0819600	32,2	conserved Plasmodium protein, unknown function
PF3D7_1104200	167,4	chromatin remodeling protein
PF3D7_1110200	152,2	pre-mRNA-processing factor 6, putative

Figure 4. Pf-DDX5 interactome (A) Co-IP MS volcano plot of enrichment for all 5 replicates for Pf-DDX5 vs control protein, protein of unknown function (Pf3D7_1423700), proteins are indicated and labeled. Each dot represents a protein, and its size corresponds to the sum of peptides from both conditions used to quantify the ratio of enrichment. *x*-axis = $\log_2$(fold-change), *y*-axis = $-\log_{10}$(p-value), horizontal red line indicates adjusted p-value = 0.05, and vertical green lines indicate absolute fold-change = 2.0. Side panels indicate proteins uniquely identified in either sample (*y*-axis = number of peptides per 100 amino acids) with a minimum of 3 total peptides. Adjusted p-value of 0.05 is displayed as a horizontal red line and fold change greater than 2 is labeled as vertical green lines. (B) Venn diagram showing total number of shared proteins found in RUF6 ChIRP-MS and DDX5 IP MS. (C) Biological processes groupings displayed represent shared significant proteins from RUF6 ChIRP-MS and DDX5 IP MS prepared for visualization using PlasmoDB (plasmodb.org) and the REViGO tool (http://revigo.irb.hr/). Proteins had a p-value cut-off at 0.05. (D) Significant proteins enriched in Pf-DDX5 sample that are common to the original ChIRP-MS selected candidate proteins.

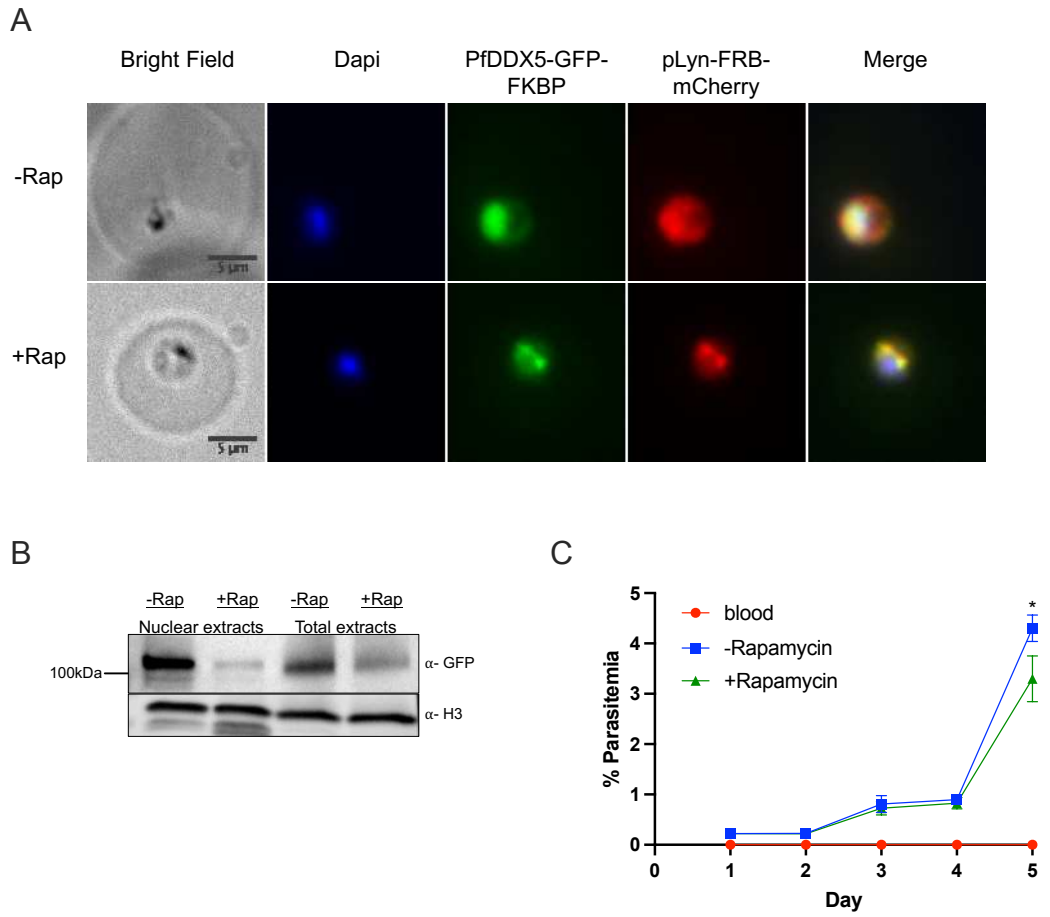


Figure 5. Knock-sideways of Pf-DDX5 (A) IFA images of Pf-DDX5-GFP-2FKBP mislocalization with plasma-membrane mislocalizer, pLyn-FRB-mCherry, following addition of rapamycin (B) Western blot displaying Pf-DDX5 protein removal from the nucleus after addition of rapamycin. Histone H3 was used as a nuclear extract control. Using image analysis tools, addition of rapamycin caused a 10x decrease in nuclear levels (normalized to H3) Total extracts show Pf-DDX5 presence before and after addition of rapamycin. (C) Growth curve over 4 days of Pf-DDX5-GFP-2FKBP parasites cultured in the presence (250nM final concentration, +Rapamycin) or absence (-Rapamycin) of rapamycin. Error bars represent standard deviations from three independent experiments.

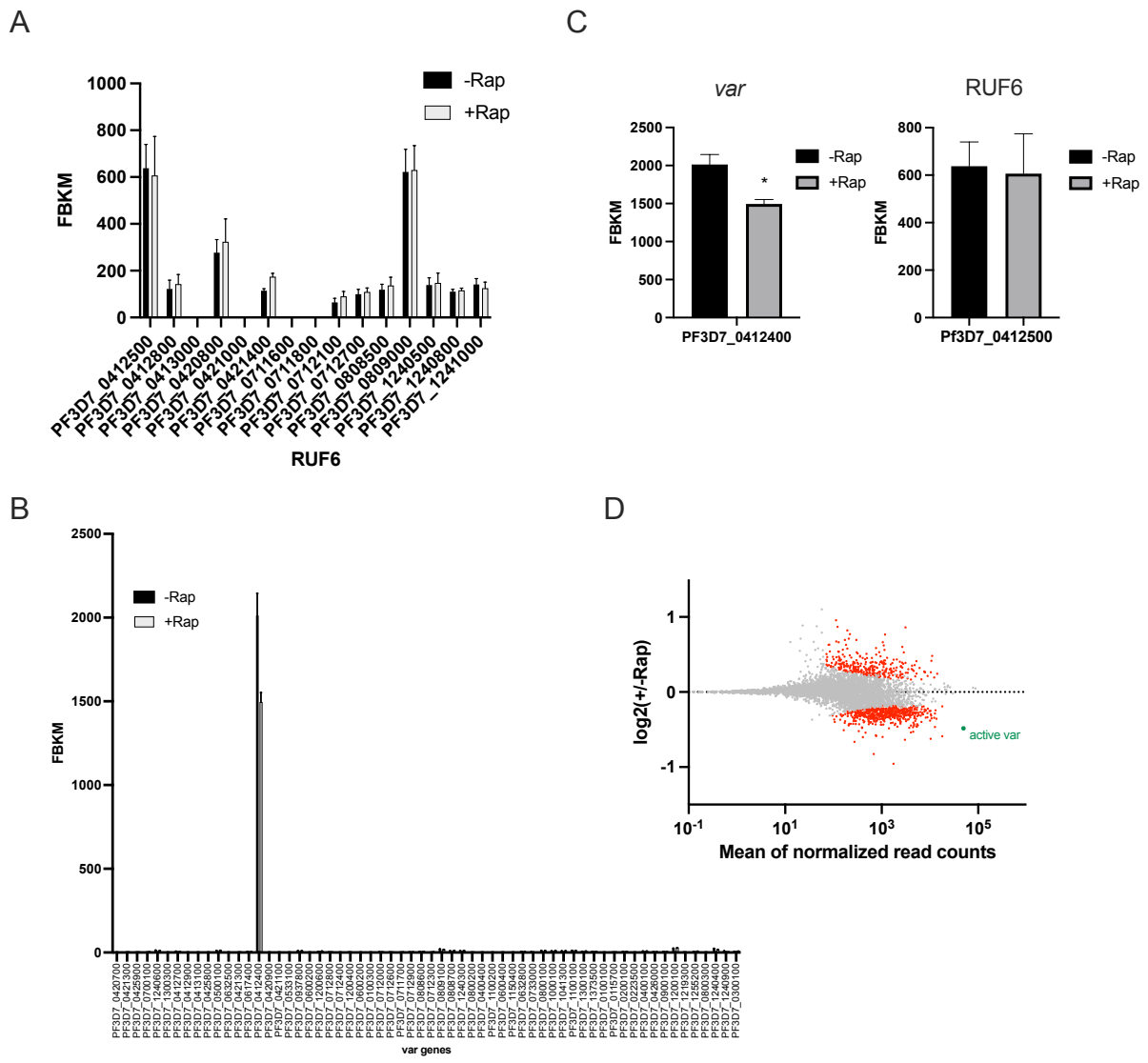


Figure 6. Pf-DDX5 is involved in *var* gene transcription (A) Transcriptional profile of RUF6 gene family and (B) *var* gene family at 12 hpi assayed by RNA sequencing for the control -Rap and treated +Rap samples. (C) Transcriptional levels of the active *var* gene and associated RUF6 member at 12hpi assayed by RNA sequencing for the control -Rap and treated +Rap samples. (D) MA plot of $\log_2$(rapamycin-treated/untreated, M) plotted over the mean abundance of each gene (A) at 12 hpi. Transcripts with a significantly higher (above *x*-axis) or lower (below *x*-axis) abundance in the presence of rapamycin are highlighted in red ($q \leq 0.05$). The active *var* gene is highlighted in green ($q = 2.59 \times 10^{-6}$). Three replicates were used for untreated and rapamycin-treated parasites. p-values were calculated with a Wald test for significance of coefficients in a negative binomial generalized linear model as implemented in DESeq2 (Love *et al*, 2014). *q* = *B*onferroni corrected p-value. Mean_SEM of three independent experiments is shown.

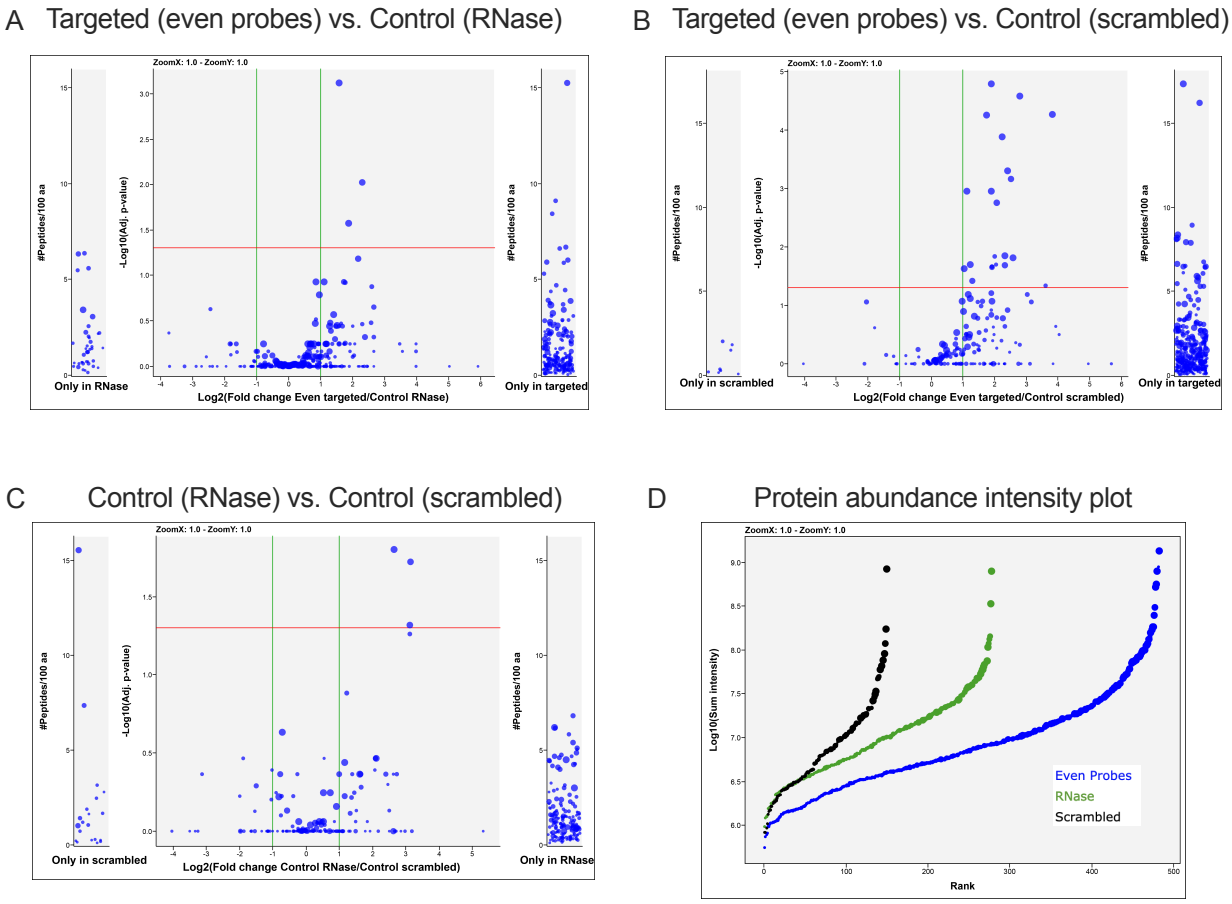


Figure S1. Identification of candidate proteins from RUF6 ChIRP-MS. Volcano plot of label-free quantitative proteomic analysis of *P. falciparum* proteins present for all four replicates in samples with: (A) targeted (even probes) compared to control (RNase treated), (B) targeted (even probes) compared to control (scrambled probes), (C) control (RNase treated) compared to control (scrambled probe). Each dot represents a protein, and its size corresponds to the sum of peptides from both conditions used to quantify the ratio of enrichment. x-axis = log2(fold-change), y axis = 210 amino acids). (D) Intensity plot for *P. falciparum* protein distribution of all four biological replicates (blue= targeted (even probes), green= control (RNase treated), and black= control (scrambled probes).

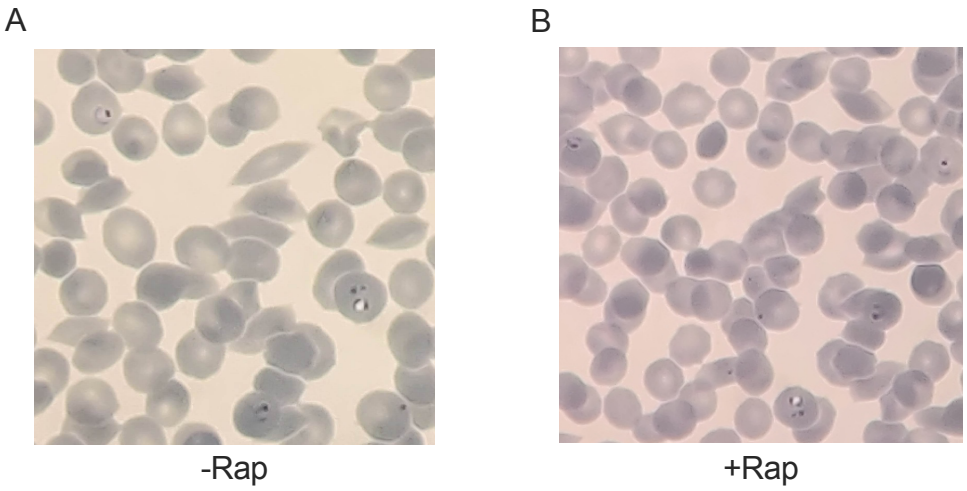


Figure S2. Giemsa stain of rings stage parasites harvested for RNA-seq analysis (12hpi). Images show ring stage parasites cultured without (-Rap) rapamycin (A) or with (+Rap) rapamycin (B).

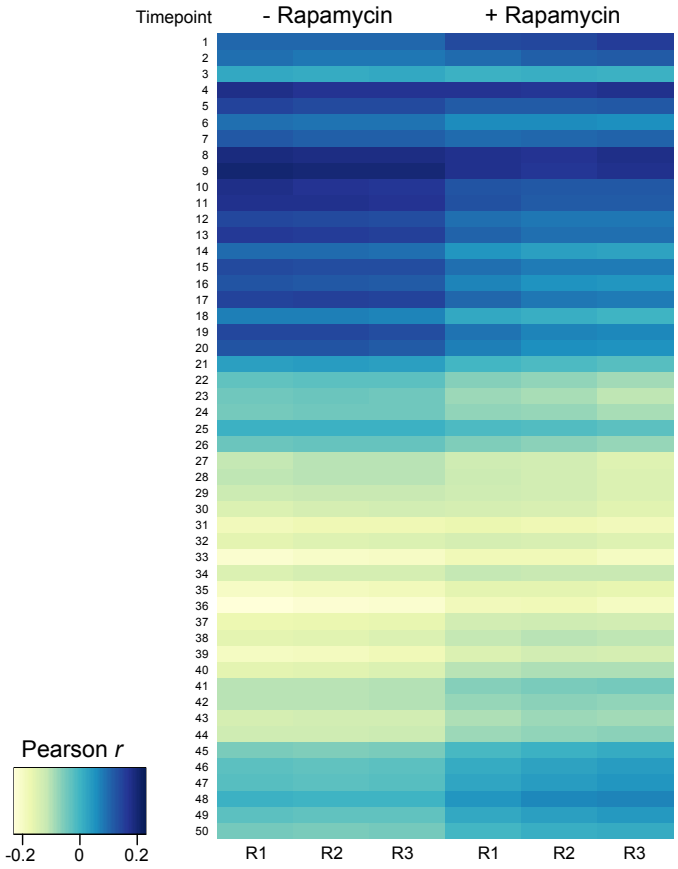


Figure S3. Heatmap of Pearson *r* correlation coefficient calculated between transcript levels in +/- rapamycin samples and transcript levels of a reference transcriptome dataset (Bozdech et al., 2003)

2.1 Discovery of RUF6 ncRNA interacting proteins

Table S1. (A) Sequences used for probes and primers for EMSA, ChIRP, and PCR/qPCR

EMSA	
Primer name	**Sequence**
Primers to amplify gDNA Pf3D7_1241000	F: 5'-AAG-CTG-CCT-CAG-TAG-CCC-3' R: 5'-TTG-CGC-CAC-CCC-CCT-C-3'
Pf3D7_1241000 probes:	
Sense:	Biotin:AAGCUGCCUCAGUAGCCCAAUCGUUAGGUAUGUUGCCUUUCCUUGUGAGAACGUUGGUUCGACUCCGCUUGCCGACAAUUUCAUAGGAAAAGUUGCAGAUCGAGCUGUUGGGUUACCCGGAGGGGGGCGGCGCAA
Antisense:	Biotin:UUGCGCCGCCCCCCUCCGGGUAACCCAACAGCUCGAUCUGCAACUUUUCCUAUGAAAUUGUCGGCAAGCGGAGUCAACCAACGUUCUCACAAGGAAAGGCAACAUACCUAACGAUUGGGCUACUGAGGCAGCUU
Sense:	AAGCUGCCUCAGUAGCCCAAUCGUUAGGUAUGUUGCCUUUCCUUGUGAGAACGUUGGUUCGACUCCGCUUGCCGACAAUUUCAUAGGAAAAGUUGCAGAUCGAGCUGUUGGGUUACCCGGAGGGGGGCGGCGCAA:Biotin
Antisense:	UUGCGCCGCCCCCCUCCGGGUAACCCAACAGCUCGAUCUGCAACUUUUCCUAUGAAAUUGUCGGCAAGCGGAGUCAACCAACGUUCUCACAAGGAAAGGCAACAUACCUAACGAUUGGGCUACUGAGGCAGCUU:Biotin
ChIRP	
Probe name	**Sequence**
RUF6 1 (odd set)	TACCTAACGACTGGGCTACTG
RUF6 2 (even set)	TCGAACCAACGTTCTCACAAG
RUF6 3 (odd set)	TGAAATTGTTGGCAAGCGGAG
RUF6 4 (even set)	CTCCGGGTAACTCAGCAGCTC
Scrambled 1	GGCCTGTTATGTCGATTCGA
Scrambled 2	GGAACGACATAGTCTGAATG
qPCR/PCR	
Primer name	**Sequence**
RUF6-A	Forward: 5'-AAGCTGCCTCAGTAGCCCA-3' Reverse: 5'-AAAAATTGCGCCACCCCC-3'
RUF6-B	Forward: 5'-AAGCTGCCCCAGTAGCCCA-3' Reverse: 5'-AAAAATTGCGCCGCCCCC-3'
fructose-bisphosphate aldolase (FBA, Pf3D7_1444800)	Forward: 5'-TGTACCACCAGCCTTACCAG-3' Reverse: 5'-TTCCTTGCCATGTGTTCAAT-3'
SLI HB:	
SLI 3 Forward primer	5'- ATGCTCACACAGCAAGATCAG-3'
SLI 3 Reverse primer	5' ATGCTCACACAGCAAGATCAG-3'
SLI 5 Forward primer	5'-TGTTTTGTGTCGTACATTTGCAT-3'
SLI 5 Reverse primer	
SLI Integration check:	5'-GTAAAATCTGTTATTTCCTCTTAGTGG-3'
SLI check reverse primer	5'-ACCTTCACCCTCTCCACTGAC-3'
SLI 3 WT locus check reverse	5'-GAGGAGCATGTAAATATATTCATTTTAC-3'
SLI 5 WT locus check reverse	5'-ATATATATATGTATATTATAATAATATTAGTAAAATCTG-3'
BSD primers for mislocalizer	Forward: 5'-GAA GAC TAC AGC GTC GCC AGC GC-3' Reverse: 5'-CTC CCA CAC ATA ACC AGA GGG CAG C-3'

Additional supplementary excel files are not shown due to their size. Below are the figure legends and summaries for each file.

Table EV1: Total proteins detected by LC-MS/MS in ChIRP samples (all *P. falciparum* proteins detected with FDR ≤ 1%, ≥ 3 peptides in all replicates)
Total number of P. falciparum proteins detected by LC-MS/MS. Protein molecular weight (MW) in kilodaltons (kDa) are provided for all proteins. Genes and their predicted products are listed. Only proteins with at least three peptides in all replicates are shown. Four replicates were used for all samples.

Table EV2: Label-free quantitative LC-MS/MS analysis of ChIRP RUF6 targeted - versus ChIRP scrambled control probes (all *P. falciparum* proteins detected with FDR ≤ 1%, ≥ 3 peptides in all replicates, ratio ≥ 2.0, $p \leq 0.05$)
Proteomics analysis of ChIRP-MS. Comparison of RUF6 targeted (even probes) to scrambled control probes. Label-free quantitative LC-MS/MS analysis of *RUF6 targeted (even probes)*- and scrambled control probes. Ratio, log2(ratio), peptides used (Pept. used), total peptides, and protein molecular weight (MW) in kilodaltons (kDa) are provided for all proteins. Adjusted *p*-value and the percent coefficient of variation (CV %) are shown for proteins present in both samples. Genes and their predicted products are listed. Only proteins with ≥2.0 fold ratio of enrichment, an adjusted *p*-value ≤ 0.05, and at least three peptides in all replicates are shown. Four replicates were used for both samples.

Table EV3: Label-free quantitative LC-MS/MS analysis of ChIRP RUF6 targeted - versus ChIRP RNase treated control (all *P. falciparum* proteins detected with FDR ≤ 1%, ≥ 3 peptides in all replicates, ratio ≥ 2.0, $p \leq 0.05$)
Proteomics analysis of ChIRP-MS. Comparison of RUF6 targeted (even probes) to RNase treated control. Label-free quantitative LC-MS/MS analysis of *RUF6 targeted (even probes)*- and RNase treated control. Ratio, log2(ratio), peptides used (Pept. used), total peptides, and protein molecular weight (MW) in kilodaltons (kDa) are provided for all proteins. Adjusted *p*-value and the percent coefficient of variation (CV %) are shown for proteins present in both samples. Genes and their predicted products are listed. Only proteins with ≥2.0 fold ratio of enrichment, an adjusted *p*-value ≤ 0.05, and at least three peptides in all replicates are shown. Four replicates were used for both samples.

Table EV4: Gene Ontology analysis (molecular function) of proteins significantly enriched in the *ChIRP-MS targeted samples*
Gene Ontology analysis of proteins that were significantly enriched in the *ChIRP-MS RUF6 targeted samples* when compared to both control samples (Tables EV2-EV3). Number of genes annotated with each "molecular function" term in the *P. falciparum* genome (Bgd count), number of significantly enriched proteins with this term (Result count), *p*-value (calculated using a Fisher's exact test), and Benjamini corrected *p*-value (FDR) are shown. Analysis was performed using the GO enrichment tool at PlasmoDB.org (Aurrecoechea et al, 2017).

Table EV5: Candidate proteins significantly enriched in the *ChIRP-MS targeted samples*
Candidate proteins that were significantly enriched in the *ChIRP-MS RUF6 targeted samples* when compared to both control samples (Tables EV2-EV3).

Table EV6: Final candidate proteins significantly enriched in the *ChIRP-MS targeted samples*

Final candidate proteins that were significantly enriched in the *ChIRP-MS RUF6 targeted samples* when compared to both control samples (Tables EV2-EV3).

Table EV7: LC-MS/MS analysis of GFP-tagged DDX5 Pf3D7_1445900 and protein of unknown function Pf3D7_1423700 (control) immunoprecipitation-MS
Proteomics analysis of GFP-tagged DDX5 Pf3D7_1445900 and protein of unknown function Pf3D7_1423700 (control) immunoprecipitation-MS. Comparison of DDX5 to control. Label-free quantitative LC-MS/MS analysis of DDX5 and control. Ratio, log2(ratio), peptides used (Pept. used), total peptides, and protein molecular weight (MW) in kilodaltons (kDa) are provided for all proteins. Adjusted p-value and the percent coefficient of variation (CV %) are shown for proteins present in both samples. Genes and their predicted products are listed. Only proteins with ≥2.0 fold ratio of enrichment, an adjusted p-value ≤ 0.05, and at least three peptides in all replicates are shown. Five replicates were used for both samples.

Table EV8: Gene Ontology analysis (biological process) of proteins significantly enriched in the *DDX5 IP MS samples*
Gene Ontology analysis of proteins that were significantly enriched in the *DDX5 IP MS samples* when compared to control sample (Table EV7). Number of genes annotated with each "biological process" term in the *P. falciparum* genome (Bgd count), number of significantly enriched proteins with this term (Result count), *p*-value (calculated using a Fisher's exact test), and Benjamini corrected *p*-value (FDR) are shown. Analysis was performed using the GO enrichment tool at PlasmoDB.org (Aurrecoechea et al, 2017).

Table EV9: Gene Ontology analysis (molecular function) of proteins significantly enriched in the *DDX5 IP MS samples*
Gene Ontology analysis of proteins that were significantly enriched in the *DDX5 IP MS samples* when compared to control sample (Table EV7). Number of genes annotated with each "molecular function" term in the *P. falciparum* genome (Bgd count), number of significantly enriched proteins with this term (Result count), *p*-value (calculated using a Fisher's exact test), and Benjamini corrected *p*-value (FDR) are shown. Analysis was performed using the GO enrichment tool at PlasmoDB.org (Aurrecoechea et al, 2017).

Table EV10: Shared proteins significantly enriched in the ChIRP RUF6 MS and DDX5 IP MS samples
Shared proteins significantly enriched in the ChIRP RUF6 MS and DDX5 IP MS samples. Results from proteomics analysis of ChIRP RUF6 targeted MS and DDX5 IP-MS.

Table EV11: Gene Ontology analysis (biological process) of shared proteins significantly enriched in the ChIRP RUF6 MS and *DDX5 IP MS samples*
Gene Ontology analysis of shared proteins significantly enriched in the ChIRP RUF6 MS and DDX5 IP MS samples (Table EV10). Number of genes annotated with each "biological function" term in the *P. falciparum* genome (Bgd count), number of significantly enriched proteins with this term (Result count), *p*-value (calculated using a Fisher's exact test), and Benjamini corrected *p*-value (FDR) are shown. Analysis was performed using the GO enrichment tool at PlasmoDB.org (Aurrecoechea et al, 2017).

Table EV12: Gene Ontology analysis (molecular function) of shared proteins significantly enriched in the ChIRP RUF6 MS and *DDX5 IP MS samples*
Gene Ontology analysis of shared proteins significantly enriched in the ChIRP RUF6 MS and DDX5 IP MS samples (Table EV10). Number of genes annotated with each "molecular function" term in the *P. falciparum* genome (Bgd count), number of significantly enriched proteins with this term (Result count), *p*-value (calculated using a Fisher's exact test), and Benjamini corrected *p*-value (FDR) are shown. Analysis was performed using the GO enrichment tool at PlasmoDB.org (Aurrecoechea et al, 2017).

Table EV13: Differential gene expression analysis at 12 hpi of rapamycin-treated over untreated DDX5-GFP-FKBP parasites (genes showing significant change)
Differential gene expression analysis at 12 hpi between untreated (-Rapamycin) and treated (+Rapamycin) DDX5-GFP-FKBP parasites. Analysis was performed for n=5,610 genes (ID and chromosome locations are given) with three replicates for untreated and rapamycin-treated DDX5-GFP-FKBP parasites. Only significantly up- or down-regulated genes are listed, and active *var* is highlighted in grey. log2(FoldChange) = Fold change of baseMean (average of the normalized read counts across all samples and replicates for this gene) in rapamycin-treated/untreated parasites (log2). *p*-values are calculated with a Wald test for significance of coefficients in a negative binomial generalized linear model as implemented in DESeq2 (Love et al, 2014). *q* = Bonferroni corrected *p*-value.

Table EV14: Gene Ontology analysis (biological process) of significantly down-regulated genes in DDX5 knocksideways
Gene Ontology analysis of genes that were significantly down-regulated upon DDX5 knocksideways (from Table EV13). Shown are number of genes annotated with each "biological process" term in the *P. falciparum* genome (Bgd count), number of significantly differentially expressed genes in the ISWI knockdown with this term (Result count), *p*-value (calculated using a Fisher's exact test), and Bonferroni corrected *p*-value (FDR). Analysis was performed using the GO enrichment tool at PlasmoDB.org (Aurrecoechea et al, 2017).

Table EV15: Gene Ontology analysis (biological function) of significantly up-regulated genes in DDX5 knocksideways
Gene Ontology analysis of genes that were significantly up-regulated upon DDX5 knocksideways (from Table EV13). Shown are number of genes annotated with each "biological process" term in the *P. falciparum* genome (Bgd count), number of significantly differentially expressed genes in the ISWI knockdown with this term (Result count), *p*-value (calculated using a Fisher's exact test), and Bonferroni corrected *p*-value (FDR). Analysis was performed using the GO enrichment tool at PlasmoDB.org (Aurrecoechea et al, 2017).

References:

Allen, T. A., Kaenel, S., Goodrich, J. A., & Kugel, J. F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol*.

Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E., & Dzikowski, R. (2015). Antisense long noncoding rnas regulate var gene activation in the malaria parasite plasmodium falciparum. *Proc Natl Acad Sci U S A*, *112*(9).

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*.

Arnold, P. R., Wells, A. D., & Li, X. C. (2020). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Frontiers in Cell and Developmental Biology*, *7*.

Barcons-Simon, A., Cordon-Obras, C., Guizetti, J., Bryant, J. M., & Scherf, A. (2020). CRISPR Interference of a Clonally Variant GC-Rich Noncoding RNA Family Leads to General Repression of var Genes in Plasmodium falciparum. *MBio*, *11*(1).

Baum, J., Papenfuss, A. T., Mair, G. R., Janse, C. J., Vlachou, D., Waters, A. P., Cowman, A. F., Crabb, B. S., & Koning-Ward, T. F. (2009). Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res*.

Birnbaum, J., Flemming, S., & Reichard, N. (2017). A genetic system to study Plasmodium falciparum protein function. *Nat Methods*, *14*, 450–456.

Boivin, V., Faucher-Giguère, L., Scott, M., & Abou-Elela, S. (2019). The cellular landscape of mid-size noncoding RNA. *Wiley Interdisciplinary Reviews. RNA*, *10*(4), 1530.

Bourgeois, C. F., Mortreux, F., & Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat Rev Mol Cell Biol*.

Broadbent, K. M., Broadbent, J. C., Ribacke, U., Wirth, D., Rinn, J. L., & Sabeti, P. C. (2015). Strand specific RNA sequencing in Plasmodium falciparum malaria identi_es developmentally regulated long non-coding RNA and circular RNA. *BMC Genomics*, *16*(454).

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

Bryant, J. M., Baumgarten, S., Dingli, F., Loew, D., Sinha, A., Claës, A., Preiser, P. R., Dedon, P. C., & Scherf, A. (2020). Exploring the virulence gene interactome with CRISPR / DC as9 in the human malaria parasite. *Molecular Systems Biology*, *16*(8).

Bryant, J. M., Regnault, C., Scheidig-Benatar, C., Baumgarten, S., Guizetti, J., & Scherf, A. (2017). Crispr/cas9 genome editing reveals that the intron is not essential for var2csa gene activation or silencing in plasmodium falciparum. *MBio*, *8*(4).

Cajigas, I., Chakraborty, A., Swyter, K. R., Luo, H., Bastidas, M., Nigro, M., Morris, E. R., Chen, S., VanGompel, M. J. W., Leib, D., Kohtz, S. J., Martina, M., Koh, S., Ay, F., & Kohtz, J. D. (2018). The Evf2 Ultraconserved Enhancer lncRNA Functionally and Spatially Organizes Megabase Distant Genes in the Developing Forebrain. *Mol Cell*.

Calderwood, M. S., Gannoun-Zaki, L., Wellems, T. E., & Deitsch, K. W. (2003). Plasmodium falciparum var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J Biol Chem*.

Cech, T. R., & Steitz, J. A. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*, *157*(1), 77–94.

Chakrabarti, K., Pearson, M., Grate, L., Sterne-Weiler, T., Deans, J., Donohue, J. P., & Ares, M. (2007). Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA*, *13*(11), 1923–1939.

Chêne, A., Vembar, S. S., Rivière, L., Lopez-Rubio, J. J., Claes, A., Siegel, T. N., Sakamoto, H., Scheidig-Benatar, C., Hernandez-Rivas, R., & Scherf, A. (2012). PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. *Nucleic Acids Res*.

Chu, C., & Chang, H. Y. (2018). ChIRP-MS: RNA-directed proteomic discovery. *Methods Mol. Biol*, *1861*, 37–45.

Chu, C., Quinn, J., & Chang, H. Y. (2012). Chromatin isolation by RNA purification (ChIRP. *J Vis Exp*.

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E., & Chang, H. Y. (2015). Systematic Discovery of Xist RNA Binding Proteins. *Cell*, *161*(2), 404–416.

Claessens, A., Harris, L. M., Stanojcic, S., Chappell, L., Stanton, A., Kuk, N., Veneziano-Broccia, P., Sterkers, Y., Rayner, J. C., & Merrick, C. J. (2018). RecQ helicases in the malaria parasite Plasmodium falciparum affect genome stability, gene expression patterns and DNA replication dynamics. *PLOS Genetics*, *14*(7), e1007490.

Clark, E. L., Hadjimichael, C., Temperley, R., Barnard, A., Fuller-Pace, F. V., & Robson, C. N. (2013). P68/DdX5 Supports β-Catenin & RNAP II during Androgen Receptor Mediated Transcription in Prostate Cancer. *PLoS ONE*, *8*(1), e54150.

Erdmann, V. A., Barciszewska, M. Z., Hochberg, A., Groot, N., & Barciszewski, J. (2001). Regulatory RNAs. *Cellular and Molecular Life Sciences*, *58*(7), 960–977.

Espinoza, C. A., Allen, T. A., Hieb, A. R., Kugel, J. F., & Goodrich, J. A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nature Structural & Molecular Biology*, *11*(9), 822–829.

Fan, Y., Shen, S., Wei, G., Tang, J., Zhao, Y., Wang, F., He, X., Guo, G., Shang, X., Yu, X., Ma, Z., He, X., Liu, M., Zhu, Q., Le, Z., Wei, G., Cao, J., Jiang, C., Zhang, Q., & Miller, L. H. (2020). Rrp6 Regulates Heterochromatic Gene Silencing via ncRNA RUF6 Decay in Malaria Parasites. *MBio*, *11*(3), e01110-20.

Fernandes, J. C. R., Acuña, S. M., Aoki, J. I., Floeter-Winter, L. M., & Muxel, S. M. (2019). Long non-coding RNAs in the regulation of gene expression: Physiology and disease. *Non-Coding RNA*, *5*(1), 17.

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

Flueck, C., Bartfai, R., Volz, J., Niederwieser, I., Salcedo-Amaya, A. M., Alako, B. T., Ehlgen, F., Ralph, S. A., Cowman, A. F., Bozdech, Z., Stunnenberg, H. G., & Voss, T. S. (2009). Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathogens*, *5*(9), 1000569.

Gage, H. L., & Merrick, C. J. (2020). Conserved associations between G-quadruplex-forming DNA motifs and virulence gene families in malaria parasites. *BMC Genomics*, *21*, 236.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., … Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, *419*(6906), 498–511.

Gazanion, E., Lacroix, L., Alberti, P., Gurung, P., Wein, S., Cheng, M., Mergny, J.-L., Gomes, A. R., & Lopez-Rubio, J.-J. (2020). Genome wide distribution of G-quadruplexes and their impact on gene expression in malaria parasites. *PLOS Genetics*, *16*(7), e1008917.

Goyal, M., Alam, A., Iqbal, M. S., Dey, S., Bindu, S., Pal, C., Banerjee, A., Chakrabarti, S., & Bandyopadhyay, U. (2012). Identification and molecular characterization of an Alba-family protein from human malaria parasite Plasmodium falciparum. *Nucleic Acids Research*, *40*(3), 1174–1190.

Guizetti, J., Barcons-Simon, A., & Scherf, A. (2016). Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite. *Nucleic Acids Res*, *44*, 9710–9718.

Guizetti, J., & Scherf, A. (2013). Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum. *Cell Microbiology*, *15*, 718–726.

Harris, L. M., Monsell, K. R., Noulin, F., Famodimu, M. T., Smargiasso, N., Damblon, C., Horrocks, P., & Merrick, C. J. (2018). G-Quadruplex DNA Motifs in the Malaria Parasite Plasmodium

falciparum and Their Potential as Novel Antimalarial Drug Targets. *Antimicrobial Agents and Chemotherapy*, *62*(3), 01828–17.

Hasenkamp, S., Merrick, C. J., & Horrocks, P. (2013). A quantitative analysis of Plasmodium falciparum transfection using DNA-loaded erythrocytes. *Mol Biochem Parasitol*, *187*, 117–120.

Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., Yang, W., Turner, L., Lavstsen, T., Theander, T. G., Peng, W., Wei, G., Jing, Q., Wakabayashi, Y., Bansal, A., Luo, Y., Ribeiro, J. M., Scherf, A., Aravind, L., … Miller, L. H. (2013). PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. *Nature*.

Jing, Q., Cao, L., Zhang, L., Cheng, X., Gilbert, N., Dai, X., Sun, M., Liang, S., & Jiang, L. (2018). Plasmodium falciparumvar Gene Is Activated by Its Antisense Long Noncoding RNA. *Frontiers in Microbiology*, *9*, 3117.

Kachaev, Z. M., Lebedeva, L. A., Shaposhnikov, A. V., Moresco, J. J., Yates, J. R., Schedl, P., & Shidlovskii, Y. V. (2019). Paip2 cooperates with Cbp80 at an active promoter and participates in RNA Polymerase II phosphorylation in *Drosophila*. *FEBS Letters*, *593*(10), 1102–1112.

Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with BurrowsWheeler transform. *Bioinformatics*, *25*, 1754–1760.

Li, Z., Yin, S., Sun, M., Cheng, X., Wei, J., Gilbert, N., Miao, J., Cui, L., Huang, Z., Dai, X., & Jiang, L. (2019). DNA helicase RecQ1 regulates mutually exclusive expression of virulence genes in Plasmodium falciparum via heterochromatin alteration. *Proceedings of the National Academy of Sciences*, *116*(8), 3177–3182.

Lopez-Rubio, J. J., Gontijo, A. M., Nunes, M. C., Issar, N., Hernandez Rivas, R., & Scherf, A. (2007). 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol*, *66*, 1296–1305.

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

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*, 550.

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

Martins, R. M., Macpherson, C. R., Claes, A., Scheidig-Benatar, C., Sakamoto, H., Yam, X. Y., Preiser, P., Goel, S., Wahlgren, M., Sismeiro, O., Coppée, J. Y., & Scherf, A. (2017). An ApiAP2 member regulates expression of clonally variant genes of the human malaria parasite Plasmodium falciparum. *Sci Rep*.

Miller, L. H., Baruch, D. I., Marsh, K., & Doumbo, O. K. (2002). The pathogenic basis of malaria. *Nature*, *415*, 673–679.

Nguyen, T. N. M., Choo, A., & Baxter, S. W. (2021). Lessons from Drosophila: Engineering Genetic Sexing Strains with Temperature-Sensitive Lethality for Sterile Insect Technique Applications. *Insects*.

Otto, T. D., Gilabert, A., Crellen, T., Böhme, U., Arnathau, C., Sanders, M., Oyola, S. O., Okouga, A. P., Boundenga, L., Willaume, E., Ngoubangoye, B., Moukodoum, N. D., Paupy, C., Durand, P., Rougeron, V., Ollomo, B., Renaud, F., Newbold, C., Berriman, M., & Prugnolle, F. (2018). Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol*, *3*, 687–697.

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

Petter, M., Lee, C. C., Byrne, T. J., Boysen, K. E., Volz, J., Ralph, S. A., Cowman, A. F., Brown, G. V., & Du_y, M. F. (2011). Expression of p. Falciparum var genes involves exchange of the histone variant h2a.z at the promoter. *PLoS Pathog*, *7*(2).

Petter, M., Selvarajah, S. A., Lee, C. C., Chin, W. H., Gupta, A. P., Bozdech, Z., Brown, G. V., & Du_y, M. F. (2013). 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*(6), 1167–1182.

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., & Grummt, I. (2015). LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. *Mol Cell. Nov*, *19;60(4):626-36*.

Poullet, P., Carpentier, S., & Barillot, E. (2007). MyProMS, a web server for management and validation of mass spectrometry-based proteomic data. *Proteomics*, *7*, 2553–2556.

Raabe, J. R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P. A., Hausler, D., & Kamakaka, R. T. (2012). Human trna genes function as chromatin insulators. *EMBO J*, *31*(2), 330–350.

Ralph, S. A., Scheidig-Benatar, C., & Scherf, A. (2005). Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. *Proc Natl Acad Sci U S A*, *102*, 5414–5419.

Reddy, B. N., Shrestha, S., & Hart, K. J. (2015). A bioinformatic survey of RNA-binding proteins in Plasmodium. *BMC Genomics*, *16*, 890.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*, 139–140.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nat Methods*, *9*, 676–682. https://doi.org/10.1038/nmeth.2019.)

Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., & Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*.

Siegel, T. N., Hon, C. C., Zhang, Q., Lopez-Rubio, J. J., Scheidig-Benatar, C., Martins, R. M., Sismeiro, O., Coppée, J. Y., & Scherf, A. (2014). Strand-speci_c RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum. *BMC Genomics*, *15*(150).

Sierra-Miranda, M., Delgadillo, D. M., Mancio-Silva, L., Vargas, M., Villegas-Sepulveda, N., Martínez-Calvillo, S., Scherf, A., & Hernandez-Rivas, R. (2012). Two long non-coding RNAs generated from subtelomeric regions accumulate in a novel perinuclear compartment in Plasmodium falciparum. *Mol Biochem Parasitol*.

Smargiasso, N., Gabelica, V., Damblon, C., Rosu, F., Pauw, E., Teulade-Fichou, M. P., Rowe, J. A., & Claessens, A. (2009). Putative DNA G-quadruplex formation within the promoters of Plasmodium falciparum var genes. *BMC Genomics*.

Su, X. Z., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. A., & Wellems, T. E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of plasmodium falciparum-infected erythrocytes. *Cell*, *82*(1), 89–100.

Tan-Wong, S. M., Dhir, S., & Proudfoot, N. J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol Cell*.

Tuteja, R. (2017). Unraveling the importance of the malaria parasite helicases. *FEBS J*.

Valot, B., Langella, O., Nano, E., & Zivy, M. (2011). MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics*, *11*, 3572–3577.

Vembar, S. S., Macpherson, C. R., Sismeiro, O., Coppée, J. Y., & Scherf, A. (2015). The pfalba1 rna-binding protein is an important regulator of translational timing in plasmodium falciparum blood stages. *Genome Biol*, *16*(1).

Wahlgren, M., Goel, S., & Akhouri, R. R. (2017). Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. *Nat Rev Microbiol*.

Wei, G., Zhao, Y., Zhang, Q., & Pan, W. (2015). Dual regulatory effects of non-coding GC-rich elements on the expression of virulence genes in malaria parasites. *Infect Genet Evol*, *36*, 490–499.

W.H.O. (2021). *World Malaria Report 2021*.

Wu, G., Xing, Z., Tran, E. J., & Yang, D. (2019). DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proceedings of the National Academy of Sciences*, *116*(41), 20453.

Zhang, M., Wang, C., Otto, T. D., Oberstaller, J., Liao, X., Adapa, U., SR, K, B., IF, C., D, M., M, B., J, L., S, S., J, R., JC, J., RHY, A., & J.H. (2018). Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis. *Science*.

Zhang, Q., Siegel, T. N., Martins, R. M., Wang, F., Cao, J., Gao, Q., Cheng, X., Jiang, L., Hon, C. C., Scheidig-Benatar, C., Sakamoto, H., Turner, L., Jensen, A. T., Claes, A., Guizetti, J., Malmquist, N. A., & Scherf, A. (2014). Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. *Nature*, *513*, 431–435.

Zhu, X., Wei, Y., & Dong, J. (2020). Long Noncoding RNAs in the Regulation of Asthma: Current Research and Clinical Implications. *Frontiers in Pharmacology*, *11*.

2.2 RUF6 ncRNA is regulated by external factors

The highlights of this section are:

- RUF6 ncRNA is downregulated upon addition of $MgCl_2$
- PfMaf1 is involved in RNA Pol III transcriptional regulation

The following results are part of an additional project to my main PhD project. Results are presented in the form of an unsubmitted manuscript. My contribution to this work was conceptualization, experimental design, performing all experiments (with the exception of the RNA Pol III inhibition assays and RNA-seq experiments), analyzing the data, and writing the manuscript. During this time, I was responsible for supervising two separate M1 students, Marta Miera Maluenda and Nathan Pinatel. Each stayed for 6 months (February 2021-July 2021 and November 2021 to May 2022).

2.2 RUF6 ncRNA is regulated by external factors

Environmental factors regulate malaria parasite virulence via the downregulation of RNA Polymerase III-transcribed ncRNA gene family

Diffendall, G[1,2]., Barcons-Simon, A[1,2,#]., Claës, A[1]., Scherf, A[1 *]

[1]Universite Paris Cité, Institut Pasteur, Biology of Host-Parasite Interactions Unit, INSERM U1201, CNRS EMR9195, Paris, France.

[2]Sorbonne Université Ecole doctorale Complexité du Vivant ED515, Paris, France.

[#]Actual address: Biomedical Center, Division of Physiological Chemistry, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich 82152, Germany

**Corresponding author: artur.scherf@pasteur.fr*

Abstract

During the dry season, in many African regions, persistent asymptomatic malaria cases have been linked to downregulated virulence gene expression resulting in low adhesion and increased circulation time of *Plasmodium falciparum*-infected erythrocytes. The underlying mechanism responsible for lower expression of the surface adhesins encoded by the *var* gene family is poorly understood. Here we discover environmental factors that trigger downregulation of *var* gene transcription. We observe that isoleucine starvation and high $MgCl_2$ concentrations in the medium inhibit RNA Polymerase III transcribed genes. Importantly, this includes a *P. falciparum*-specific regulatory ncRNA gene family (encoded by the RUF6 gene family) that is a key regulator in *var* gene activation. We identified a homologous gene to the highly conserved eukaryotic Maf1, as a negative effector of RUF6 ncRNA transcription. Elevated $MgCl_2$ concentrations led to a shift of cytoplasmic PfMaf1 to the nuclear compartment. We used an inducible protein degradation system to show that external stimuli depend on PfMaf1 to trigger lower expression of RUF6 genes. Our results uncover a TOR-independent signaling pathway that inhibits Pol III transcription and alters

parasite virulence. This work provides novel molecular understanding that is highly relevant for malaria pathogenesis of subclinical parasite persistence in the dry season.

Keywords: RNA Pol III, Maf1, var genes, RUF6 ncRNA, malaria virulence,

Introduction

The parasite *Plasmodium falciparum* is responsible for the deadliest form of human malaria that annually affects over 200 million people, with 627,000 fatal cases, majorly African children under the age of 5 (W.H.O., 2021). The disease is transmitted to its human host during a blood meal by the parasites' vector, *Anopheles* mosquitoes. Disease symptoms are seen while the parasite multiplies asexually within the host red blood cells (RBCs). During this time, variant surface antigens (VSAs) are exported and displayed on the surface of the infected RBCs (iRBCs). During the trophozoite and schizont stage, the VSA that is linked to immune evasion and pathogenesis, termed *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Leech et al., 1984), mediate adhesion to the vascular endothelium within the host, thereby preventing mature iRBCs from traveling to, and being cleared by, the spleen. During the ~48-hour asexual intraerythrocytic developmental cycle (IDC), the parasite develops through different morphological stages: ring, trophozoite, and schizont. During schizont stage 16-32 merozoites are formed, that, upon bursting of the iRBC, invade new RBCs thus enabling the cycle to restart.

P. falciparum relies largely on mutually exclusive expression of PfEMP1 surface adhesins, that are encoded by the *var* gene family (reviewed in Scherf et al., 1998; Smith et al., 1995). This 60-member gene family is conserved among species from the *Laverania* subgenus of *Plasmodium*, which includes the only species that can infect humans (Larremore et al., 2015; Otto et al., 2018). Most of the members of this gene family are located at subtelomeres of the 14 chromosomes and approximately 24 out of 60 *var* genes are in central regions of 5 chromosomes. The mutually exclusive expression of *var* genes, while under tight epigenetic regulation by the parasite (Guizetti and Scherf, 2013), is additionally controlled by a family of ncRNA, termed RUF6. *P. falciparum* 3D7 encodes 15 RUF6 genes dispersed over several chromosomes with a location adjacent to central *var* genes (Gardner et al., 2002). RUF6 ncRNA have been observed to associate with the active *var* gene in trans and overexpression or transcriptional repression disrupts the monoallelic expression of *var* genes by up-regulating multiple *var* genes or down-regulating the entire *var* gene

family, respectively (Barcons-Simon et al., 2020; Guizetti et al., 2016). During prolonged dry seasons in many endemic African regions subclinical *P. falciparum* persistence have been linked to reduced *var* gene transcription resulting in low adhesion and increased circulation time of infected erythrocytes (Andrade et al., 2020). Environmental factors and pathways, triggering lower expression of the surface adhesins, are poorly understood in malaria parasites.

In almost all eukaryotes, the target of rapamycin complex (TORC) pathway has been shown to control many cellular processes including signaling of external factors (reviewed in (Wullschleger et al., 2006)) Specifically, TORC participates in the regulation of the RNA Pol III activity via the phosphorylation of a specific inhibitor termed Maf1. When nutrients are available and mTOR kinase is active, Maf1 is hyperphosphorylated resulting in an inactive cytoplasmic state. Stress-induced Maf1 dephosphorylation results in nuclear localization and inhibition of RNA polymerase III (Pluta et al., 2001; Rollins et al., 2007; Vannini et al., 2010). This pathway responds to a range of positive and negative external stimuli, most notably the presence of amino acids, that drive or inhibit cellular growth.

The majority of the TORC pathway components are lost in the apicomplexan lineage (McLean & Jacobs-Lorena, 2017; Serfontein et al., 2010; van Dam et al., 2011). While *P. falciparum* encodes none of the core TORC1 components, it does contain the only known RNA Pol III repressor, Maf1 (McLean & Jacobs-Lorena, 2017). In contrast to other eukaryotes, plasmodial Maf1 seemingly affects asexual blood stage proliferation making classical gene KO experiments unattainable (McLean & Jacobs-Lorena, 2017; Zhang et al., 2018). A transposon insertion into the promoter region of PfMaf1 gene generates a deregulated transcription profile leading to an improved recovery from amino acid starvation and other stress factors (McLean & Jacobs-Lorena, 2017).

The biology of *P. falciparum* Pf-Maf1 is still poorly studied and its link to RNA Pol III transcribed regulatory ncRNA has not been investigated. We set out to explore if external factors can trigger changes in the transcription of RUF6 ncRNA. Our previous bioinformatic analysis predicted that RUF6 may be transcribed by RNA Pol III (Guizetti et al., 2016) and in this work we further corroborate this idea showing that RUF6 transcription is sensitive to RNA Pol III inhibitor. We identify isoleucine starvation and increased $MgCl_2$ concentrations as external factors that do inhibit RNA Pol III transcribed genes including RUF6. We provide experimental evidence for PfMaf1

being a key player in the signal pathway that inhibits RUF6 transcription leading to reduced *var* gene expression. This finding is highly relevant for exploring the underlying mechanism of subclinical *P. falciparum* persistence in malaria patients.

Results:

RUF6 ncRNA is transcribed by RNA Polymerase III

Previous bioinformatical analysis highlighted conserved promoter elements, namely A- and B-box, within the RUF6 genes, suggesting transcription via RNA Pol III (Guizetti et al., 2016). Here we performed RNA Pol III inhibition assays with synchronized wild type parasites to test whether RUF6 genes are transcribed by this polymerase (Fig. 1A) (Dieci et al., 2013; Guizetti et al., 2016). We observed a similar downregulation in the levels of tRNA and RUF6 transcripts after treatment with 50 μM of an RNA Pol III inhibitor (CAS 577784-91-9, Calbiochem) for 20 hours. No transcript level changes were observed for a RNA Pol II transcribed control gene, ubiquitin-conjugating enzyme (PF3D7_0812600) (Fig. 1B). Our data provide direct experimental support for RNA Pol III-dependent transcription of RUF6 genes, corroborating DNA sequence predictions.

RNA Pol III-transcribed genes are downregulated upon isoleucine deprivation and $MgCl_2$ addition

To investigate external factors that may trigger changes in RNA Pol III transcription in *P. falciparum*, we chose to use Ile-deprived culture medium for nutrient deprivation, as was described earlier (Babbitt et al., 2012; McLean & Jacobs-Lorena, 2017). We additionally included the addition of magnesium chloride ($MgCl_2$). We chose 3mM $MgCl_2$ in our culture medium since this concentration does not affect growth (Fig. 3C), similar to an earlier report (Hess et al., 1995) where they showed that culture concentrations for *P. falciparum* support a wide range from 0.5 to 5.0mM without affecting parasitic growth after 24 hours. We evaluated the transcript levels for tRNA valine and RUF6 ncRNA at various time points (6, 18, and 24 hpi) using *P. falciparum* 3D7 cultivated either without isoleucine or with the addition of $MgCl_2$ (Fig. 2A, Fig. S1). We observed that both conditions reduced RNA Pol III dependent transcription. 3mM $MgCl_2$ in the culture medium significantly decreased transcription levels at 18 hpi for RNA Pol III-transcribed genes, tRNA valine and RUF6 ncRNA while deprivation of isoleucine resulted in slightly less decrease

of RUF6 and tRNA transcription. At 18hpi, the RNA Pol II-transcribed active *var* gene, Pf3D7_1240900, was significantly decreased (70%) upon addition of $MgCl_2$.

Given the greater decrease in RUF6 ncRNA transcription at 3mM $MgCl_2$, we used this culture condition for a global assessment of transcriptional changes. RNA-seq was done on control and 3mM $MgCl_2$ cultured parasites at 12 and 24 hpi for two consecutive cycles. At 24 hpi in the first cycle, 39 RNA Pol III transcribed genes were downregulated including all 15 of the RUF6 ncRNA, 8 tRNAs, 4 spliceosomal RNAs, and 1 snoRNA (Fig. 2B). In the second cycle at 12 hpi, 53 genes were downregulated (logFC < 1.95, FDR < 0.05) including all 15 RUF6 ncRNA, 6 tRNAs, 1 spliceosomal RNA, and 3 snoRNA (Fig. 2C). Most importantly, the active *var* gene was significantly downregulated. Additionally, 12 genes were up-regulated including a cAMP-dependent protein kinase, Pf3D7_0934800. This demonstrates that 3mM $MgCl_2$ primarily inhibits RNA Pol III transcribed genes, including the entire RUF6 gene family that acts as activator of the major virulence gene family of *P. falciparum.*

Validation of an inducible protein degradation system for PfMaf1

To assess PfMaf1 localization and perform functional studies, *P. falciparum* parasites were transfected as previously described by (Birnbaum et al., 2017) modified with the addition of a 3HA tag and a ligand-controlled destabilization domain (ddFKBP), pSLI-Maf1-FKBP (Fig. 3A). The ddFKBP domain enables rapid and reversible protein stabilization by reversible binding of a synthetic ligand, termed Shield-1. Upon removal of Shield-1, the ddFKBP domain is not stabilized, leading to degradation of the protein of interest, PfMaf1. pSLI-Maf1-FKBP transfected parasite clonal populations were validated for proper integration. Confirmation that the system was working was shown by western blot for addition and removal of Shield-1 (Fig. 3B). We observed that upon removal of Shield-1, PfMaf1 levels were shown to decrease by approximately 57% in total extracts. Parasite growth rate was assessed over the course of 5 days for 4 different conditions: in the presence or absence of PfMaf1 (+/- Shield) and magnesium chloride (-/+$MgCl_2$, Fig. 3C) on two pSLI-Maf1-FKBP clones. No significant difference in growth rate was observed after one cycle for all of the conditions. However, after two cycles conditions where $MgCl_2$ was added, where Shield was removed, and when both $MgCl_2$ was added and Shield removed, resulted in a significantly lower growth rate. These results suggest that the level of decreased PfMaf1 and

3mM $MgCl_2$ significantly alter growth in culture. Transposon mutagenesis of PfMaf1 reported an effect on parasite growth (Zhang et al., 2018), and attempts for classical gene KO studies were not successful (McLean & Jacobs-Lorena, 2017). Here we have generated an inducible PfMaf1 protein degradation system that allows to explore its role in RNA Pol III transcription under culture conditions and in particular the RUF6 ncRNA-dependent transcription of the *var* gene family.

Nuclear PfMaf1 levels increase upon addition of $MgCl_2$

The localization of PfMaf1 was investigated to determine if the protein is predominantly cytoplasmic under regular culture conditions as has been described in other eukaryotic systems (Orioli et al., 2016). Parasites were tightly synchronized, split into control and addition of $MgCl_2$, and harvested at 18hpi. PfMaf1 localizes to both the cytoplasm and nucleus in control culture conditions during asexual blood stage development. Western blot analysis indicates that the majority of PfMaf1 is in the cytoplasmic fraction (Fig. 4A). Importantly, upon addition of $MgCl_2$ (3mM final concentration), we observed 1.64 times increase in nuclear PfMaf1 levels (Fig. 4A). Immunofluorescence (IF) was done at 18hpi for control and addition of $MgCl_2$ parasites (Fig. 4B). PfMaf1 appeared to cluster in foci near the nuclear periphery in both culture conditions.

PfMaf1 is involved in regulating RNA Pol III-transcribed genes including RUF6

We made use of the PfMaf1-FKBP transfected parasites to investigate if the downregulation of RNA Pol III-transcribed genes triggered by 3mM $MgCl_2$ is dependent on PfMaf1. We used the parasite growth conditions described earlier for establishing the growth curve (Fig. 3C). Parasites were harvested at 18 hpi. and RNA was extracted and prepared for RT-qPCR (Fig. 4C). We observed the previously observed down-regulation of RNA Pol III-transcribed tRNA valine and RUF6 upon addition of $MgCl_2$ at 18 hpi in two pSLI-Maf1-FKBP clones. Notably, when PfMaf1 was decreased (-Shield) and $MgCl_2$ was added, Pol III transcription did not decrease. In fact, no other condition, besides +Shield + $MgCl_2$, showed a significant decrease in transcription of RUF6 genes when compared to the control (+Shield, - MgCl2). As observed earlier in Fig. 2, the reduction of RUF6 ncRNA leads to lower transcription of *var* genes. These results suggest that PfMaf1 is necessary for the decrease in RNA Pol III-transcribed genes observed upon addition of $MgCl_2$.

Discussion

While it is now established that RUF6 ncRNA is involved in the regulation of *var* gene expression, internal and external factors that regulate this ncRNA gene family have not been explored yet. Here we present experimental evidence that PfMaf1, a well-known repressor of RNA Pol III in other organisms, is involved in regulating *P. falciparum* virulence gene expression. Based on DNA sequence elements typically found in RNA Pol III transcribed genes termed A and B box, RUF6 was first predicted to be transcribed by RNA Pol III (Fig. 1A). Here we show that RNA Pol III inhibitors reduce RUF6 transcription (Fig. 1B), supporting the concept that RNA Pol III transcribed ncRNA regulates RNA Pol II transcribed virulence genes.

Given the observed link between mTOR signaling pathway and RNA Pol III repressor Maf1 in yeast and other eukaryotes (Michels, 2011), we investigated if plasmodial PfMaf1 can be activated by external factors and modulate parasite virulence. We identified two factors in the *P. falciparum* by depleting (-Ile) or supplementing (+$MgCl_2$) the culture medium, that influence RNA Pol III-transcribed genes (Fig. 2). Isoleucine starvation and $MgCl_2$ concentrations both affected RNA Pol III activity. $MgCl_2$ yielded repeatedly higher levels of inhibition and was therefore chosen as the external stimulus to explore the pathway that regulates the ncRNA RUF6. Importantly, the transcription levels of RUF6 ncRNA was downregulated upon exposure to 3mM $MgCl_2$ levels when both RUF6 and *var* gene transcription occurs (Fig. 2B). Expectedly, because of RUF6 downregulation, RNA Pol II transcribed *var* gene transcription was significantly reduced. RNA-seq experiments further validated that $MgCl_2$ significantly affects RNA Pol III-transcribed genes, but not RNA Pol II (Fig. 2B, C), except for *var* genes. Downregulated ncRNA including: tRNA, spliceosomal RNA, and snoRNA all have the TFIIIB subunit, typical of type 2 and 3 RNA Pol III promoters.

We further went on to investigate if PfMaf1 activity is linked to its cellular localization. While the activation of Maf1 is linked to dephosphorylation followed by nuclear re-localization in other organisms. Several phosphorylation sites of PfMaf1 have been reported (https://plasmodb.org), but their impact on the nuclear PfMaf1 location in *P. falciparum*, especially in response to external factors, has not been studied yet. We show that PfMaf1 localizes predominantly to the cytoplasm

in control conditions and increases its nuclear occupancy upon $MgCl_2$ addition (Fig. 4A). The increased nuclear levels, upon $MgCl_2$ addition, correlate with the decrease in RNA Pol III-transcribed genes (Fig. 4B). IFA analysis revealed that Pf-Maf1 clusters in distinct foci throughout the IDC (data not shown) and localizes near the nuclear periphery (Fig. 4B). This might be PfMaf1 storage bodies that enable re-localization to the nucleus upon unfavorable conditions. In contrast to other eukaryotic studies of Maf1, *P. falciparum* expresses a basic level of this repressor in the nucleus. This pathogen may need to dampen RNA Pol III transcription levels to maintain tRNA and other ncRNA at physiological levels for optimal parasite proliferation. This could explain why PfMaf1 gene is essential during long-term asexual blood stage development (McLean & Jacobs-Lorena, 2017).

What firstly drew our attention to PfMaf1 is the highly conserved 150 amino acid N-terminal domain found only within *Laverania* species, the only species that contain *var* genes and RUF6 (Fig. S1). We hypothesize that this conserved region has evolved to target PfMaf1 to RUF6 genes and regulate its transcription. Removing this region by genome editing was not successful in our hands, possible because this may have caused changes in the expression level of Pf-Maf1 thereby interfering with its essential function.

Transmission of malaria by mosquitos is seasonal in many African endemic regions, where malaria is most prevalent. The year is divided into a rainy season and a dry season. Malaria cases are restricted to the rainy season when its mosquito vectors are present. Because of the lack of transmitting mosquito vectors, the dry season poses a great challenge for *P. falciparum* parasites. Parasites thus need to survive within the infected individuals, without killing them, until transmission is possible with the return of mosquitoes in the rainy season. A recent remarkable study on clinical *P. falciparum* isolates taken during the dry and rainy season observed a significant reduction in the level of expression of *var* genes, which could explain the subclinical persistent parasites (Andrade et al., 2020). The underlying mechanism remains poorly understood. But the transition between the rainy and dry season most likely comes along with changes in the food sources and overall nutrient levels and possibly the microbiome in the host. The effect on nutrient availability from the host has been demonstrated to have a profound effect on malaria parasite using a rodent malaria model (Mancio-Silva et al., 2017). Diet changes may activate parasite

intracellular pathways; for example, it has been shown that calorie restriction leads to an increase in intracellular magnesium ion production (Abraham et al., 2016). In this work, we have identified two external factors that impact RUF6 ncRNA transcript level during in vitro culture. It remains to be explored if Ile and $MgCl_2$ concentration vary in the plasma of malaria patients during the rainy and dry season. Evidently, a systematic plasma metabolic composition analysis may identify other factors that could modulate *var* gene expression.

In conclusion, this work reveals a pathway that links external plasma factors with the nuclear activation of a repressor of RNA Pol III transcribed genes, including a *P. falciparum*-specific ncRNA gene family that regulates *var* gene transcription (see schematic model in Fig. 5). Importantly, this work uncovers novel molecular mechanism that is highly relevant for our understanding of factor leading to subclinical parasite persistence in the dry season. Importantly, our work provides a testable hypothesis that RNA Pol III transcription inhibition via the PfMaf1 repressor modulates malaria pathogenesis.

Acknowledgments

This work was supported by the Laboratoire d'Excellence (LabEx) ParaFrap [ANR-11-LABX-0024], ERC AdG PlasmoSilencing (670301). We thank Miguel Santos, Marta Miera Maluenda, and Natan Pinatel for technical assistance with this project.

Author contributions

Conceptualization A.S, GD, AB; Methodology: GD, AB, AC; Investigation GD, AB, AC; Analysis: GD, AB; Writing: GD and A.S; Funding acquisition: A.S

Materials and Methods:

Parasite culture and synchronization:

Asexual blood stage 3D7 *P. falciparum* parasites were cultured as previously described in (Lopez-Rubio et al., 2009). Parasites were cultured in human RBCs (obtained from the Etablissement Français du Sang with approval number HS 2016-24803) in RPMI-1640 culture medium (Thermo Fisher 11875) supplemented with 10% v/v Albumax I (Thermo Fisher 11020), hypoxanthine (0.1 mM final concentration, C.C.Pro Z-41-M) and 10 mg gentamicin (Sigma G1397) at 4% hematocrit and under 5% O_2, 3% CO_2 at 37 °C. Parasite development was monitored by Giemsa staining. Parasites were synchronized by sorbitol (5%, Sigma S6021) lysis at ring stage, plasmagel (Plasmion, Fresenius Kabi) enrichment of late stages 24 h later, and an additional sorbitol lysis 6 h after plasmagel enrichment. The 0 h time point was considered to be 3 h after plasmagel enrichment. Parasites were harvested at 1–5% parasitemia.

Polymerase III inhibition assay:

Parasites were treated with 50 μM of RNA Pol III inhibitor CAS 577784-91-9 (Calbiochem, Merck) after sorbitol treatment at 3 ±3 hpi and RNA was harvested at 24 hpi in parallel with untreated and control samples. The control was treated with the same volume of DMSO added to the inhibitor treated flasks of stock solution.

External stimuli induction (+$MgCl_2$ & Ile-deprivation):

Synchronized 0 ±3 hpi parasites were divided into control addition of magnesium chloride ($MgCl_2$), and isoleucine-deficient and harvested at 6 hpi, 18 hpi and 24 hpi. Parasites exposed to an addition of $MgCl_2$ were supplemented with 2.5 mM $MgCl_2$, for a final concentration of [3mM] including [0.5mM] found in RPMI. Isoleucine-deficient medium consisted of 10.3 g/liter RPMI 1640 isoleucine (Ile) Drop-out medium (United States BioLogicals; catalog no. R9014), supplemented with 2.0 g/liter $NaHCO_3$, 6.0 g/liter HEPES, 10% v/v Albumax I (Thermo Fisher 11020), hypoxanthine (0.1 mM final concentration, C.C.Pro Z-41-M) and 10 mg gentamicin (Sigma G1397). Parasites were then harvested with 0.075% Saponin lysis at ~2-5% parasitemia for RNA, genomic DNA, and protein extraction at 6 hpi, 18 hpi and 24 hpi.

RNA isolation and reverse transcription-quantitative PCR (RT-qPCR):

RNA was harvested from synchronized parasite cultures after saponin lysis in 0.075% saponin in PBS, followed by one wash in PBS and resuspension in the QIAzol reagent. Total RNA was extracted using an miRNeasy minikit and performing on-column DNase treatment (Qiagen). Reverse transcription from total RNA was achieved using SuperScript VILO (Thermo Fisher Scientific) and random hexamer primers. cDNA levels were quantified by quantitative PCR in the CFX384 real time PCR detection system (BioRad) using Power SYBR Green PCR Master Mix (Applied Biosystems) and primers from a previous study (Guizetti et al., 2016). Starting quantity means of three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA. Transcript levels shown by using the following primers: RUF6, valine, alanine, *var* 58, *var* con, and Ap2G were normalized to the reference gene, fructose-bisphosphate aldolase (PF3D7_1444800). The starting quantity means from three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA.

Stranded RNA sequencing and analysis:

Infected RBCs containing synchronized (12 and 24 hpi ± 3 h) parasites were lysed in 0.075% saponin (Sigma S7900) in DPBS at 37°C. The parasite cell pellet was washed once with DPBS and then resuspended in 700 µL QIAzol reagent (Qiagen 79306). Total RNA was subjected to rRNA depletion to ensure ncRNA and mRNA capture using the RiboCop rRNA Depletion Kit (Lexogen) prior to strand-specific RNA-seq library preparation using the TruSeq Stranded RNA LT Kit (Illumina) with the KAPA HiFi polymerase (Kapa Biosystems) for the PCR amplification. Multiplexed libraries were subjected to 150 bp paired-end sequencing on a NextSeq 500 platform (Illumina). Sequenced reads (150 bp paired end) were mapped to the *P. falciparum* genome (Gardner et al., 2002) (plasmoDB.org, version 3, release 57) using "bwa mem" (Li & Durbin, 2009) allowing a read to align only once to the reference genome (option "–c 1"). Alignments were subsequently filtered for duplicates and a mapping quality ≥ 20 using samtools (Li & Durbin, 2009). Three biological replicates for -$MgCl_2$ and +$MgCl_2$ samples were analyzed for both timepoints.

Generation of PfMaf1 strains:

All cloning was performed using KAPA HiFi DNA Polymerase (Roche 07958846001), In-Fusion HD Cloning Kit (Clontech 639649), and XL10-Gold Ultracompetent *E. coli* (Agilent Technologies 200315). Transgenic pSLI parasites were generated as previously described in (Birnbaum et al.,

2017) with the following modifications: GFP was replaced with a 3HA tag and a ddFKBP domain was added after the protein of interest, PfMaf1. For localization and knock down studies, the last 500-1000 bp of target gene, PfMaf1 Pf3D7_0416500, was cloned into pSLI-3HA-ddFKBP. Each sequence started with an in-frame stop codon but the stop codon at the end of the gene was removed. 50 µg of plasmid DNA was transfected into ring stage 3D7 *P. falciparum* parasites using the protocol described elsewhere (Hasenkamp et al., 2013). Transfected parasites were selected with constant drug selection pressure of 4 nM WR99210 (Jacobus Pharmaceuticals) to obtain a cell line containing the episomal plasmid. A second drug selection using 400 µg/ml of G418 was done to select for integrants. Once parasites emerged, gDNA of each integration cell line was collected using a commercial kit (DNeasy Blood & Tissue Kit) and checked by PCR to show that integration occurred at the correct locus. Both genome and vector specific primers for the 5' and 3' region were used so that the PCR product would cover the plasmid/genome junction. Vector primers used were the same as described (Birnbaum et al., 2017). Once proper size gel bands from PCR were seen, parasites were cloned by limiting dilution, and the targeted genomic locus was sequenced to confirm tag and FKBP integration.

Western blot analysis:

Shield-1 was removed for one full cycle before harvesting parasites for western blot. $MgCl_2$ was added at 0hpi. iRBCs were washed once with Dulbecco's phosphate-buffered saline (DPBS, Thermo Fisher 14190) at 37°C and lysed with 0.075% saponin (Sigma S7900) in DPBS at 37°C. Parasites were washed once with DPBS, resuspended in 1 ml cytoplasmic lysis buffer (25 mM Tris–HCl pH 7.5, 10 mM NaCl, 1.5 mM $MgCl_2$, 1% IGEPAL CA-630, and 1× protease inhibitor cocktail ["PI", Roche 11836170001]) at 4°C, and incubated on ice for 30 min. Cells were further homogenized with a chilled glass douncer, and the cytoplasmic lysate was cleared with centrifugation (13,500 *g*, 10 min, 4°C). The pellet (containing the nuclei) was resuspended in 100 µl nuclear extraction buffer (25 mM Tris–HCl pH 7.5, 600 mM NaCl, 1.5 mM $MgCl_2$, 1% IGEPAL CA-630, PI) at 4°C and sonicated for 10 cycles with 30 s (on/off) intervals (5 min total sonication time) in a Diagenode Pico Bioruptor at 4°C. This nuclear lysate was cleared with centrifugation (13,500 *g*, 10 min, 4°C). Protein samples were supplemented with NuPage Sample Buffer (Thermo Fisher NP0008) and NuPage Reducing Agent (Thermo Fisher NP0004) and denatured for 5 min at 95°C. Proteins were separated on a 4-15% TGX (Tris-Glycine eXtended) (Bio-Rad) and

transferred to a PVDF membrane. The membrane was blocked for 1 h with 5% milk in PBST (PBS, 0.1% Tween 20) at 25°C. HA-tagged proteins and histone H3 were detected with anti-HA (Abcam 9110, 1:1,000 in 5% milk-PBST) and anti-H3 (Abcam ab1791, 1:1,000 in 5% milk-PBST) primary antibodies, respectively, followed by donkey anti-rabbit secondary antibody conjugated to horseradish peroxidase ("HRP", Sigma GENA934, 1:5,000 in 5% milk-PBST). Aldolase was detected with anti-aldolase-HRP (Abcam ab38905, 1:5,000 in 5% milk-PBST). HRP signal was developed with SuperSignal West Pico chemiluminescent substrate (Thermo Fisher 34080) and imaged with a ChemiDoc XRS+ (Bio-Rad).

Immunofluorescence assay:

pSLI-Maf1-FKBP parasites were used with rat anti-HA (Roche 3F10) antibodies. 10 µl of iRBCs were washed with PBS and fixed with for 30 min in 0.0075% Glutaraldehyde/4% PFA/PBS. After PBS washing, parasites were permeabilized with 0.1% TritonX100/PBS for 10-15 min before quenching free aldehyde groups with $NaBH_4$ solution for 10 min. Next, parasites were blocked with 3% BSA–PBS for 30 min. Primary antibody incubation lasted for 1 h before three washes with PBS, and secondary antibody incubation for 30-60 min, Alexa Fluor 488-conjugated anti-mouse IgG (Invitrogen) diluted 1:2,000 in 4% BSA–PBS. After three final washes in PBS, cells were mounted in Vectashield containing DAPI for nuclear staining. Images were captured using a Nikon Eclipse 80i microscope with a CoolSnap HQ2 camera (Photometrics). NIS elements 3.0 software (Nikon) was used for acquisition and Fiji software (http://fiji.sc/) for analysis.

Flow cytometry:

Two different pSLI-Maf1-FKBP parasite clones were tightly synchronized and diluted to 0.2% parasitemia (5% hematocrit) at ring stage. One cycle before, Shield-1 was removed from half of the culture. Parasites were tightly synchronized and split was culture was further split for addition of $MgCl_2$ (3.0 mM final concentration) to half of the culture that had Shield-1 and half that had Shield-1 removed. The growth curve was performed in a 96-well plate (200 µl culture per well) with three technical replicates per condition per clone. At 0 h, 24 h, 48 h, 72 h, and 96 h, 5 µl of the culture was stained in 95 µl of D-PBS (Gibco) supplemented with 2× Sybr Green I (Ozyme; stock = 10,000×) for 30 min at room temperature, diluted 20-fold in D-PBS (final volume = 200 µl), and the Sybr Green fluorescence measured in a Guava easyCyte Flow Cytometer (EMD

Millipore). 10000 events were counted in duplicate to establish an accurate parasitemia value for each culture. Data were analyzed using the InCyte software (EMD Millipore).

Statistical analysis:

All statistical analyses were performed using GraphPad Prism version 9.1.0 (216) for Mac. To test for a normal distribution of the data, the Shapiro-Wilk normality test was used. To test for significance between two groups, a two-sided independent-samples t test was used. Gene ontology enrichments were calculated using the build-in tool at https://plasmoDB.org.

2.2 RUF6 ncRNA is regulated by external factors

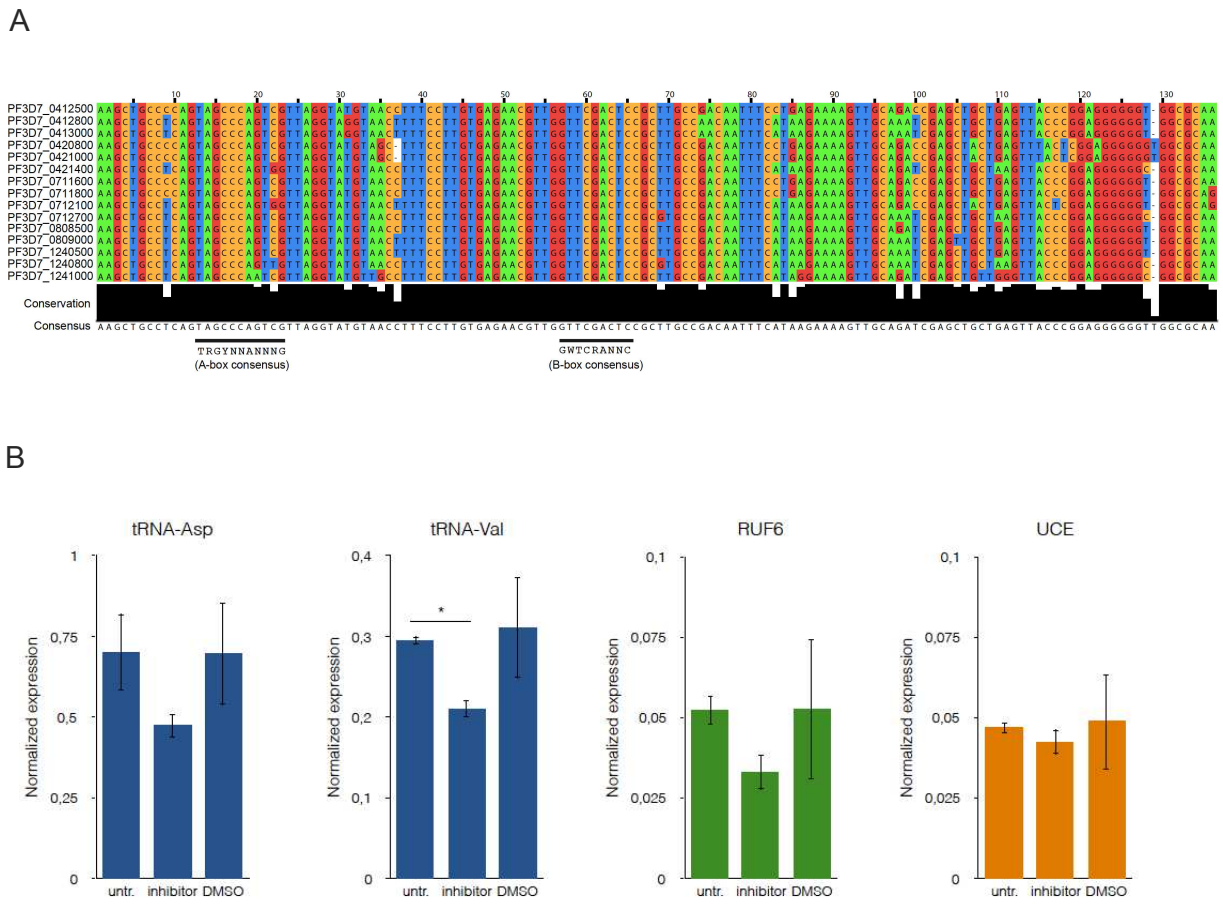


Figure 1. RNA polymerase III inhibition assay Transcript levels as quantified by RT-qPCR in synchronized wild type parasites at 24 hpi for tRNA-Asp, tRNA-Val, RUF6 and housekeeping control gene coding for ubiquitin-conjugating enzyme, in untreated cultures and in the presence of RNA Pol III inhibitor and DMSO. Expression is normalized to fructose-bisphosphate aldolase (PF3D7_1444800) transcript levels. Mean SEM of two independent experiments are shown. Statistical significance was determined by two-tailed Student's t-test (*** $p<0.001$).

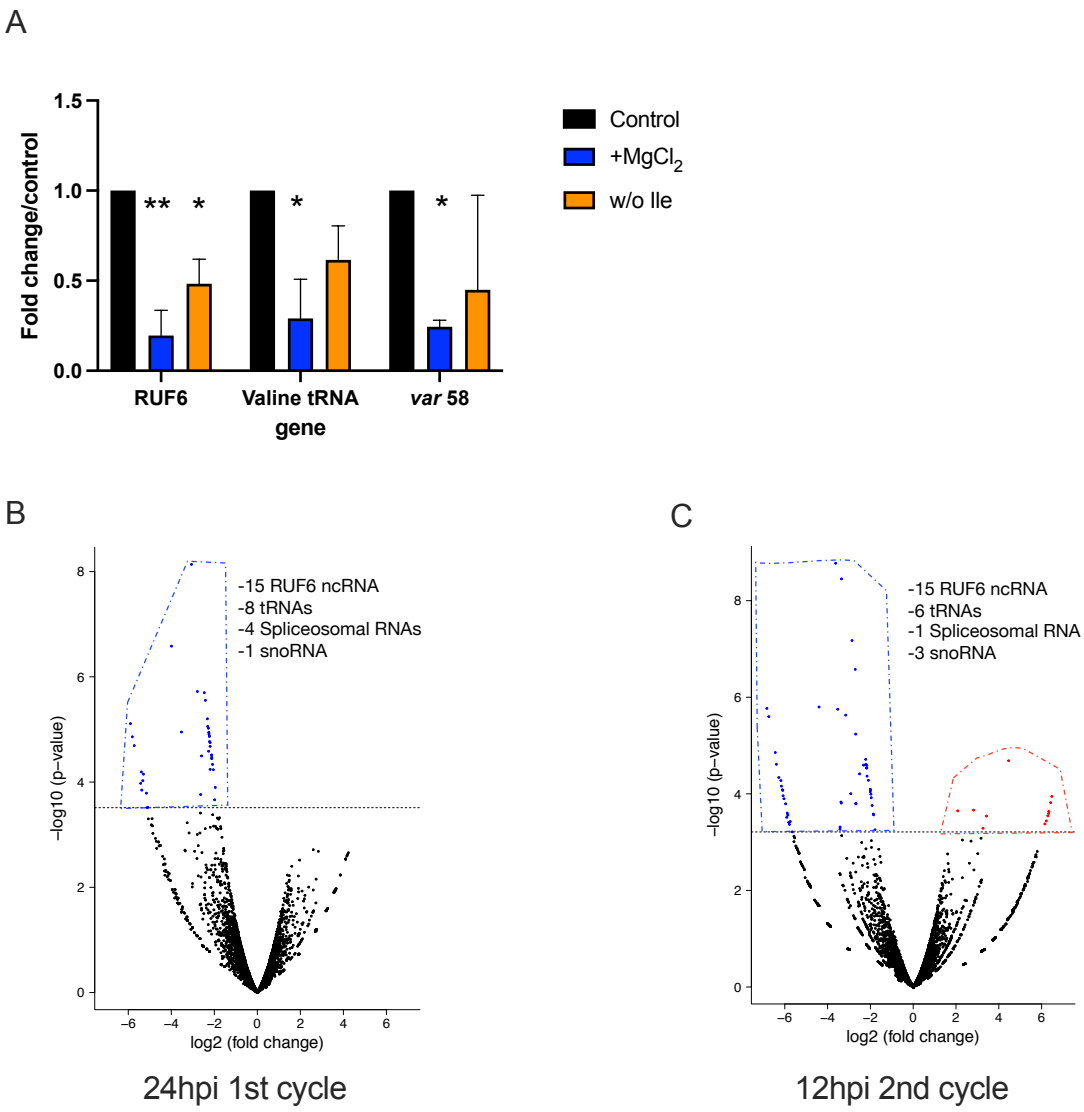


Figure 2. Transcriptional differences from isoleucine deprivation and $MgCl_2$ addition (A) RT-qPCR results on wild type 3D7 parasites harvested at 18 hpi for control parasites and parasites in the absence of isoleucine (w/o Ile) or presence of $MgCl_2$ (+$MgCl_2$). Primers used were for all RUF6 ncRNA, tRNA valine (Pf3D7_0312600), and active *var* 58 (Pf3D7_1240900). Results are normalized to an RNA Pol II-transcribed reference gene FBA (Fructose-bisphosphate aldolase Pf3D7_1444800) and presented as fold change/control. Error bars are displayed from 3 biological replicates. (B, C) Volcano plot showing log2 (fold change, FC) against −log10 (p-value) of transcripts identified by RNA-Seq analysis of 3D7 control and addition of $MgCl_2$ at 24 hpi first cycle (B) and 12 hpi second cycle (C). Expressed transcripts from three replicates between control and addition of $MgCl_2$ that are significantly up-regulated are highlighted in red while significantly down-regulated genes are highlighted in blue (FDR corrected p-value of ≤(FDR corrected p-value <0.05) and a FC ≥± 1.95). Black dots indicate non-significant transcripts with a FC ≤ 2.0.

2.2 RUF6 ncRNA is regulated by external factors

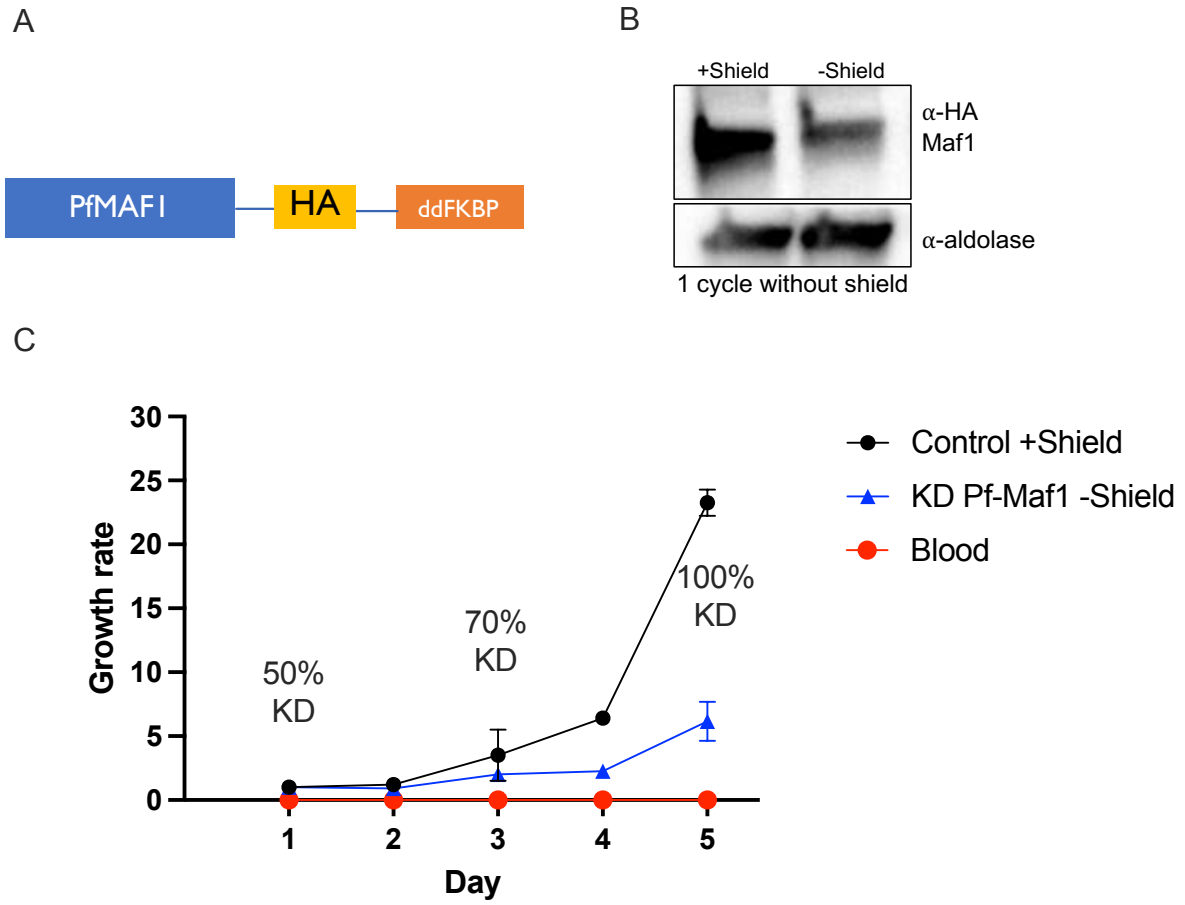


Figure 3. Validation of PfMaf1 knockdown system (A) Illustration of plasmid design for PfMaf1 with a 3HA tag followed by a ddFKBP domain to allow for knockdown studies. (B) Western blot analysis for nuclear and cytoplasmic PfMaf1 localization in pSLI-Maf1-FKBP transfected parasites with Shield1 (control) and without addition of Shield1 (w/o shield) harvested at 18hpi. (C) Growth curve over 5 days of clonal pSLI-Maf1-FKBP parasites in the presence (+S black) or absence (-S blue) of Shield. Uninfected red blood cells (red) serve as reference of background. Error bars indicate standard deviation of three technical replicates in blood from two different clones.

2.2 RUF6 ncRNA is regulated by external factors

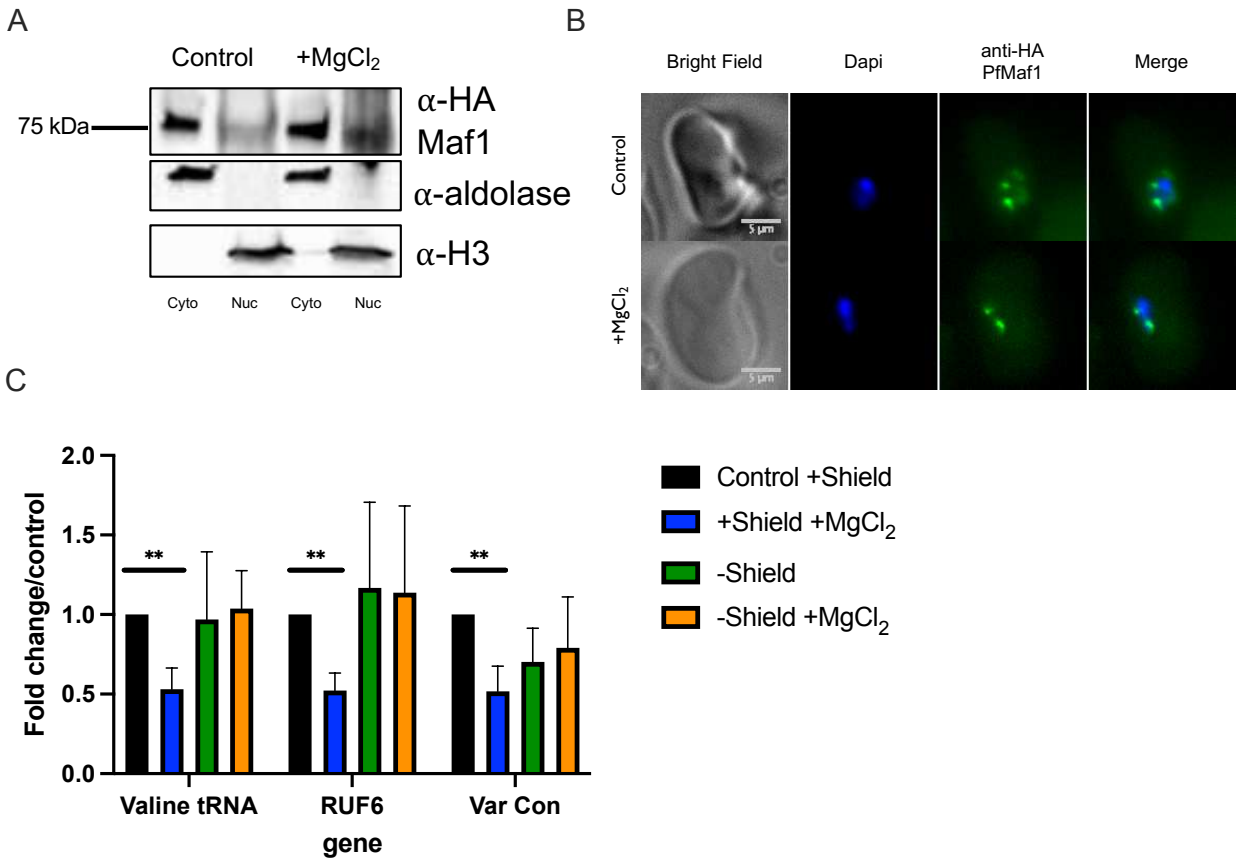


Figure 4. PfMaf1 involvement in regulating RNA Pol III-transcribed genes (A) Western blot analysis for PfMaf1 expression in pSLI-Maf1-FKBP transfected parasites without addition of $MgCl_2$ (control) and with addition of $MgCl_2$ (+$MgCl_2$) harvested at 18hpi. (B) Representative immunofluorescence images show brightfield, Dapi, GFP, and Dapi-GFP merge for PfMaf1 control and addition of $MgCl_2$ at 18pi (C) RT-qPCR results for 4 conditions: in the presence (+S black) or absence (-S green) of Shield and absence of $MgCl_2$ or presence (+Shield +Mg blue) or absence (-S +$MgCl_2$ orange) of Shield, PfMaf1, and presence (+$MgCl_2$) of $MgCl_2$. Error bars are displayed from 3 biological replicates.

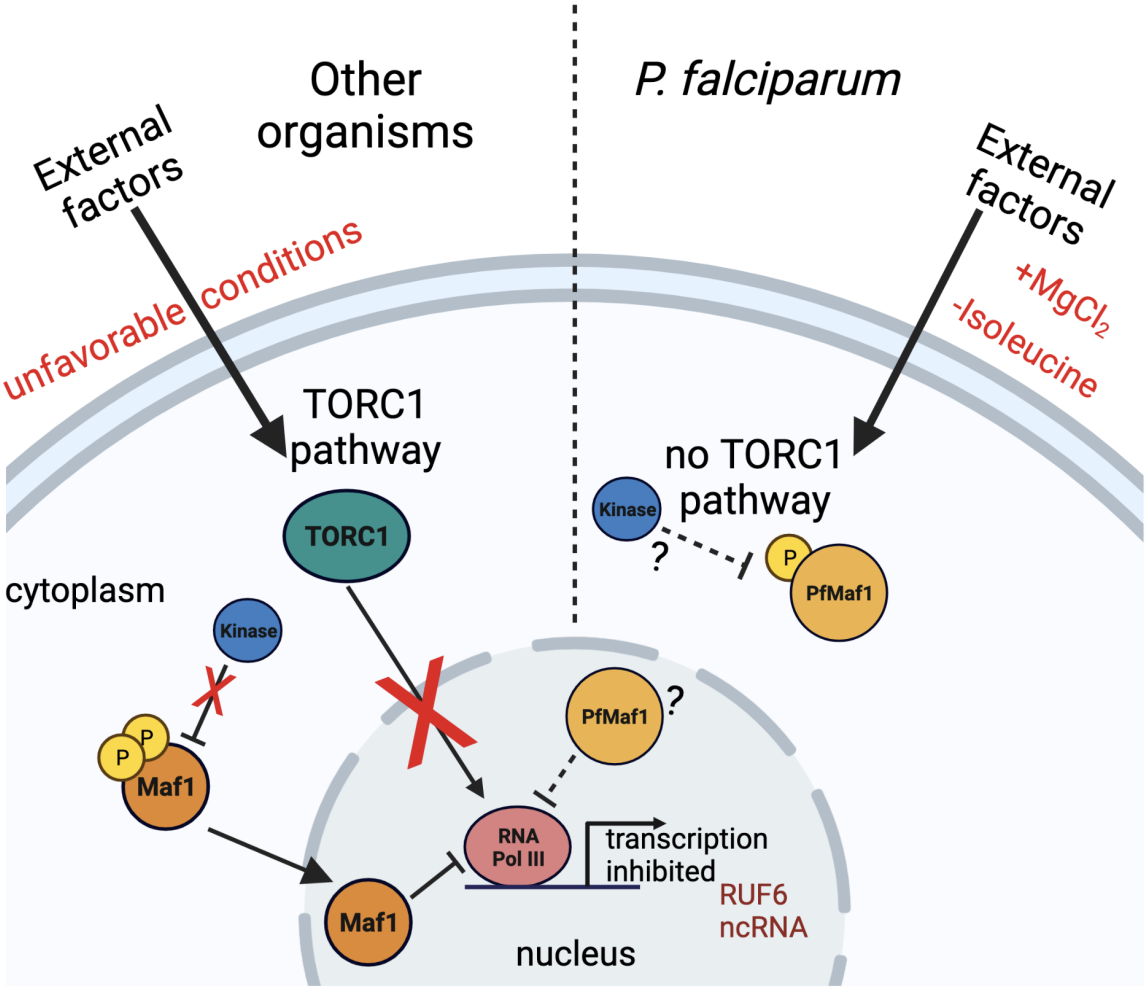
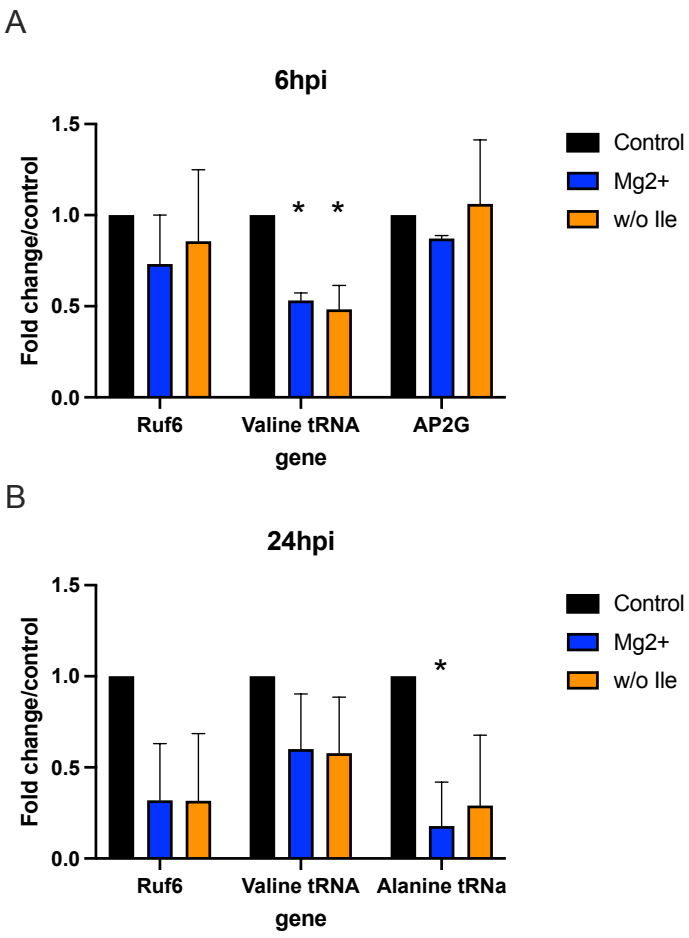


Figure 5. Illustration showing the common TORC1 pathway in other organisms compared to *P. falciparum*. In other organisms, TORC1 pathway represses Maf1 in favorable conditions. External factors, like nutrient depletion, lead to the activation of Maf1 by repressing TORC1 and kinases that phosphorylate Maf1. Unphosphorylated Maf1 re-localizes to the nucleus where it represses RNA Pol III leading to reduced transcription. In *P. falciparum,* Maf1 is one of the few conserved components of TORC1 pathway. Its phosphorylation state and interacting kinases remain unknown.

2.2 RUF6 ncRNA is regulated by external factors



Supplementary Figure 1. Transcriptional differences from isoleucine deprivation and $MgCl_2$ addition (A) RT-qPCR results on wild type 3D7 parasites harvested at 6 (A) and 24 (B) hpi for control parasites and parasites in the absence of isoleucine (w/o Ile) or presence of $MgCl_2$ (+$MgCl_2$). Primers used were for all RUF6 ncRNA, tRNA valine (Pf3D7_0312600), RNA Pol II-transcribed Ap2G (Pf3D7_14082000, and tRNA Alanine (Pf3D7_0411500). Results are normalized to an RNA Pol II-transcribed reference gene FBA (Fructose-bisphosphate aldolase Pf3D7_1444800) and presented as fold change/control. Error bars are displayed from 3 biological replicates.

```
PVP01_0525900.1-p1            ------GDMGDN--------GDIGDIGDI------------------------------     70
PVX_089915.1-p1               ------GDMGDN--------GDIGDIGDI------------------------------     70
PVL_000204600-t42_1-p1        ------GDMGDN--------SDIGDIGGI------------------------------     74
PCYB_053080-t26_1-p1          ------GGD---------------------------------------------------     56
PcyM_0526100-t36_1-p1         ------GGD---------------------------------------------------     56
C922_02518-t30_1-p1           ------GNNGNN--------GNNGNSGDS------------------------------     72
AK88_01180-t30_1-p1           ------GKE---------------------------------------------------     60
PKA1H_050014500.1-p1          ------GKE---------------------------------------------------     60
PKNH_0508900.1-p1             ------GKE---------------------------------------------------     60
PKNOH_S04361600-t35_1-p1      ------GKE---------------------------------------------------     60
PCOAH_00012030-t30_1-p1       ------SKE---------------------------------------------------     60
VWU50958.1                    IDNL-VKNKIPNSFSSTTFQVNT-----------------------NIPTECTKNEQTT     89
PmUG01_05034000.1-p1          TT----------NNSSANYYNGSGVTAVCREE---------------DTHKECLS-----     82
Pf7G8-2_000109400.1-p1        SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PF3D7_0416500.1-p1            SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfSN01_040021300-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfTG01_040021900-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfSD01_070032500-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfNF54_040022200.1-p1         SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfNF135_040022000.1-p1        SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfKH02_040021500-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfKH01_040021900-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfKE01_040023400-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfHB3_040020500-t41_1-p1      SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfGA01_040021000-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfCD01_040022000-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
Pf7G8_040021400-t41_1-p1      SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfNF166_040022500.1-p1        SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfML01_040022600-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfGN01_040021900-t41_1-p1     SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfDd2_040021700-t41_1-p1      SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfIT_040020900-t41_1-p1       SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PPRFG01_0413800.1-p1          SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PfGB4_040021800-t41_1-p1      SSELDINNKLDEENNMSPFYDNRNIISQINENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PGAL8A_00090000.1-p1          ------------------------------------------------------------     45
PRELSG_0511800.1-p1           ------------------------------------------------------------     45
PRCDC_0414000.1-p1            SSELDINNKLDEENNMSPFYDNRNIISQMNENNVKHLDPLNPFDNENNISKEIHNNKEIE     119
PRG01_0421200-t36_1-p1        SSELDINNKLDEENNMSPFYDNRNIISQMNEINVKHLDPLNPFDNENNISKEIHNNKEIE     119
PADL01_0415700-t36_1-p1       SSELDINNKQDEENNMNPLYGNRNNMSQMTENNVNPLDSLNPFDDENNISKEIHNIKEIE     119
PGSY75_0416500-t31_1-p1       SSELDINNKQDEENNMNPLYDNRNNMSQMTENNVNLLDSLNPFDDENNMSKEIHNNKEIE     119
PGABG01_0414300-t36_1-p1      SSELDINNKQDEENNMNPLYGNRNNMSQMTENNVNLLDSLNPFDDENNMSKEIHNNKEIE     119
PBILCG01_0409800-t36_1-p1     SSEFDINNKLDEENNMNPFYGNRNIISQMNENNVNHLDPLNPSDNENNVLKEIHNNKEIE     120
PBLACG01_0414600-t36_1-p1     SSELDINNKLDEENNMSPFYGNRNIISQMKENNVNTLDSLNPFEDENNMSKEIHNNKEIE     119
VEV55807.1                    --------------------------------NI---D----------------------     48
YYE_04226-t30_1-p1            --------------------------------NI---D----------------------     48
PCHAS_0727600.1-p1            ------------------------------------------------------------     44
CAD2089355.1                  ------------------------------------------------------------     44
YYG_04102-t30_1-p1            ------------------------------------------------------------     44
CAD2101887.1                  ------------------------------------------------------------     44
CAD2101855.1                  ------------------------------------------------------------     44
CAD2089196.1                  ------------------------------------------------------------     44
PBANKA_0718500.1-p1           NSE----------------------------HC---N----------------------     59
PY17X_0718700.1-p1            ------------------------------------------------------------     51
PY02861-t26_1-p1              ------------------------------------------------------------     51
PYYM_0718600.1-p1             ------------------------------------------------------------     51
```

Supplementary Figure 2. Maf1 alignment for *Plasmodium* Alignment sequences for N-term region, conserved sequences are within the *Laverania* species.

Supplementary Material:

Target	Forward primer	Reverse primer
Alanine	5'-GGGCAGGTGGTGTAGTGG-3'	5'-TGCTGGACAGACGGGGAATT-3'
AP2-G	5'-TGGGAAGAGAGCATGCAATGA-3'	5'-TCGCTTTCTTGTCCATGCAAC-3'
FBA	5'-TGTACCACCAGCCTTACCAG-3'	5'-TTCCTTGCCATGTGTTCAAT-3'
RUF6	5'-AAGCTGCCCCAGTAGCCCA-3'	5'-AAAAATTGCGCCGCCCCC-3'
UCE	5'-TAACAGCCCAGCGAATCAAG-3'	5'-CGGCATCTTCTTCAGCTTTCTG-3'
Valine	5'-GCGGGCATGGTCTAGTGG-3'	5'-ACTACGGGCACCGAGGATC-3'
var 58	5'-CAAAATGGTAGTGATGGTGGTCG-3'	5'-CCCCCTGCTTTATTATCTTTCGTC-3'
Var Con	5'-GCACGAACTTTTGCA-3' 5'-GCACGCAGTTTTGCA-3'	5'-GCCCCATTCGTCGAACC-3' 5'-GCCCCATTCCTCGAACC-3'

Supplementary Table 1. qPCR analysis primer pairs

References:

Abraham, K. J., Chan, J. N., Salvi, J. S., Ho, B., Hall, A., Vidya, E., Guo, R., Killackey, S. A., Liu, N., Lee, J. E., Brown, G. W., & Mekhail, K. (2016). Intersection of calorie restriction and magnesium in the suppression of genome-destabilizing RNA-DNA hybrids. *Nucleic Acids Res. Oct*, *14;44(18):8870-8884*.

Andrade, C. M., Fleckenstein, H., & Thomson-Luque, R. (2020). Increased circulation time of Plasmodium falciparum underlies persistent asymptomatic infection in the dry season. *Nat Med*, *26*, 1929–1940.

Babbitt, S. E., Altenhofen, L., Cobbold, S. A., Istvan, E. S., Fennell, C., Doerig, C., Llinás, M., & Goldberg, D. E. (2012). *Plasmodium falciparum* responds to amino acid starvation by entering into a hibernatory state. *Proceedings of the National Academy of Sciences*, *109*(47).

Barcons-Simon, A., Cordon-Obras, C., Guizetti, J., Bryant, J. M., & Scherf, A. (2020). CRISPR Interference of a Clonally Variant GC-Rich Noncoding RNA Family Leads to General Repression of var Genes in Plasmodium falciparum. *MBio*, *11*(1).

Birnbaum, J., Flemming, S., & Reichard, N. (2017). A genetic system to study Plasmodium falciparum protein function. *Nat Methods*, *14*, 450–456.

Dieci, G., Conti, A., Pagano, A., & Carnevali, D. (2013). Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. *Biochim Biophys Acta*, *1829*, 296–305.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., … Barrell, B. (2002).

Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, *419*(6906), 498–511.

Guizetti, J., Barcons-Simon, A., & Scherf, A. (2016). Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite. *Nucleic Acids Res*, *44*, 9710–9718.

Hasenkamp, S., Merrick, C. J., & Horrocks, P. (2013). A quantitative analysis of Plasmodium falciparum transfection using DNA-loaded erythrocytes. *Mol Biochem Parasitol*, *187*, 117–120.

Hess, F. I., Kilian, A., Söllner, W., Nothdurft, H. D., Pröll, S., & Löscher, T. (1995). Plasmodium falciparum and Plasmodium berghei: Effect of Magnesium on the Development of Parasitemia. *Experimental Parasitology*, *80*, 196–193.

Larremore, D. B., Sundararaman, S. A., Liu, W., Proto, W. R., Clauset, A., Loy, D., Speede, S., Plenderleith, L. J., Sharp, P. M., Hahn, B. H., Rayner, J. C., & Buckee, C. O. (2015). Ape parasite origins of human malaria virulence genes. *Nat Commun*, *6*(8368).

Leech, J. H., Barnwell, J. W., Aikawa, M., Miller, L. H., & Howard, R. J. (1984). Plasmodium falciparum malaria: Association of knobs on the surface of infected erythrocytes with a histidine-rich protein and the erythrocyte skeleton. *J Cell Biol*, *98*(4), 1256–1264.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with BurrowsWheeler transform. *Bioinformatics*, *25*, 1754–1760.

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

Mancio-Silva, L., Slavic, K., Grilo Ruivo, M. T., Grosso, A. R., Modrzynska, K. K., Vera, I. M., Sales-Dias, J., Gomes, A. R., MacPherson, C. R., Crozet, P., Adamo, M., Baena-Gonzalez, E., Tewari, R., Llinás, M., Billker, O., & Mota, M. M. (2017). Nutrient sensing modulates malaria parasite virulence. *Nature*, *547*(7662), 213–216.

McLean, K. J., & Jacobs-Lorena, M. (2017). Plasmodium falciparum Maf1 Confers Survival upon Amino Acid Starvation. *mBio*, *8*(2), 02317–16.

Michels, A. A. (2011). MAF1: A new target of mTORC1. *Biochemical Society Transactions*, *39*(2), 487–491.

Orioli, A., Praz, V., Lhôte, P., & Hernandez, N. (2016). Human MAF1 targets and represses active RNA polymerase III genes by preventing recruitment rather than inducing long-term transcriptional arrest. *Genome Research*, *26*(5), 624–635.

Otto, T. D., Gilabert, A., Crellen, T., Böhme, U., Arnathau, C., Sanders, M., Oyola, S. O., Okouga, A. P., Boundenga, L., Willaume, E., Ngoubangoye, B., Moukodoum, N. D., Paupy, C., Durand, P., Rougeron, V., Ollomo, B., Renaud, F., Newbold, C., Berriman, M., & Prugnolle, F. (2018). Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol*, *3*, 687–697.

Pluta, K., Lefebvre, O., Martin, N. C., Smagowicz, W. J., Stanford, D. R., Ellis, H., SR, AK, S., A, B., & M. (2001). Maf1p, a negative effector of RNA polymerase III in Saccharomyces cerevisiae. *Mol Cell Biol*, *21*, 5031–5040.

Rollins, J., Veras, I., Cabarcas, S., Willis, I., & Schramm, L. (2007). Human Maf1 negatively regulates RNA Polymerase III transcription via the TFIIB family members Brf1 and Brf2. *International Journal of Biological Sciences*, 292–302.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., & Lanzer, M. (1998). Antigenic variation in malaria: In situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum. *EMBO J*, *17*, 5418–5426.

Serfontein, J., Nisbet, R. E. R., Howe, C. J., & de Vries, P. J. (2010). Evolution of the TSC1/TSC2-TOR Signaling Pathway. *Science Signaling*, *3*(128).

Smith, J. D., Chitnis, C. E., Craig, A. G., Roberts, D. J., Hudson-Taylor, D. E., Peterson, D. S. P., R., N., I., C., & Miller, L. H. (1995). Switches in expression of plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell*, *82*(1), 101–110.

van Dam, T. J. P., Zwartkruis, F. J. T., Bos, J. L., & Snel, B. (2011). Evolution of the TOR Pathway. *Journal of Molecular Evolution*, *73*(3–4), 209–220.

Vannini, A., Ringel, R., Kusser, A. G., Berninghausen, O., Kassavetis, G. A., & Cramer, P. (2010). Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell*, *143*(59), 70.

W.H.O. (2021). *World Malaria Report 2021*.

Wullschleger, S., Loewith, R., & Hall, M. N. (2006). TOR Signaling in Growth and Metabolism. *Cell*, *124*(3), 471–484.

Zhang, M., Wang, C., Otto, T. D., Oberstaller, J., Liao, X., Adapa, U., SR, K, B., IF, C., D, M., M, B., J, L., S, S., J, R., JC, J., RHY, A., & J.H. (2018). Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis. *Science*.

Part 3

3 General discussion and outlook

Contents:

In Part 3 I summarize and contextualize the main results presented. I also aim to discuss the concepts that still remain unclear that can serve as a basis for future research perspectives.

3.1 Establishing ChIRP-MS in *P. falciparum*

We initially set out to establish the ChIRP-MS protocol in *P. falciparum*. We chose this technique, specifically, because it allows for the identification of ncRNA-binding proteins in their native cellular context. This technique allowed us to isolate and identify specific RUF6 ncRNA-associated proteins that were not detected using nuclear extracts in RUF6 ncRNA affinity purification assays. The ChIRP-MS method appears to detect interactions that cannot be reconstituted once the nucleus is disrupted.

Using validated biotinylated tiling oligo probes for RUF6, we were able to identify RUF6-associated proteins. We included two controls: scrambled probes and RNase A-treated samples, to exclude or minimize non-specific protein candidates. Binding proteins were detected with label free quantitative proteomics (LC-MS/MS). One challenging aspect of proteomic analysis is the amount of input material needed to overcome non-specific background hits. To overcome this obstacle, we used 4.0E10 parasites in ChIRP-MS for each of the four replicates. Quantitative analysis identified proteins that were uniquely enriched in our RUF6 target samples compared to controls ($p< 0.05$). Therefore, our top ranked candidate proteins were uniquely found in our target samples and absent in one or both control groups. After validation of one of the top ranked candidates as a RUF6 ncRNA-associated protein, this method should allow further studies of the interactome of other key regulatory ncRNA in malaria parasites; such as the antisense RNA of the GDV1 gene that inhibits the activation of the master regulator Ap2G of sexual commitment (Filarsky *et al*., 2018). The method could also serve to pull down mRNA specific proteins that bind to modified bases, an emerging area in malaria parasites termed epitranscriptome.

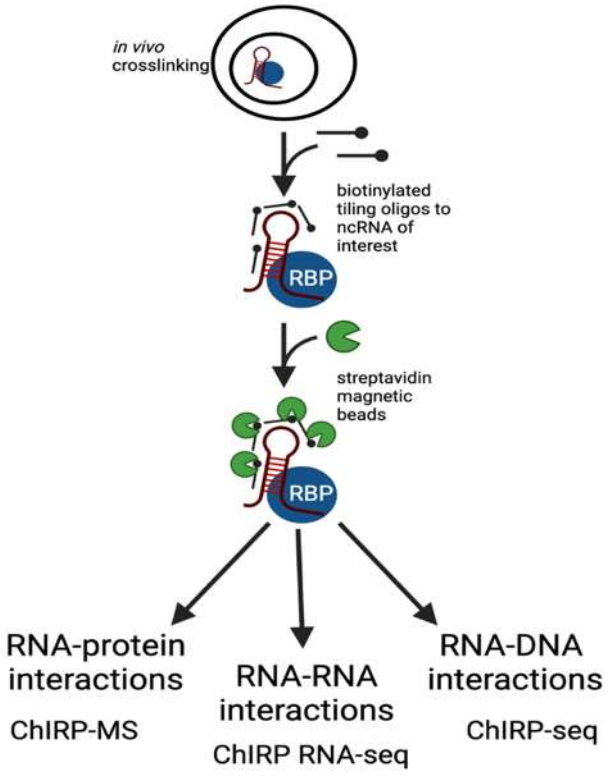


Figure 3.1 Various potential applications of the ChIRP protocol are shown to identify interactions of ncRNA with proteins, DNA and other RNA molecules.

In addition to identifying protein interactions, the ChIRP protocol can also be used to identify RNA or DNA interactions. ChIRP-seq and ChIRP-RNA-seq are being optimized in our lab for use in future applications related to the biology of RNA. One example is to investigate the interaction of the subtelomeric repeat TARE6 lncRNA with the *P. falciparum* genome. So far, I have designed and validated probes that are specific to TARE 6 by RT-qPCR following the ChIRP protocol. Furthermore, purified DNA that was bound to the RUF6 ncRNA is ready to be prepared for sequence analysis to identify potential regions of the *var* gene locus that interact with the RUF6 protein complex. Optimization of the protocol is still ongoing in the laboratory.

3.2 Identifying RUF6 ncRNA-binding proteins

The rationale of this study was to gain biological insight into the RUF6 ncRNA-mediated activation of the *var* gene family. Previous studies have explored its role in the monoallelic expression of *var* genes (Barcons-Simon et al., 2020; Guizetti et al., 2016; Wei et al., 2015). Still, it remains unknown how the RUF6 ncRNA is targeted to an active *var* gene expression site and which proteins it recruits and interacts with to promote efficient singular *var* gene transcription.

Using ChIRP-MS we were able to establish a comprehensive RUF6 ncRNA protein interactome including several proteins with predicted RNA-binding domains and several proteins of unknown function. We validated a homolog of human DEAD box protein, DDX5, as a regulator of *var* gene transcription. DDX5 has a canonical role in unwinding RNA as well as DNA (G. Wu et al., 2019), is involved in transcriptional regulation and elongation, and can interact directly with RNA Pol II (Clark et al., 2013). IFA localization studies and Western blot analysis confirmed the nuclear presence of Pf-DDX5 during ring stage and RIP RT-qPCR confirmed its association with RUF6 ncRNA.

Co-IP/MS of Pf-DDX5 identified several proteins originally found by our RUF6-ChIRP protocol, validating the ChIRP-MS technique as a method to identify ncRNA interactomes in *P. falciparum*. Shared proteins had functions related to RNA binding and involved in processes related to initiation and regulation of transcription from RNA Pol II. These findings corroborate our previously proposed mode of function, namely the co-localization of RUF6 ncRNA with actively transcribed *var* genes (Guizetti *et al*, 2016) and that RUF6 transcription is linked to *var* gene activation (Barcons-Simon et al., 2020; Guizetti et al., 2016). Most importantly, our findings strengthen the notion that Pf-DDX5 is associated with RUF6 ncRNA at the *var* gene RNA Pol II promoter. The combination of CHIRP-MS and Pf-DDX5 Co-IP-MS provides the first comprehensive catalogue of proteins that interacts with RUF6 ncRNA.

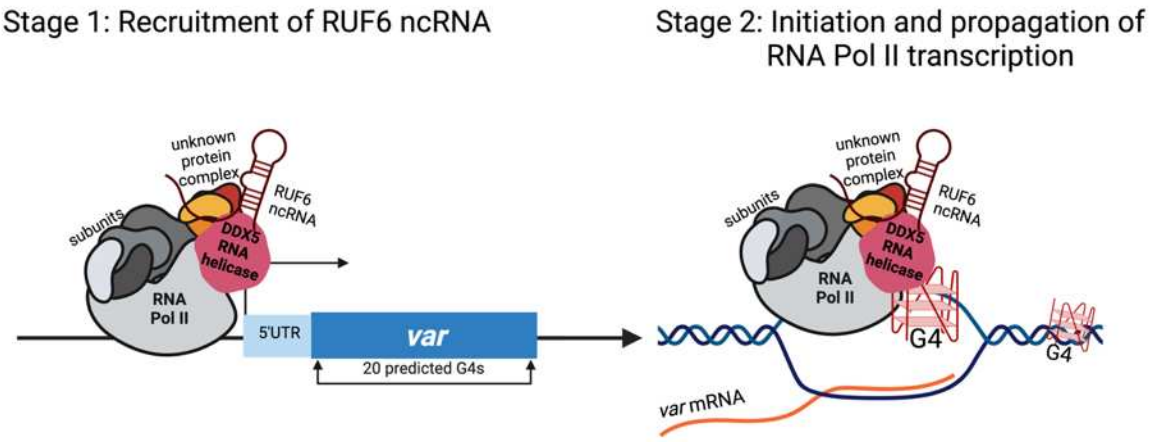


Figure 3.2 Hypothetical model of PfDDX5 and RUF6 ncRNA function at *var* gene. *Var* genes are highly enriched for G-quadruplexes (Gazanion et al., 2020). The recruitment of Pf-DDX5 via RUF6 to the active *var* gene transcription complex may enhance the transcriptional RNA Pol II activity for this type of gene.

Inducible displacement of nuclear Pf-DDX5 resulted in the significant down-regulation of the active *var* gene without disrupting monoallelic expression, identifying Pf-DDX5 as a novel regulator of transcriptional activation of *var* genes. We hypothesize that Pf-DDX5 helps unwind DNA G-quadruplexes (G. Wu et al., 2019), in the *var* gene promotor and exon I (Fig. A1) (Gazanion et al., 2020), to allow for proper transcription by RNA Pol II.

Our data show that RUF6 ncRNA recruit proteins with predicted RNA-binding activities such as Pf-DDX5 but also a number of candidate genes with unknown function. Given that RUF6 and *var* genes are restricted to Laveranian species and no homologous genes exist in model eukaryotes, it is somehow expected to find RUF6-associated proteins that have evolved with *RUF6/var* genes. Exploring the function of those genes may reveal functions that are still poorly understood.

3.3 Identification of external factors that regulate transcription of RUF6 ncRNA

While it is now established that RUF6 ncRNA is involved in the regulation of *var* gene expression, environmental factors that regulate the ncRNA have not been explored. Here we show that PfMaf1, a well-known repressor of RNA Pol III in other organisms, is involved in regulating transcription of RNA Pol III-transcribed RUF6 in response to external stimuli.

What initially drew our attention to PfMaf1 specifically is the highly conserved 150AA N-terminal region found only within *Laverania* species, the only species that contain *var* genes and RUF6. We hypothesize that this conserved region has a role specific to regulating RUF6. We showed that *P. falciparum* parasites exposed to different external stimuli has an effect on RNA Pol III-transcribed genes, importantly RUF6 ncRNA. A similar RNA Pol III down-regulation has been shown in other systems to be essential to cell survival by conserving energy in times of reduced nutrient availability (Willis and Moir, 2018). Additionally, we observed that *var* gene transcription decreased and suspect that the RUF6 downregulation is responsible, since *var* genes are transcribed by RNA Pol II. We showed that PfMaf1 is needed for the downregulation of Pol III-transcribed genes in response to the external stimulus, [3mM] $MgCl_2$.

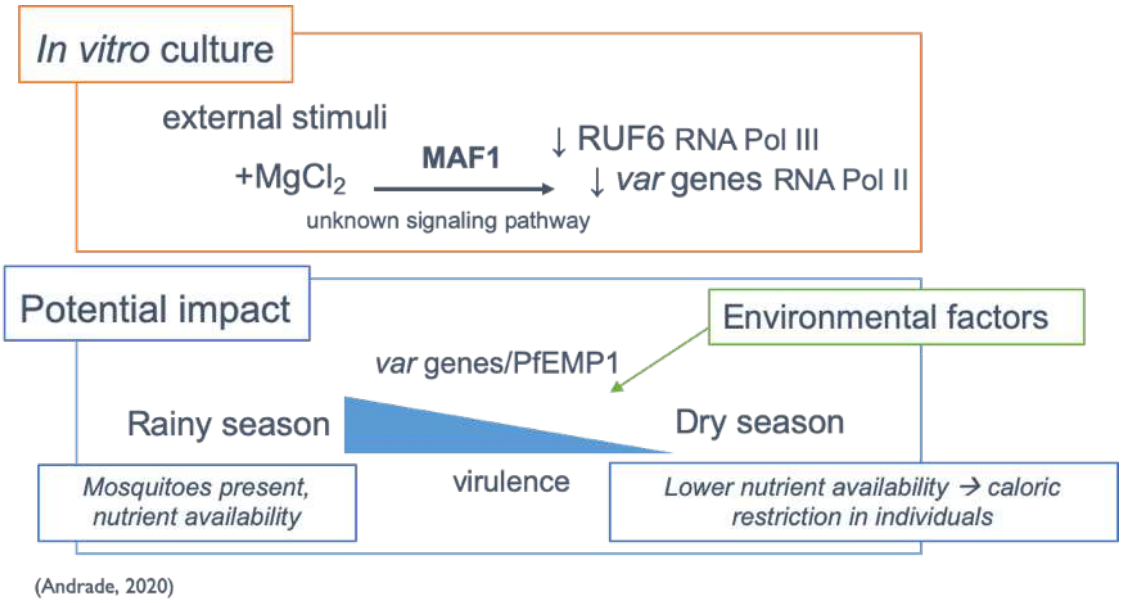


Figure 3.3 Potential impact of the *in vitro* culture established pathway of PfMaf1-dependent RNA Pol III gene transcription on *P. falciparum* virulence in malaria endemic regions during rainy and dry season. We hypothesize the established link between external factors and PfMaf1 and the control of RNA Pol III transcription of the ncRNA RUF6 could decrease parasite virulence by downregulating *var* genes (Andrade et al., 2020). This hypothesis can be tested by analyzing the RNA Pol III transcript levels and in particular RUF6 in malaria patients from dry and rainy seasons.

In *P. falciparum,* particularly, lower nutrient abundancy occurs during the dry season in African regions, where malaria is most prevalent. We hypothesize that parasites can sense the transition from the rainy to the dry season through changes in nutrient availability. During the dry season, the parasite from asymptomatic patients show reduced expression of *var* genes (Andrade et al., 2020). Changes in diet in a rodent malaria model *P. berghei* (which do not contain *var* genes) demonstrated the impact on changes in diet of the host on asexual blood stage development (Mancio-Silva et al., 2017). Overall, it is possible that PfMaf1 is activated in response to external stimuli, leading to its re-localization to the nucleus, where it reduces expression of RUF6 ncRNA and consequently, *var* genes, necessary for the parasites to 'bridge' transmission seasons. This is a new concept and testable hypothesis that could link RUF6 ncRNA regulation via PfMaf1 to reduced parasite virulence during the dry season.

3.5 Conclusions and future perspectives

Based on the results of the work done during the course of my thesis, the previous model of the interaction between RUF6 ncRNA and *var* gene regulation (Fig. 1.26) can updated by new molecular details important for the transcription of *var* genes. We now identify a RUF6-binding protein interactome which suggest a direct association with RNA Pol II. Further work will be done in order to characterize other interacting proteins in the complex, in particular proteins of unknown function may reveal answers to open questions. Additionally, ChIP-seq will be completed to study the PfDDX5 association with *var* gene loci and possibly with DNA G quadruplexes.

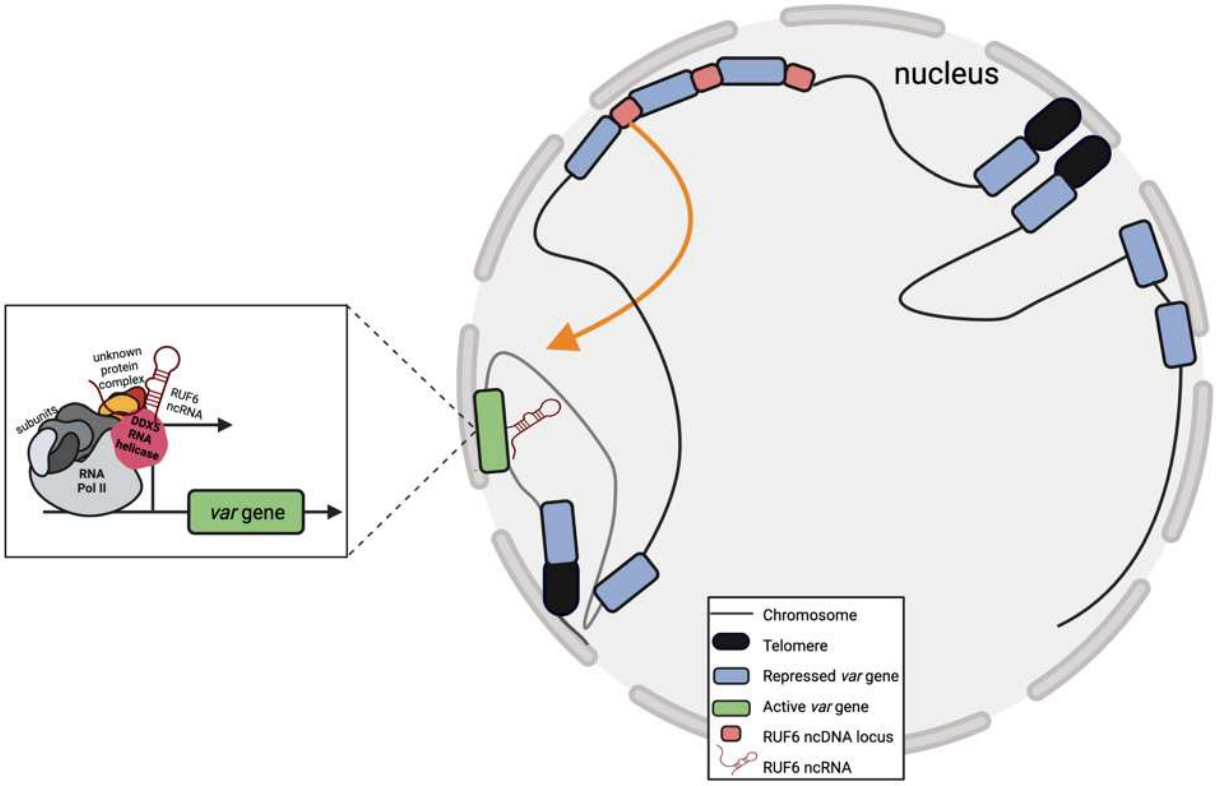


Figure 3.4 Illustration of nuclear organization of the active *var* gene at the expression site and interacting partners. Molecular insight into the interaction of ncRNA RUF6 with RNA Pol II; the helicase PfDDX5 is shown. The specific recruitment of RUF6 to the active expression site remains unknown. Further characterization of the interactome may reveal specific proteins that target RUF6 to the expression site.

Regarding the additional research project, the investigation of the regulation of RUF6 by PfMaf1, immunoprecipitation experiments will be carried out to validate the involvement of nuclear PfMaf1 with RNA Pol III. So far, we have developed and validated a TFIIIC antibody in our lab (Patty Chen). IP-MS studies will also be done in order to determine the phosphorylation state of cytoplasmic and nuclear PfMaf1. In parallel we are in the process of deleting the N-terminal region of PfMaf1, the conserved region only found amongst *Laverania* species. We hope to confirm that this region is specific for regulating RUF6.

The thesis opens many new research avenues that can no address experimentally important aspects of virulence gene regulation and the role of environmental factors in modulating parasite pathogenesis.

Appendix A

Additional preliminary data for results section 2.1

There are 20 predicted DNA G quadruplexes in the active *var* gene from the clone used in RNA-seq (Fig. A1). We set out to determine if PfDDX5 was binding to these specific DNA regions using ChIP-seq. We harvested rings and trophozoite stage parasites (12 and 24 hpi). We did not see any peaks spanning the predominantly transcribed RUF6 ncRNA (Fig. A2). However, we did see peaks throughout the active *var* gene (Fig. A2). Peak calling and further analysis are ongoing to determine if the observed peaks are significant as well as if they are over the predicted G4 sites.

QGRS sequences found (overlaps not included)

Position	Length	QGRS	G-Score
3	26	GGCGCCAGGTAGTACTGGTACGCAGG	19
223	29	GGTAGTGGTAGTGGTGTTGCTGCTCGCGG	12
770	30	GGTGGACTGCAAATCGAGAAACGGTGTGGG	5
1561	14	GGCAGTGGTGGTGG	18
1816	16	GGTCAGGAGGGTGAGG	20
1863	18	GGTAGAAACTGGAGGTGG	14
2356	17	GGTGGCTGTGGTCCCGG	18
2625	11	GGAGGAGGCGG	21
2673	14	GGACGCGGTGGTGG	18
2928	27	GGGGGGATTAACGAAGTGGGCGAGTGG	12
3228	27	GGGGGATTATAGGGATATATTATTTGG	11
3365	29	GGTGGAGTCAAAACGGTGAACATATATGG	11
3538	20	GGTGGTGAAGGTGCCGAAGG	15
3745	19	GGTAGCGGTTATGGGGAGG	20
4030	29	GGAGAAGGTACAATAAAATTTAATGGAGG	6
4278	30	GGCTTGTGAAGGTAAAGGTATATTTGAAGG	15
4372	25	GGAAATAGGGAAACCGTCAGAGGGG	10
5246	24	GGACGGTCGGCGTCGCGTTTATGG	11
6631	23	GGGGATAGTGGAACCGATAGTGG	11
7066	23	GGTGGAAATGGTTCATATAGTGG	12

Figure A1. Predicted *var* gene DNA G quadruplexes 20 predicted G4s in Pf3D7_0412400.

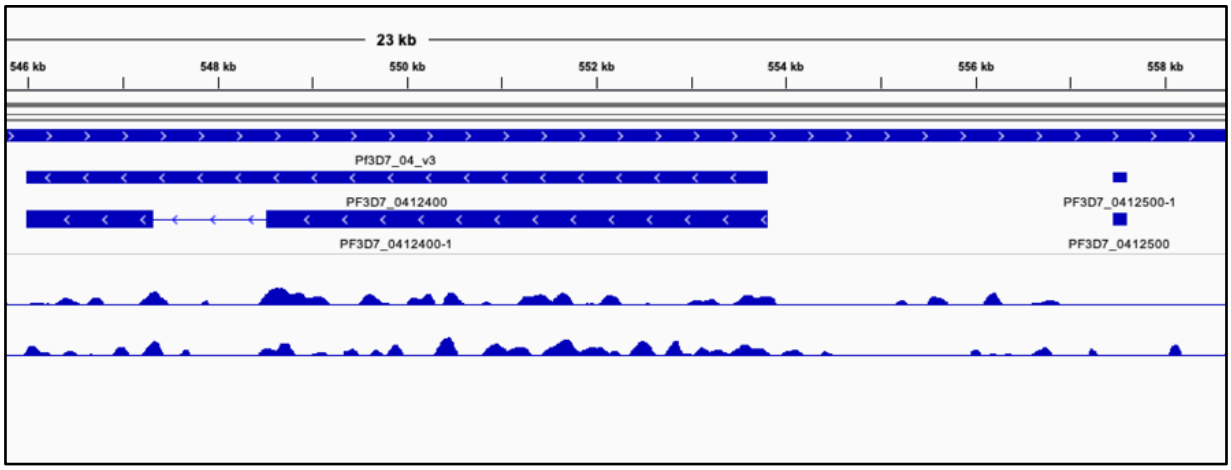


Figure A2. ChIP-seq for 12 and 24 hpi ChIP-seq data show enrichment of PfDDX5 in clonal parasites at 12 (above) and 24 (below) hpi relative to regions of chromosome 4. Genome location is indicated at the top of the panel. Pf3D7_0412400 is the active *var* gene and Pf3D7_0412500 is the predominantly transcribed RUF6 ncRNA. The x-axis is DNA sequence, with genes represented by blue boxes and labeled with white arrowheads to indicate transcription direction. The y-axis is enrichment (ChIP/Input /genomic DNA). One replicate was used for the PfDDX5 ChIP-seq.

Materials and Methods:

Chromatin immunoprecipitation and next generation sequencing

ChIP-seq was performed as previously described (Bryant et al., 2020) with some modification using ring or trophozoite stage parasites (12 or 24 hpi). Sonicated chromatin chromatin (500 ng DNA content) was combined with 25 µL of α-GFP trap beads (Chromotek) . After overnight incubation, subsequent washing, cross-link reversion and DNA extraction were carried out as described before (Bryant et al., 2020). Sequencing libraries were produced with the immunoprecipitated DNA using the MicroPlex Library Preparation Kit v2 (Diagenode) with the KAPAHiFi polymerase (Kapa Biosystems) for the PCR amplification. For each ChIP sample a control DNA corresponding to the ChIP input was processed in parallel. Multiplexed libraries were subjected to 150 bp paired-end sequencing on a NextSeq 500 (Illumina). Fastq files were obtained by demultiplexing the data using bcl2fastq (Illumina) prior to downstream analysis. So far, only one biological replicate has been analyzed for one clone at each time point.

Appendix B

Additional supplementary figure for results section 2.2

Percent parasitemia was lower than expected on day 5 of the growth curve for the 'Environmental factors regulate malaria parasite virulence via the downregulation of RNA Pol III-transcribed ncRNA gene family' manuscript. This can be explained by the high amount of unburst schizonts. There appears to be less merozoites formed in schizonts without PfMaf1. This result can be explained by the high energetic cost the parasite might be using without PfMaf1 to regulate RNA Pol III-transcription (Willis & Moir, 2018).

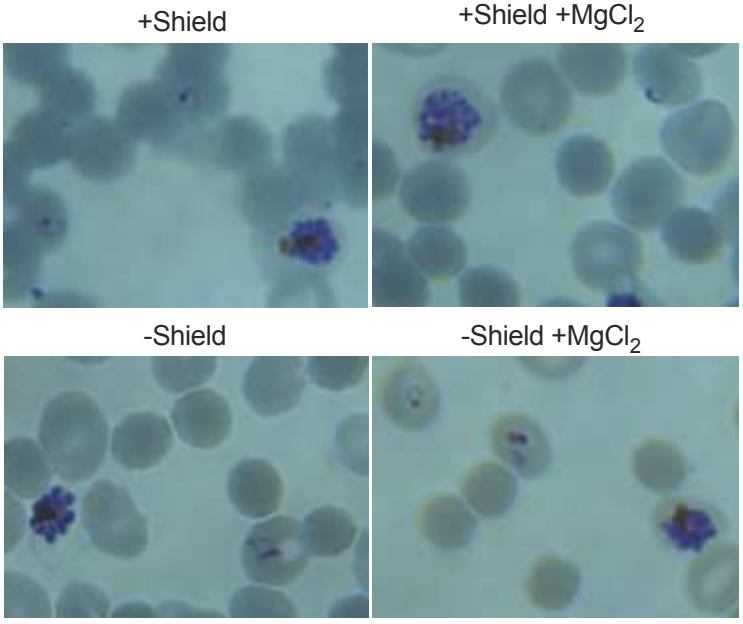


Figure B1. Day 5 of growth curve
Giemsa staining showing stage and morphology of parasites from control (+Shield) versus parasites with addition of $MgCl_2$ (+Shield +$MgCl_2$), degradation of PfMaf1 (-Shield), and combined addition of $MgCl_2$ and degradation of PfMaf1 (-Shield +$MgCl_2$

Appendix B

Appendix C

Insight into the DNA cytosine epigenetic landscape in *Plasmodium falciparum*

The highlights of this appendix are:

- Pf-DNMT2 has a dual function in methylating tRNA and DNA
- DNMT2 and oxygen levels are critical determinants to the DNA cytosine epigenetic landscape

The results presented in this appendix are from the submitted manuscript under the title "Variable oxygen environments and DNMT2 determine the DNA cytosine epigenetic landscape of *Plasmodium falciparum*". My contribution to this work was setting up and harvesting the parasites for the three replicates in Figure 3.

Appendix C

Appendix C

Variable oxygen environments and DNMT2 determine the DNA cytosine epigenetic landscape of *Plasmodium falciparum*

Elie Hammam[1*], Samia Miled[2*], Frédéric Bonhomme[3], Gretchen Diffendall[1,5], Benoit Arcangioli[4], Paola B. Arimondo[3], Artur Scherf[1,#]

[1] Institut Pasteur, Université de Paris, Biology of Host Parasite Interaction Unit, INSERM 1201, CNRS EMR 9195, 75724 Paris, France

[2] Université de Paris, CNRS, Institut Jacques Monod, 75013 Paris, France,

École Technique Supérieure du Laboratoire, 75013 Paris, France

[3] Institut Pasteur, Université de Paris, Epigenetic Chemical Biology, UMR 3523, CNRS, 75724 Paris, France

[4] Institut Pasteur, Université de Paris, Genome Dynamics Unit, UMR 3525 CNRS, 75724 Paris, France

[5] Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France.

[*]co-first authors

[#]Correspondance to: artur.scherf@pasteur.fr

Abstract

DNA cytosine methylation and its oxidized products are important epigenetic modifications in mammalian cells. Although 5-methydeoxyctidine (5mdC) was detected in the human malaria parasite *Plasmodium falciparum*, the presence of oxidized 5mdC forms remain to be characterized. Here we establish a protocol to optimize nuclease-based DNA digestion for the extremely AT-rich genome of *P. falciparum* (>80% A+T) for quantitative LC-MS/MS analysis of 5-hydroxymethylcytosine

(5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). We demonstrate the presence of 5hmC, 5fC and 5caC cytosine modifications in a DNMT2-only organism and observe striking ratio changes between 5mdC and 5hmC during the 48-hour blood stage parasite development. Parasite-infected red blood cells cultured in different physiological oxygen concentrations revealed a shift in the cytosine modifications distribution towards the oxidized 5hmC, 5fC and 5caC forms. In the absence of the canonical C5-DNA methyltransferase (DNMT1 and DNMT3A/B) in *P. falciparum,* we show that all cytosine modifications depend on the presence of DNMT2. In addition, we identified a hydroxylase (PF3D7_0829400) in *P. falciparum* that is linked to the non-canonical 5hmC biogenesis pathway in *P. falciparum* and produces 5hmC DNA cytosines when expressed in *Saccharomyces cerevisiae*. We conclude that DNMT2 and oxygen levels are critical determinants that shape the dynamic DNA cytosine epigenetic landscape in this human pathogen.

Introduction

Since its discovery in 1948(Hotchkiss, 1948), 5-methydeoxyctidine (5mdC) in DNA, also known as the 5th base of the genome, is now the best characterized epigenetic mark in higher eukaryotes, playing very important functions in different cellular processes such as cell differentiation and X chromosome inactivation(Holliday & Pugh, 1975; Riggs, 1975). DNA cytosine methylation is catalyzed by members of the DNA methyltransferases family (DNMT1 and DNMT3A/B)(Aapola et al., 2000; Bachman et al., 2001; Baubec et al., 2015) and 5mdC is mostly found in CpG rich regions of gene promoters (CpG islands or CGIs)(Doi et al., 2009; Jones, 2012; Sharif et al., 2007). The presence of 5mdC within CGIs of promoter regions is often linked with the inactivation of the corresponding gene(Guy et al., 2011; Meng et al., 2015). This occurs through the impairment of the binding of transcription factors and the recruitment of methyl CpG binding proteins (MBDs) creating a transcriptionally repressive environment(Bienvenu & Chelly, 2006).

Besides 5mdC, another major DNA cytosine modification in eukaryotes is 5-hydroxymethylcytosine (5hmC)(Shi et al., 2017). 5hmC is generated as the first oxidative product of the TET (Ten eleven translocation)-dependent DNA cytosine demethylation pathway(X. Wu & Zhang, 2017). TET enzymes (namely TET1, TET2 and TET3 in mammalian cells(Ito et al., 2011)) progressively oxidize 5mdC into 5hmC and further into 5fC (5-formylcyotsine) and 5caC (5-carboxylcytosine)(Ito et al., 2011). These

latter can be actively removed by different mechanisms including the activity of a thymine DNA glycosylase (TDG) coupled with the base excision repair machinery (BER) or passively lost upon DNA replication(X. Wu & Zhang, 2017). It is now well established that 5hmC is on its own an independent epigenetic mark in particular in the brain(Mellén et al., 2012; Sun et al., 2014). In addition, it has been suggested that 5fC and 5caC may also have regulatory functions especially when present at the promoters of highly expressed genes(Neri et al., 2015; Song & He, 2013). Importantly, TET proteins require cofactors for their activity such as alpha-ketoglutarate, iron and oxygen (O_2)(Tahiliani et al., 2009). The latter is a very important factor that modulates the activity of TET enzymes such as in tumours(Matuleviciute et al., 2021; Thienpont et al., 2016).

DNA 5mdC and its oxidized form has also been reported for the extremely AT-rich genome (>80%) of the protozoan human malaria parasite *P. falciparum*(Hammam et al., 2020; Ponts et al., 2013). In two previous publications bisulfite sequence analysis and affinity capture of methylated cytosine (hmeDIP-seq) was performed on genomic DNA of *P. falciparum* to study 5mdC(Ponts et al., 2013) and 5hmC(Hammam et al., 2020) sites. Both studies reported that less then 10% of methylated cytosines (5mdC and 5hmC) are at CpG sites and the vast majority exist in a CHH context. Furthermore, the hmeDIP-seq analysis showed 5hmC enrichment in gene bodies and low levels at 5', 3' and intergenic regions in 2100 genes of the sexual blood stages. This enrichment positively correlated with higher steady state transcript levels in *P. falciparum*[23]. Likewise, the lack of defined cytosine methylation patterns has been reported for *Drosophila melanogaster* and *Schistosoma mansoni* (Raddatz et al., 2013).

Malaria parasites belong to a small group of eukaryotes denoted as "DNMT2-only organisms" (absence of DNMT1 and DNMT3A/B genes) that includes a wide range of phylogenetically diverse group of organisms such as *D. melanogaster, S. mansoni, Schizosaccharomyces pombe* and *Entameoba histolytica*(Jeltsch et al., 2017). Similarly, to what is observed in other eukaryotes, plasmodial DNMT2 methylates $tRNA^{Asp}$ at position C38(Hammam et al., 2021). DNMT2-mediated tRNA methylation is not essential during standard *in vitro* culture conditions but was shown to maintain the pathogen's cellular homeostasis including sexual commitment in the presence of different stressors.

It is still debated if tRNA methylation is the only activity or could DNMT2 enzymes be promiscuous in their substrates(Jeltsch et al., 2017). Although, the TET enzymatic machinery needed for the generation of oxidative forms of 5mdC are absent in DNMT2-only model organisms, in malaria parasites relatively high levels of an oxidative cytosine form termed 5hmC-like (0.2-0.4%) was

reported but only very low levels of 5mdC (0.05%) in schizont blood stage preparations (essentially free of human white blood cell contamination) using ultra performance liquid chromatography-tandem mass spectrometry (LC/MS-MS)(Hammam et al., 2020).

In this study we identified the enzymatic machinery in an DNMT2-only organism that catalyzes malaria parasite DNA cytosine methylation and its oxidative forms. Before investigating the extremely AT-rich genome of *P. falciparum* (>80% A+T) we established a nuclease-based DNA digestion protocol that optimizes for 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). LC/MS-MS was used for the identification and quantification of 5mdC and its oxidative derivatives under different developmental blood stages and oxygen levels (5-20%) encountered during the complex life cycle. We show that DNA cytosine 5mdC/5hmC ratio is dynamically regulated throughout the 48-hour asexual blood stage parasite progression with a sharp increase (>10-fold) of 5mdC at the trophozoite stage. Importantly, hyperoxic culture conditions led to a dramatic shift of DNA cytosine modifications towards the oxidative derivatives 5hmC and 5caC in trophozoites. Deletion of *P. falciparum* DNMT2 erased all DNA cytosine modifications in mutant parasite genomic DNA. The identification of a non-canonical 5hmC biogenesis pathway in *P. falciparum* is supported by the identification of a hydroxylase (PF3D7_0829400) that is linked to the generation of cytosine oxidative products in *P. falciparum* and when expressed in *Saccharomyces cerevisiae*. This work reports dual specificities for DNMT2 as methyltransferase for DNA (this work) and tRNA (Hammam et al., 2021). In addition, we report evidence for a non-canonical pathway for 5hmC, which points to an adaptative evolutionary process of DNA cytosine modifications in malaria parasites. Our findings expand the plasmodial epigenetic marks to three cytosine 5mdC oxidative modifications and establish a direct link between variable oxygen environments encountered during the life cycle and the DNA cytosine modification pathway in malaria parasites.

Materials and methods

P. falciparum and mESC *in vitro* culture

P. falciparum 3D7-wildtype and Pf-DNMT2KO clones 1 and 2 parasites were maintained in culture and synchronized as previously described(Trager & Jensen, 1976). Parasites were routinely cultured at 5% normoxic O_2 levels and for specific experiments at 20% hyperoxic O_2 levels keeping CO_2 levels

well controlled at 5%. For genomic DNA preparation, parasites were first lysed using 0.15% saponin and parasite pellets were used to extract gDNA for digestion and hydrolysis.
mESC (mouse embryonic stem cells) were cultured in standard mouse ES medium supplemented with LIF (1000 U/mL, Life Technologies) to prevent differentiation, and incubated at 37 °C, 5% CO2 as describe earlier(Sambi et al., 2017).

Genomic DNA extraction

gDNA from parasites and mESC pellets were prepared using phenol-chloroform extraction. Briefly, 100 µL of DNA Extraction Buffer and 15µL of Proteinase K from Thermo Fisher (# EO0491) were added to 100ng of gDNA. The tubes were mixed and incubated at 55°C overnight. Equal volume of Phenol/Chloroform (pH 8.0) was added to the digested sample. The tubes were then spun for 5 min at 15.000 rpm separating the phases into the aqueous and the organic. For each sample, 650 µL 70% EtOH and 200 µL of 7.5 M ammonium acetate ($AcNH_4$), were added to each microcentrifuge tube and vortexed. The tubes were then placed in a −20°C freezer overnight to precipitate DNA. Once extracted, gDNA was treated twice with a mixture of RNAse A/ RNAse T1 at 10 ug/µL for 1 h with phenol extraction and ethanol precipitation steps between the two treatments to ensure the total elimination of RNA.

gDNA Enzymatic Digestion Protocols

Benzonase. 100 ng of gDNA was dissolved in 100 µL of 10 mM Tris-HCl buffer pH 7.9 containing 10 mM $MgCl_2$, 50 mM NaCl, 5 µM BHT from Sigma (B1378-100 G), and 3 mM deferoxamine from Sigma (# BP987). Next, 3 U of benzonase (in 20 mM Tris-HCl pH 8. 0.2mM $MgCl_2$, and 20 mM NaCl), 4 mU phosphodiesterase I from Sigma (# P3243-1VL), 3 U DNAse I, 2 mU of phosphodiesterase II from Sigma (# P9041-10 UN) and 2 U of alkaline phosphatase from Biolabs (# M0290) were added and the mixture was incubated at 37°C. Finally, the mixture was transferred in a microspin filter from Agilent Technologies (# 5190-5275) and recovered after spinning at 20°C for 10 min at 14 000 rpm.

Nuclease P1. We used a modified protocol from (Wang et al; 2011)(Wang et al., 2011). 100 ng of gDNA was dissolved in a 10 µL of buffer containing 0.3 M AcONa pH 5.6, 10 mM $ZnCl_2$, 3m M deferoxamine, and 1 mM EHNA. Next, 4 U of Nuclease P1 from Sigma (# N8630-1VL) (in 30 mM AcONa pH 5.3, 5 mM $ZnCl_2$ and 50 mM NaCl) and 5 mU phosphodiesterase II from Sigma (# P9041-

10UN) were added and the mixture was incubated at 37°C. After 4 h, 4 U of alkaline phosphatase from Biolabs (# M0290) and 5 mU of phosphodiesterase I from Sigma (# P3243-1VL) were added and the mixture was incubated for additionally 2h. Finally, the mixture was transferred in a microspin filter from Agilent Technologies (# 5190-5275) and recovered after spinning at 20°C for 10 min at 14 000 rpm.

NEB Nucleoside Digestion kit. 100 ng of gDNA was digested according to the Biolabs New England manufacture using the Nucleoside Digestion kit (# M0649), with little modification on time incubation, extended to overnight at 37°C. Finally, the mixture was transferred in a microspin filter from Agilent Technologies (# 5190-5275) and recovered after spinning at 20°C for 10 min at 14 000 rpm.

DNA Degradase Plus kit. 100 ng of gDNA was digested according to the Zymo Research manufacture using DNA Degradase Plus™ kit (# E 2020), with little modification on time incubation, extended to overnight at 37°C. Finally, the mixture was transferred in a microspin filter (3 kDa) and recovered after spinning at 20°C for 10 min at 14 000 rpm.

Nuclease S1. We used modified protocol from (Globisch et al., 2010)(Globisch et al., 2010). 100 ng of gDNA was heated to 100°C for 5 min to denature the DNA and rapidly cooled on ice. Buffer A (10 µL, 300 mM ammonium acetate, 100 µM $CaCl_2$, 1 mM $ZnSO_4$, pH 5.7) and nuclease S1 from Sigma (# N5661-50 KU) (80 U, *Aspergillus oryzae*) were added to the mixture and incubated for 3 h at 37°C. Addition of buffer B (12 mL, 500 mM Tris-HCl, 1 mM EDTA), Antarctic phosphatase from NEB (# M0289S-10 units), snake venom phosphodiesterase I (0.2 units, *Crotalus adamanteus* venom) from Sigma (# P3243-1VL) and incubation for further 3 h at 37°C completed the digestion. Finally, the mixture was transferred in a microspin filter from Agilent Technologies (# 5190-5275) and recovered after spinning at 20°C for 10 min at 14 000 rpm.

Quantification of DNA cytosine modifications by LC-MS/MS

Analysis of global levels of 5mdC, 5hmdC, 5fdC and 5cadC (herein abbreviated 5mdC, 5hmC, 5fC and 5caC) was performed on a Q exactive mass spectrometer (Thermo Fisher Scientific). It was equipped with an electrospray ionisation source (H-ESI II Probe) coupled with an Ultimate 3000 RS HPLC (Thermo Fisher Scientific). DNA digestions were injected onto a Thermo Fisher Hypersil Gold aQ chromatography column (100 mm * 2.1 mm, 1,9 um particle size) heated at 30°C. The flow rate was set at 0.3 mL/min and run with an isocratic eluent of 1% acetonitrile in

water with 0.1% formic acid for 10 minutes. Parent ions were fragmented in positive ion mode with 10% normalized collision energy in parallel-reaction monitoring (PRM) mode. MS2 resolution was 17,500 with an AGC target of 2e5, a maximum injection time of 50 ms and an isolation window of 1.0 m/z. The inclusion list contained the following masses: dC (228.1), 5mdC (242.1), 5hmdC (258.1), 5fdC (256.1) and 5cadC (272.1). Extracted ion chromatograms of base fragments (±5ppm) were used for detection and quantification (112.0506 Da for dC; 126.0662 Da for 5mdC; 142.0609 for 5hmdC; 140.0449 for 5fdC and 156.0399 for 5cadC). Calibration curves were previously generated using synthetic standards in the ranges of 0.2 to 50 pmoles injected for dC and 0.02 to 10 pmoles for 5mdC, 5hmdC, 5fdC and 5cadC. Results are expressed as a percentage of total dC

Yeast culture and strain construction

Standard methods were used for cultivation and manipulation of yeast strains(Sherman, 2002). Epitope-tagging was performed as described previously(Andresen et al., 2004). Yeast strains expressing *P. falciparum* Dnmt2^{pf}-SNAP and/ or Tet-like-CLIP were constructed, and the coding sequences of the SNAP-tag and CLIP-tag amplified using respectively the primers:
5'-GCGGGTACCATGGACAAAGACTGCGAAATG-3' and
5'-GCGCTCGAGTCATTAGCCCAGCCCAGGCTTGCCCAG-3', and
5'-GCGGGTACCGACAAAGACTGCGAAATG-3' and
5'-CGCCTCGAGTCATTAACCCAGCCCAGGCTTGCCCAG-3'

SNAP immunofluorescence assay and image acquisition

For the evaluation of the staining procedure, the commercially available fluorescent ligands SNAP-Cell TMR-Star (# S9105S) and CLIP-Cell Starter kit (# E9200) were used. For image acquisition, an epifluorescence microscope (Leica Microsystems, Germany) equipped with an oil immersion lens (1.4 NA; 100 x; DM200 LED; Leica Microsystems), a N3 filter cube (excitation: 546/12 nm; emission: 600/40 nm) and a GFP filter cube (excitation: 450/90 nm; emission: LP 590 nm) was employed.

DNA content quantification by flow cytometry

Parasites cultured at 5% or 20% O_2 were fixed in 4% PFA + 0,0075% glutaraldehyde in 1x PBS. Fixed cells were then stained with 2X SYBR-Green and the mean fluorescence intensity was measured using a BD Fortessa FACS instrument and analyzed using FlowJo software.

Statistics

Statistical analysis was carried out a 2-way ANOVA significance test, unless otherwise specified.

Results

An optimized mass spectrometry-based protocol for the detection of DNA cytosine modifications in AT-rich gDNA

In order to accurately investigate the levels of DNA cytosine epigenetic marks (5mdC and its oxidative forms (5hmdC, 5fdC and 5cadC) see figure 1A) in the highly AT-rich genomic DNA (gDNA) of *P. falciparum* using LC-MS/MS, we compared the gDNA digestion and hydrolysis protocol using five different methods: Nuclease P1, Benzonase, NEB Nucleoside Digestion kit, Degradase Plus kit and Nuclease S1 (Figure1B and 1C). All methods showed similar levels of 5mdC and oxidative cytosine forms in gDNAs from mouse Embryonic Stem Cells (mESC) (Figure1B). However, in the case of malaria parasites, the detected levels of 5hmC varied dramatically dependent on the method used to digest and hydrolyze gDNA. The levels of 5hmC in *Plasmodium* gDNA samples treated with Nuclease P1 are up to 8-fold higher compared to all other hydrolysis methods (Figure1C and Figure S1). Thus, our data show that Nuclease P1 (NP1) digestion is by far the most efficient enzyme to detect 5hmC in an extremely AT-rich organism. Therefore, the NP1 hydrolysis method was adopted to explore cytosine methylation marks in *P. falciparum*.

DNA cytosine modifications are dynamically regulated during the 48-hour blood stage cycle of *P. falciparum*

Parasites were synchronized and harvested at the ring (10-15 hours post invasion; hpi), trophozoite (24-30hpi) and schizont stages (35-40hpi) to determine how the levels of 5mdC and its oxidative derivatives vary across the intraerythrocytic developmental cycle (IDC) of *P. falciparum*. Using the LC-MS/MS protocol on DNA digested with Nuclease P1, we show that the levels of DNA cytosine modifications

in malaria parasites dynamically change as the parasite transitions from the early ring stage to the mature schizonts in the infected red blood cells (iRBCs) (Figure 2A). While 5mdC levels are very low in the ring (0.077% ± 0.015%) and schizont stages (0.13% ± 0.064%), we observe a drastic increase (10-fold) from the ring stage transition to trophozoites (0.727% ± 0.265%) (Figure 2A). Surprisingly, 5hmC levels exhibit the opposite pattern. 5hmC is by far the predominant cytosine modification in rings (0.533% ± 0.352%) and schizonts (0.697% ±0.072%) and decreases by 1.6-fold at the trophozoite stage (0.333% ± 0.162%). In addition, the levels of 5fC and 5caC are extremely low in all three stages of the IDC with 5caC being only detectable in the schizont stages (0.03% ± 0.01%).

Along with LC-MS/MS, we used a validated *in vitro* enzymatic assay(Ceccaldi et al., 2011) in order to quantify DNA methylation activity (DNMT) in nuclear extracts prepared from the three stages of the IDC. Human recombinant DNMT3aC is used as internal positive control and the data is presented as relative fluorescence units (RFU) (sample)-RFU (background signal) (Figure 2B). Our results show that DNA methylase activity is very low in ring stages (5000 ± 200 RFU) with a sharp increase in activity (8-fold) in trophozoites (40000 ± 200 RFU). The high activity signal is maintained in schizonts (Figure 2B). The DNMT activity assay data corroborates our LC-MS/MS findings showing the sharp increase of DNA methylation in the trophozoite stages of *P. falciparum*. Altogether, our data shows that DNA cytosine methylation is very dynamic in malaria blood stages and suggests a tight developmental regulation of 5mdC and the oxidative forms throughout *P. falciparum* IDC.

The DNA cytosine methylation epigenetic landscape is modulated by variable oxygen environments

We next wanted to evaluate how different physiological oxygen environments might affect the levels of DNA cytosine modifications in the parasites asexual blood stages. *P. falciparum* is exposed to fluctuating oxygen levels during its complex life cycle in the human host (liver and blood) and *Anopheles* mosquito vector (midgut and salivary glands). While circulating parasite infected erythrocytes (IE) are exposed to different O_2 levels in patients that vary between 5% in venous blood and 13% in arterial blood(Archer et al., 2018), these levels can reach up to 20-21% in the mosquito vector, where the tracheal system enables transport of gaseous oxygen from the atmosphere directly to the inner organs(Ha et al., 2017).

We analyzed wild-type and DNMT2KO *P. falciparum in vitro* cultures grown at the hyperoxic (atmospheric) 20% O_2 condition 12% O_2 and 5% O_2. Synchronous ring stage parasites were exposed

to different O_2 levels and trophozoites, schizonts and rings were harvested for each condition during the next 72 hours. Giemsa-stained blood smears showed no morphological differences between trophozoites, schizonts and rings parasites cultured in normoxic or atmospheric O_2 conditions (Figure S2). In addition, SYBR green-based DNA quantification showed that O_2 levels do not impact parasite DNA content of blood stages (Figure S2) as suggested by previous studies(Archer et al., 2018).

Importantly, LC-MS/MS-based DNA modifications quantification shows a shift of 5mdC towards its oxidative forms in 3D7-WT trophozoite stage parasites cultured under hyperoxic 12% and 20% O_2 levels (Figure 3). Exposure of the parasites to 12% O_2 led to a significant 2.2-fold decrease in 5mdC levels and 1.7-fold increase in 5hmdC levels when compared to the normoxic culture condition (5%). Furthermore, exposure of parasites to higher O_2 levels (20%) led to a sharp 12-fold decrease in 5mdC levels accompanied with a significant 2.9 -old increase in the levels of 5hmdC and in the abundance of 5caC in 3D7-WT parasites (Figure 3 and Figure S5 for rings and schizont data) indicating an increased oxidation of 5hmC.

Given that *P. falciparum* belongs to a small group of DNMT2-only organisms(Jeltsch et al., 2017; Raddatz et al., 2013), we investigated if the loss of Pf-DNMT2 impacts the DNA cytosine methylation repertoire of the parasite. To this end, we used trophozoites genomic DNA from two Pf-DNMT2KO that we generated previously(Hammam et al., 2021), cultured at 5%, 12% and 20% O_2 levels. Importantly, DNMT2KO parasites show no detectable amounts of any of 5hmdC, 5fdC or 5cadC in all O2 culture conditions (Figure 3 and Figure S3) and 5mdC levels are also very low (below the detection limits of the LC-MS/MS machine) and thus not quantifiable (Figure S3).

Altogether, the LC-MS/MS data shows that DNA cytosine methylation in *P. falciparum* is determined by DNMT2 and strongly supports a direct influence of oxygen O_2 concentration on the ratio of 5mdC and its oxidative derivatives, which may allow the malaria parasite to adapt its epigenetic cytosine landscape to variable host environments. Importantly, the LC-MS/MS data also shows that the presence of 5hmC, 5fC and 5caC in *P. falciparum* is completely dependent on 5mdC as previously shown for other model systems(Carell et al., 2018; Ito et al., 2011; Tahiliani et al., 2009; X. Wu & Zhang, 2017).

Identification of a non-canonical TET-like 5mdC hydroxylation pathway in *P. falciparum*

Our results strongly infer the presence of a yet unknown TET-like pathway in *P. falciparum* that triggers the oxidation of 5mdC into 5hmC and further into 5fC and 5caC (Figure 1A). We conducted an in-silico analysis by using the sequence of the DNA hydroxylase called JBP1 identified in African

Trypanosomes(Cliffe et al., 2009, p. 1) using HHPred (https://toolkit.tuebingen.mpg.de/tools/hhpred). We identified one hit with a similar active domain: PF3D7_0829400. The episomal overexpression of the Pf-TET-like candidate gene in *P. falciparum* was confirmed by western blot analysis (figure 4A) and shows the dual cytoplasmic and nuclear localization of the protein. LC-MS/MS analysis on gDNA extracted from 3D7-WT parasites and parasites overexpressing the TET candidate show a clear shift of the DNA cytosine marks towards oxidative forms 5hmC and 5fC, which indicates that this candidate gene encodes a 5mdC hydroxylase (Figure 4B). To confirm the function of PF3D7-0829400 in DNA 5mdC oxidation we carried out the expression of plasmodial DNMT2 and TET-like gene in a yeast species (*S. cerevisiae*) that has no modified DNA cytosines(Capuano et al., 2014). 5mdC was detectable in DNMT2+ yeast cells (up to 1.2%) however, 5hmdC, 5fdC and 5cadC were only detectable in double positive DNMT2+/TET+ yeast cells. In addition, non-transformed DNMT2-/TET- yeast cells showed no trace of detectable cytosine methylation. Taken together, our data strongly supports the identification of a non-canonical plasmodial TET enzyme.

Discussion

P. falciparum is among the most AT-rich genomes known to date with an overall A+T composition of 80.6%, which rises to ~90% in introns and intergenic regions(Gardner et al., 2002b). Given this unusual base composition, we compared different nuclease digestion procedures into single nucleosides prior LC-MS/MS analysis of cytosine modifications in genomic DNA. We observed striking differences in the digestion efficacy of 5hmdC between the distinct nucleases. With Nuclease P1 (NP1) 5hmdC levels are up to 8-fold higher compared to all others digestion methods, highlighting an important difference between AT-rich genomes and organisms with a balanced nucleotide composition. AT-rich DNA sequences display features that could account for the observed low efficacy of digestion of the oxidized DNA cytosine modifications. For example, AT-rich DNA stretches can form local curvature(Hagerman, 1985) that may exert an influence on the interaction of nucleases with DNA. Furthermore, 5hmC was shown to enhance DNA flexibility(Ngo et al., 2016), a feature that could change the affinity of certain type of nucleases for 5hmdC containing DNA.

We thus used nuclease P1 for digestion of gDNA to investigate the plasmodial DNA cytosine modifications landscape during different developmental blood stages using LC-MS/MS. Our

results show several established DNA cytosine methylation concepts that have not been observed in other eukaryotes. For example, 5mdC is present at rather constant levels in of 4–5% in all mouse tissues whereas the levels of 5hmC are relatively low and vary widely with peak steady-state levels of 5hmC of up to 1.3% in brain tissues (Carell et al., 2018). In malaria parasites, the ratio between 5mdC and the oxidized cytosine forms are inversed when compared to higher eukaryotes (Figure 1A/B). Furthermore, we demonstrate dynamic changes of 5mdC levels (10-fold) during the 48-hours blood stage parasite cycle. For most of the time, we detect only very low levels of 5mdC (0.07-0.1%) and high levels of 5hmC (0.5-0.9%). But during the short period of parasite proliferation (trophozoite stage), the 5mdC/5hmC ratio is inverted (Figure 2A). While ring stages circulate in the blood, trophozoite stages start to be sequestered in capillaries of host organs. It is tempting to speculate that the change in parasite tropism during blood stage development is linked to increased DNA methyltransferase activity. The levels of 5fC and 5caC remain very low in all three stages (standard *in vitro* culture condition). The question arises as to whether the elevated 5mdC levels in the trophozoite stage results from the re-location of PfDNMT2 to the nucleus. In our initial characterization of PfDNMT2(Hammam et al., 2021), Western blot analysis detected PfDNMT2 at similar levels in cytoplasmic and nuclear extracts of ring, trophozoite and schizont stages. Those data do not support the hypothesis that shuffling to the nucleus is the reason for higher 5mdC in trophozoites. More studies are needed to explore the link between the DNA replication that occurs during trophozoite stage and elevated 5mdC levels.

Previous work has reported relatively low 5mdC levels in *P. falciparum* genomic DNA, whereas about 10-fold higher oxidized form was detected using highly specific anti 5hmC antibodies[23]. The same study pointed to a cytosine base modification with unknown biochemical identity based on , LC-MS/MS analysis (Hammam et al., 2020). The digestion protocol used in this report was based on the benzonase nuclease. It is important to note that the unusual base reported earlier was also detectable in our nuclease digestion protocols (Figure S1). We observed that this cytosine form was predominant in *P. falciparum* gDNA hydrolyzed with benzonase but present only at very low levels when the DNA was digested with Nuclease P1 (Figure S1). We conclude that the choice of nuclease for the digestion of AT-rich DNA is important to maximize the liberation of oxidative cytosine nucleosides.

Experimental evidence from the cancer field demonstrated that reduction of oxygen levels in tumors is associated with a reduced ability to produce 5hmC(Thienpont et al., 2016). Since *P.*

falciparum is exposed to fluctuating oxygen concentrations (5-20%) during the life cycle, we investigated if malaria parasite cytosine modification marks respond to physiological changes in oxygen levels. Our results demonstrate that host-dependent oxygen concentrations result in important alterations in the cytosine modification landscape of *P. falciparum*. We observed a shift towards oxidative derivatives of 5mdC at higher oxygen levels (Figure 3). DNA cytosine modifications are now considered to function as distinct epigenetic marks with distinct profiles in tissues/organs in higher eukaryotes(Carell et al., 2018). A recent study used anti 5hmC antibodies to immunoprecipitate DNA regions enriched for this cytosine mark (hmeDIP-seq) of gDNA of *P. falciparum.* The presence of hydroxymethylated cytosine in gene bodies positively correlated with transcript levels in blood stages(Hammam et al., 2020). We hypothesize that the observed dynamic changes of cytosine modifications arose as an evolutionary adaptation to protect malaria parasites against changing environments encountered in the human host and the mosquito vector.

TET enzymes, which catalyze the oxidative cytosine forms depend on oxygen as co-substrate(Ito et al., 2011; Matuleviciute et al., 2021), may sense changes in oxygen levels and change the basal cytosine modification level toward oxidative cytosine forms. For example, TET1 activity as well as 5hmC levels are decreased in low oxygen conditions leading to the inhibition of the differentiation of mouse embryonic stem cells (mESC)(Burr et al., 2018). Although TET-like methylcytosine hydroxylase activity was previously detected in nuclear extracts of *P. falciparum* using a TET activity kit (Abcam) no obvious ortholog of mammalian TET enzymes had been identified(Hammam et al., 2020). In this study we identified a *P. falciparum* hydroxylase gene (PF3D7_0829400) that strongly support the existence of a non-canonical 5mdC to 5hmC/5fC/5caC hydroxylation pathway in *P. falciparum* (see Fig. 4). Two independent functional experiments strongly support the idea that malaria parasites have evolved a noncanonical TET pathway.

An obvious question is if the identified oxidative cytosine forms (5hmC, 5fC, 5caC) depend on 5mdC as a precursor or are formed by oxidative lesions. *P. falciparum* is part of a small number of organisms called DNMT2-only, given the absence of genes coding for the canonical DNA methyltransferase genes DNMT1 and DNMT3(Raddatz et al., 2013). A recent study reported a *P. falciparum* Pf-DNMT2 KO parasite line, and this work confirmed the well-established function of DNMT2 as tRNA aspartic acid (GTC) cytosine methyltransferase(Hammam et al., 2021). These KO parasites show no growth defect but are significantly more sensitive to various stressor and

respond to specific stress situations by 8-fold higher transmission stages or higher drug sensitivity, respectively. This study also observed a reduction of 5mdC in gDNA of KO parasites, however, given the very low level reported for 5mdC (0.06%) in the investigated schizont stage, it remained unclear if the observed reduction in 5mdC can be attributed to the loss of Pf-DNMT2 or variations in the experimental procedure(Hammam et al., 2021). DNMT2-dependent DNA methylation is still disputed with contradictory results from distinct DNMT2-only organisms(Jeltsch et al., 2017). Here we demonstrate that DNMT2 has a dual function using tRNA and DNA as substrate in malaria parasites. Quantitative LC-MS/MS analysis demonstrate that the levels of 5mdC and its oxidative derivatives are drastically decreased to background levels in the Pf-DNMT2KO clones when compared to the 3D7 control strain (Figure 3 and Figures S3 and S4). Furthermore, the identified oxidative cytosine forms are also reduced to not detectable levels, inferring that 5mdC is the precursor for 5hmC, 5fC and 5caC cytosine modifications. The very low background of 5hmC detected in preparations of *P. falciparum* gDNA was recently reported to originate from minor contaminations of human white blood cells in cultured blood stage parasites(Hammam et al., 2021).

In conclusion, this work demonstrates that plasmodial Pf-DNMT2 has a dual function in methylating tRNA and DNA. The 5mdC methylation of gDNA is developmentally regulated during the blood stage development and the 5mdC distribution *versus* oxidized forms can vary dramatically depending on environmental oxygen concentrations. Expanding the number of potential epigenetic DNA marks in this pathogen may open new avenues to understand the regulatory circuits that act during the life cycle stage.

Acknowledgements

This work was supported by the Laboratoire d'Excellence (LabEx) ParaFrap [ANR-11-LABX-0024], the Agence Nationale pour la Recherche [Project EpiKillMal, ANR-20-CE18-0006] and [Project 5metdC, ANR-18-CE12-0011-01]. This work was supported by DIM1Health 2019 equipment grant “EpiK” from Région Ile de France.

Appendix C

Author contributions

Conceptualization A.S, E.H, S.M and B.A; Methodology: E.H, S.M, G.D and F.B; Investigation E.H and S.M; Analysis: E.H, S.M and F.B; Writing: E.H, S.M and A.S; Funding acquisition: A.S and P.B.

Data availability

LC-MS/MS raw data is shown in supplementary tables S1 and S2.

References

1. Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J. Biol. Chem. 175, 315–332 (1948).
2. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. Science 187, 226–232 (1975).
3. Riggs, A. D. X inactivation, differentiation, and DNA methylation. Cytogenetic and Genome Research 14, 9–25 (1975).
4. Aapola, U. et al. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65, 293–298 (2000).
5. Bachman, K. E., Rountree, M. R. & Baylin, S. B. Dnmt3a and Dnmt3b Are Transcriptional Repressors That Exhibit Unique Localization Properties to Heterochromatin. Journal of Biological Chemistry 276, 32282–32287 (2001).
6. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243–247 (2015).
7. Sharif, J. et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450, 908–912 (2007).

8. Doi, A. et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet. 41, 1350–1353 (2009).
9. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492 (2012).
10. Guy, J., Cheval, H., Selfridge, J. & Bird, A. The Role of MeCP2 in the Brain. Annual Review of Cell and Developmental Biology 27, 631–652 (2011).
11. Meng, H. et al. DNA Methylation, Its Mediators and Genome Integrity. International Journal of Biological Sciences 11, 604–617 (2015).
12. Bienvenu, T. & Chelly, J. Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. Nature Reviews Genetics 7, 415–426 (2006).
13. Shi, D.-Q., Ali, I., Tang, J. & Yang, W.-C. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. Front Genet 8, 100 (2017).
14. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. 18, 517–534 (2017).
15. Ito, S. et al. Tet Proteins Can Convert 5-methydeoxyctidine to 5-Formylcytosine and 5-Carboxylcytosine. Science 333, 1300–1303 (2011).
16. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. Cell 151, 1417–1430 (2012).
17. Sun, W., Zang, L., Shu, Q. & Li, X. From development to diseases: The role of 5hmC in brain. Genomics 104, 347–351 (2014).

18. Neri, F. et al. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. Cell Reports 10, 674–683 (2015).
19. Song, C.-X. & He, C. Potential functional roles of DNA demethylation intermediates. Trends in Biochemical Sciences 38, 480–484 (2013).
20. Tahiliani, M. et al. Conversion of 5-methydeoxyctidine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935 (2009).
21. Matuleviciute, R., Cunha, P. P., Johnson, R. S. & Foskolou, I. P. Oxygen regulation of TET enzymes. FEBS J febs.15695 (2021) doi:10.1111/febs.15695.
22. Thienpont, B. et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. Nature 537, 63–68 (2016).
23. Hammam, E. et al. Discovery of a new predominant cytosine DNA modification that is linked to gene expression in malaria parasites. Nucleic Acids Research 48, 184–199 (2020).
24. Ponts, N. et al. Genome-wide Mapping of DNA Methylation in the Human Malaria Parasite Plasmodium falciparum. Cell Host & Microbe 14, 696–706 (2013).
25. Raddatz, G. et al. Dnmt2-dependent methylomes lack defined DNA methylation patterns. Proc. Natl. Acad. Sci. U.S.A. 110, 8627–8631 (2013).
26. Jeltsch, A. et al. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. RNA Biol 14, 1108–1123 (2017).
27. Hammam, E. et al. Malaria Parasite Stress Tolerance Is Regulated by DNMT2-Mediated tRNA Cytosine Methylation. mBio e02558-21 (2021) doi:10.1128/mBio.02558-21.
28. Trager, W. & Jensen, J. Human malaria parasites in continuous culture. Science 193, 673–675 (1976).

29. Sambi, M. et al. Acellular Mouse Kidney ECM can be Used as a Three-Dimensional Substrate to Test the Differentiation Potential of Embryonic Stem Cell Derived Renal Progenitors. Stem Cell Rev and Rep 13, 513–531 (2017).
30. Wang, J. et al. Quantification of Oxidative DNA Lesions in Tissues of Long-Evans Cinnamon Rats by Capillary High-Performance Liquid Chromatography−Tandem Mass Spectrometry Coupled with Stable Isotope-Dilution Method. Anal. Chem. 83, 2201–2209 (2011).
31. Globisch, D. et al. Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. PLoS ONE 5, e15367 (2010).
32. Sherman, F. Getting started with yeast. in Methods in Enzymology vol. 350 3–41 (Elsevier, 2002).
33. Andresen, M., Schmitz-Salue, R. & Jakobs, S. Short Tetracysteine Tags to β-Tubulin Demonstrate the Significance of Small Labels for Live Cell Imaging. MBoC 15, 5616–5622 (2004).
34. Ceccaldi, A. et al. C5-DNA Methyltransferase Inhibitors: From Screening to Effects on Zebrafish Embryo Development. ChemBioChem 12, 1337–1345 (2011).
35. Archer, N. M. et al. Resistance to Plasmodium falciparum in sickle cell trait erythrocytes is driven by oxygen-dependent growth inhibition. Proc Natl Acad Sci USA 115, 7350–7355 (2018).
36. Ha, Y.-R., Ryu, J., Yeom, E. & Lee, S.-J. Comparison of the tracheal systems of Anopheles sinensis and Aedes togoi larvae using synchrotron X-ray microscopic computed tomography (respiratory system of mosquito larvae using SR-μCT): HA ET AL . Microsc Res Tech 80, 985–993 (2017).

37. Carell, T., Kurz, M. Q., Müller, M., Rossa, M. & Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. Angew. Chem. Int. Ed. 57, 4296–4312 (2018).

38. Cliffe, L. J. et al. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. Nucleic Acids Res. 37, 1452–1462 (2009).

39. Capuano, F., Mülleder, M., Kok, R., Blom, H. J. & Ralser, M. Cytosine DNA Methylation Is Found in Drosophila melanogaster but Absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe , and Other Yeast Species. Anal. Chem. 86, 3697–3702 (2014).

40. Gardner, M. J. et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419, 498–511 (2002).

41. Hagerman, P. J. Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. Biochemistry 24, 7033–7037 (1985).

42. Ngo, T. T. M. et al. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. Nat Commun 7, 10813 (2016).

43. Burr, S. et al. Oxygen gradients can determine epigenetic asymmetry and cellular differentiation via differential regulation of Tet activity in embryonic stem cells. Nucleic Acids Research 46, 1210–1226 (2018).

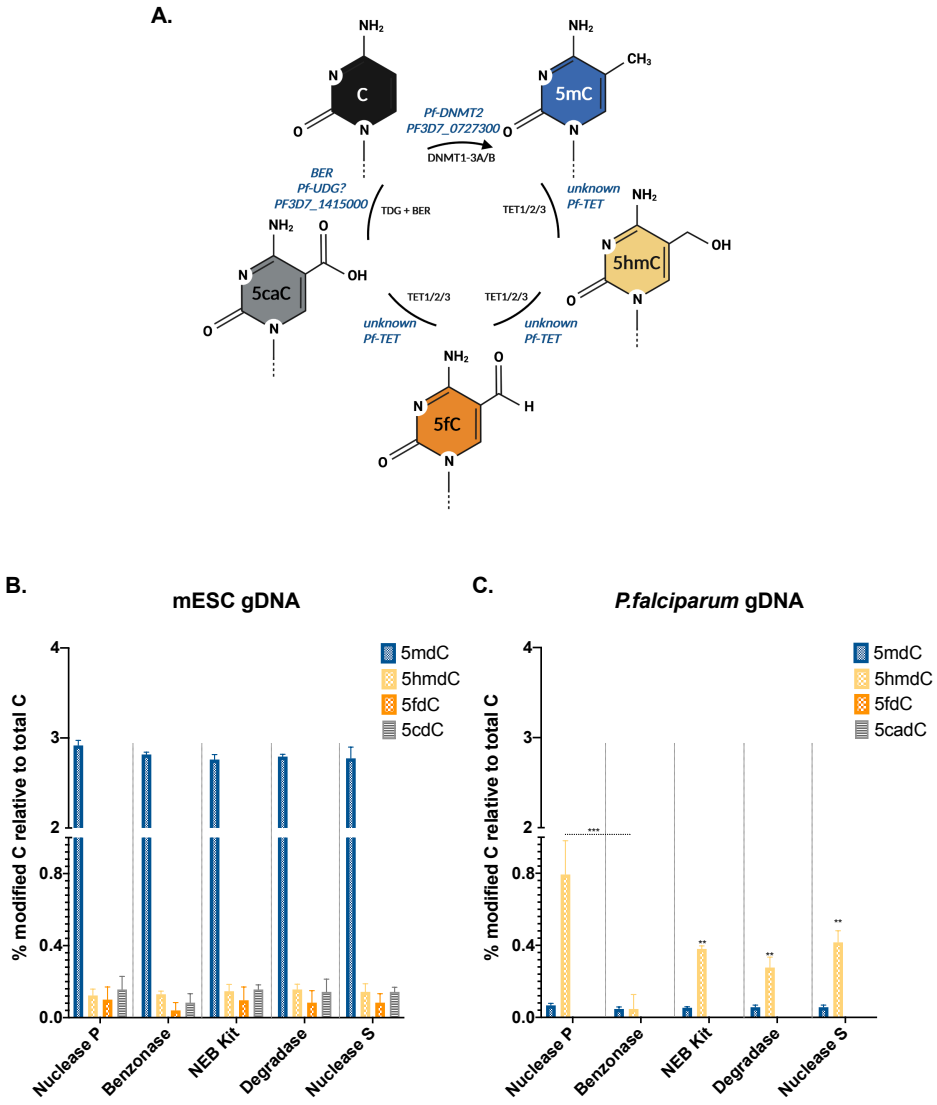


Figure 1: Comparative cytosine modification analysis by LC-MS/MS of gDNA from *P. falciparum* and mouse embryonic stem cells (mESC) using different digestion methods. (A)Schematic representation summarizing the DNA methylation/demethylation pathway as established in model organisms (black font) compared to what is known about the DNA methylation/demethylation machinery in *P. falciparum* (blue font). Pf-DNMT2: *P. falciparum* DNA methyltransferase 2; TET: ten eleven translocation enzymes; BER: base excision repair machinery; TGD/UDG: thymine/uracil-DNA glycosylase. Genomic DNA from mESC cells (B) and *P. falciparum* (C) was digested using five different protocols: Nuclease P1(NP1), Benzonase, NEB Nucleoside Digestion kit (NEB kit), Degradase kit (Degradase) and Nuclease S. Quantitative data analysis is shown as percentage of modified deoxycytidines relative to the total number of deoxycytidines in the tested sample and represents the mean (± SD) of two independent replicates for each condition (n=2). Statistical analysis was carried out using a 2way ANOVA test. P Value (Benzonase/NP1)=0,0007(***) ; P Value (NEB kit/NP1)=0,0048(**) ; P Value (Degradase/NP1)=0,002(**) and P Value (Nuclease S/NP1)=0,0065(**).

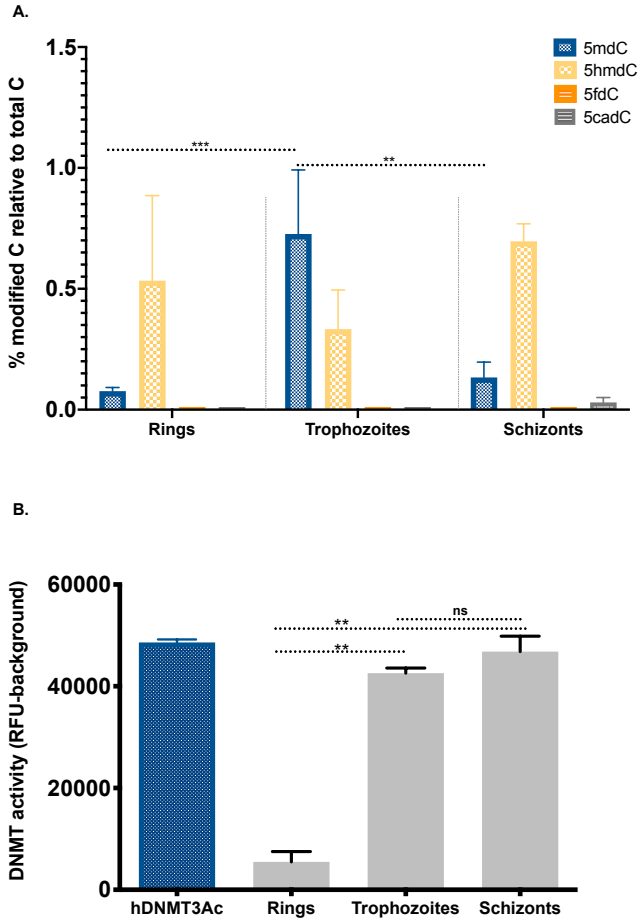


Figure 2: DNA cytosine methylation is dynamically regulated during the 48-hour *P. falciparum* asexual blood stage development. (A) LC-MS/MS-based quantification of 5mdC, 5hmdC, 5fdC and 5cadC in gDNA of synchronized ring, trophozoite and schizont stages. Data is shown as percentage of modified nucleoside relative to the total number of deoxycytidines in each sample and represents the mean of three independent replicates for each stage (n=3). (B) DNA methyltransferase (DNMT) activity assay using nuclear extracts from synchronous rings, trophozoite and schizonts. Data is shown as the RFU (sample)-RFU (background) and represents the mean (± SD) of two independent replicates for each stage (n=2). Human recombinant DNMT3aC was used as internal positive control. Statistical analysis was carried out using a 2way ANOVA for the LC-MS/MS data and a t-test for the DNMT activity data. P Value <0.05 = * ; P Value <0.005 =**; P Value <0.001=***.

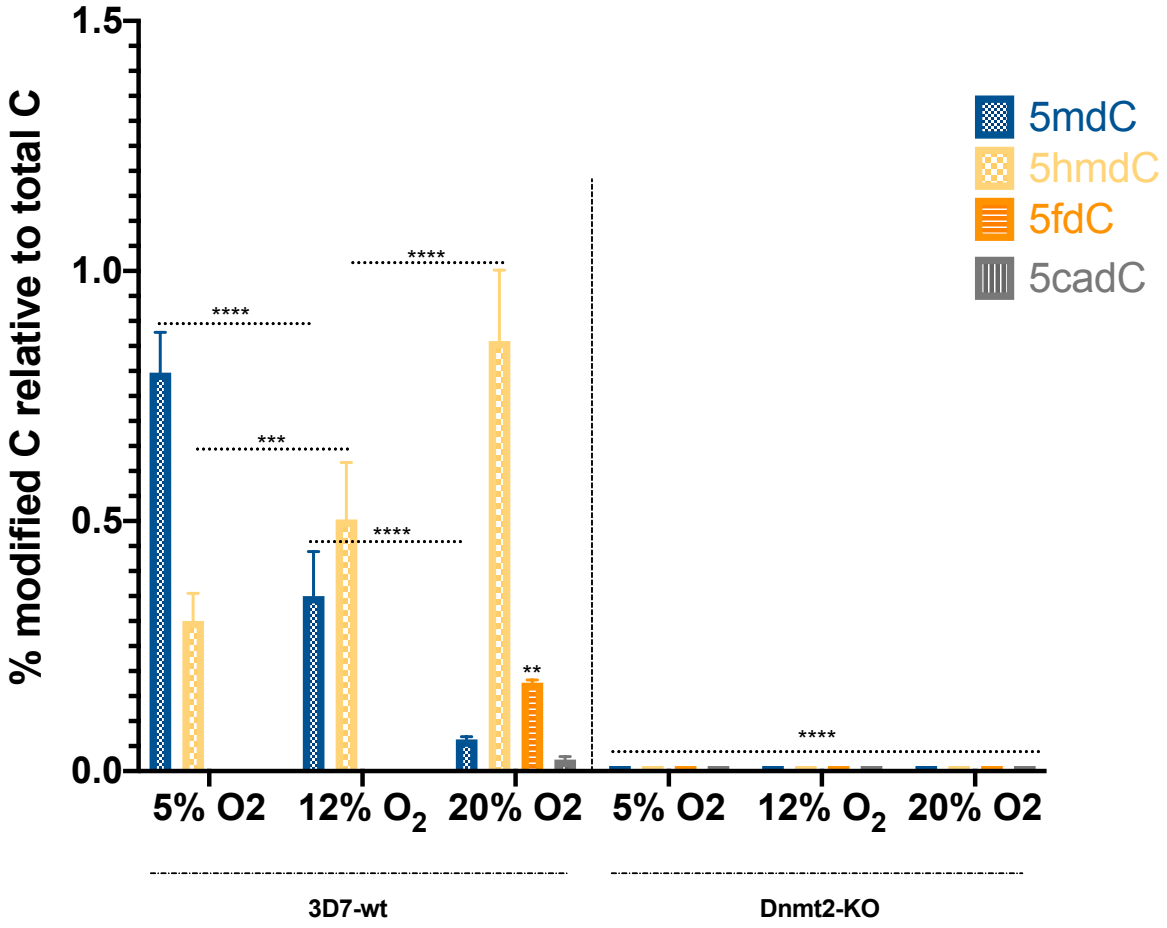


Figure 3: Effect of variable oxygen concentrations on gDNA cytosine modifications in 3D7-WT and DNMT2KO *P. falciparum*. DNA was isolated from *in vitro* cultured parasites (trophozoite stage) propagated at different oxygen levels. Data is shown as percentage of modified deoxycytidines relative to the total number of deoxycytidines in each sample and represents the mean (± SD) of three independent replicates n=3. Statistical analysis was carried out using a 2-way ANOVA test. P Value <0.0001 = ****; P Value <0.005 =**.

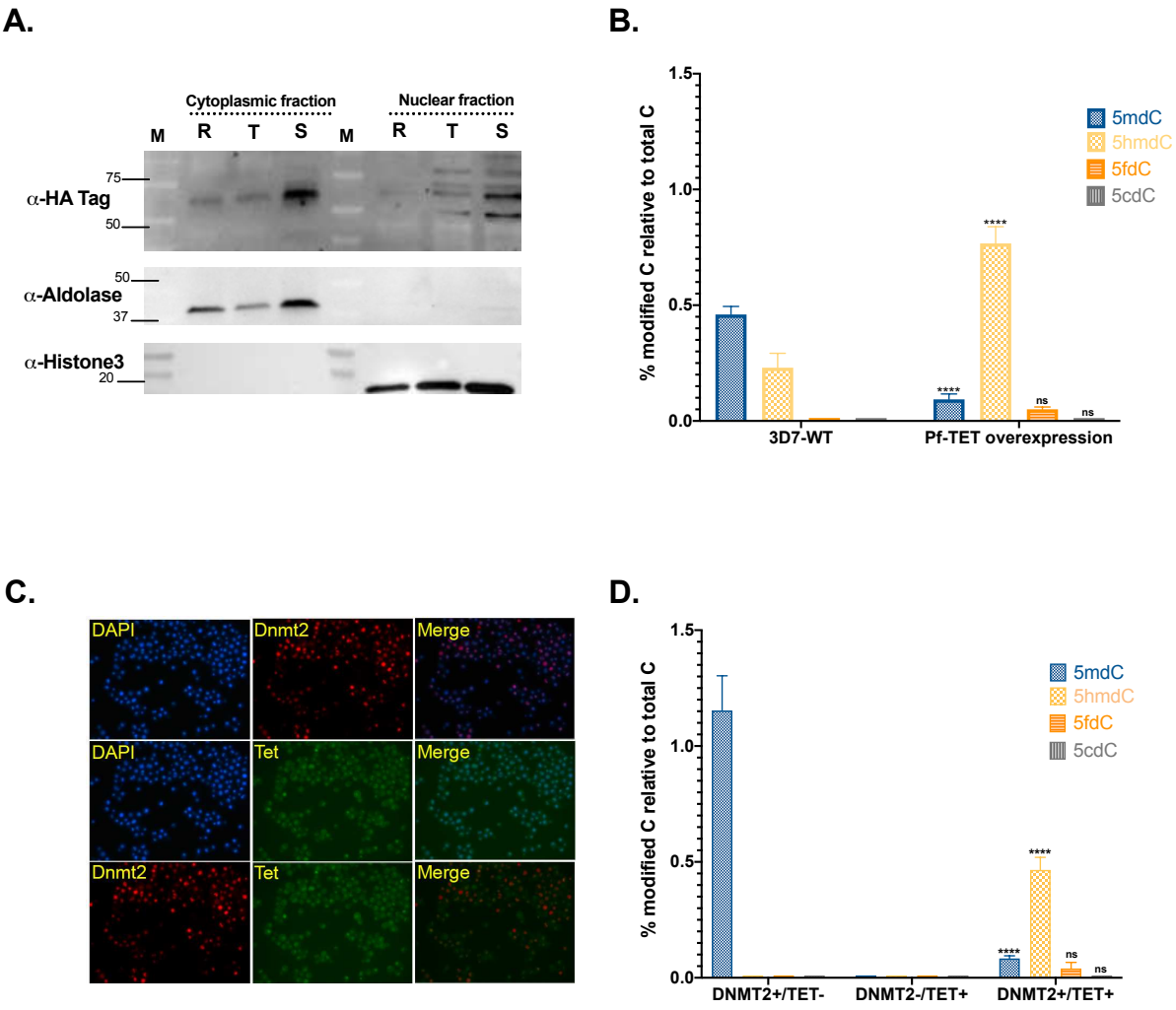


Figure 4: Identification of a non-canonical 5mdC oxidation pathway in *P. falciparum*. A. Western blot detection of Pf-TET episomal expression in nuclear and cytoplasmic fractions from rings (R), trophozoites (T) and schizonts (S), using anti-HA antibodies (expected size around 57 KDa). Anti-aldolase and anti-H3 antibodies were used as cytoplasmic and nuclear markers, respectively. B. LC-MS/MS-based quantification of 5mdC, 5hmdC, 5fdC and 5cadC in gDNA from 3D7-wild type (WT) parasites *versus* gDNA from parasites overexpressing Pf-TET. Data is shown as percentage of modified deoxycytidines relative to the total number of deoxycytidines in each sample and represents the mean (± SD) of three independent replicates of each strain(n=3). (C) Pf-DNMT2-SNAP and Pf-TET-CLIP expression in *S. Cerevisiae* cells. Scale bar= 10 µm. DIC: differential interference contrast; Hoechst: nuclear staining. (D) LC-MS/MS-based quantification of 5mdC, 5hmdC, 5fdC and 5cadC in non-transformed yeast (DNMT2-/TET-), yeast only expressing DNMT2 (DNMT2+/TET-) and double transformed yeast (DNMT2+/TET+). Statistical analysis was carried out using 2way ANOVA. P Value <0.0001 = ****; P Value <0.005 =**.

Supplementary figures legends

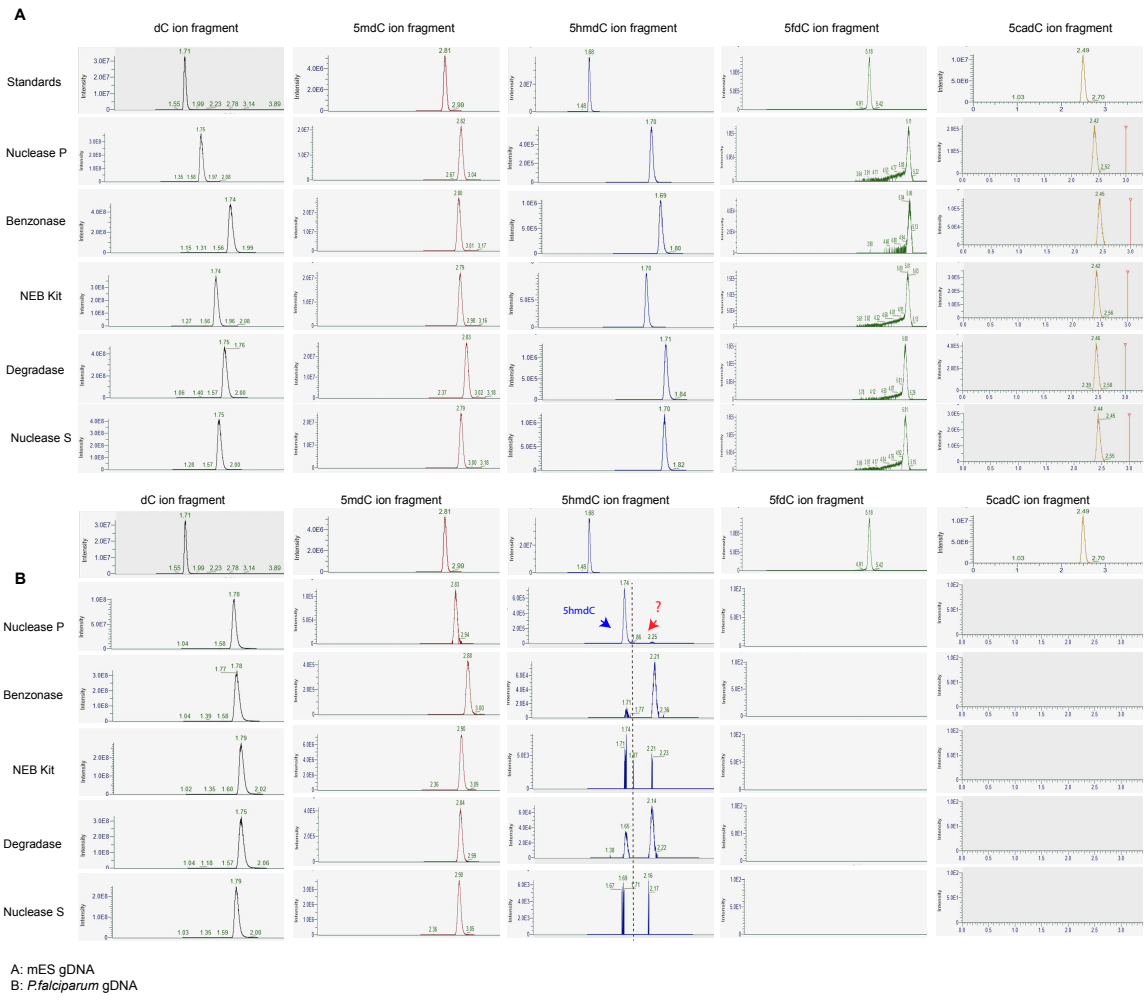


Figure S1: LC-MS/MS chromatogram of modified cytosines. A) mESC gDNA and B) *P. falciparum* gDNA according to the digestion protocol. Blue arrows show the expected 5hmdC peak while the red arrows indicate the shifted 5hmdC peak observed earlier (Hammam et al., 2020). deoxycytidine (dC), 5-methyl-2'-deoxycytidine (5mdC), 5-hydroxymethyl-2'-deoxycytidine (5hmdC), 5-carboxy-2'-deoxycytidine (5cadC), and 5'-formyl-2'-deoxycytidine (5fdC).

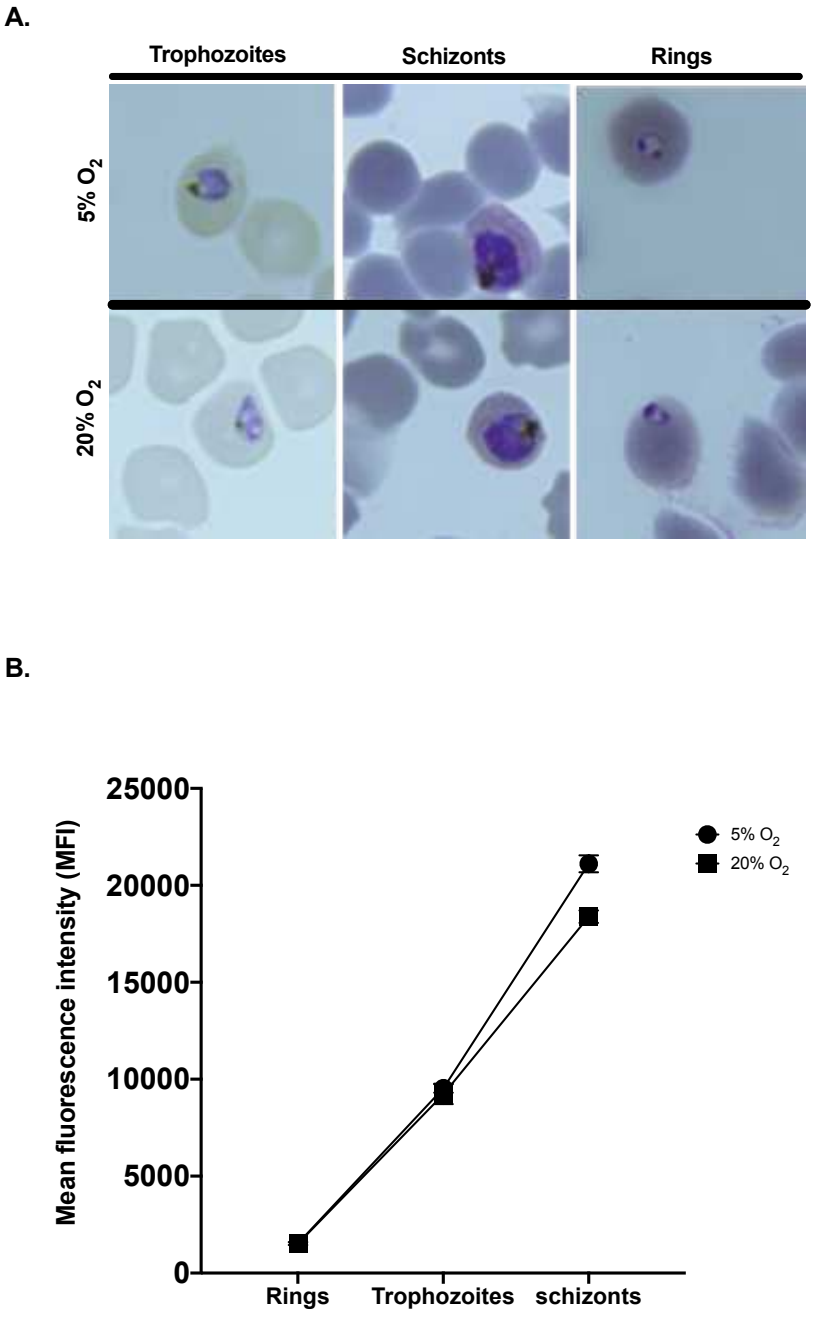


Figure S2: *In vitro* cultured parasite growth under normal (5% O_2) and high oxygen concentrations (20% O_2). Giemsa staining (A) and DNA quantification (B) during the 48h blood stage cycle do not reveal any developmental changes between parasites cultured at 5% *vs* 20% O_2.

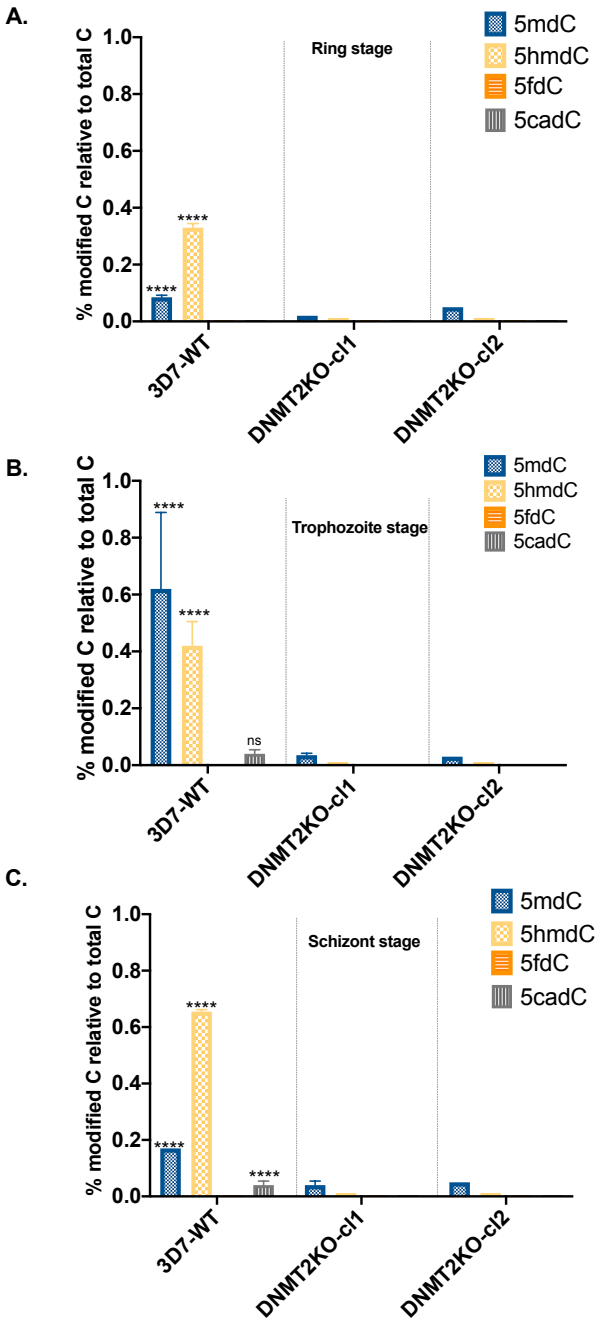


Figure S3: LC-MS/MS quantification of 5mdC, 5hmdC, 5fdC and 5cadC in gDNA of 3D7-WT and DNMT2KO parasite from ring (A), trophozoite (B) and schizont stages (C). Nucleoside quantification of 5mdC, 5hmdC, 5fdC and 5cadC in gDNA from 3D7-wild type (WT) parasites versus gDNA from the two Pf-DNMT2KO clones cl1 and cl2. Data is shown as percentage of modified deoxycytidines relative to the total number of deoxycytidine in each sample and represents the mean (+-SD) of two independent replicates of each strain.

Appendix C

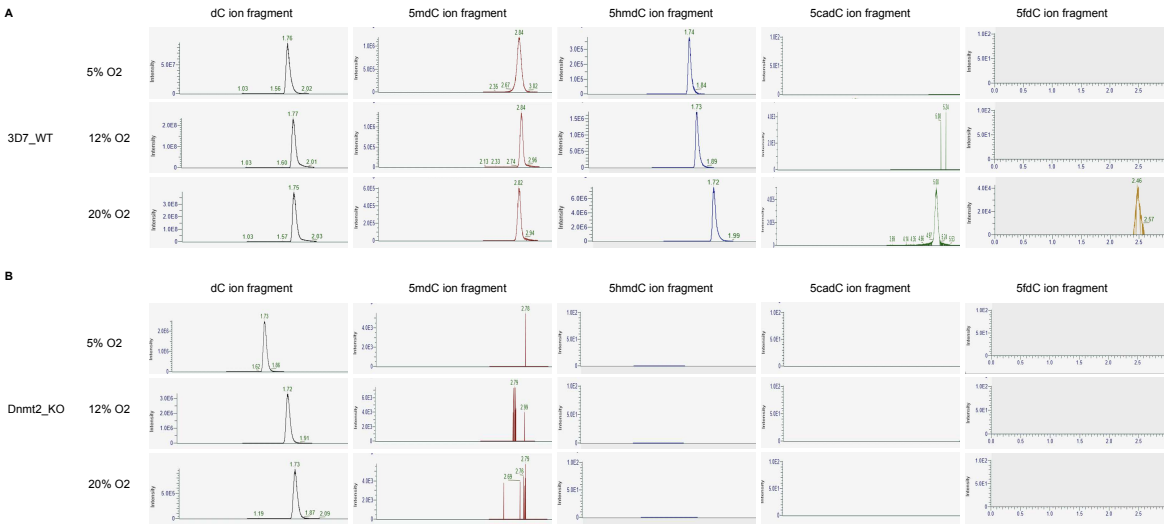


Figure S4: LC-MS/MS chromatogram of dC, 5mdC, 5hmdC, 5cadC, and 5fdC. parasites cultured at 5%, 12% or 20%. A) 3D7-WT and B) DNMT2-KO. This experiment was done in 3 different replicates (Figure 3), here one replicate experiment is shown.

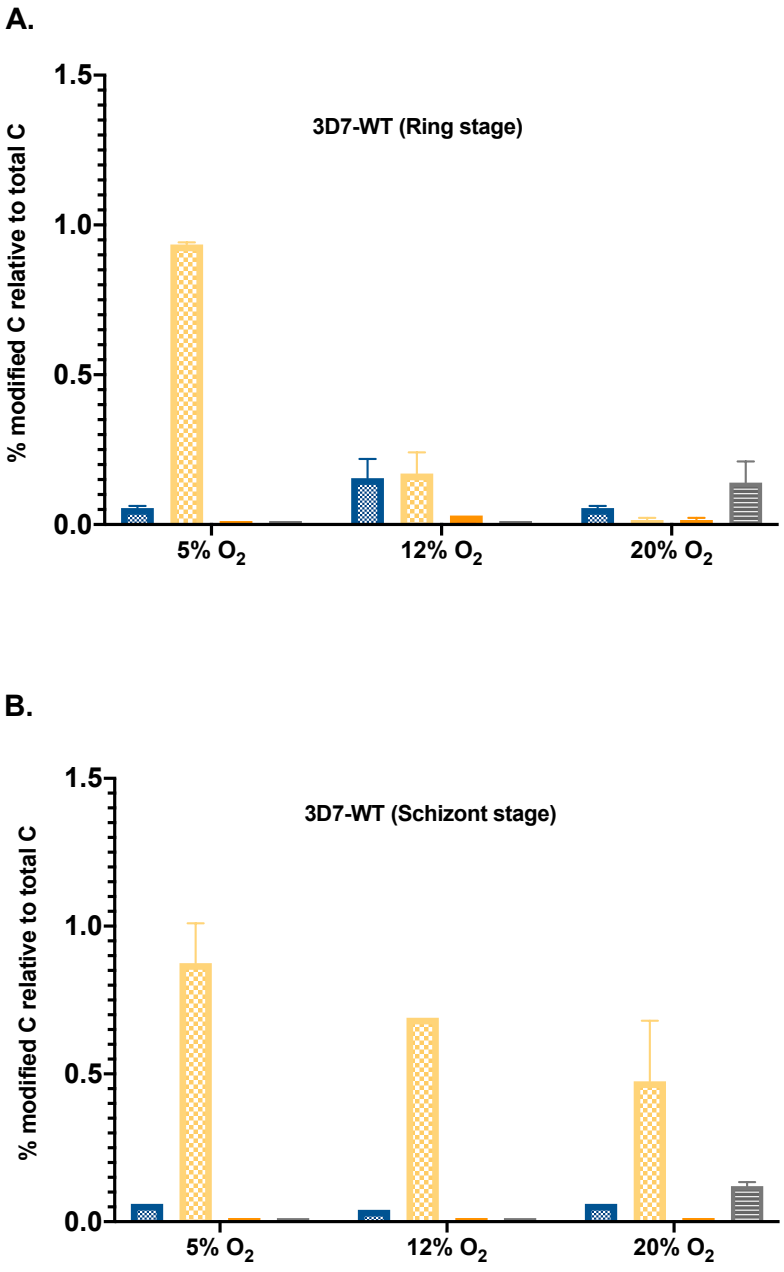


Figure S5: LC-MS/MS quantification of 5mdC, 5hmdC, 5fdC and 5cadC in gDNA of 3D7-WT parasites from ring (A) or Schizonts stages (B). Data is shown as percentage of modified deoxycytidines relative to the total number of deoxycytidine in each sample and represents the mean (+-SD) of two independent replicates of each strain.

Appendix C

Appendix C

References

[Aird *et al.*, 2014] Aird WC, Mosnier LO, Fairhurst RM. Plasmodium falciparum picks (on) EPCR. Blood. 2014 Jan 9;123(2):163-7.

[Amit-Avraham *et al.*, 2015] Amit-Avraham I., Pozner G., Eshar S., Fastman Y., Kolevzon N., Yavin E., Dzikowski R. 2015. Antisense long noncoding RNAs regulate var gene activation in the malaria parasite Plasmodium falciparum. Proc. Natl. Acad. Sci. U.S.A. 112:e982–E991.

[Andrade *et al.*, 2020] Andrade, C.M., Fleckenstein, H., Thomson-Luque, R. *et al.* Increased circulation time of *Plasmodium falciparum* underlies persistent asymptomatic infection in the dry season. Nat Med 26, 1929–1940 (2020).

[Ariey and Ménard, 2019] Ariey, F. and Ménard, D. (2019). An update on artemisinin resistance. Methods Mol Biol, 2013:141–149.

[Ariey *et al.*, 2014] Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A. C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C. M., Bout, D. M., Ménard, S., Rogers, W. O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J. C., Fairhurst, R. M., Benoit-Vical, F., Mercereau-Puijalon, O., and Ménard, D. (2014). A molecular marker of artemisinin- resistant plasmodium falciparum malaria. Nature, 505(7481):50–55.

[Ashley *et al.*, 2014] Ashley, E. A., Dhorda, M., Fairhurst, R. M., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J. M., Mao, S., Sam, B., Sopha, C., Chuor, C. M., Nguon, C., Sovannaroth,

S., Pukrittayakamee, S., Jittamala, P., Chotivanich, K., Chutasmit, K., Suchatsoonthorn, C., Runcharoen, R., Hien, T. T., Thuy-Nhien, N. T., Thanh, N. V., Phu, N. H., Htut, Y., Han, K. T., Aye, K. H., Mokuolu, O. A., Olaosebikan, R. R., Folaranmi, O. O., Mayxay, M., Khanthavong, M., Hongvanthong, B., Newton, P. N., Onyamboko, M. A., Fanello, C. I., Tshefu, A. K., Mishra, N., Valecha, N., Phyo, A. P., Nosten, F., Yi, P., Tripura, R., Borrmann, S., Bashraheil, M., Peshu, J., Faiz, M. A., Ghose, A., Hossain, M. A., Samad, R., Rahman, M., Hasan, M., Islam, A., Miotto, O., Amato, R., MacInnis, B., Stalker, J., Kwiatkowski, D., Bozdech, Z., Jeeyapant, A., Cheah, P., Sakulthaew, T., Chalk, J., Intharabut, B., Silamut, K., Lee, S. J., Vihokhern, B., Kunasol, C., Imwong, M., Tarning, J., Taylor, W. J., Yeung, S., Woodrow, C. J., Flegg, J. A., Das, D., Smith, J., Venkatesan, M., Plowe, C. V., Stepniewska, K., Guerin, P. J., Dondorp, A. M., Day, N. P., White, N. J., and Tracking Resistance to Artemisinin Collaboration, T. (2014). Spread of artemisinin resistance in plasmodium falciparum malaria. N Engl J Med, 371(5):411–423.

[Ay *et al.*, 2014] Ay, F., Bunnik, E. M., Varoquaux, N., Bol, S. M., Prudhomme, J., Vert, J. P., Noble, W. S., and Le Roch, K. G. 2014. Three-dimensional modeling of the p. falciparum genome during the erythrocyte cycle reveals a strong connection between genome architecture and gene expression. Genome Res, 24(6): 974-988.

[Azzalin *et al*., 2007] Azzalin, C.M., Reichenbach,P.,Khoriauli,L.,Giulotto,E.,andLingner, J. (2007). Telomeric repeat containing rna and rna surveillance factors at mammalian chromosome ends. Science, 318(5851):798–801.

[Balaji *et al.*, 2005] Balaji, S., Babu, M. M., Iyer, L. M., and Aravind, L. (2005). Discovery of the principal specific transcription factors of apicomplexa and their implication for the evolution of the ap2-integrase dna binding domains. Nucleic Acids Res, 33(13):3994–4006.

[Balk *et al.*, 2013] Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric rna-dna hybrids affect telomere-length dynamics and senescence. Nat Struct Mol Biol, 20(10):1199–1205.

[Balu *et al.*, 2011] Balu B, Maher SP, Pance A, Chauhan C, Naumov AV, Andrews RM, Ellis PD, Khan SM, Lin JW, Janse CJ, et al. CCR4-associated factor 1 coordinates the expression of Plasmodium falciparum egress and invasion proteins. Eukaryot Cell 2011, 10:1257–1263.

[Barcons-Simon *et al.*, 2020] Barcons-Simon A, Cordon-Obras C, Guizetti J, Bryant JM, Scherf A. 2020. CRISPR Interference of a Clonally Variant GC-Rich Noncoding RNA Family Leads to General Repression of *var* Genes in Plasmodium falciparum. mBio. 11(1):e03054-19.

[Bargieri *et al.*, 2014] Bargieri, D., Lagal, V., Andenmatten, N., Tardieux, I., Meissner, M., and Ménard, R. (2014). Host cell invasion by apicomplexan parasites: the junction co- nundrum. PLoS Pathog, 10(9):e1004273.

[Bártfai *et al.*, 2010] Bártfai, R., Hoeijmakers, W. A., Salcedo-Amaya, A. M., Smits, A. H., Janssen-Megens, E., Kaan, A., Treeck, M., Gilberger, T. W., Françoijs, K. J., and Stun- nenberg, H. G. (2010). H2a.z demarcates intergenic regions of the plasmodium falci- parum epigenome that are dynamically marked by h3k9ac and h3k4me3. PLoS Pathog, 6(12):e1001223.

[Bartoloni and Zammarchi, 2012] Bartoloni, A. and Zammarchi, L. (2012). Clinical aspects of uncomplicated and severe malaria. Mediterr J Hematol Infect Dis, 4(1):e2012026.

[Baum *et al.*, 2009] Baum, J., Papenfuss, A. T., Mair, G. R., Janse, C. J., Vlachou, D., Waters, A. P., Cowman, A. F., Crabb, B. S., and de Koning-Ward, T. F. (2009). Molecular genetics and comparative genomics reveal rnai is not functional in malaria parasites. Nucleic Acids Res, 37(11):3788–3798.

[Baumgarten *et al.*, 2019] Baumgarten, S., Bryant, J. M., Sinha, A., Reyser, T., Preiser, P. R., Dedon, P. C., and Scherf, A. (2019). Transcriptome-wide dynamics of extensive m^6a mrna methylation during plasmodium falciparum blood-stage development. Nat Microbiol.

[Beltran *et al.*, 2008] Beltran, M., Puig, I., Peña, C., García, J. M., Alvarez, A. B., Peña, R., Bonilla, F., and de Herreros, A. G. (2008). A natural antisense transcript regu- lates zeb2/sip1 gene expression during snail1-induced epithelial-mesenchymal transition. Genes Dev, 22(6):756–769.

[Bertschi *et al.*, 2017] Bertschi NL, Toenhake CG, Zou A, Niederwieser I, Henderson R, Moes S, Jenoe P, Parkinson J, Bartfai R, Voss TS. Malaria parasites possess a telomere repeat-binding protein that shares ancestry with transcription factor IIIA. Nat Microbiol. 2017 Mar 13;2:17033.

References

[Birnbaum *et al.*, 2017] Birnbaum, J., Flemming, S., Reichard, N. *et al.* 2017. A genetic system to study *Plasmodium falciparum* protein function. Nat Methods 14, 450–456.

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

[Borst, 2003] Borst, P. (2003). Mechanisms of antigenic variation: an overview. Antigenic Variation, ed. A Craig, A Scherf. London: Academic Press:1–16.

[Bowman *et al.*, 1999] Bowman, S., Lawson, D., Basham, D., Brown m D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C., and Barrell, B. G. 1999. The complete nucleotide sequence of chromosome 3 of plasmodium falciparum. Nature, 400(6744):532-538.

[Bowman *et al.*, 1999] Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Wood- ward, J. R., Newbold, C., and Barrell, B. G. (1999). The complete nucleotide sequence of chromosome 3 of plasmodium falciparum. Nature, 400(6744):532–538.

[Bozdechetal.,2003] Bozdech, Z.,Llinás,M.,Pulliam,B.L.,Wong,E.D.,Zhu,J.,andDeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of plasmod- ium falciparum. PLoS Biol, 1:E5.

[Branco and Pombo, 2007] Branco, M. R. and Pombo, A. (2007). Chromosome organization: new facts, new models. Trends Cell Biol, 17(3):127–134.

References

[Brancucci *et al.*, 2012] Brancucci N.M., Witmer K., Schmid C.D., Flueck C., Voss T.S. 2012. Identification of a cis-acting DNA-protein interaction implicated in singular var gene choice in Plasmodium falciparum. Cell Microbiol. 14:1836–1848.

[Broadbent *et al.*, 2011] Broadbent, K. M., Park, D., Wolf, A. R., Van Tyne, D., Sims, J., Ribacke, U., Volkman, S., Duraisingh, M., Wirth, D., Sabeti, P., and Rinn, J. L. (2011). A global transcriptional analysis of plasmodium falciparum malaria reveals a novel family of telomere-associated lncrnas. Genome Biol, 12(6):R56.

[Broadbent *et al.*, 2015] Broadbent, K. M., Broadbent, J. C., Ribacke, U., Wirth, D., Rinn, J. L., and Sabeti, P. C. (2015). Strand-specific rna sequencing in plasmodium falciparum malaria identifies developmentally regulated long non-coding rna and circular rna. BMC Genomics, 16:454.

[Brolin *et al.*, 2009] Brolin, K. J., Ribacke, U., Nilsson, S., Ankarklev, J., Moll, K., Wahlgren, M., and Chen, Q. (2009). Simultaneous transcription of duplicated var2csa gene copies in individual plasmodium falciparum parasites. Genome Biol, 10:R117.

[Bryant *et al.*, 2017] Bryant, J. M., Regnault, C., Scheidig-Benatar, C., Baumgarten, S., Guizetti, J., and Scherf, A. (2017). Crispr/cas9 genome editing reveals that the intron is not essential for var2csa gene activation or silencing in plasmodium falciparum. MBio, 8(4).

[Bull *et al.*, 1998] Bull, P. C., Lowe, B. S., Kortok, M., Molyneux, C. S., Newbold, C. I., and Marsh, K. 1998. Parasite antigens on the infected red blood cell surface are targets for naturally acquired immunity to malaria. Nat Med, 4(3):358-360.

[Bumgarner *et al.*, 2012] Bumgarner, S. L., Neuert, G., Voight, B. F., Symbor-Nagrabska, A., Grisafi, P., van Oudenaarden, A., and Fink, G. R. 2012. Single-cell analysis reveals that noncoding rnas contribute to clonal heterogeneity by modulating transcription factor recruitment. Mol Cell, 45(4):470-482.

[Bunnik *et al.*, 2014] Bunnik, E. M., Polishko, A., Prudhomme, J., Ponts, N., Gill, S. S., Lonardi, S., and Le Roch, K. G. (2014). Dna-encoded nucleosome occupancy is asso- ciated with

transcription levels in the human malaria parasite plasmodium falciparum. BMC Genomics, 15:347.

[Calderón *et al.*, 2013] Calderón, F., Wilson, D. M., and Gamo, F. J. 2013. Antimalarial drug discovery: recent progress and future directions. Prog Med Chem, 52:97-151.

[Calderónetal.,2013] Calderón,F.,Wilson,D.M.,andGamo,F.J.(2013).Antimalarialdrug discovery: recent progress and future directions. Prog Med Chem, 52:97–151.

[Calderwood *et al.*, 2003] Calderwood M.S., Gannoun-Zaki L., Wellems T.E., Deitsch K.W. 2003. Plasmodium falciparum var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. J. Biol. Chem. 278:34125–34132.

[Carrieri *et al.*, 2012] Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., Forrest, A. R., Carninci, P., Biffo, S., Stupka, E., and Gustincich, S. (2012). Long non-coding antisense rna controls uchl1 translation through an embedded sineb2 repeat. Nature, 491(7424):454–457.

[Cesana *et al.*, 2011] Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding rna controls muscle differentiation by functioning as a competing endogenous rna. Cell, 147(2):358–369.

[Chakrabarti *et al.*, 2007] Chakrabarti, K., Pearson, M., Grate, L., Sterne-Weiler, T., Deans, J., Donohue, J. P., and Ares, M. (2007). Structural rnas of known and unknown func- tion identified in malaria parasites by comparative genomics and rna analysis. RNA, 13(11):1923–1939.

[Chen and Shyu, 2011] Chen, C. Y. and Shyu, A. B. (2011). Mechanisms of deadenylation-dependent decay. Wiley Interdiscip Rev RNA, 2:167–183.

[Chookajorn *et al.*, 2007] Chookajorn, T., Dzikowski, R., Frank, M., Li, F., Jiwani, A. Z., Hartl, D. L., and Deitsch, K. W. (2007). Epigenetic memory at malaria virulence genes. Proc Natl Acad Sci U S A, 104(3):899–902.

References

[Chookajorn *et al.*, 2008] Chookajorn T., Ponsuwanna P., Cui L. 2008. Mutually exclusive var gene expression in the malaria parasite: multiple layers of regulation. Trends Parasitol. 24:455–46.

[Chu *et al.*, 2011] Chu, C., Qu, K., Zhong, FL, Artandi, SE, Chang, HY. 2011. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44:667-678.

[Chu *et al.*, 2012] Chu, C., Quinn, J., Chang, HY. 2012. Chromatin isolation by RNA purification (ChIRP). J Vis Exp.

[Chu *et al.*, 2018] Chu C., Chang H.Y. (2018) ChIRP-MS: RNA-Directed Proteomic Discovery. In: Sado T. (eds) X-Chromosome Inactivation. Methods in Molecular Biology, 1861.

[Cottrell *et al.*, 2014] Cottrell, G., Musset, L., Hubert, V., Le Bras, J., Clain, J., and Atovaquone-Proguanil, T. F. S. G. (2014). Emergence of resistance to atovaquone- proguanil in malaria parasites: insights from computational modeling and clinical case reports. Antimicrob Agents Chemother, 58(8):4504–4514.

[Cottrell *et al.*, 2014] Cottrell, G., Musset, L., Hubert, V., Le Bras, J., Clain, J., and Atovaquone-Proguanil, T. F. S. G. 2014. Emergence of resistance to atovaquone-proguanil in malaria parasites: insights from computational modeling and clinical case reports. Antimicrob Agents Chemother, 58(8):4504-4514.

[Cowman *et al.*, 2017] Cowman, A. F., Tonkin, C. J., Tham, W. H., and Duraisingh, M. T. (2017). The molecular basis of erythrocyte invasion by malaria parasites. Cell Host Mi- crobe, 22(2):232–245.

[Cox, 2010] Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. Parasit Vectors, 3(1):5.

References

[Cui *et al.*, 2007] Cui, L., Miao, J., Furuya, T., Li, X., Su, X. Z., and Cui, L. (2007). Pfgcn5-mediated histone h3 acetylation plays a key role in gene expression in plasmodium fal- ciparum. Eukaryot Cell, 6(7):1219–1227.

[Cui *et al.*, 2008a] Cui, L., Fan, Q., Cui, L., and Miao, J. (2008a). Histone lysine methyltransferases and demethylases in plasmodium falciparum. Int J Parasitol, 38(10):1083–1097.

[Cui *et al.*, 2008b] Cui, L., Miao, J., Furuya, T., Fan, Q., Li, X., Rathod, P. K., Su, X. Z., and Cui, L. (2008b). Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro plasmodium falciparum development. Eukaryot Cell, 7:1200–1210.

[Dastidar *et al.*, 2012] Dastidar, E. G., Dayer, G., Holland, Z. M., Dorin-Semblat, D., Claes, A., Chêne, A., Sharma, A., Hamelin, R., Moniatte, M., Lopez-Rubio, J. J., Scherf, A., and Doerig, C. (2012). Involvement of plasmodium falciparum protein kinase ck2 in the chro- matin assembly pathway. BMC Biol, 10:5.

[David *et al.*, 1983] David, P. H., Hommel, M., Miller, L. H., Udeinya, I. J., and Oligino, L. D. (1983). Parasite sequestration in plasmodium falciparum malaria: spleen and an- tibody modulation of cytoadherence of infected erythrocytes. Proc Natl Acad Sci U S A, 80(16):5075–5079.

[Day and Tuite, 1998] Day, D. A. and Tuite, M. F. (1998). Post-transcriptional gene regu- latory mechanisms in eukaryotes: an overview. J Endocrinol, 157(3):361–371.

[De Santa *et al.*, 2010] De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C. L., and Natoli, G. (2010). A large fraction of extragenic rna pol ii transcription sites overlap enhancers. PLoS Biol, 8(5):e1000384.

[Deitsch *et al.*, 2009] Deitsch, K. W., Lukehart, S. A., and Stringer, J. R. (2009). Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat Rev Microbiol, 7:493–503.

[Deitschm *et al.*, 2009] Deitschm K. W., Lukehart, S. A., and Stringer, J. R. 2009. Common strategies for antigenic variation by bacterial, fungal, and protozoan pathogens. Nat Rev Microbiol, 7:493-503.

[Dieci *et al.*, 2013] Dieci, G, Conti, A, Pagano, A, Carnevali, D. 2013. Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. Biochim Biophys Acta 1829:296–305.

[Dondorp *et al.*, 2010] Dondorp, A. M., Yeung, S., White, L., Nguon, C., Day, N. P., Socheat, D., and von Seidlein, L. (2010). Artemisinin resistance: current status and scenarios for containment. Nat Rev Microbiol, 8(4):272–280.

[Donze *et al.*, 2001] Donze, D., and Kamakaka, R. T. 2001. Rna polymerase iii and rna polymerase ii promoter complexes are heterochromatin barriers in saccharomyces cerevisiae. EMBO J, 20(3):520-531.

[Droll *et al.*, 2018] Droll, D., Wei, G., Guo, G., Fan, Y., Baumgarten, S., Zhou, Y., Xiao, Y., Scherf, A., and Zhang, Q. (2018). Disruption of the rna exosome reveals the hidden face of the malaria parasite transcriptome. RNA Biol, 15(9):1206–1214.

[Duraisingh *et al.*, 2005] Duraisingh, M. T., Voss, T. S., Marty, A. J., Duffy, M. F., Good, R. T., Thompson, J. K., Freitas-Junior, L. H., Scherf, A., Crabb, B. S., and Cowman, A. F. (2005). Heterochromatin silencing and locus repositioning linked to regulation of viru- lence genes in plasmodium falciparum. Cell, 121(1):13–24.

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

[Dzikowski *et al.*, 2007] Dzikowski R., Li F., Amulic B., Eisberg A., Frank M., Patel S., Wellems T.E., Deitsch K.W. 2007. Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. EMBO Rep. 8:959–965.

References

[Ebersole *et al.*, 2011] Ebersole, T., Kim, J. H., Samoshkin, A., Kouprina, N., Pavlicek, A., White, R., and Larionov, V. 2011. Trna genes protect a reporter gene from epigenetic silencing in mouse cells. Cell Cycle, 10(16):2779-2791.

[Epp *et al.*, 2009] Epp, C., Li, F., Howitt, C. A., Chookajorn, T., and Deitsch, K. W. (2009). Chromatin associated sense and antisense noncoding rnas are transcribed from the var gene family of virulence genes of the malaria parasite plasmodium falciparum. RNA, 15(1):116–127.

[Faghihi and Wahlestedt, 2009] Faghihi, M. A. and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol, 10(9):637–643.

[Faghihi *et al.*, 2008] Faghihi,M.A.,Modarresi,F.,Khalil,A.M.,Wood,D.E.,Sahagan,B.G., Morgan, T. E., Finch, C. E., St Laurent, G., Kenny, P. J., and Wahlestedt, C. (2008). Expres- sion of a noncoding rna is elevated in alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med, 14(7):723–730.

[Fan *et al.*, 2020] Fan Y, Shen S, Wei G, Tang J, Zhao Y, Wang F, He X, Guo G, Shang X, Yu X, Ma Z, He X, Liu M, Zhu Q, Le Z, Wei G, Cao J, Jiang C, Zhang Q. Rrp6 Regulates Heterochromatic Gene Silencing via ncRNA RUF6 Decay in Malaria Parasites. mBio. 2020 Jun 2;11(3):e01110-20.

[Figueiredo *et al.*, 2000] Figueiredo, L. M., Pirrit, L. A., Scherf, A., and Pirritt, L. A. (2000). Genomic organisation and chromatin structure of plasmodium falciparum chromosome ends. Mol Biochem Parasitol, 106(1):169–174.

[Figueiredo *et al.*, 2002] Figueiredo L.M., Freitas-Junior L.H., Bottius E., Olivo-Marin J.C., Scherf A. 2002. A central role for Plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation. EMBO J. 21:815–824.

[Filarsky *et al.*, 2018] Filarsky, M., Fraschka, S. A., Niederwieser, I., Brancucci, N. M., 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(6381):1259–1263.

References

[Fire *et al.*, 1998] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded rna in caenorhabditis elegans. Nature, 391(6669):806–811.

[Flueck *et al.*, 2009] Flueck, C., Bartfai, R., Volz, J., Niederwieser, I., Salcedo-Amaya, A. M., Alako, B. T., Ehlgen, F., Ralph, S. A., Cowman, A. F., Bozdech, Z., Stunnenberg, H. G., and Voss, T. S. (2009). Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog, 5(9):e1000569.

[Flueck *et al.*, 2010] Flueck, C., Bartfai, R., Niederwieser, I., Witmer, K., Alako, B. T., Moes, S., Bozdech, Z., Jenoe, P., Stunnenberg, H. G., and Voss, T. S. (2010). A major role for the plasmodium falciparum apiap2 protein pfsip2 in chromosome end biology. PLoS Pathog, 6(2):e1000784.

[Francia and Striepen, 2014] Francia, M. E. and Striepen, B. (2014). Cell division in apicomplexan parasites. Nat Rev Microbiol, 12(2):125–136.

[Francia and Striepen, 2014] Francia, M. E., and Striepen, B. 2014. Cell division in apicomplexan parasites. Nat Rev Microbiol, 12(2):125-136.

[Frank *et al.*, 2006] Frank M., Dzikowski R., Costantini D., Amulic B., Berdougo E., Deitsch K. 2006. Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite Plasmodium falciparum. J. Biol. Chem. 281:9942–9952.

[Frank *et al.*, 2007] Frank, M., Dzikowski, R., Amulic, B., and Deitsch, K. W. (2007). Variable switching rates of malaria virulence genes are associated with chromosomal position. Mol Microbiol, 64(6):1486–1498.

[Fraschka *et al.*, 2016] Fraschka, S., Henderson, R. W., and Bártfai, R. (2016). H3.3 demar- cates gc-rich coding and subtelomeric regions and serves as potential memory mark for virulence gene expression in plasmodium falciparum. Sci Rep, 6:31965.

[Freitas-Junior *et al.*, 2000] Freitas-Junior, L. H., Bottius, E., Pirrit, L., Deitsch, K. W., Schei- dig, C., Guinet, F., Nehrbass, U., Wellems, T. E., and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of p. falci- parum. Nature, 407(6807):1018–1022.

[Freitas-Junior *et al.*, 2005] Freitas-Junior, L. H., Hernandez-Rivas, R., Ralph, S. A., Montiel-Condado, D., Ruvalcaba-Salazar, O. K., Rojas-Meza, A. P., Mancio-Silva, L., Leal-Silvestre, R. J., Gontijo, A. M., Shorte, S., and Scherf, A. (2005). Telomeric heterochromatin propaga- tion and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell, 121(1):25–36.

[Frénal *et al.*, 2017] Frénal, K., Dubremetz, J. F., Lebrun, M., and Soldati-Favre, D. (2017). Gliding motility powers invasion and egress in apicomplexa. Nat Rev Microbiol, 15(11):645–660.

[Fried *et al.*, 1998] Fried, M., Nosten, F., Brockman, A., Brabin, B. J., and Duffy, P. E. 1998. Maternal antibodies block malaria. Nature, 395(6705):851-852.

[Fujioka and Aikawa, 2002] Fujioka, H. and Aikawa, M. (2002). Structure and life cycle. Chem Immunol, 80:1–26.

[Gardner *et al.*, 1996] Gardner, J. P., Pinches, R. A., Roberts, D. J., and Newbold, C. I. (1996). Variant antigens and endothelial receptor adhesion in plasmodium falciparum. Proc Natl Acad Sci U S A, 93(8):3503–3508.

[Gardner *et al.*, 1998] Gardner, M. J., Tettelin, H., Carucci, D. J., Cummings, L. M., Aravind, L., Koonin, E. V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D. C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G. G., Clayton, R., White, O., Smith, H. O., Fraser, C. M., Adams, M. D., Venter, J. C., and Hoffman, S. L. (1998). Chromosome 2 sequence of the human malaria parasite plasmodium falciparum. Science, 282(5391):1126–1132.

[Gardner *et al.*, 2002] Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A.,

Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002). Genome sequence of the human malaria parasite plasmodium falciparum. Nature, 419(6906):498–511.

[Geisler and Coller, 2013] Geisler, S. and Coller, J. (2013). Rna in unexpected places: long non-coding rna functions in diverse cellular contexts. Nat Rev Mol Cell Biol, 14(11):699– 712.

[Goel *et al.*, 2015] Goel, S., Palmkvist, M., Moll, K., Joannin, N., Lara, P., Akhouri, R. R., Moradi, N., Öjemalm, K., Westman, M., Angeletti, D., Kjellin, H., Lehtiö, J., Blixt, O., Ideström, L., Gahmberg, C. G., Storry, J. R., Hult, A. K., Olsson, M. L., von Heijne, G., Nilsson, I., and Wahlgren, M. (2015). Rifins are adhesins implicated in severe plasmodium falciparum malaria. Nat Med, 21(4):314–317.

[Gomes *et al.*, 2013] Gomes, A. Q., Nolasco, S., & Soares, H. (2013). Non-coding RNAs: multi-tasking molecules in the cell. *International journal of molecular sciences*, *14*(8), 16010–16039.

[Grewal and Jia, 2007] Grewal, S. I. and Jia, S. (2007). Heterochromatin revisited. Nat Rev Genet, 8(1):35–46.

[Guizetti and Scherf, 2013] Guizetti, J. and Scherf, A. (2013). Silence, activate, poise and switch! mechanisms of antigenic variation in plasmodium falciparum. Cell Microbiol, 15(5):718–726.

[Guizetti *et al.*, 2016] Guizetti, J, Barcons-Simon, A, Scherf, A. 2016. Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite. Nucleic Acids Res 44:9710–9718.

[Gunasekera *et al.*, 2007] Gunasekera, A. M., Myrick, A., Militello, K. T., Sims, J. S., Dong, C. K., Gierahn, T., Le Roch, K., Winzeler, E., and Wirth, D. F. (2007). Regulatory mo- tifs uncovered among gene expression clusters in plasmodium falciparum. Mol Biochem Parasitol, 153(1):19–30.

References

[Hall *et al.*, 2002] Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D., Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K., Buckee, C. O., Bur- rows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Corton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J., Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K. D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P., Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Rutter, S., Rutherford, K. M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J. E., Craig, A., Newbold, C., and Barrell, B. G. (2002). Sequence of plasmodium falciparum chromosomes 1, 3-9 and 13. Nature, 419(6906):527–531.

[Hammam *et al.*, 2020] Hammam, E., Ananda, G., Sinha, A., Scheidig-Benatar, C., Bohec, M., Preiser, P., Dedon, P., Scherf, A., Vembar, S. (2020) Discovery of a new predominant cytosine DNA modification that is linked to gene expression in malaria parasites, *Nucleic Acids Research*, 48(10):184–199.

[Hammam *et al.*, 2021] Hamman, E., Miled, S., Bonhomme, F., Diffendall, G., Arcangioli, B., Arimondo, P.B., Scherf, A. (2021) Variable oxygen environments and DNMT2 determine the DNA cytosine epigenetic landscape of Plasmodium falciparum, unpublished.

[Hasenkamp *et al.*, 2013] Hasenkamp, S., Merrick, C. J. & Horrocks, P. 2013. A quantitative analysis of Plasmodium falciparum transfection using DNA-loaded erythrocytes. *Mol Biochem Parasitol* 187, 117-120.

[Hastings *et al.*, 1997] Hastings, M. L., Milcarek, C., Martincic, K., Peterson, M. L., and Munroe, S. H. (1997). Expression of the thyroid hormone receptor gene, erbaalpha, in b lymphocytes: alternative mrna processing is independent of differentiation but corre- lates with antisense rna levels. Nucleic Acids Res, 25(21):4296–4300.

References

[Henikoff and Smith, 2015] Henikoff, S. and Smith, M. M. (2015). Histone variants and epigenetics. Cold Spring Harb Perspect Biol, 7(1):a019364.

[Herrera-Solorio *et al.*, 2019] Herrera-Solorio, A. M., Vembar, S. S., MacPherson, C. R., Lozano-Amado, D., Meza, G. R., Xoconostle-Cazares, B., Martins, R. M., Chen, P., Vargas, M., Scherf, A., & Hernández-Rivas, R. (2019). Clipped histone H3 is integrated into nucleosomes of DNA replication genes in the human malaria parasite *Plasmodium falciparum*. *EMBO reports*, *20*(4), e46331.

[Hess *et al.*, 1995] Hess, F. I., Kilian, A., Söllner, W., Nothdurft, H. D., Pröll, S., and Löscher, T., 1995. *Plasmodium falciparum* and *Plasmodium berghei*: Effect of Magnesium on the Development of Parasitemia. Experimental Parasitology, 80, 196-193.

[Hiller *et al.*, 2004] Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estraño, C., and Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science, 306(5703):1934–1937.

[Hirota *et al.*, 2008] Hirota, K., Miyoshi,T., Kugou,K., Hoffman, C.S., Shibata,T., and Ohta, K. (2008). Stepwise chromatin remodelling by a cascade of transcription initiation of non- coding rnas. Nature, 456(7218):130–134.

[Hoeijmakers *et al.*, 2012] Hoeijmakers, W.A., Flueck,C.,Françoijs,K.,Smits,A.H.,Wetzel, J., Volz, J. C., Cowman, A. F., Voss, T. S., Stunnenberg, H. G., and Bártfai, R. (2012). Plasmodium falciparum centromeres display a unique epigenetic makeup and cluster prior to and during schizogony. Cell Microbiology, 14(9):1391–1401.

[Hoeijmakers *et al.*, 2013] Hoeijmakers, W. A., Salcedo-Amaya, A. M., Smits, A. H., Françoijs, K. J., Treeck, M., Gilberger, T. W., Stunnenberg, H. G., and Bártfai, R. (2013). H2a.z/h2b.z double-variant nucleosomes inhabit the at-rich promoter regions of the plasmodium falciparum genome. Molecular Microbiology, 87(5):1061–1073.

[Horrocks *et al.*, 2004] Horrocks, P., Pinches, R., Christodoulou, Z., Kyes, S. A., and New- bold, C. I. (2004). Variable var transition rates underlie antigenic variation in malaria. Proc Natl Acad Sci U S A, 101(30):11129–11134.

[Hyman *et al.*, 2002] Hyman, R. W., Fung, E., Conway, A., Kurdi, O., Mao, J., Miranda, M., Nakao, B., Rowley, D., Tamaki, T., Wang, F., and Davis, R. W. (2002). Sequence of plasmodium falciparum chromosome 12. Nature, 419(6906):534–537.

[Issar *et al.*, 2008] Issar, N., Roux, E., Mattei, D., and Scherf, A. (2008). Identification of a novel post-translational modification in plasmodium falciparum: protein sumoylation in different cellular compartments. Cell Microbiology, 10:1999–2011.

[Jeltsch *et al.*, 2017] Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, Nellen W, Schaefer M, Helm M. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. RNA Biol. 2017 Sep 2;14(9):1108-1123.

[Jiang *et al.*, 2013] Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., Yang, W., Turner, L., Lavstsen, T., Theander, T. G., Peng, W., Wei, G., Jing, Q., Wakabayashi, Y., Bansal, A., Luo, Y., Ribeiro, J. M., Scherf, A., Aravind, L., Zhu, J., Zhao, K., and Miller, L. H. (2013). Pfsetvs methylation of histone h3k36 represses virulence genes in plasmodium falciparum. Nature, 499(7457):223–227.

[Jones, 2012] Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet, 13(7):484–492.

[Josling and Llinás, 2015] Josling, G. A. and Llinás, M. (2015). Sexual development in plasmodium parasites: knowing when it's time to commit. Nat Rev Microbiology, 13(9):573–587.

[Kelly *et al.*, 2006] Kelly, J. M., McRobert, L., and Baker, D. A. (2006). Evidence on the chromosomal location of centromeric DNA in plasmodium falciparum from etoposide- mediated topoisomerase-ii cleavage. Proc Natl Acad Sci U S A, 103(17):6706–6711.

[Kensche *et al.*, 2016] Kensche, P. R., Hoeijmakers, W. A., Toenhake, C. G., Bras, M., Chappell, L., Berriman, M., and Bártfai, R. (2016). The nucleosome landscape of plasmodium falciparum

reveals chromatin architecture and dynamics of regulatory sequences. Nucleic Acids Res, 44(5):2110–2124.

[Khalil *et al.*, 2009] Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., and Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A, 106(28):11667–11672.

[Khattab and Meri, 2011] Khattab, A. and Meri, S. (2011). Exposure of the plasmodium falciparum clonally variant stevor proteins on the merozoite surface. Malar J, 10:58.

[Khattab *et al.*, 2008] Khattab, A., Bonow, I., Schreiber, N., Petter, M., Schmetz, C., and Klinkert, M. Q. (2008). Plasmodium falciparum variant stevor antigens are expressed in merozoites and possibly associated with erythrocyte invasion. Malar J, 7:137.

[Kim *et al.*, 2010] Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff- Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., and Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature, 465(7295):182–187.

[Kino *et al.*, 2010] Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). Noncoding rna gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal, 3(107):ra8.

[Kirkland *et al.*, 2013] Kirkland, J. G., Raab, J. R., and Kamakaka, R. T. (2013). Tfiiic bound DNA elements in nuclear organization and insulation. Biochim Biophys Acta, 1829(3- 4):418–424.

[Knapp *et al.*, 1991] Knapp, B., Nau, U., Hundt, E., and Küpper, H. A. (1991). Demonstration of alternative splicing of a pre-mrna expressed in the blood stage form of plasmodium falciparum. J Biol Chem, 266(11):7148–7154.

[Kornberg and Lorch, 1999] Kornberg, R. D. and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell, 98(3):285–294.

[Kornberg, 1974] Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science, 184(4139):868–871.

[Kouzarides, 2007] Kouzarides, T. (2007). Chromatin modifications and their function. Cell, 128(4):693–705.

[Kraemer *et al.*, 2007] Kraemer, S. M., Kyes, S. A., Aggarwal, G., Springer, A. L., Nelson, S. O., Christodoulou, Z., Smith, L. M., Wang, W., Levin, E., Newbold, C. I., Myler, P. J., and Smith, J. D. (2007). Patterns of gene recombination shape var gene repertoires in plasmodium falciparum: comparisons of geographically diverse isolates. BMC Genomics, 8:45.

[Kriek *et al.*, 2003] Kriek, N., Tilley, L., Horrocks, P., Pinches, R., Elford, B. C., Ferguson, D. J., Lingelbach, K., and Newbold, C. I. (2003). Characterization of the pathway for transport of the cytoadherence-mediating protein, pfemp1, to the host cell surface in malaria parasite-infected erythrocytes. Mol Microbiology, 50(4):1215–1227.

[Kriek *et al.*, 2003] Kriek, N., Tilley, L., Horrocks, P., Pinches, R., Elford, B. C., Ferguson, D. J., Lingelbach, K., and Newbold, C. I. 2003. Characterization of the pathway for transport of the cytoadherence-mediating protein, pfemp1, to the host cell surface in malaria parasite-infected erythrocytes, Mol Microbiology, 50(4):1215-1227.

[Kumar *et al.*, 2008] Kumar, K., Singal, A., Rizvi, M. M., and Chauhan, V. S. (2008). High mobility group box (hmgb) proteins of plasmodium falciparum: Dna binding proteins with pro-inflammatory activity. Parasitol Int, 57(2):150–157.

[Kyes *et al.*, 1999] Kyes, S. A., Rowe, J. A., Kriek, N., and Newbold, C. I. (1999). Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with plasmodium falciparum. Proc Natl Acad Sci U S A, 96(16):9333–9338.

[Kyes *et al.*, 2000] Kyes, S., Pinches, R., and Newbold, C. (2000). A simple rna analysis method shows var and rif multigene family expression patterns in plasmodium falciparum. Mol Biochem Parasitol, 105(2):311–315.

References

[Kyes *et al.*, 2003] Kyes S.A., Christodoulou Z., Raza A., Horrocks P., Pinches R., Rowe J.A., Newbold C.I. 2003. A well-conserved Plasmodium falciparum var gene shows an unusual stage-specific transcript pattern. Mol. Microbiology. 48:1339–1348.

[Lanzer *et al.*, 1992a] Lanzer, M., de Bruin, D., and Ravetch, J. (1992a). Transcription map- ping of a 100 kb locus of plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates. EMBO J, 11(5):1949–1955.

[Lanzer *et al.*, 1992b] Lanzer, M., de Bruin, D., and Ravetch, J. V. (1992b). A sequence element associated with the plasmodium falciparum kahrp gene is the site of developmentally regulated protein-DNA interactions. Nucleic Acids Res, 20(12):3051–3056.

[Larremore *et al.*, 2015] Larremore, D. B., Sundararaman, S. A., Liu, W., Proto, W. R., Clauset, A., Loy, D., Speede, S., Plenderleith, L. J., Sharp, P. M., Hahn, B. H., Rayner, J. C., and Buckee, C. O. (2015). Ape parasite origins of human malaria virulence genes. Nat Commun, 6:8368.

[Lavazec *et al.*, 2006] Lavazec, C., Sanyal, S., and Templeton, T. J. (2006). Hypervariability within the rifin, stevor and pfmc-2tm superfamilies in plasmodium falciparum. Nucleic Acids Res, 34(22):6696–6707.

[Lavazec *et al*., 2007] Lavazec, C., Sandals., and Templeton. (2007).Expression switching in the stevor and pfmc-2tm superfamilies in plasmodium falciparum. Mol Microbiology, 64(6):1621–1634.

[Lavstsen *et al.*, 2003] Lavstsen, T., Salanti, A., Jensen, A. T., Arnot, D. E., and Theander, T. G. (2003). Sub-grouping of plasmodium falciparum 3d7 var genes based on sequence analysis of coding and non-coding regions. Malar J, 2:27.

[Le Roch *et al.*, 2003] LeRoch, K.G., Zhou,Y., Blair,P.L., Grainger,M., Moch,J.K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., and Winzeler, E. A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Science, 301(5639):1503–1508.

[Lee *et al.*, 2004] Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D., and Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet, 36(8):900–905.

[Lee, 2009] Lee,J.T.(2009).Lessonsfromx-chromosomeinactivation:longncrnaasguides and tethers to the epigenome. Genes Dev, 23(16):1831–1842.

[Leech *et al.*, 1984a] Leech, J. H., Barnwell, J. W., Aikawa, M., Miller, L. H., and Howard, R. J. (1984a). Plasmodium falciparum malaria: association of knobs on the surface of infected erythrocytes with a histidine-rich protein and the erythrocyte skeleton. J Cell Biol, 98(4):1256–1264.

[Leech *et al.*, 1984b] Leech, J. H., Barnwell, J. W., Miller, L. H., and Howard, R. J. (1984b). Identification of a strain-specific malarial antigen exposed on the surface of plasmodium falciparum-infected erythrocytes. J Exp Med, 159(6):1567–1575.

[Lefevre *et al.*, 2008] Lefevre, P., Witham, J., Lacroix, C. E., Cockerill, P. N., and Bonifer, C. (2008). The lps-induced transcriptional upregulation of the chicken lysozyme locus involves ctcf eviction and noncoding rna transcription. Mol Cell, 32(1):129–139.

[Legrand *et al.*, 2019] Legrand, J.M.D., Chan, AL., La, H.M. *et al.* 2019. DDX5 plays essential transcriptional and post-transcriptional roles in the maintenance and function of spermatogonia. Nat Commun 10, 2278.

[Lemieux *et al.*, 2013] Lemieux, J. E., Kyes, S. A., Otto, T. D., Feller, A. I., Eastman, R. T., Pinches, R. A., Berriman, M., Su, X. Z., and Newbold, C. I. (2013). Genome-wide profiling of chromosome interactions in plasmodium falciparum characterizes nuclear architecture and reconfigurations associated with antigenic variation. Mol Microbiology, 90(3):519– 537.

[Li *et al.*, 2008] Li,F.,Sonbuchner,L.,Kyes,S.A.,Epp,C.,andDeitsch,K.W.(2008).Nuclear non-coding rnas are transcribed from the centromeres of plasmodium falciparum and are associated with centromeric chromatin. J Biol Chem, 283(9):5692–5698.

[Li *et al.*, 2016] Li,W., Notani,D., and Rosenfeld, M.G. (2016).Enhancers as non-coding rna transcription units: recent insights and future perspectives. Nat Rev Genet, 17(4):207–223.

References

[Lin *et al.*, 2003] Lin, S. P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A. C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the dlk1-gtl2 imprinted cluster on mouse chromosome 12. Nat Genet, 35(1):97–102.

[Lomvardas *et al.*, 2006] Lomvardas,S., Barnea,G., Pisapia,D.J., Mendelsohn,M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell, 126(2):403–413.

[López-Barragán *et al.*, 2011] López-Barragán, M.J., Lemieux, J., Quiñones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., and Su, X. Z. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of plasmodium falciparum. BMC Genomics, 12:587.

[López-Estraño *et al.*, 2007] López-Estraño, C., Gopalakrishnan, A. M., Semblat, J. P., Fergus, M. R., Mazier, D., and Haldar, K. (2007). An enhancer-like region regulates hrp3 promoter stage-specific gene expression in the human malaria parasite plasmodium falciparum. Biochim Biophys Acta, 1769(7-8):506–513.

[Lopez-Rubio *et al.*, 2007] Lopez-Rubio, J. J., Gontijo, A. M., Nunes, M. C., Issar, N., Hernandez Rivas, R., and Scherf, A. (2007). 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. Mol Microbiology, 66(6):1296–1305.

[Lopez-Rubio *et al.*, 2009] Lopez-Rubio, J. J., Mancio-Silva, L., and Scherf, A. (2009). Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe, 5(2):179–190.

[Lubell *et al.*, 2014] Lubell, Y., Dondorp, A., Guérin, P.J., Drake, T., Meek, S., Ashley, E., Day, N., White, N. J., and White, L. J. (2014). Artemisinin resistance–modelling the potential human and economic costs. Malar J, 13:452.

[Lunyak *et al.*, 2007] Lunyak, V.V., Prefontaine, G.G., Núñez, E., Cramer, T., Ju, B.G., Ohgi, K. A., Hutt, K., Roy, R., García-Díaz, A., Zhu, X., Yung, Y., Montoliu, L., Glass, C. K., and Rosenfeld, M. G. (2007). Developmentally regulated activation of a sine b2 repeat as a domain boundary in organogenesis. Science, 317(5835):248–251.

[Lyon, 1986] Lyon, M. F. (1986). X chromosomes and dosage compensation. Nature, 320(6060):313.

[Lyons *et al.*, 2013] Lyons, D. B., Allen, W. E., Goh, T., Tsai, L., Barnea, G., and Lomvardas, S. (2013). An epigenetic trap stabilizes singular olfactory receptor expression. Cell, 154(2):325–336.

[Ma *et al.*, 2013] Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding rnas. RNA Biol, 10(6):925–933.

[Magklara *et al.*, 2011] Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z. A., Kheradpour, P., Mountoufaris, G., Carey, C., Barnea, G., Kellis, M., and Lomvardas, S. (2011). An epigenetic signature for monoallelic olfactory receptor expression. Cell, 145(4):555–570.

[Maharana *et al.*, 2018] Maharana, S., Wang, J., Papadopoulos, D. K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T. M., Jahnel, M., Marrone, L., Chang, Y. T., Sterneckert, J., Tomancak, P., Hyman, A. A., and Alberti, S. (2018). Rna buffers the phase separation behavior of prion-like rna binding proteins. Science, 360(6391):918–921.

[Maier *et al.*, 2009] Maier, A. G., Cooke, B. M., Cowman, A. F., and Tilley, L. (2009). Malaria parasite proteins that remodel the host erythrocyte. Nat Rev Microbiology, 7(5):341–354.

[Mancio-Silva and Scherf, 2013] Mancio-Silva L., Scherf A. 2013. In situ fluorescence visualization of transcription sites and genomic loci in blood stages of plasmodium falciparum. Methods Mol. Biol. 923:335–351.

[Mancio-Silva *et al.*, 2010] Mancio-Silva, L., Zhang, Q., Scheidig-Benatar, C., and Scherf, A. (2010). Clustering of dispersed ribosomal DNA and its role in gene regulation and chromosome-end associations in malaria parasites. Proc Natl Acad Sci U S A, 107(34):15117–15122.

References

[Mandava *et al.*, 2008] Mandava, V., Janzen, C. J., and Cross, G. A. (2008). Trypanosome h2bv replaces h2b in nucleosomes enriched for h3 k4 and k76 trimethylation. Biochem Biophys Res Commun, 368(4):846–851.

[Mantel *et al.*, 2013] Mantel, P. Y., Hoang, A. N., Goldowitz, I., Potashnikova, D., Hamza, B., Vorobjev, I., Ghiran, I., Toner, M., Irimia, D., Ivanov, A. R., Barteneva, N., and Marti, M. (2013). Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system. Cell Host Microbe, 13(5):521–534.

[Mao *et al.*, 2011] Mao, Y. S., Sunwoo, H., Zhang, B., and Spector, D. L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding rnas. Nat Cell Biol, 13(1):95–101.

[Marck *et al.*, 2006] Marck C., Kachouri-Lafond R., Lafontaine I., Westhof E., Dujon B., Grosjean H. 2006. The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications. Nucleic Acids Res. 34:1816–1835.

[Mariner *et al.*, 2008] Mariner, P. D., Walters, R. D., Espinoza, C. A., Drullinger, L. F., Wagner, S. D., Kugel, J. F., and Goodrich, J. A. (2008). Human alu rna is a modular transacting repressor of mrna transcription during heat shock. Mol Cell, 29(4):499–509.

[Mariner *et al.*, 2008] Mariner, P. D., Walters, R. D., Espinoza, C. A., Drullinger, L. F., Wagner, S. D., Kugel, J. F., and Goodrich, J. A. 2008. Human alu rna is a modular transacting repressor of mrna transcription during heat shock. Mol Cell, 29(4):499-509.

[Markenscoff-Papadimitriou *et al.*, 2014] Markenscoff-Papadimitriou, E., Allen, W. E., Colquitt, B. M., Goh, T., Murphy, K. K., Monahan, K., Mosley, C., Ahituv, N., and Lomvardas, S. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. Cell, 159(3):543–557.

References

[Marti *et al.*, 2004] Marti, M., Good, R. T., Rug, M., Knuepfer, E., and Cowman, A. F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science, 306(5703):1930–1933.

[Martianov *et al.*, 2007] Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non- coding interfering transcript. Nature, 445(7128):666–670.

[Marty *et al.*, 2006] Marty, A.J., Thompson, J.K., Duffy, M.F., Voss, T.S., Cowman, A.F., and Crabb, B. S. (2006). Evidence that plasmodium falciparum chromosome end clusters are cross-linked by protein and are the sites of both virulence gene silencing and activation. Mol Microbiology, 62(1):72–83.

[Maston *et al.*, 2006] Maston, G. A., Evans, S. K., and Green, M. R. (2006). Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet, 7:29–59.

[McLean *et al.*, 1982] McLean, S. A., Pearson, C. D., and Phillips, R. S. (1982). Plasmodium chabaudi: antigenic variation during recrudescent parasitaemias in mice. Exp Parasitol, 54(3):296–302.

[Memczak *et al.*, 2013] Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F., and Rajewsky, N. (2013). Circular rnas are a large class of animal rnas with regulatory potency. Nature, 495(7441):333–338.

[Merkenschlager and Nora, 2016] Merkenschlager, M. and Nora, E. P. (2016). Ctcf and cohesin in genome folding and transcriptional gene regulation. Annu Rev Genomics Hum Genet, 17:17–43.

[Merrick and Duraisingh, 2007] Merrick, C. J. and Duraisingh, M. T. (2007). Plasmodium falciparum sir2: an unusual sirtuin with dual histone deacetylase and adp- ribosyltransferase activity. Eukaryot Cell, 6(11):2081–2091.

[Miao *et al.*, 2006] Miao, J., Fan, Q., Cui, L., Li, J., Li, J., and Cui, L. (2006). The malaria parasite plasmodium falciparum histones: organization, expression, and acetylation. Gene, 369:53–65.

[Miao *et al.*, 2010] Miao, J., Fan, Q., Cui, L., Li, X., Wang, H., Ning, G., Reese, J. C., and Cui, L. (2010). The myst family histone acetyltransferase regulates gene expression and cell cycle in malaria parasite plasmodium falciparum. Mol Microbiology, 78(4):883–902.

[Miller and Su, 2011] Miller, L. H. and Su, X. (2011). Artemisinin: discovery from the chinese herbal garden. Cell, 146(6):855–858.

[Miller *et al.*, 1994] Miller, L. H., Good, M. F., and Milon, G. (1994). Malaria pathogenesis. Science, 264(5167):1878–1883.

[Miller *et al.*, 2002] Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K. (2002). The pathogenic basis of malaria. Nature, 415(6872):673–679.

[Miller *et al.*, 2013] Miller, L., Ackerman, H., Su, Xz., Wellems, T. (2013). Malaria biology and disease pathogenesis: insights for new treatments. Nat Med 19:156–167.

[Mok *et al.*, 2008] Mok, B., Ribacke, U., Rasti, N., Kironde, F., Chen, Q., Nilsson, P., and Wahlgren, M. (2008). Default pathway of var2csa switching and translational repression in plasmodium falciparum. PLoS One, 3(4):e1982.

[Mourier *et al.*, 2008] Mourier, T., Carret, C., Kyes, S., Christodoulou, Z., Gardner, P. P., Jeffares, D. C., Pinches, R., Barrell, B., Berriman, M., Griffiths-Jones, S., Ivens, A., Newbold, C., and Pain, A. (2008). Genome-wide discovery and verification of novel structured rnas in plasmodium falciparum. Genome Res, 18(2):281–292.

[Muhia *et al.*, 2003] Muhia, D. K., Swales, C. A., Eckstein-Ludwig, U., Saran, S., Polley, S. D., Kelly, J. M., Schaap, P., Krishna, S., and Baker, D. A. (2003). Multiple splice variants encode a novel adenylyl cyclase of possible plastid origin expressed in the sexual stage of the malaria parasite plasmodium falciparum. J Biol Chem, 278(24):22014–22022.

[Müller *et al.*, 2018] Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A. E., Sebra, R. P., and Siegel, T. N. (2018). Genome organization and DNA accessibility control antigenic variation in trypanosomes. Nature, 563(7729):121–125.

[Mundwiler-Pachlatko and Beck, 2013] Mundwiler-Pachlatko, E. and Beck, H. P. (2013). Maurer's clefts, the enigma of plasmodium falciparum. Proc Natl Acad Sci U S A, 110(50):19987–19994.

[Munroe and Lazar, 1991] Munroe, S. H. and Lazar, M. A. (1991). Inhibition of cerba mrna splicing by a naturally occurring antisense rna. J Biol Chem, 266(33):22083–22086.

[Nacer *et al.*, 2015] Nacer, A., Claes, A., Roberts, A., Scheidig-Benatar, C., Sakamoto, H., Ghorbal, M., Lopez-Rubio, J. J., and Mattei, D. (2015). Discovery of a novel and conserved plasmodium falciparum exported protein that is important for adhesion of pfemp1 at the surface of infected erythrocytes. Cell Microbiology, 17(8):1205–1216.

[Nergadze *et al.*, 2009] Nergadze, S. G., Farnung, B. O., Wischnewski, H., Khoriauli, L., Vitelli, V., Chawla, R., Giulotto, E., and Azzalin, C. M. (2009). Cpg-island promoters drive transcription of human telomeres. RNA, 15(12):2186–2194.

[Ng *et al.*, 2018] Ng, C. S., Sinha, A., Aniweh, Y., Nah, Q., Babu, I., Gu, C., Chionh, Y. H., Dedon, P. C., and Preiser, P. R. (2018). trna epitranscriptomics and biased codon are linked to proteome expression in plasmodium falciparum. Mol Syst Biol, 14(10):e8009.

[Niang *et al.*, 2014] Niang, M., Bei, A. K., Madnani, K. G., Pelly, S., Dankwa, S., Kanjee, U., Gunalan, K., Amaladoss, A., Yeo, K. P., Bob, N. S., Malleret, B., Duraisingh, M. T., and Preiser, P. R. (2014). Stevor is a plasmodium falciparum erythrocyte binding protein that mediates merozoite invasion and rosetting. Cell Host Microbe, 16(1):81–93.

[Noble *et al.*, 2013] Noble R, Christodoulou Z, Kyes S, Pinches R, Newbold CI, Recker M. The antigenic switching network of Plasmodium falciparum and its implications for the immuno-epidemiology of malaria. Elife. 2013 Sep 17;2:e01074.

[Noma *et al.*, 2006] Noma, K., Cam, H. P., Maraia, R. J., and Grewal, S. I. (2006). A role for tfiiic transcription factor complex in genome organization. Cell, 125(5):859–872.

[Olins and Olins, 1974] Olins, A. L. and Olins, D. E. (1974). Spheroid chromatin units (v bodies). Science, 183(4122):330–332.

[Olivieri *et al.*, 2008] Olivieri, A., Silvestrini, F., Sanchez, M., and Alano, P. (2008). A 140-bp at-rich sequence mediates positive and negative transcriptional control of a plasmodium falciparum developmentally regulated promoter. Int J Parasitol, 38(3-4):299–312.

[Orioli *et al.*, 2012] Orioli, A., Pascali, C., Pagano, A., Teichmann, M., and Dieci, G. (2012). Rna polymerase iii transcription control elements: themes and variations. Gene, 493(2):185–194.

[Ørom *et al.*, 2010] Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., Guigo, R., and Shiekhattar, R. (2010). Long noncoding rnas with enhancer-like function in human cells. Cell, 143(1):46– 58.

[Osta *et al.*, 2002] Osta, M., Gannoun-Zaki, L., Bonnefoy, S., Roy, C., and Vial, H. J. (2002). A 24 bp cis-acting element essential for the transcriptional activity of plasmodium falciparum cdp-diacylglycerol synthase gene promoter. Mol Biochem Parasitol, 121(1):87–98.

[Otto *et al.*, 2010] Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., Newbold, C., and Llinás, M. (2010). New insights into the blood-stage transcriptome of plasmodium falciparum using rna-seq. Mol Microbiology, 76(1):12–24.

[Otto *et al.*, 2018a] Otto, T. D., Böhme, U., Sanders, M., Reid, A., Bruske, E. I., Duffy, C. W., Bull, P. C., Pearson, R. D., Abdi, A., Dimonte, S., Stewart, L. B., Campino, S., Kekre, M., Hamilton, W. L., Claessens, A., Volkman, S. K., Ndiaye, D., Amambua-Ngwa, A., Diakite, M., Fairhurst, R. M., Conway, D. J., Franck, M., Newbold, C. I., and Berriman, M. (2018a). Long read assemblies of geographically dispersed plasmodium falciparum isolates reveal highly structured subtelomeres. Wellcome Open Res, 3:52.

[Otto *et al.*, 2018b] Otto, T. D., Gilabert, A., Crellen, T., Böhme, U., Arnathau, C., Sanders, M., Oyola, S. O., Okouga, A. P., Boundenga, L., Willaume, E., Ngoubangoye, B., Moukodoum, N. D., Paupy, C., Durand, P., Rougeron, V., Ollomo, B., Renaud, F., Newbold, C., Berriman, M., and Prugnolle, F. (2018b). Genomes of all known members of a plasmodium subgenus reveal paths to virulent human malaria. Nat Microbiology, 3(6):687– 697.

[Paddon *et al.*, 2013] Paddon, C. J., Westfall, P., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S., and Newman, J. D. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. Nature, 496(7446):528–532.

[Pandey *et al.*, 2008] Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding rna mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell, 32(2):232–246.

[Panneerselvam *et al.*, 2011] Panneerselvam, P., Bawankar, P., Kulkarni, S., and Patankar, S. (2011). In silico prediction of evolutionarily conserved gc-rich elements associated with antigenic proteins of plasmodium falciparum. Evol Bioinform Online, 7:235–255.

[Patankar *et al.*, 2001] Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M., and Wirth, D. F. (2001). Serial analysis of gene expression in plasmodium falciparum reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. Mol Biol Cell, 12(10):3114–3125.

[Patel *et al.*, 2009] Patel,V.,Mazitschek,R.,Coleman,B.,Nguyen,C.,Urgaonkar,S.,Cortese, J., Barker, R. H., Greenberg, E., Tang, W., Bradner, J. E., Schreiber, S. L., Duraisingh, M. T., Wirth, D. F., and Clardy, J. (2009). Identification and characterization of small molecule inhibitors of a class i histone deacetylase from plasmodium falciparum. J Med Chem, 52(8):2185–2187.

References

[Pérez-Toledo *et al.*, 2009] Pérez-Toledo, K., Rojas-Meza, A. P., Mancio-Silva, L., Hernández-Cuevas, N. A., Delgadillo, D. M., Vargas, M., Martínez-Calvillo, S., Scherf, A., and Hernandez-Rivas, R. (2009). Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. Nucleic Acids Res, 37(8):2596–2606.

[Petter *et al.*, 2007] Petter, M., Haeggström, M., Khattab, A., Fernandez, V., Klinkert, M. Q., and Wahlgren, M. (2007). Variant proteins of the plasmodium falciparum rifin fam- ily show distinct subcellular localization and developmental expression patterns. Mol Biochem Parasitol, 156(1):51–61.

[Petter *et al.*, 2008] Petter, M., Bonow, I., and Klinkert, M. Q. (2008). Diverse expression patterns of subgroups of the rif multigene family during plasmodium falciparum gametocytogenesis. PLoS One, 3(11):e3779.

[Petter *et al.*, 2011] Petter, M., Lee, C. C., Byrne, T. J., Boysen, K. E., Volz, J., Ralph, S. A., Cowman, A. F., Brown, G. V., and Duffy, M. F. (2011). Expression of p. falciparum var genes involves exchange of the histone variant h2a.z at the promoter. PLoS Pathog, 7(2):e1001292.

[Petter *et al.*, 2013] Petter, M., Selvarajah, S. A., Lee, C. C., Chin, W. H., Gupta, A. P., Bozdech, Z., Brown, G. V., and Duffy, M. F. (2013). 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(6):1167–1182.

[PlasmoDB] PasmoDB.

[Poinar, 2005] Poinar, G. (2005). Plasmodium dominicana n. sp. (plasmodiidae: Haemosporida) from tertiary dominican amber. Syst Parasitol, 61(1):47–52.

[Policarpi *et al.*, 2017] Policarpi, C., Crepaldi, L., Brookes, E., Nitarska, J., French, S., Coatti, A., and Riccio, A. (2017). Enhancer sines link pol iii to pol ii transcription in neurons. Cell Rep, 21(10):2879–2894.

[Pollack *et al.*, 1982] Pollack, Y., Katzen, A. L., Spira, D. T., and Golenser, J. (1982). The genome of plasmodium falciparum. i: Dna base composition. Nucleic Acids Res, 10(2):539– 546.

[Pollack *et al.*, 1991] Pollack,Y.,Kogan,N.,andGolenser,J.(1991).Plasmodium falciparum: evidence for a dna methylation pattern. Exp Parasitol, 72(4):339–344.

[Ponting *et al.*, 2009] Ponting, C. P., Oliver, P. L., and Reik, W. (2009). Evolution and functions of long noncoding rnas. Cell, 136(4):629–641.

[Ponts *et al.*, 2010] Ponts, N., Harris, E. Y., Prudhomme, J., Wick, I., Eckhardt-Ludka, C., Hicks, G. R., Hardiman, G., Lonardi, S., and Le Roch, K. G. (2010). Nucleosome landscape and control of transcription in the human malaria parasite. Genome Res, 20(2):228–238.

[Ponts *et al.*, 2011] Ponts, N., Harris, E. Y., Lonardi, S., and Le Roch, K. G. (2011). Nucleosome occupancy at transcription start sites in the human malaria parasite: a hard-wired evolution of virulence. Infect Genet Evol, 11(4):716–724.

[Ponts *et al.*, 2013] Ponts, N., Fu, L., Harris, E. Y., Zhang, J., Chung, D. W., Cervantes, M. C., Prudhomme, J., Atanasova-Penichon, V., Zehraoui, E., Bunnik, E. M., Rodrigues, E. M., Lonardi, S., Hicks, G. R., Wang, Y., and Le Roch, K. G. (2013). Genome-wide mapping of dna methylation in the human malaria parasite plasmodium falciparum. Cell Host Microbe, 14(6):696–706.

[Poullet *et al.*, 2007] Poullet P., Carpentier S., and Barillot E. 2007. myProMS, a web server for management and validation of mass spectrometry-based proteomic data. Proteomics, 7: 2553-2556.

[Quinn and Chang, 2015] Quinn, J. J. and Chang, H. Y. (2015). Unique features of long non-coding rna biogenesis and function. Nat Rev Genet, 17(1):47–62.

[Raab *et al.*, 2012] Raab, J. R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P. A., Haussler, D., and Kamakaka, R. T. (2012). Human trna genes function as chromatin insulators. EMBO J, 31(2):330–350.

References

[Raabe *et al.*, 2010] Raabe, C. A., Sanchez, C. P., Randau, G., Robeck, T., Skryabin, B. V., Chinni, S. V., Kube, M., Reinhardt, R., Ng, G. H., Manickam, R., Kuryshev, V. Y., Lanzer, M., Brosius, J., Tang, T. H., and Rozhdestvensky, T. S. (2010). A global view of the nonprotein- coding transcriptome in plasmodium falciparum. Nucleic Acids Res, 38(2):608–617.

[Ralph *et al.*, 2005a] Ralph,S.A.,Bischoff,E.,Mattei,D.,Sismeiro,O.,Dillies,M.A.,Guigon, G., Coppee, J. Y., David, P. H., and Scherf, A. (2005a). Transcriptome analysis of antigenic variation in plasmodium falciparum–var silencing is not dependent on antisense rna. Genome Biol, 6(11):R93.

[Ralph *et al.*, 2005b] Ralph, S. A., Scheidig-Benatar, C., and Scherf, A. (2005b). Antigenic variation in plasmodium falciparum is associated with movement of var loci between subnuclear locations. Proc Natl Acad Sci U S A, 102(15):5414–5419.

[Rasmussen and Helin, 2016] Rasmussen, K. D. and Helin, K. (2016). Role of tet enzymes in dna methylation, development, and cancer. Genes Dev, 30(7):733–750.

[Raulet *et al.*, 1985] Raulet, D. H., Garman, R. D., Saito, H., and Tonegawa, S. (1985). Devel- opmental regulation of t-cell receptor gene expression. Nature, 314(6006):103–107.

[Recker *et al.*, 2011] Recker, M., Buckee, C. O., Serazin, A., Kyes, S., Pinches, R., Christodoulou, Z., Springer, A. L., Gupta, S., and Newbold, C. I. (2011). Antigenic variation in plasmodium falciparum malaria involves a highly structured switching pattern. PLoS Pathog, 7(3):e1001306.

[Regev-Rudzki *et al.*, 2013] Regev-Rudzki, N., Wilson, D. W., Carvalho, T. G., Sisquella, X., Coleman, B. M., Rug, M., Bursac, D., Angrisano, F., Gee, M., Hill, A. F., Baum, J., and Cowman, A. F. (2013). Cell-cell communication between malaria-infected red blood cells via exosome-like vesicles. Cell, 153(5):1120–1133.

[Reynolds *et al.*, 2016] Reynolds, D., Hofmeister, B. T., Cliffe, L., Alabady, M., Siegel, T. N., Schmitz, R. J., and Sabatini, R. (2016). Histone h3 variant regulates rna polymerase ii transcription termination and dual strand transcription of sirna loci in trypanosoma brucei. PLoS Genet, 12(1):e1005758.

[Riechmann and Meyerowitz, 1998] Riechmann, J. L. and Meyerowitz, E. M. (1998). The ap2/erebp family of plant transcription factors. Biol Chem, 379(6):633–646.

[Rinn and Chang, 2012] Rinn, J. L. and Chang, H. Y. (2012). Genome regulation by long noncoding rnas. Annu Rev Biochem, 81:145–166.

[Rovira-Graells *et al.*, 2012] Rovira-Graells,N.,Gupta,A.P.,Planet,E.,Crowley,V.M.,Mok, S., Ribas de Pouplana, L., Preiser, P. R., Bozdech, Z., and Cortés, A. (2012). Transcriptional variation in the malaria parasite plasmodium falciparum. Genome Res, 22(5):925–938.

[Rowley and Corces, 2018] Rowley,M.J.andCorces,V.G.(2018).Organizationalprinciples of 3d genome architecture. Nat Rev Genet, 19(12):789–800.

[Royo and Cavaillé, 2008] Royo, H. and Cavaillé, J. (2008). Non-coding rnas in imprinted gene clusters. Biol Cell, 100(3):149–166.

[RTSS Clinical Trials Partnership, 2015] RTSS Clinical Trials Partnership, R. (2015). Efficacy and safety of rts malaria vaccine with or without a booster dose in infants and children in africa: final results of a phase 3, individually randomised, controlled trial. Lancet, 386(9988):31–45.

[Ruiz *et al.*, 2018] Ruiz, J. L., Tena, J. J., Bancells, C., Cortés, A., Gómez-Skarmeta, J. L., and Gómez-Díaz, E. (2018). Characterization of the accessible genome in the human malaria parasite plasmodium falciparum. Nucleic Acids Res, 46(18):9414–9431.

[Ruvalcaba-Salazar *et al.*, 2005] Ruvalcaba-Salazar, O. K., del Carmen Ramírez-Estudillo, M., Montiel-Condado, D., Recillas-Targa, F., Vargas, M., and Hernández-Rivas, R. (2005). Recombinant and native plasmodium falciparum tata-binding-protein binds to a specific tata box element in promoter regions. Mol Biochem Parasitol, 140(2):183–196.

[Sachs and Malaney, 2002] Sachs,J. and Malaney, P. (2002).The economic and social burden of malaria. Nature, 415(6872):680–685.

[Saito *et al.*, 2017] Saito, F., Hirayasu, K., Satoh, T., Wang, C. W., Lusingu, J., Arimori, T., Shida, K., Palacpac, N. M. Q., Itagaki, S., Iwanaga, S., Takashima, E., Tsuboi, T., Kohyama, M., Suenaga,

T., Colonna, M., Takagi, J., Lavstsen, T., Horii, T., and Arase, H. (2017). Immune evasion of plasmodium falciparum by rifin via inhibitory receptors. Nature, 552(7683):101–105.

[Salcedo-Amaya *et al.*, 2009] Salcedo-Amaya, A. M., van Driel, M. A., Alako, B. T., Trelle, M. B., van den Elzen, A. M., Cohen, A. M., Janssen-Megens, E. M., van de Vegte-Bolmer, M., Selzer, R. R., Iniguez, A. L., Green, R. D., Sauerwein, R. W., Jensen, O. N., and Stunnenberg, H. G. (2009). Dynamic histone h3 epigenome marking during the intraerythrocytic cycle of plasmodium falciparum. Proc Natl Acad Sci U S A, 106(24):9655–9660.

[Salmena *et al.*, 2011] Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A cerna hypothesis: the rosetta stone of a hidden rna language. Cell, 146(3):353–358.

[Sam-Yellowe *et al.*, 2004] Sam-Yellowe, T. Y., Florens, L., Johnson, J. R., Wang, T., Drazba, J. A., Le Roch, K. G., Zhou, Y., Batalov, S., Carucci, D. J., Winzeler, E. A., and Yates, J. R. (2004). A plasmodium gene family encoding maurer's cleft membrane proteins: structural properties and expression profiling. Genome Res, 14(6):1052–1059.

[Santos *et al.*, 2017] Santos JM, Josling G, Ross P, Joshi P, Orchard L, Campbell T, Schieler A, Cristea IM, Llinás M. Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I Transcription Factor. Cell Host Microbe. 2017 Jun 14;21(6):731-741.e10

[Santos-Rosa *et al.*, 2009] Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T., Saksouk, N., Cote, J., and Kouzarides, T. (2009). Histone h3 tail clipping regulates gene expression. Nat Struct Mol Biol, 16(1):17–22.

[Saraf *et al.*, 2016] Saraf, A., Cervantes, S., Bunnik, E. M., Ponts, N., Sardiu, M. E., Chung, D. W., Prudhomme, J., Varberg, J. M., Wen, Z., Washburn, M. P., Florens, L., and Le Roch, K. G. (2016). Dynamic and combinatorial landscape of histone modifications during the intraerythrocytic developmental cycle of the malaria parasite. J Proteome Res, 15(8):2787– 2801.

[Sasaki *et al.*, 2009] Sasaki, Y. T., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). Menepsilon/beta noncoding rnas are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci U S A, 106(8):2525–2530.

[Scherf *et al.*, 1998] Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., and Lanzer, M. (1998). Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in plasmodium falciparum. EMBO J, 17(18):5418–5426.

[Scherf *et al.*, 2001] Scherf, A., Figueiredo, L. M., and Freitas-Junior, L. H. (2001). Plasmodium telomeres: a pathogen's perspective. Curr Opin Microbiol, 4:409–414.

[Scherf *et al.*, 2008] Scherf, A., Lopez-Rubio, J. J., and Riviere, L. (2008). Antigenic variation in plasmodium falciparum. Annu Rev Microbiol, 62:445–470.

[Scherf *et al.*, 2017] Scherf, A., Malquist, N. A., Martins, R. M., Vembar, S. S., and Lopez- Rubio, J. J. (2017). Gene regulation: New insights and possible intervention strategies. Advances in Malaria Research, ed. D Gaur, CE Chitnis, VS Chauhan. Wiley Blackwell.

[Schieck *et al.*, 2007] Schieck, E., Pfahler, J. M., Sanchez, C. P., and Lanzer, M. (2007). Nu- clear run-on analysis of var gene expression in plasmodium falciparum. Mol Biochem Parasitol, 153(2):207–212.

[Schneider *et al.*, 2004] Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone h3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol, 6(1):73–77.

[Schoeftner and Blasco, 2008] Schoeftner, S. and Blasco, M. A. (2008).Developmentallyregulated transcription of mammalian telomeres by dna-dependent rna polymerase ii. Nat Cell Biol, 10(2):228–236.

[Schramm *et al.*, 2002] Schramm L., Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 16:2593–2620.

[Schulz *et al.*, 2016] Schulz, D., Zaringhalam, M., Papavasiliou, F. N., and Kim, H. S. (2016). Base j and h3.v regulate transcriptional termination in trypanosoma brucei. PLoS Genet, 12(1):e1005762.

[Schwartz, 2016] Schwartz, S. (2016). Cracking the epitranscriptome. RNA, 22(2):169–174.

[Scott *et al.*, 2006] Scott, K. C., Merrett, S. L., and Willard, H. F. (2006). A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. Curr Biol, 16(2):119–129.

[Selsing, 2006] Selsing, E. (2006). Ig class switching: targeting the recombinational mecha- nism. Curr Opin Immunol, 18(3):249–254.

[Serizawa *et al.*, 2003] Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., and Sakano, H. (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. Science, 302(5653):2088–2094.

[Shamovsky *et al.*, 2006] Shamovsky, I., Ivannikov, M., Kandel, E. S., Gershon, D., and Nudler, E. (2006). Rna-mediated response to heat shock in mammalian cells. Nature, 440(7083):556–560.

[Sharma and Shukla, 2017] Sharma, L. and Shukla, G. (2017). Placental malaria: A new insight into the pathophysiology. Front Med (Lausanne), 4:117.

[Sharma *et al.*, 2017] Sharma, L., and Shukla, G. 2017. Placental malaria: A new insight into the pathophysiology. Front Med (Lausanne), 4:117.

[Shevtsov and Dundr, 2011] Shevtsov, S. P. and Dundr, M. (2011). Nucleation of nuclear bodies by rna. Nat Cell Biol, 13(2):167–173.

[Shi *et al.*, 2017] Shi, D. Q., Ali, I., Tang, J., and Yang, W. C. (2017). New insights into 5hmc dna modification: Generation, distribution and function. Front Genet, 8:100.

[Shykind *et al.*, 2004] Shykind, B. M., Rohani, S. C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., Axel, R., and Barnea, G. (2004). Gene switching and the stability of odorant receptor gene choice. Cell, 117(6):801–815.

[Siegel *et al.*, 2009] Siegel, T. N., Hekstra, D. R., Kemp, L. E., Figueiredo, L. M., Lowell, J. E., Fenyo, D., Wang, X., Dewell, S., and Cross, G. A. (2009). Four histone variants mark the

boundaries of polycistronic transcription units in trypanosoma brucei. Genes Dev, 23(9):1063–1076.

[Siegel *et al.*, 2014] Siegel, T. N., Hon, C. C., Zhang, Q., Lopez-Rubio, J. J., Scheidig-Benatar, C., Martins, R. M., Sismeiro, O., Coppée, J. Y., and Scherf, A. (2014). Strand-specific rna- seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in plasmodium falciparum. BMC Genomics, 15:150.

[Sierra-Miranda *et al.*, 2012] Sierra-Miranda, M., Delgadillo, D. M., Mancio-Silva, L., Vargas, M., Villegas-Sepulveda, N., Martínez-Calvillo, S., Scherf, A., and Hernandez-Rivas, R. (2012). Two long non-coding rnas generated from subtelomeric regions accumulate in a novel perinuclear compartment in plasmodium falciparum. Mol Biochem Parasitol, 185(1):36–47.

[Singh *et al.*, 2004] Singh, N., Preiser, P., Rénia, L., Balu, B., Barnwell, J., Blair, P., Jarra, W., Voza, T., Landau, I., and Adams, J. H. (2004). Conservation and developmental control of alternative splicing in maebl among malaria parasites. J Mol Biol, 343(3):589–599.

[Skog *et al.*, 2008] Skog, J., Würdinger, T., van Rijn, S., Meijer, D. H., Gainche, L., Sena- Esteves, M., Curry, W. T., Carter, B. S., Krichevsky, A. M., and Breakefield, X. O. (2008). Glioblastoma microvesicles transport rna and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol, 10(12):1470–1476.

[Sleutelsetal.,2002] Sleutels, F., Zwart, R., and Barlow, D.P. (2002).The non-coding air rna is required for silencing autosomal imprinted genes. Nature, 415(6873):810–813.

[Smith *et al.*, 1995] Smith, J. D., Chitnis, C. E., Craig, A. G., Roberts, D. J., Hudson-Taylor, D. E., Peterson, D. S., Pinches, R., Newbold, C. I., and Miller, L. H. (1995). Switches in expression of plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell, 82(1):101–110.

[Sorber *et al.*, 2011] Sorber K, Dimon MT, DeRisi JL. RNA-Seq analysis of splicing in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. Nucleic Acids Res. 2011 May;39(9):3820-35.

References

[Sorber *et al.*, 2011] Sorber, K., Dimon, M. T., and DeRisi, J. L. (2011). Rna-seq analysis of splicing in plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. Nucleic Acids Res, 39(9):3820–3835.

[Spitale *et al.*, 2011] Spitale, R. C., Tsai, M. C., and Chang, H. Y. (2011). Rna templating the epigenome: long noncoding rnas as molecular scaffolds. Epigenetics, 6(5):539–543.

[Stanne and Rudenko, 2010] Stanne, T. M. and Rudenko, G. (2010). Active vsg expression sites in trypanosoma brucei are depleted of nucleosomes. Eukaryot Cell, 9(1):136–147.

[Straimer *et al.*, 2015] Straimer, J., Gnädig, N. F., Witkowski, B., Amaratunga, C., Duru, V., Ramadani, A. P., Dacheux, M., Khim, N., Zhang, L., Lam, S., Gregory, P. D., Urnov, F. D., Mercereau-Puijalon, O., Benoit-Vical, F., Fairhurst, R. M., Ménard, D., and Fidock, D. A. (2015). Drug resistance. k13-propeller mutations confer artemisinin resistance in plas- modium falciparum clinical isolates. Science, 347(6220):428–431.

[Struhl and Segal, 2013] Struhl, K. and Segal, E. (2013). Determinants of nucleosome positioning. Nat Struct Mol Biol, 20(3):267–273.

[Su *et al.*, 1995] Su, X. Z., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. A., and Wellems, T. E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of plasmodium falciparum-infected erythrocytes. Cell, 82(1):89–100.

[Sullivan *et al.*, 2006] Sullivan, W. J., Naguleswaran, A., and Angel, S. O. (2006). Histones and histone modifications in protozoan parasites. Cell Microbiol, 8(12):1850–1861.

[Sun *et al.*, 2013] Sun, S., Del Rosario, B. C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J. T. (2013). Jpx rna activates xist by evicting ctcf. Cell, 153(7):1537–1551.

[Sunwoo *et al.*, 2009] Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S., and Spector, D. L. (2009). Men epsilon/beta nuclear-retained non-coding rnas are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res, 19(3):347–359.

References

[Swamy *et al.*, 2011] Swamy, L., Amulic, B., and Deitsch, K. W. (2011). Plasmodium falciparum var gene silencing is determined by cis dna elements that form stable and heritable interactions. Eukaryot Cell, 10(4):530–539.

[Toenhake *et al.*, 2018] 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(4):557–569.e9.

[Toenhake *et al.*, 2019] *Toenhake* CG, Bártfai R. What functional genomics has taught us about transcriptional regulation in malaria parasites. Brief Funct Genomics. 2019 Sep 24;18(5):290-301.

[Tonkin *et al.*, 2009] Tonkin, C. J., Carret, C. K., Duraisingh, M. T., Voss, T. S., Ralph, S. A., Hommel, M., Duffy, M. F., Mancio-Silva, L., Scherf, A., Ivens, A., Speed, T. P., Beeson, J. G., and Cowman, A. F. (2009). Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in plasmodium falciparum. PLoS Biol, 7(4):e84.

[Treeck *et al.*, 2011] Treeck, M., Sanders, J. L., Elias, J. E., and Boothroyd, J. C. (2011). The phosphoproteomes of plasmodium falciparum and toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. Cell Host Microbe, 10(4):410– 419.

[Trelle *et al.*, 2009] Trelle, M. B., Salcedo-Amaya, A. M., Cohen, A. M., Stunnenberg, H. G., and Jensen, O. N. (2009). Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite plasmodium falciparum. J Proteome Res, 8(7):3439–3450.

[Tsai *et al.*, 2010] Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E., and Chang, H. Y. (2010). Long noncoding rna as modular scaffold of histone modification complexes. Science, 329(5992):689–693.

[Tu, 2011] Tu, Y. (2011). The discovery of artemisinin (qinghaosu) and gifts from chinese medicine. Nat Med, 17(10):1217–1220.

[Tun *et al.*, 2015] Tun, K. M., Imwong, M., Lwin, K. M., Win, A. A., Hlaing, T. M., Hlaing, T., Lin, K., Kyaw, M. P., Plewes, K., Faiz, M. A., Dhorda, M., Cheah, P. Y., Pukrittayakamee, S., Ashley, E. A., Anderson, T. J., Nair, S., McDew-White, M., Flegg, J. A., Grist, E. P., Guerin, P., Maude, R. J., Smithuis, F., Dondorp, A. M., Day, N. P., Nosten, F., White, N. J., and Woodrow, C. J. (2015). Spread of artemisinin-resistant plasmodium falciparum in myanmar: a cross-sectional survey of the k13 molecular marker. Lancet Infect Dis, 15(4):415–421.

[Upadhyay *et al.*, 2005] Upadhyay, R., Bawankar, P., Malhotra, D., and Patankar, S. (2005). A screen for conserved sequences with biased base composition identifies noncoding rnas in the a-t rich genome of plasmodium falciparum. Mol Biochem Parasitol, 144(2):149–158.

[Valot *et al.*, 2011] Valot, B., Langella, O., Nano, E. and Zivy, M. 2011. MassChroQ: a versatile tool for mass spectrometry quantification. Proteomics, (11)3572–3577.

[van Dooren and Striepen, 2013] van Dooren, G. G. and Striepen, B. (2013). The algal past and parasite present of the apicoplast. Annu Rev Microbiol, 67:271–289.

[van Dooren *et al.*, 2002] van Dooren, G. G., Su, V., D'Ombrain, M. C., and McFadden, G. I. (2002). Processing of an apicoplast leader sequence in plasmodium falciparum and the identification of a putative leader cleavage enzyme. J Biol Chem, 277(26):23612–23619.

[Vannini and Cramer, 2012] Vannini, A. and Cramer, P. (2012). Conservation between the rna polymerase i, ii, and iii transcription initiation machineries. Mol Cell, 45(4):439–446.

[Vembar *et al.*, 2014] Vembar, S. S., Scherf, A., and Siegel, T. N. (2014). Noncoding rnas as emerging regulators of plasmodium falciparum virulence gene expression. Curr Opin Microbiol, 20:153–161.

[Vembar *et al.*, 2015] Vembar, S. S., Macpherson, C. R., Sismeiro, O., Coppée, J. Y., and Scherf, A. (2015). The pfalba1 rna-binding protein is an important regulator of translational timing in plasmodium falciparum blood stages. Genome Biol, 16(1):212.

[Vembar *et al.*, 2016] Vembar, S. S., Droll, D., and Scherf, A. (2016). Translational regulation in blood stages of the malaria parasite plasmodium spp.: systems-wide studies pave the way. Wiley Interdiscip Rev RNA, 7(6):772–792.

[Verma and Surolia, 2014] Verma, G. and Surolia, N. (2014). The dimerization domain of pfcenp-c is required for its functions as a centromere protein in human malaria parasite plasmodium falciparum. Malar J, 13:475.

[Volz *et al.*, 2012] Volz, J. C., Bártfai, R., Petter, M., Langer, C., Josling, G. A., Tsuboi, T., Schwach, F., Baum, J., Rayner, J. C., Stunnenberg, H. G., Duffy, M. F., and Cowman, A. F. (2012). Pfset10, a plasmodium falciparum methyltransferase, maintains the active var gene in a poised state during parasite division. Cell Host Microbe, 11(1):7–18.

[Voss *et al.*, 2003] Voss, T. S., Kaestli, M., Vogel, D., Bopp, S., and Beck, H. P. (2003). Identification of nuclear proteins that interact differentially with plasmodium falciparum var gene promoters. Mol Microbiol, 48(6):1593–1607.

[Voss *et al.*, 2006] Voss, T. S., Healer, J., Marty, A. J., Duffy, M. F., Thompson, J. K., Bee- son, J. G., Reeder, J. C., Crabb, B. S., and Cowman, A. F. (2006). A var gene promoter controls allelic exclusion of virulence genes in plasmodium falciparum malaria. Nature, 439(7079):1004–1008.

[Wahlgren *et al.*, 1992] Wahlgren, M., Carlson, J., Helmby, H., Hedlund, I., and Treutiger, C. J. (1992). Molecular mechanisms and biological importance of plasmodium falciparum erythrocyte rosetting. Mem Inst Oswaldo Cruz, 87 Suppl 3:323–329.

[Wahlgren *et al.*, 2017] Wahlgren M, Goel S, Akhouri RR. (2017). Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. Nat Rev Microbiol. 15(8):479-491.

[Wahlgren *et al.*, 2017] Wahlgren, M., Goel, S., and Akhouri, R. R. (2017). Variant surface antigens of plasmodium falciparum and their roles in severe malaria. Nat Rev Microbiol, 15(8):479–491.

[Wang *et al.*, 1998] Wang, F., Nemes, A., Mendelsohn, M., and Axel, R. (1998). Odorant receptors govern the formation of a precise topographic map. Cell, 93(1):47–60.

References

[Wang *et al.*, 2010] Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., Chen, N., Sun, F., and Fan, Q. (2010). Creb up-regulates long non-coding rna, hulc expression through interaction with microrna-372 in liver cancer. Nucleic Acids Res, 38(16):5366–5383.

[Wang *et al.*, 2011] Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M. U., Ohgi, K. A., Glass, C. K., Rosenfeld, M. G., and Fu, X. D. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by erna. Nature, 474(7351):390–394.

[Wei *et al.*, 2014] Wei C., Xiao T., Zhang P., Wang Z., Chen X., Zhang L., Yao M., Chen R., Wang H. 2014. Deep profiling of the novel intermediate-size noncoding RNAs in intraerythrocytic Plasmodium falciparum. PLoS One. 9:e92946.

[Wei *et al.*, 2015] Wei, G, Zhao, Y, Zhang, Q, Pan, W. 2015. Dual regulatory effects of non-coding GC-rich elements on the expression of virulence genes in malaria parasites. Infect Genet Evol 36:490–499.

[Weiner *et al.*, 2011] Weiner,A.,Dahan-Pasternak,N.,Shimoni,E.,Shinder,V.,vonHuth,P., Elbaum, M., and Dzikowski, R. (2011). 3d nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite plasmodium falciparum. Cell Microbiol, 13(7):967–977.

[Westenberger *et al.*, 2009] Westenberger, S., Cui, L., Dharia, N., Winzeler, E., and Cui, L. (2009). 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, 10:610.

[White and Suvorova, 2018] White, M. W. and Suvorova, E. S. (2018). Apicomplexa cell cycles: Something old, borrowed, lost, and new: (trends in parasitology 34, 759-771; 2018). Trends Parasitol, 34(11):1012–1013.

[White *et al.*, 2018] White, M. W., and Suvorova, E. S., 2018. Apicomplex cell cycles: Something old, borrowed, lost, and new. Trends Parasitol, 34(11):1012-1013.

References

[WHO 2021] WHO. 2021. World Malaria Report 2021

[Williams *et al.*, 2010] Williams, A., Spilianakis, C. G., and Flavell, R. A. (2010). Interchromosomal association and gene regulation in trans. Trends Genet, 26(4):188–197.

[Williamson *et al.*, 2011] Williamson, C. M., Ball, S. T., Dawson, C., Mehta, S., Beechey, C. V., Fray, M., Teboul, L., Dear, T. N., Kelsey, G., and Peters, J. (2011). Uncoupling antisense-mediated silencing and dna methylation in the imprinted gnas cluster. PLoS Genet, 7(3):e1001347.

[Wu *et al.*, 2008] Wu, J., Sieglaff, D. H., Gervin, J., and Xie, X. S. (2008). Discovering regu- latory motifs in the plasmodium genome using comparative genomics. Bioinformatics, 24(17):1843–1849.

[Xue *et al.*,2008] Xue,X.,Zhang,Q.,Huang,Y.,Feng,L., and Pan, W.(2008). No mi rna were found in plasmodium and the ones identified in erythrocytes could not be correlated with infection. Malar J, 7:47.

[Yamazaki *et al.*, 2018] Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y. S., Fox, A. H., Bond, C. S., Nakagawa, S., Pierron, G., and Hirose, T. (2018). Functional domains of neat1 architectural lncrna induce paraspeckle assembly through phase separation. Mol Cell, 70(6):1038–1053.e7.

[Yang and Kuroda, 2007] Yang P.K., Kuroda M.I. 2007. Noncoding RNAs and intranuclear positioning in monoallelic gene expression. Cell. 128:777–786.

[Yang *et al.*, 2011] Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., Dorestein, P. C., and Rosenfeld, M. G. 2011. ncrna-and pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. Cell, 147(4):773-788.

[Yap *et al.*, 2010] Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M. J., and Zhou, M. M. (2010). Molecular interplay of the noncoding rna anril and methylated histone h3 lysine 27 by polycomb cbx7 in transcriptional silencing of ink4a. Mol Cell, 38(5):662–674.

[Yeoh *et al.*, 2019] Yeoh, L. M., Lee, V. V., McFadden, G. I., and Ralph, S. A. (2019). Alternative splicing in apicomplexan parasites. MBio, 10(1).

[Young *et al.*, 2005] Young, J. A., Fivelman, Q. L., Blair, P. L., de la Vega, P., Le Roch, K. G., Zhou, Y., Carucci, D. J., Baker, D. A., and Winzeler, E. A. (2005). The plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. Mol Biochem Parasitol, 143(1):67–79.

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

[Zanghì *et al.*, 2018] Zanghì, G., Vembar, S. S., Baumgarten, S., Ding, S., Guizetti, J., Bryant, J. M., Mattei, D., Jensen, A. T. R., Rénia, L., Goh, Y. S., Sauerwein, R., Hermsen, C. C., Franetich, J. F., Bordessoulles, M., Silvie, O., Soulard, V., Scatton, O., Chen, P., Mecheri, S., Mazier, D., and Scherf, A. (2018). A specific pfemp1 is expressed in p. falciparum sporozoites and plays a role in hepatocyte infection. Cell Rep, 22(11):2951–2963.

[Zhang and Tycko, 1992] Zhang, Y. and Tycko, B. (1992). Monoallelic expression of the human h19 gene. Nat Genet, 1:40–44.

[Zhang *et al.*, 2011] Zhang, Q., Huang, Y., Zhang, Y., Fang, X., Claes, A., Duchateau, M., Namane, A., Lopez-Rubio, J. J., Pan, W., and Scherf, A. (2011). A critical role of perinuclear filamentous actin in spatial repositioning and mutually exclusive expression of virulence genes in malaria parasites. Cell Host Microbe, 10(5):451–463.

[Zhang *et al.*, 2014] Zhang, Q., Siegel, T. N., Martins, R. M., Wang, F., Cao, J., Gao, Q., Cheng, X., Jiang, L., Hon, C. C., Scheidig-Benatar, C., Sakamoto, H., Turner, L., Jensen, A. T., Claes, A., Guizetti, J., Malmquist, N. A., and Scherf, A. (2014). Exonuclease- mediated degradation of nascent rna silences genes linked to severe malaria. Nature, 513(7518):431–435.

[Zhao *et al.*, 2019] Zhao, X., Cai, Y., & Xu, J. (2019). Circular RNAs: Biogenesis, Mechanism, and Function in Human Cancers. International journal of molecular sciences, 20(16), 3926.

References

[Zhu *et al.*, 2018] Zhu, L., Tripathi, J., Rocamora, F. M., Miotto, O., van der Pluijm, R., Voss, T. S., Mok, S., Kwiatkowski, D. P., Nosten, F., Day, N. P. J., White, N. J., Dondorp, A. M., Bozdech, Z., and Tracking Resistance to Artemisinin Collaboration, I. (2018). The origins of malaria artemisinin resistance defined by a genetic and transcriptomic background. Nat Commun, 9(1):5158.

[Zuccala and Baum, 2011] Zuccala, E. S. and Baum, J. (2011). Cytoskeletal and membrane remodelling during malaria parasite invasion of the human erythrocyte. Br J Haematol, 154(6):680–689.

References