Clonally variant non-coding RNA family and its role in Plasmodium falciparum antigenic variation

Anna Barcons Simon

Thèse de doctorat en parasitologie moléculaire de Sorbonne Université

Ecole Doctorale Complexité du Vivant (ED515)

Unité de Biologie des Interactions Hôte-Parasite

présentée par

Anna Barcons Simon

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

Clonally variant non-coding RNA family and its role in *Plasmodium falciparum* antigenic variation

Membres du jury :

Dr. Antonin Morillon	Président du jury
Pr. Nicolai Siegel	Rapporteur
Pr. Shruthi S. Vembar	Rapporteur
Dr. Lucy Glover	Examinateur
Dr. Julien Guizetti	Examinateur
Pr. Artur Scherf	Directeur de thèse

Présentée et soutenue publiquement à Paris, le 25 Octobre 2019

Abstract

Antigenic variation is an immune evasion mechanism used by the malaria parasite *Plasmodium falciparum* to establish prolonged infection. It prevents parasite clearance by switching the display of variant surface antigens encoded by the 60-member *var* gene family. Blood stage parasites express *var* genes in a mutually exclusive manner, with a single *var* active at a given time, ensuring that the immune system remains naive to a parasite sub-population. Despite multiple factors known to be involved in default *var* gene silencing, such as histone modifications and nuclear architecture, a specific activation factor has not been identified yet. In this thesis work, a polymerase III-transcribed family of GC-rich ncRNA was explored as a candidate regulatory factor of *var* gene activation. The GC-rich ncRNA family consists of 15 homologous members positioned adjacent to chromosome-central *var* genes. RNA FISH analysis revealed that the GC-rich ncRNA is targeted to the *var* expression site *in trans*. Furthermore, overexpression of one GC-rich ncRNA member overrules singular *var* expression and induces transcription of a specific subset of *var* genes. We developed a dead Cas9 system aiming to target transcription of all GC-rich members. Strikingly, transcriptional blocking of all GC-rich members by CRISPR interference results in down-regulation of *var* transcription to background levels. Our data support a role of this ncRNA in the activation process of *var* genes. Moreover, we show that the GC-rich ncRNA gene family is clonally variant, indicating that this may determine switch rates and switch order of the *var* gene family. We developed a robust ChIRP (chromatin isolation by RNA purification) protocol that allowed us to identify potential chromatin binding regions and interacting partners of this ncRNA family. The validation of several promising candidate proteins obtained my mass spectrometry is currently ongoing. Additionally, we investigated whether the GC-rich genes display long-range interactions using a chromosome conformation capture technique. We obtained evidence that the gene loci of these ncRNAs display long-range interactions between them and can thus play a role in nuclear spatial foci organization. This

study provides the first functional link between Pol III and Pol II transcription in the process of mutually exclusive expression of virulence genes. Moreover, exploring this ncRNA is a key step in unveiling the process of immune evasion and pathogenesis of *P. falciparum*.

Résumé

La variation antigénique est un mécanisme d'évasion immunitaire utilisé par le parasite du paludisme *Plasmodium falciparum* pour établir une infection prolongée. Il empêche la clairance parasitaire en modifiant l'expression des antigènes variables de surface, codés par la famille de gènes appelés «*var*», comprenant 60 membres. Au stade sanguin, les parasites expriment les gènes *var* d'une manière mutuellement exclusive, avec un seul gène *var* actif à un moment donné, permettant ainsi au parasite d'échapper au système immunitaire qui reste naïf envers une sous-population de parasites. Malgré les multiples facteurs connus pour être impliqués dans l'inactivation des gènes *var* par défaut, tels que les modifications des histones et l'architecture nucléaire, un facteur d'activation spécifique n'a pas encore été identifié. Dans ce travail de thèse, une famille d'ARNnc riche en GC, transcrite par l'ARN polymérase III, a été explorée comme facteur régulateur candidat à l'activation des gènes *var*. La famille des ARNnc riches en GC se compose de 15 membres homologues, positionnés de façon adjacente aux gènes *var* situés au centre du chromosome. L'analyse d'expériences de FISH de l'ARN a révélé que l'ARNnc riche en GC cible le site d'expression du gène *var en trans*. De plus, la surexpression d'un membre des ARNnc riches en GC surpasse l'expression mutuellement exclusive des gènes *var* et induit la transcription d'un sous-ensemble spécifique de gènes *var*. Nous avons développé un système de dead Cas9 ciblant la transcription de tous les membres riches en GC. Il est frappant de constater que le blocage transcriptionnel de tous les membres riches en GC par interférence CRISPR entraîne une régulation à la baisse de la transcription des gènes *var* à un niveau basal. Nos données confirment le rôle de cet ARNnc dans le processus d'activation des gènes *var*. De plus, nous montrons que la famille de gènes de l'ARNnc riche en GC est une variante clonale, ce qui indique que cela peut déterminer les taux de commutation et l'ordre de commutation de la famille de gènes *var*. Nous avons développé un protocole ChIRP (chromatin isolation by RNA purification) robuste qui nous a permis d'identifier des régions potentielles de liaison à la chromatine et

des partenaires interagissant avec cette famille d'ARNnc. La validation de plusieurs protéines candidates prometteuses obtenues par spectrométrie de masse est en cours. De plus, nous avons cherché à savoir si les loci géniques de ces ARNnc présentent des interactions à longue distance, à l'aide d'une technique de capture de conformation chromosomique. Nous avons obtenu la preuve que les loci géniques de ces ARNnc présentent des interactions à longue distance entre eux et peuvent donc jouer un rôle dans l'organisation spatiale des foci nucléaires. Cette étude fournit le premier lien fonctionnel entre la transcription de l'ARN Pol III et Pol II dans le processus d'expression mutuellement exclusive des gènes de virulence. De plus, l'exploration de cet ARNnc est une étape clé dans le dévoilement du processus d'évasion immunitaire et de la pathogenèse de *P. falciparum*.

Acknowledgments

In the first place, I am extremely grateful to my supervisor Artur Scherf for providing me the opportunity to do my PhD in his unit and supporting me during the process. I thank Artur for allowing me to develop as a scientist and for pushing me to always look at the bigger picture of my research. I really appreciate the independence and trust he has given me and the fact his door has always been open to discuss. I am thankful for his encouragement when experiments didn't work and for allowing me to present my work in international conferences.

I am also really thankful to Julien, who was my mentor in the lab during the early days of my PhD and managed to transmit me his enthusiasm and excitement for research. I thank Julien for everything I learnt from him and for starting with the project of this PhD. I also really appreciated Julien's presence in the lab and have missed him since he left to start his own group in Germany. A heartfelt gratitude is for Shruthi who introduced me to NGS and also helped to shape my scientific thinking during the beginning of the PhD. It was a pleasure to share time with her inside and outside the lab and I am happy I can keep her as a great friend since she moved back to India to start her own group. Additionally, I would like to thank Shruthi for being a reviewer of this thesis and for the insightful comments she provided and Julien for being part of the jury and also coming to the defense.

Thanks also to Lucy Glover and Antonin Morillon for accepting to be part of my thesis jury. I want to thank Nicolai Siegel for being a reviewer of this thesis, providing valuable feedback and also allowing me to continue my scientific career in his lab next year. I am also grateful for my mid-term thesis committee members Alain Jacquier, Benoit Arcangioli and Valérie Soulard for encouragement and advice.

It has been a pleasure working at the Biology of Host-Parasite Interactions Unit these years and have enjoyed the time with present and past members. I would like to acknowledge in particular Carlos for the scientific and non-scientific discussions and for the team

work we shared, which I really appreciate. I am grateful for the people with whom we shared quite some evenings in an otherwise empty lab like Gigliola, Flore and Shuai, it was so great to have you around and get a bit distracted of work by visiting your offices or having you in mine. I also want to thank Elie for sharing the little fish box office with me and for the multiple laughs we had during these years. I thank Anne for always being available and helpful with the paper work and for being such a great colleague to have around. I also thank Doro for always replying to all my questions and for making the best springerle cookies even off season. To all other people in the lab, Aurélie, Patty, Kiki, Jess, Sebastian, Gretchen, Sabine, Liliana, Nick, Denise, thank you for contributing to make the lab such a nice place to work in. I also thank Valentina Libri and Bernd Jangla from the CRT platform for the help they provided.

Thanks to the nice people I met in the Institut Pasteur with whom I enjoyed some time outside work like Soraya and Aysegul for nice afterwork drinks and Soraya's amazing shrimp curry and Ioanna for some concerts.

Besides the people from lab and Pasteur, these PhD years have been great thanks to many friends in Paris, Barcelona and around the world. To my almost entirely italian bunch of friends in Paris, thanks for all the fun moments we shared together during our respective PhD journeys: Flavia, Francesco, Chiaretta, Claudio, Anna, Diego, Alberto, Silvia, Tommaso, Titta, Federica, Lorenzo, Luca, it has been great having you all around. I also thank Rocco, Maria and Núria for climbing sessions around here and Chiara for the fun months as roommates. I am also grateful to Mirco who was with me during the beginning of the PhD and has remained a good friend in Paris.

A particular thank you goes to Laura, despite the distance we managed to see each other and travel together during these years and cannot thank her enough for being the best adventure partner I could dream of. Thanks to my other friends living far away, Martin, Lennart, Violeta, Elvi, I am happy we managed to do some trips together during these years and despite the distance we keep in good contact.

And of course, thanks to my favorite people back home, Berta and Alexia, despite the distance you have always been the closest and I am happy we are still sharing great times together. To friends I always saw when returning to Barcelona, Irene, Ester, Aina, Jud, thanks for making weekends back home so enjoyable and for visiting me in Paris.

I want also to express my gratitude for the unconditional love and support from my parents, in particular for taking care of me the days I was in Barcelona after my knee surgery

and for always asking me how are my experiments and PhD going despite they still don't really know what I am working on. I also thank the rest of my family for their love, in particular my three nephews, Lluc, Anuk and Huc as seeing them was the highlight of several visits home.

Finally, I want to thank Alessandro for his love and support during this time. Maybe I should thank the PhD itself for making me move to Paris and thus meeting Ale. I really appreciate the lonely beers he had in the corner bar while waiting for me when I was synchronizing the parasites late in the evening and also the multiple lab pickups by van to bring me climbing on Fridays or after weekend time in the lab. A special thank you for those times he came with me to lab at 3 am to read a book by the bus stop light while I had a parasite time point. Cannot wait for the next chapter with him and Roc as a team of three in Germany.

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 variable open reading frame
TAD	topologically associating domains
TAS	telomere-associated sequences
TF	transcription factor
TSS	transcription start site
UTR	untranslated region
3C	chromosome conformation capture

Part 1

Introduction

Contents

In this first part, I contextualize my thesis work and introduce background information about both the pathogen studied, the human malaria parasite *Plasmodium falciparum*, and the particular topics of PhD work on antigenic variation and non-coding RNA. This introductory part is not an exhaustive description of the subject but rather a presentation aimed for an easier understanding of the body of work. It concludes with a section highlighting the purpose and outline of the thesis.

1.1 Introduction to malaria

Malaria is a deadly vector borne disease caused by protozoan parasites of the phylum *Apicomplexa* and genus *Plasmodium* infecting mammals, reptiles and birds. The parasites are transmitted to the host during a blood meal of the vector, female mosquitoes from the genus *Anopheles*. Human malaria is an ancient disease and detailed description of its symptoms already appear in the Chinese Canon of Medicine, Nei Ching, from c. 2700 BC, clay tablets from Mesopotamia from c. 2000 BC and Egyptian papyri from c. 1500 BC [Cox, 2010]. However, most likely the disease has affected humans since much longer than that, perhaps throughout the entire history of our species. Currently, the oldest proof of its causative agent *Plasmodium* is parasite DNA isolated from a mosquito fossil dated to 30 million years ago [Poinar, 2005]. The name malaria was given by Francesco Torti in 1718 believing that the disease was air borne and came from the "bad air" -*mal aria* in Italian- of swamps. It was not until 1880 that Alphonse Laveran discovered the cause of the disease by identifying the protozoan parasites in a blood smear from a malaria patient. Finally, in 1897 Ronald Ross proved that avian malaria parasites were transmitted by mosquitoes and suggested that it was also the case for human malaria. In fact, one year later Giovanni Battista Grassi together with other Italian malariologists were the first to demonstrate that human malaria was transmitted by anopheline mosquitoes [Cox, 2010].

1.1.1 Malaria epidemiology

Malaria caused 435000 deaths in 2017 and there were an estimated 219 million cases according to the latest WHO World malaria report. Despite the control strategies deployed in endemic regions which led to an unprecedented period of success in global malaria control until 2015, the unceasing emergence of drug resistant parasites and the lack of an effective vaccine, the number of malaria cases increased in 8 million over 2015. 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 92% of cases and 93% of deaths, followed by Southeast Asia and the Eastern Mediterranian region with 5% and 2% of cases respectively. Young children under 5 years old are at highest risk of malaria and accounted for 61% of the global malaria deaths in 2017 [WHO, 2018].

There are five species of *Plasmodium* parasites known to infect humans: *P. falciparum*,

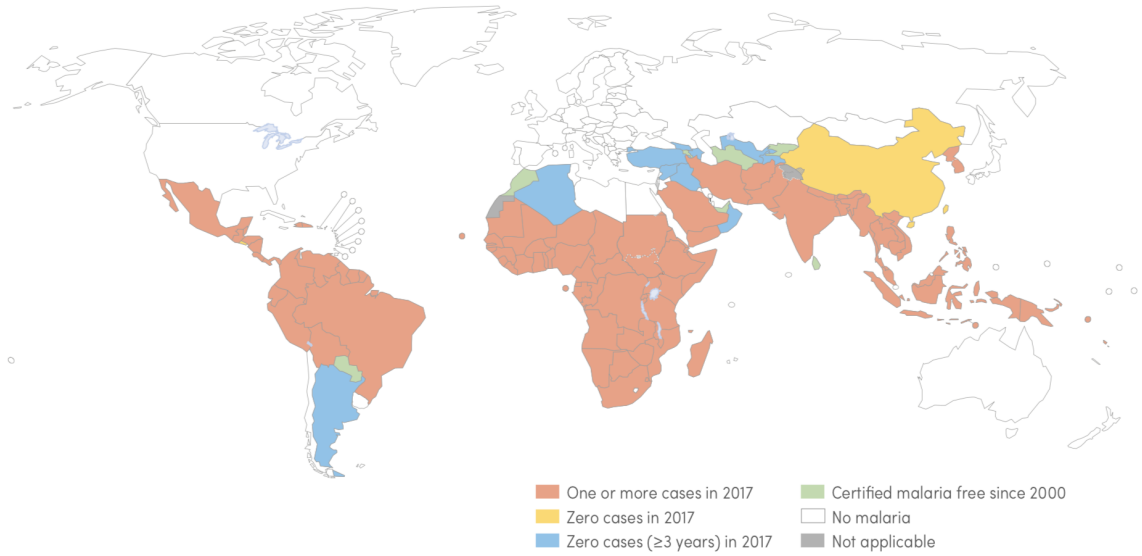


Figure 1.1: Worldwide distribution of countries with malaria cases in 2000 and their status in 2017. Countries without malaria cases during the last three years are considered malaria free. Taken from World malaria report 2018 [WHO, 2018].

P. vivax, *P. malariae*, *P. knowlesi* and *P. ovale*. However, only the first two pose the greatest threat and are more prevalent. Particularly, *P. falciparum* is the most virulent and deadliest malaria parasite accounting for over 99.7% of estimated malaria cases in Africa [WHO, 2018]. This parasite is responsible for the vast majority of severe malaria cases which can lead to fatal disease if untreated.

1.1.1.1 Current drugs and control strategies for malaria

Available tools for malaria control basically include mosquito nets, insecticides and drugs, both prophylactic and treatment. However, the lack of an 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*. 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 together with other antimalarials as Artemisinin Combination Therapies (ACT) since the combination of drugs with different targets and mechanisms of action considerably reduces the likelihood of parasite resistance [WHO, 2018].

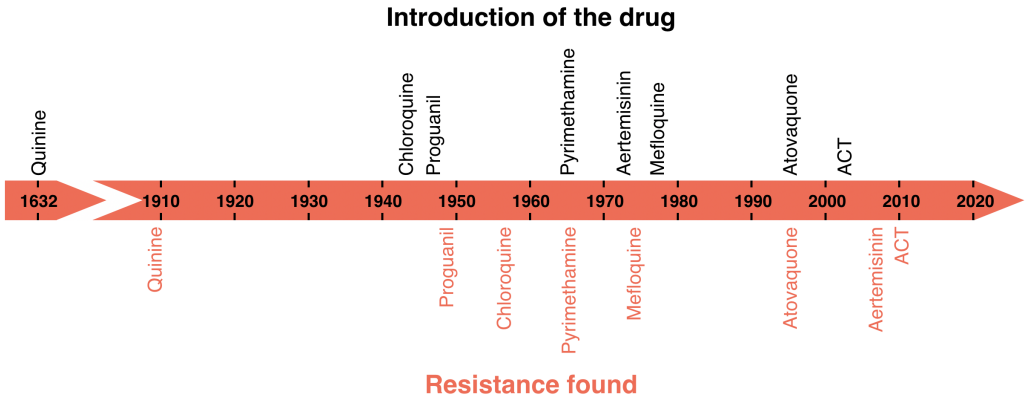


Figure 1.2: Timeline of antimalarial drugs resistance. Top panel indicates the year of the introduction of each drug, and lower panel, the year of appearance of resistance. Modified from [Calderón et al., 2013].

Nowadays, ACT are the first line treatment for *P. falciparum* malaria as recommended by WHO. 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, 2018, Zhu et al., 2018, Dondorp et al., 2009, Ashley et al., 2014]. First *P. falciparum* resistance to artemisinin was reported in Cambodia in 2009 [Dondorp et al., 2009] and few years after already appeared in other regions of Southeast Asia [Ashley et al., 2014, Tun et al., 2015]. Specifically, 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 of artemisinin resistance in Africa is a concerning public health threat and would have a

devastating effect in malaria burden, but also in the economical situation of the countries affected [Dondorp et al., 2010, Lubell et al., 2014, Ariey and Ménard, 2019].

The above reported also accentuates the need for an effective malaria vaccine which would be key to the disease eradication. Currently, the most promising vaccine candidate is RTS,S/AS01 and has been in clinical phase III trials. This vaccine 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). Unfortunately, the RTS,S/AS01 vaccine candidate provides only modest protection, 28% in children and 18% in young infants, which decreases over time [RTSS Clinical Trials Partnership, 2015].

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.

1.1.2 *Plasmodium* biology

Plasmodium spp. akin to other parasites such as *Toxoplasma*, *Cryptosporidium* and *Theileria* belong to the phylum *Apicomplexa*. This large phylum of alveolates comprises exclusively intracellular parasites [White and Suvorova, 2018]. Most apicomplexan parasites possess a peculiar organelle, a non-photosynthetic plastid named apicoplast, 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 wide spectrum 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]. The name *Apicomplexa* comes from an apical complex structure characteristic of the extracellular invasive stages of these parasites, the zoites. These stages are polarized and one pole contains the apical complex which comprises two types of vesicules, rhoptries and micronemes, that facilitate parasite entrance into the host cell [Bargieri et al., 2014] (Fig. 1.3). 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].

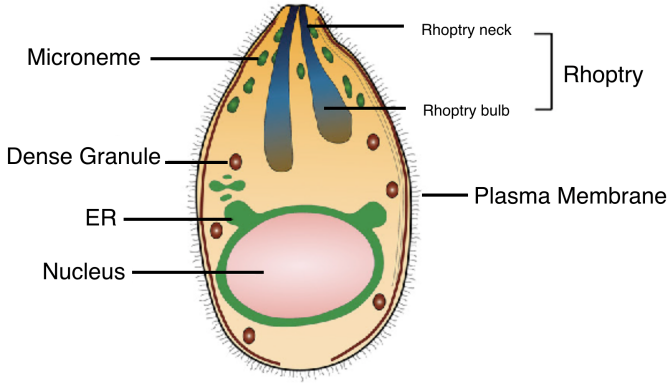


Figure 1.3: Structure of a merozoite, the *Plasmodium* zoite. Intracellular organization of a *P. falciparum* merozoite, showing the apical complex with the two particular type of organelles: micronemes and rhoptries. Modified from [Cowman et al., 2017].

1.1.2.1 *P. falciparum* life cycle

All malaria parasites have a complex but similar life cycle that involves 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 as well as in the length of each parasite cycle, which corresponds to the periodicity of the symptoms [Fujioka and Aikawa, 2002].

P. falciparum is transmitted through the bites of infected female anopheline mosquitoes, primarily *Anopheles gambiae*. During a blood meal, the mosquito vector injects sporozoites into the skin that travel through the blood stream into the liver within 30 minutes. Sporozoites multiply within the hepatocyte, exoerythrocytic schizogony, to yield thousands of merozoites which are released in the blood stream after 7-14 days. Merozoites invade red blood cells and undergo asexual reproduction within the erythrocyte before it bursts after ~48 hours. Upon bursting, newly formed merozoites reinvade other erythrocytes perpetuating the infection. Some parasites fail to progress and divide, differentiating by gamecytogenesis into female and male gametocytes which will be taken by the *Anopheles* mosquito during a blood meal to complete the sexual cycle in the mosquito midgut. Sporozoites are formed by asexual multiplication or sporogony and travel to the salivary glands of the mosquito ready to be injected into the human host during a blood meal [Miller et al., 2002,Josling and Llinás, 2015]. An overview of the *P. falciparum* life cycle is shown in Fig. 1.4.

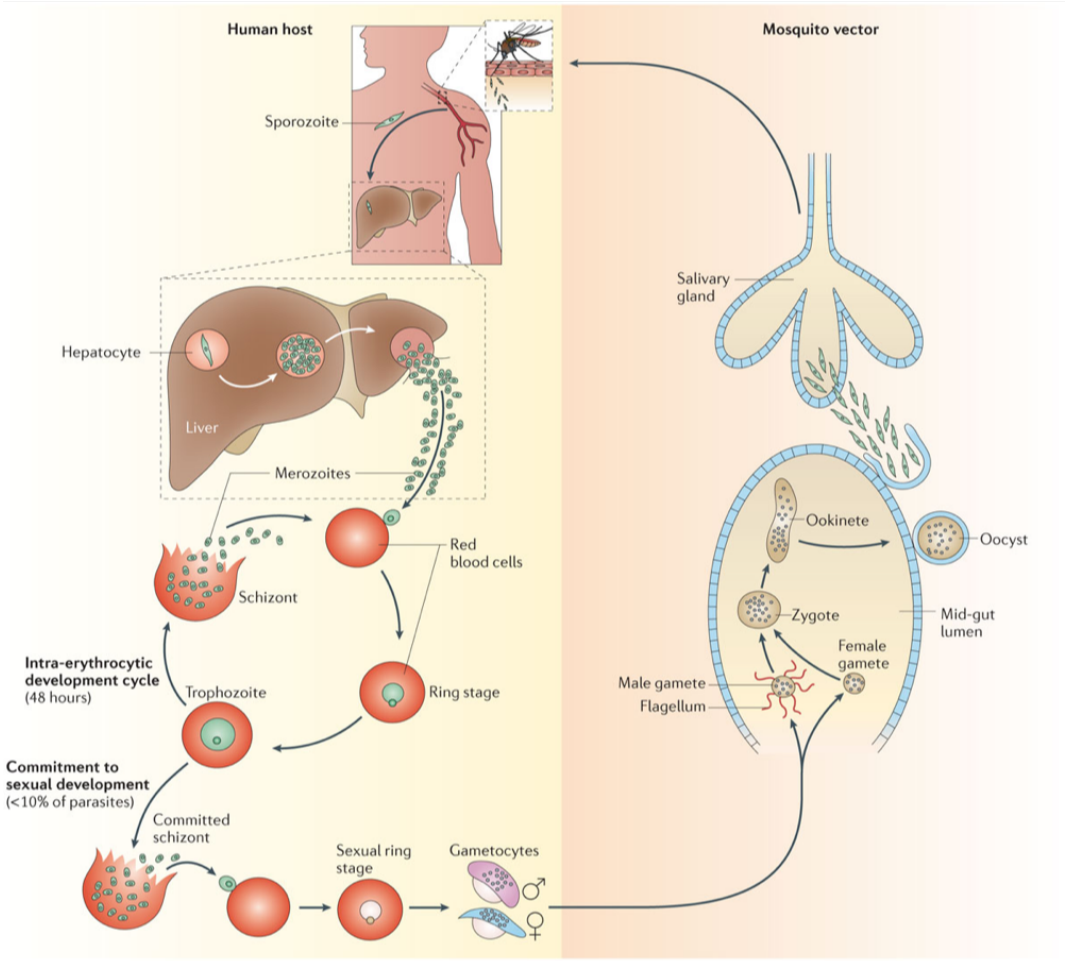


Figure 1.4: Scheme of *P. falciparum* life cycle. After transmission into the skin of the human host during a mosquito bite, sporozoites reach the liver and invade hepatocytes. Within the hepatocyte one sporozoite generates tens of thousands of hepatic merozoites, which enter the bloodstream and invade erythrocytes. The parasite asexual blood cycle causes the clinical symptoms of malaria and is initiated with 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 few (<5%) 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 [Josling and Llinás, 2015].

The intraerythrocytic cycle

Erythrocyte invasion by merozoites begins when the merozoite contacts with low affinity and reversibly the erythrocyte [Bannister and Dluzewski, 1990], followed by reorientation of the polar apical end of the merozoite towards the erythrocyte, the surface of which then wraps around the merozoite [Cowman and Crabb, 2006]. Subsequently, a tight junction is formed between the apical end of the merozoite and the erythrocyte membrane and moves through the merozoite surface until its posterior end, fueled by the parasite actin-myosin motor [Keeley and Soldati, 2004]. With the erythrocyte membrane invagination, the merozoite forms a parasitophorous vacuole which establishes an environment for parasite development isolated from the host erythrocyte cytoplasm [Zuccala and Baum, 2011, Cowman et al., 2017]. A scheme of erythrocyte invasion is shown in Fig. 1.5.

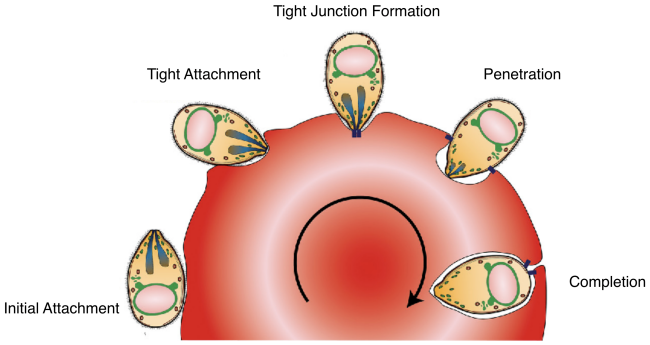


Figure 1.5: Phases of merozoite invasion into the host red blood cell. Erythrocyte invasion takes approximately one minute. This process involves an initial reversible attachment, followed by apical reorientation and tight attachment. During the tight attachment, a pore is formed between the merozoite and erythrocyte mediated by the PfRh5/PfRipr/CyRPA protein complex. Then, a tight junction is formed by ligand-receptor interactions between AMA1 on the merozoite surface and RON2 on the erythrocyte membrane. This tight junction, powered by the actin-myosin motor of the parasite, moves from the apical to posterior pole of the merozoite. When reaching the latter, the adhesive proteins forming the junction are proteolytically removed allowing the membrane to reseal. This mechanism allows the parasite to invade the host cell and create the parasitophorous vacuole without penetrating the erythrocyte membrane. Modified from [Cowman et al., 2017].

During this ~48 hour intraerythrocytic asexual cycle the haploid parasite develops through morphologically different stages while enlarging its cytoplasm within the parasitophorous vacuole (Fig. 1.6). After invasion, the ring stage parasite, ~0-24 hours post invasion (hpi),

develops through trophozoite, ~24-36 hpi, and schizont, ~36-48 hpi, stages. During schizont stage, mitotic divisions followed by nuclear membrane division and cytoplasm segmentation result in the formation of 16-32 daughter merozoites. Schizont bursting and erythrocyte rupture, release the merozoites in the blood stream enabling the cycle to restart by invading new erythrocytes.

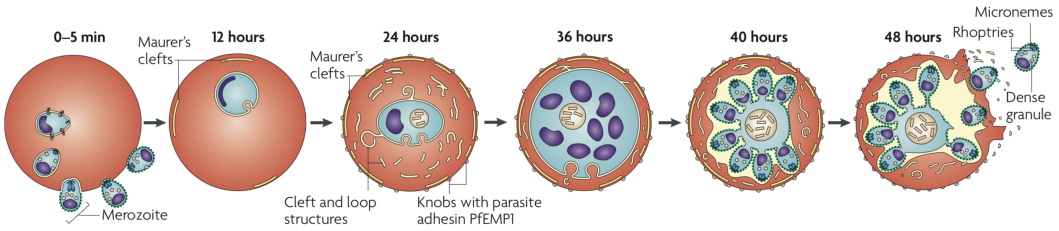


Figure 1.6: Stages of the intraerythrocytic asexual cycle. After erythrocyte invasion by a merozoite, the parasite develops through the ring (~0-24 hpi), trophozoite (~24-36 hpi) and schizont (~36-48 hpi) stages. After 24 hpi, membranous structures can be found in the erythrocyte cytoplasm and knobs 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. ~48 hpi the infected erythrocyte bursts, releasing 16–32 daughter merozoites. Taken from [Maier et al., 2009].

Blood stage parasites remodel the host erythrocyte by secreting and exposing proteins on the surface of the infected erythrocyte in order to acquire nutrients and to evade the immune system [Maier et al., 2009]. To this end, over 10% of the parasite proteins are first exported to the host erythrocyte cytosol prior to displaying a subset of them on the surface of the infected red blood cell (iRBC). This protein trafficking involves parasite-derived membranous structures, named Maurer's clefts, found in the erythrocyte cytosol [Mundwiler-Pachlatko and Beck, 2013]. The mature parasite iRBC membrane becomes distorted by a particular type of protrusions called knobs [Leech et al., 1984a], constituted primarily of knob-associated histidine-rich protein (KAHRP), in which exported surface antigens involved in cytoadhesion like *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) are anchored (Fig. 1.6). Indeed, iRBC containing ring stage parasites appear in peripheral blood circulation while iRBC containing mature stage parasites are sequestered in different organs and do not circulate [David et al., 1983]. 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]. Moreover, the parasite uses antigenic variation of these surface antigens as a mechanism to escape the host immune system and establish chronic infection [Scherf et al., 1998]. Further characterization of the molecules involved in cytoadhesion and immune evasion can be found in section 1.3.

1.1.3 *P. falciparum* malaria pathology and pathogenesis

Malaria is an acute febrile illness with at least a week of incubation prior to the appearance of its symptoms since the intraerythrocytic cycle is accountable for the disease pathogenesis. *P. falciparum* is responsible for the most virulent form of disease although its infection outcome can be very diverse, varying from asymptomatic to lethal. Likewise, *P. falciparum* malaria has a wide spectrum of clinical symptoms such as fever, chills, headache, vomiting, diarrhea, muscular and abdominal pain [Bartoloni and Zammarchi, 2012]. The characteristic waves of fever associated with malaria coincide with the bursting of red blood cells during the parasite intraerythrocytic cycle. In a minority of cases (<0.5% of infections [Sachs and Malaney, 2002, WHO, 2018]) mild malaria progresses to severe disease with symptoms like anaemia, acute renal failure, pulmonary edema, generalized seizures, circulatory collapse, prostration, followed by coma and death [Bartoloni and Zammarchi, 2012]. Each of these complications is 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 membrane. 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 unrousable coma.

Naturally acquired immunity to *P. falciparum* plays a crucial role in protecting individuals living in endemic countries from severe malaria. Indeed, young children, pregnant women and travellers are at highest risk of developing severe malaria [WHO, 2018]. Pregnant woman are at risk of a particular pathology termed pregnancy-associated malaria (PAM), caused by iRBC sequestration in the placenta which can lead to infant death due to low birthweight [Sharma and Shukla, 2017].

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

Eukaryotic regulation of gene expression involves several layers of transcriptional, post-transcriptional and translational control. Nevertheless, regulation at the level of transcription is accepted to play a key role. Transcription is primarily controlled by the binding of transcription factors to regulatory DNA elements. Essentially, typical eukaryotic genes contain a promoter, comprising the core promoter and its *cis*-acting proximal regulatory elements, and distal elements such as enhancers, silencers and insulators (Fig. 1.7) [Maston et al., 2006]. The packaging of DNA into chromatin regulates the accessibility of transcription factors and polymerases to these DNA elements and constitutes a second layer of transcriptional control. In its turn, the state of chromatin condensation is modified by epigenetic factors such as histone modifications and nucleosome repositioning [Kornberg and Lorch, 1999]. Additionally, chromatin is also organized spatially within the nuclear architecture which also contributes to transcriptional regulation [Branco and Pombo, 2007].

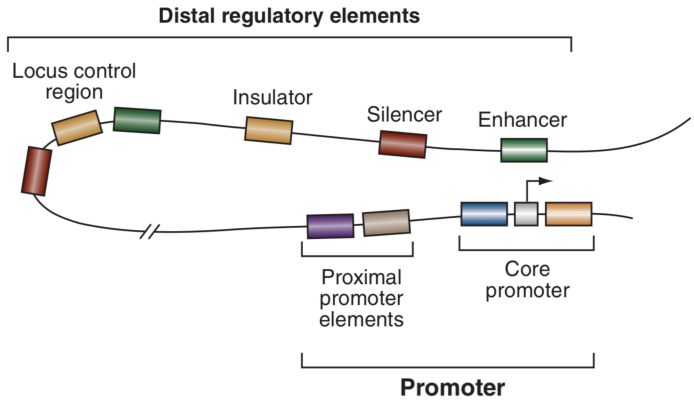


Figure 1.7: Typical eukaryotic gene regulatory region. The core promoter and proximal promoter elements compose the promoter. Upstream distal regulatory elements can be enhancers, silencers insulators, and other control regions, which are located up to 1 Mb from the promoter and enter in contact with it via DNA looping. Taken from [Maston et al., 2006].

Coordinated regulation of gene expression is crucial for *P. falciparum* development and survival within the host environment. Indeed, transcriptomic analysis of both sexual [Young et al., 2005] and asexual stages [Bozdech et al., 2003,Le Roch et al., 2003] of the parasite have shown functionally related genes to be specifically expressed and tightly regulated in the

different stages (Fig. 1.8). Transcription in *P. falciparum* akin to most eukaryotes is predominantly monocistronic [Lanzer et al., 1992a]. The parasite genome organization and different layers of gene regulation including epigenetic regulation but also specific transcription factors, alternative splicing and post-transcriptional regulation are described in the following sections.

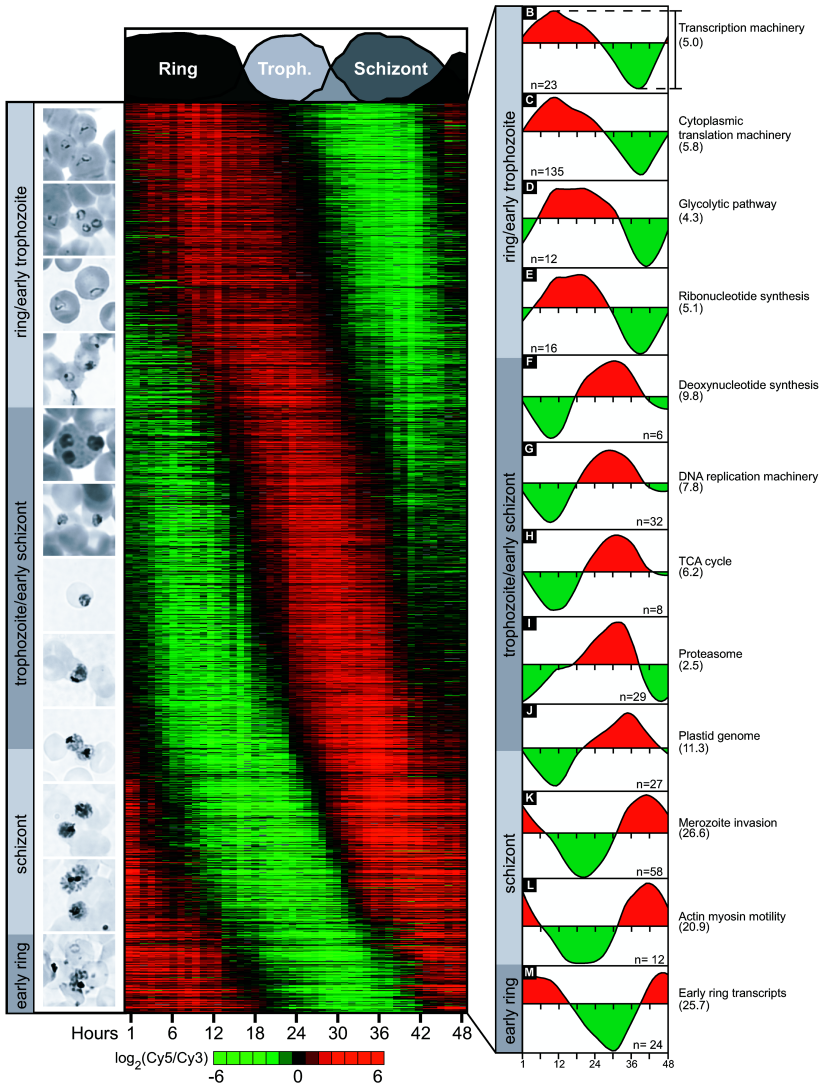


Figure 1.8: *P. falciparum* intraerythrocytic cycle transcriptome. Transcriptional profiles for 2712 genes grouped by biochemical processes and functions (B-M) and ordered along the y-axis by phase of expression throughout the 48 hour cycle. Average expression profile for each group is represented and the mean peak-to-trough amplitude is shown in parentheses. Taken from [Bozdech et al., 2003].

1.2.1 *P. falciparum* genome organization and features

Sequencing of *P. falciparum* genome was completed in 2002 [Gardner et al., 2002]. The genome has a total size of 23.3 Mb and, in addition to the nuclear genome, includes both a 35 kb circular apicoplastic and a 6 kb mitochondrial genome. The haploid nuclear genome comprises 14 chromosomes which range from 0.64 to 3.3 Mb in size. *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 of which nearly 60% lack homologues outside the *Plasmodium* genus.

P. falciparum chromosomes comprise two distinct regions: an internal region, where house-keeping genes are located, and the chromosome ends comprising the telomeric DNA and the telomere-associated sequences (TAS) [Scherf et al., 2001]. The telomeric DNA consists of degenerate G-rich repeats primarily of the motif GGGTT(T/C)A protected by the shelterin/telosome complex. TAS adjacent to the telomere comprise 6 polymorphic repetitive non-coding elements termed telomere-associated repetitive elements (TAREs 1-6) which are variable in length (15-30 kb) [Figueiredo et al., 2000]. Following the TAREs, the subtelomeric region contains 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]. Schematics of chromosome organization can be found in Fig. 1.9.

Genome sequencing of *P. falciparum* allowed for the computational prediction of putative centromeres as 2.6 kb regions particularly AT-rich [Bowman et al., 1999, Gardner et al., 2002]. Experimental evidence described tandem repeat-rich centromeric regions of 2.3 to 2.5 kb with ~97.3% AT-content in all chromosomes except 10 [Kelly et al., 2006]. Interestingly, these centromeres had a well defined size range and lacked repeat conservation amongst different chromosomes. Only two homologues for centromere associated proteins have been described in *P. falciparum*: PfCENH3 and PfCENP-C [Miao et al., 2006, Verma and Surolia, 2014]. Non-coding RNAs (ncRNAs) from centromeric regions have been reported [Li et al., 2008] and have been suggested to facilitate the loading of histone variant PfCENH3 to centromeres [Vembar et al., 2014]. Centromeres have been reported to cluster together during mitosis and cytokinesis in mature parasites and to dissociate after invasion [Hoeijmakers et al., 2012]. However, the lack of comprehensive information on centromere associated proteins leads to a limited understanding of the chromosome segregation mechanism.

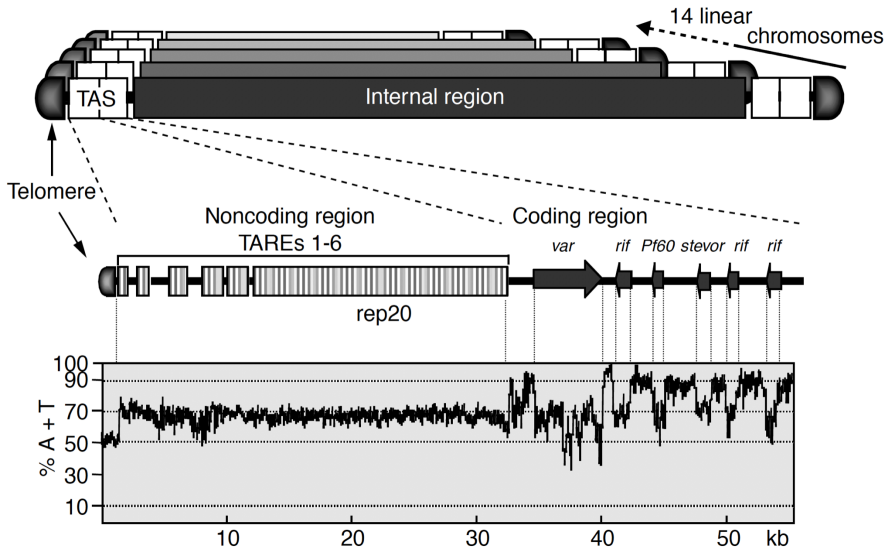


Figure 1.9: Chromosome organization in *P. falciparum*. Top panel: The 14 lineal chromosomes are composed of an internal region and the chromosome ends. Middle panel: End of chromosome 3 is represented as an example including the telomere and TAS. The latter comprise both a non-coding region, composed of 6 TAREs, and a coding region, which is the locus of gene families associated to virulence like *var*, *rif* and *stevor*. Lower panel: AT-content plotted for the different regions of the TAS. Taken from [Scherf et al., 2001].

1.2.2 Transcriptional regulation by epigenetics and chromatin structure

The uniqueness of *P. falciparum* life cycle encompassing multiple stages in two different hosts requires a highly dynamic type of gene regulation, in which epigenetic mechanisms play a major role. Different epigenetic factors contributing to the tight regulation of gene expression including chromatin remodeling, DNA methylation and nuclear architecture are briefly described below. Long ncRNAs participation in *P. falciparum* epigenetic regulation is further detailed in section 1.5.2.

1.2.2.1 Chromatin structure

Eukaryotic DNA is wound around histone proteins forming nucleosomes, which are tightly packed into chromatin [Olins and Olins, 1974, Kornberg, 1974]. One nucleosome is constituted by an octamer comprising two copies of each core histone (H2A, H2B, H3 and H4)

wrapped by 147 bp of DNA. A nucleosome bound to the linker histone H1 constitutes the chromatosome. *P. falciparum* encodes for the four canonical core histones and four histone variants, H2A.Z, H2Bv, H3.3 and CenH3, [Miao et al., 2006] though no linker histone H1 has been identified for *Apicomplexa* [Sullivan et al., 2006].

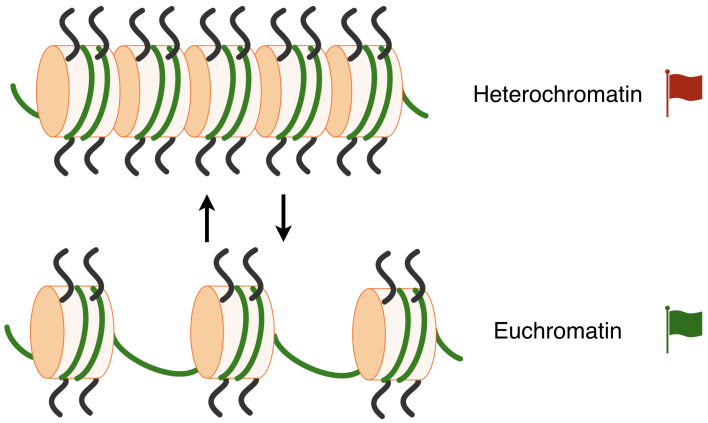


Figure 1.10: Chromatin states. Chromatin is classified in heterochromatin and euchromatin depending of its compaction level. Euchromatin is less compact and is described as a 11 nm fiber which resemble 'beads on a string', where the beads are the nucleosomes and the string is the DNA. The euchromatic state is open and permissive, thereby associated with active gene transcription. In contrast, 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.

Chromatin can be found in two different states: euchromatin, which is uncondensed, open and accessible for transcriptional machinery, and heterochromatin, which is condensed, inaccessible and transcriptionally inactive (Fig. 1.10). Furthermore, facultative chromatin can switch between euchromatic and heterochromatic states. Typically, eukaryotes exhibit lower nucleosome occupancy upstream of transcriptionally active genes whereas nucleosome occupancy is increased in regulatory regions of repressed genes [Lee et al., 2004, Struhl and Segal, 2013]. Nucleosome occupancy in *P. falciparum* has been reported to be low in intergenic regions, similar to other eukaryotes, and initially did not seem to correlate with gene expression during the intraerythrocytic asexual cycle [Westenberger et al., 2009, Ponts et al., 2010]. Furthermore, transcription start sites (TSSs) and core promoters are globally nucleosome-free during the asexual stages allowing the parasite for a general transcriptionally active state, albeit TSSs and promoters of genes expressed during sexual stages

do exhibit high nucleosome occupancy [Ponts et al., 2011]. However, a growing body of evidence suggests that nucleosome occupancy is indeed associated with transcriptional levels also for the majority of genes expressed within the asexual stages [Bunnik et al., 2014, Kensche et al., 2016]. Additionally, transcription-coupled eviction of nucleosomes has recently been detected on strong TSSs in a study that also demonstrated that regulatory DNA elements function can be predicted by nucleosome positioning and dynamics [Kensche et al., 2016].

1.2.2.2 Histone post-translational modifications

Histone post-translational modifications (PTMs) are covalent modifications in various residues of the N-terminal tails of core histones. There are 8 types of PTMs which are associated with diverse functions such as regulation of gene expression, chromatin condensation, DNA repair and replication (Table 1.1). Amongst those, acetylation and methylation of lysine (K) residues are the most common and play a major role in gene regulation: acetylation is generally associated with gene activation while methylation can be associated with both activation or silencing, depending on the affected residues [Kouzarides, 2007].

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

Histone PTMs are dynamic and regulated by histone modifying enzymes that add (histone writers) and remove (histone erasers) these modifications. For instance, histone acetylation is controlled by histone acetyltransferases (HATs) and histone deacetylases (HDACs) that deposit and remove acetyl groups from histones respectively, whereas histone methyl-

transferases (HMTs) and demethylases (HDMs) have a comparable function in adding and removing methyl groups. Finally, histone readers are proteins that recognize specific PTMs and orchestrate the different functions in DNA transcription. Reader proteins contain binding domains that allow for the recognition of the modified residues of histone tails. Furthermore, many histone modifying enzymes also contain these reader domains [Kouzarides, 2007].

The first evidence for the role of histone PTMs in plasmodial gene regulation comes from studying antigenic variation by the Scherf laboratory [Freitas-Junior et al., 2005]. Subsequently, mass spectrometry and antibody-based analyses have demonstrated that *P. falciparum* has a large repertoire of 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]. A recent study identified 232 histone PTMs throughout the intraerythrocytic cycle, of which 88 had never been detected in any other species [Saraf et al., 2016]. However, only a few histone PTMs modifications have been studied in depth: histone H3 lysine 9 acetylation (H3K9ac), lysine 9 mono-, di, and tri-methylation (H3K9me1, H3K9me2, H3K9me3) and lysine 4 methylation (H3K4me3) (Fig. 1.11). H3K9ac and H3K4me3 are predominantly euchromatic marks whereas H3K9me1/2/3 are primarily heterochromatic [Lopez-Rubio et al., 2007, Chookajorn et al., 2007, Cui et al., 2007, Cui et al., 2008b]. A recent study revealed a highly dynamic landscape for histone PTMs throughout the life cycle of *P. falciparum* unveiling specific PTMs subsets associated with asexual stages and with sexual differentiation [Coetzee et al., 2017]. Several orthologs for eukaryotic histone modifying enzymes have been identified in *P. falciparum* so far: three HAT (PfGCN5, PfMYST, PfHAT), four HDACs (PfHDAC1, PfHDAC2, PfHDAC3 and PfSir2), ten HKMTs (PfSET1-10) and two HKDMs (PfJmjC1 and PfJmjC2) [Cui et al., 2007, Fan et al., 2004, Patel et al., 2009, Miao et al., 2010, Andrews et al., 2008, Duraisingh et al., 2005, Freitas-Junior et al., 2005, Merrick and Duraisingh, 2007, Cui et al., 2008a, Volz et al., 2012, Jiang et al., 2013, Chen et al., 2016].

The best characterized histone reader in *P. falciparum* is the heterochromatin protein 1 (PfHP1) which binds specifically to H3K9me3. Genome-wide enrichment of HP1 and H3K9me3 are highly correlated and associated with repressive subtelomeric arrays and central chromosomal clusters of virulence genes [Flueck et al., 2009, Pérez-Toledo et al., 2009]. Additionally, PfHP1 is found in invasion related genes which lack H3K9me3. Notably, unlike in other eukaryotes, PfHP1 is absent from centromeric regions [Flueck et al., 2009]. In fact, *P. falciparum* epigenome is mostly euchromatic with heterochromatin restricted to

these repressive virulence genes subtelomeric and central clusters [Salcedo-Amaya et al., 2009].

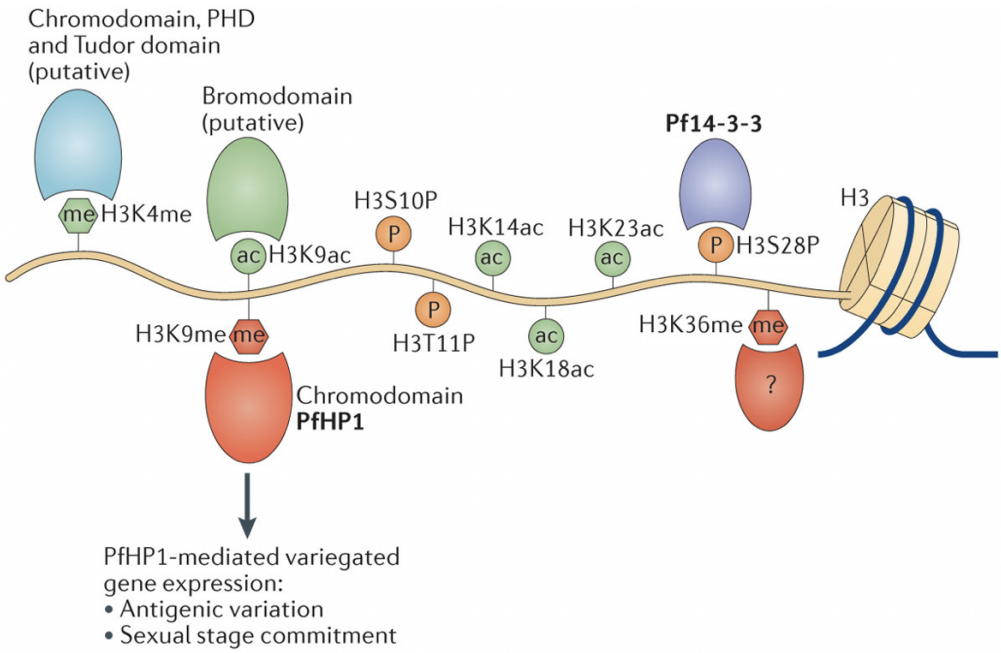


Figure 1.11: Histone H3 PTMs and their associated recognition domains in *P. falciparum*. Silencing marks are shown in red and marks associated with active and poised states, in green. Reader protein PfHP1 binds to H3K9me, and Pf14-3-3 to H3S28P. Other putative reader proteins have been bioinformatically predicted in *P. falciparum* genome containing recognition domains such as bromodomains, chromodomains, Tudor domains and plant homeodomains (PHDs). Taken from [Doerig et al., 2015].

Additionally, a novel type of irreversible PTM consisting in proteolysis of histone H3 tails previously identified in yeast [Santos-Rosa et al., 2009] has recently been detected in *P. falciparum*. Clipped histone 3 was found to be enriched in the 5' UTR of genes involved in DNA replication [Herrera-Solorio et al., 2019].

1.2.2.3 Histone variants

Histone variants constitute a fundamental level of chromatin alteration and play key roles in establishing and maintaining epigenetic states, chromosome segregation, transcriptional regulation and DNA repair among other functions. Replacement of canonical histones by

histone variants is a dynamic process that results in the generation of functionally specialized chromatin domains [Henikoff and Smith, 2015].

In *P. falciparum*, PfH2A.Z variant is found in euchromatic intergenic regions together with H3K4me3 and H3K9ac during the intraerythrocytic cycle [Bártfai et al., 2010]. Moreover, PfH2A.Z has been shown to dimerize with PfH2B.Z (PfH2Bv) forming a double-variant type of nucleosome, enriched in AT-rich promoters and correlating with their strength rather than temporal activity [Hoeijmakers et al., 2013, Bártfai et al., 2010]. Interestingly, the latter is an apicomplexan specific histone H2B variant that has evolved independently of the H2Bv from trypanosomid parasites despite both dimerize with H2A.Z [Dalmasso et al., 2011, Mandava et al., 2008, Siegel et al., 2009]. PfH3.3 variant is found in euchromatic GC-rich coding regions and subtelomeric repetitive sequences of the *P. falciparum* genome [Fraschka et al., 2016]. Finally, centromeric histone variant PfCenH3 occupies the centromeres, which are surprisingly devoid of pericentric heterochromatin in contrast with model eukaryotes [Hoeijmakers et al., 2012, Lopez-Rubio et al., 2009, Grewal and Jia, 2007]. In addition, centromeric regions are also enriched with PfH2A.Z variant [Hoeijmakers et al., 2012]

1.2.2.4 DNA methylation

Methylation of DNA in the 5th carbon of cytosines (5mC) has been broadly studied in eukaryotes. This modification frequently occurs in the context of CpG islands, is stably inherited and has been associated with different mechanisms such as genomic imprinting, regulation of gene expression and chromosome stability. However, the most widely accepted role for DNA methylation is gene silencing [Jones, 2012]. DNA methyltransferases (DNMTs) enzymes catalyze the transfer of the methyl group to DNA. Despite 5mC being the most abundant modified base of DNA, further oxidation of 5mC by ten-eleven-translocatin (TET) enzymes into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) increases the DNA methylome repertoire [Rasmussen and Helin, 2016]. Amongst those, 5hmC has been implicated in the regulation of many cellular and developmental processes in different organisms [Shi et al., 2017]. Modifications on the 5C of DNA and their related enzymes are summarized in Fig. 1.12.

Despite several attempts to demonstrate it, the presence of 5mC in *P. falciparum* still remains unclear. Initially, a first study reported the absence of DNA methylation by mass

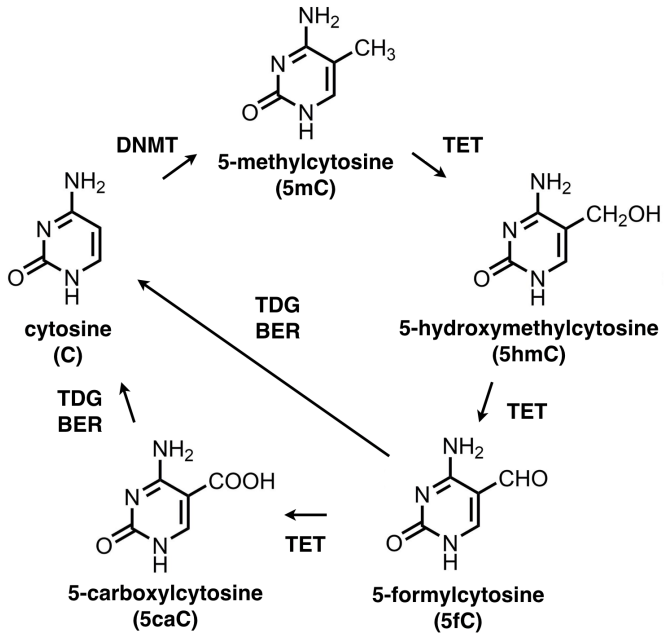


Figure 1.12: DNA methylation in cytosines. DNMTs catalyze the methylation of C5 and TET enzymes catalyze the oxidation of 5mC to 5hmC. Further TET-mediated oxidation result in the successive conversion of 5hmC into 5fC and 5caC. Removal of 5fC and 5caC by thymine-DNA glycosylase (TDG) and reparation by the base excision repair (BER) pathway, leads to the generation of an unmodified cytosine.

spectrometry (MS) [Pollack et al., 1982] but some years later very low levels of 5mC were detected using methylase-sensitive restriction analyses [Pollack et al., 1991]. Subsequently, two further studies had again discrepancy with their results: Choi et al. could not detect 5mC by MS [Choi et al., 2006] while Ponts et al. detected the highest levels described so far and reported the genome-wide distribution of methylation using bisulfite sequencing (BS-seq) [Ponts et al., 2013]. Furthermore, a recent study reported that a putative *P. falciparum* DNMT2 homolog methylates tRNA instead of DNA using a truncated recombinant DNMT2 [Govindaraju et al., 2017], pointing once more towards the lack of DNA methylation in the parasite. Interestingly, mass spectrometry analysis of DNA from two other apicomplexan parasites, *Cryptosporidium parvum* and *Toxoplasma gondii*, resulted in lack of detectable 5mC [Gissot et al., 2008]. Recently, the existence of a 5hmC-like modification and its genome-wide distribution has been determined (Hammam et al, unpublished data). Indeed, the previously reported detection and genome-wide distribution of 5mC [Ponts et al.,

2013] could be due to this 5hmC-like modification since the technique used, BS-seq, cannot differentiate between 5mC and 5hmC.

1.2.2.5 Nuclear organization

Eukaryotic chromosomes are dynamically and hierarchically organized within the nucleus into different large compartments, which in turn are constituted by smaller regions named topologically associating domains (TADs). In higher eukaryotes the CCCTC-binding factor (CTCF) forms DNA loop domains in a process that also involves cohesin, participating in establishing the 3D genome organization [Rowley and Corces, 2018]. Spatial chromatin organization mediates interactions between regulatory DNA elements that are essential for gene regulation, such as bringing together a distant enhancer to its target promoter (Fig. 1.7) [Merkenschlager and Nora, 2016].

Likewise, in *P. falciparum*, genome organization within the nucleus appears to be crucial for gene regulation. The nucleolus is the most prominent domain and is formed by transcription of ribosomal DNA (rDNA) into rRNA by RNA polymerase I (Pol I). Despite, rDNA genes are spread throughout different chromosomes, they cluster together forming the nucleolus in pre-replicative stages of the intraerythrocytic cycle. Another domain comprises RNA Pol II and Pol III transcription sites, which cluster in few loci during the early stages and more during later stages (Fig. 1.13) [Mancio-Silva et al., 2010]. These foci probably harbor co-regulated genes transcription independently of their chromosomal location [Iengar and Joshi, 2009].

Nuclear architecture in this parasite plays a key role in the regulation of clonally variant gene families associated with virulence. Most of the genes belonging to these families are located in subtelomeric regions, with some also in internal chromosomal clusters [Gardner et al., 2002]. Dynamic nuclear repositioning of these loci is critical for activation and silencing of virulence genes. *P. falciparum* telomeres form clusters that are tethered to the nuclear periphery [Freitas-Junior et al., 2000]. Subtelomeric and central virulence genes localize to the nuclear periphery forming 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]. For the virulence *var* gene family, the single active member segregates away from these repressive clusters into a permissive site allowing its transcription [Ralph et al., 2005b, Kyes et al., 2007].

The effect of nuclear positioning in gene regulation of clonally variant multigene families is further detailed in section 1.3.

Furthermore, *P. falciparum* genome organization within the nuclear architecture is remodeled throughout the parasite life cycle correlating with transcriptional changes and adding an additional layer of complexity to this 3D structure. In the transition between ring and trophozoite stages, the number of nuclear pores, widely associated with active transcriptional regions and export of messenger RNA (mRNA) to the cytoplasm, increases while chromatin becomes more open by nucleosome eviction and also displays more inter-chromosomal contacts [Weiner et al., 2011]. These changes are correlated with the higher transcriptional activity characteristic of the trophozoite stage and are later reversed when the parasite enters schizogony. Rearrangements on the genome organization have also been observed in other developmental stages of the parasite. For instance, in sporozoites the genes associated with sporozoite mobility and hepatocyte invasion display long-range interactions while erythrocyte invasion genes are associated with repressed virulence clusters [Bunnik et al., 2018].

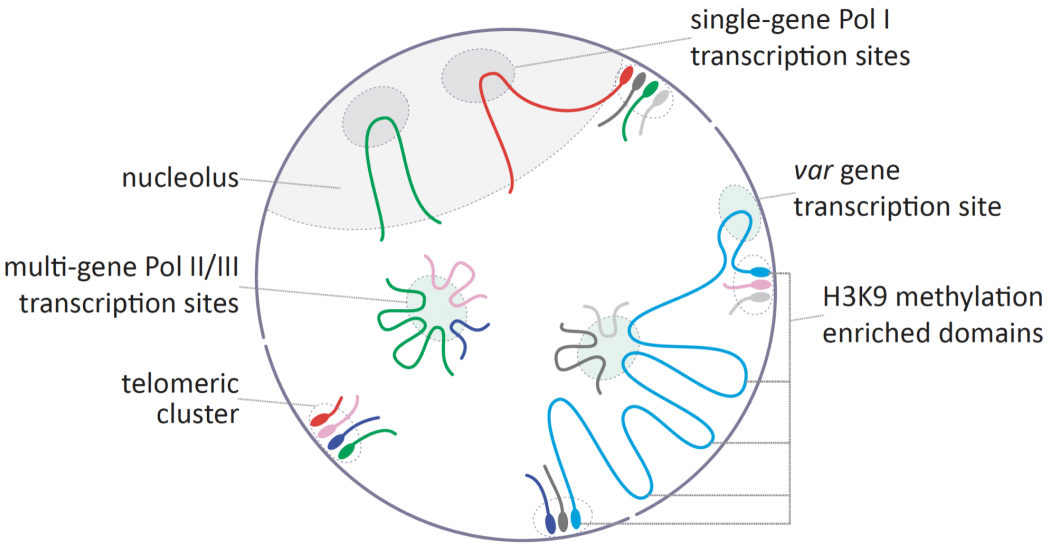


Figure 1.13: Nuclear organization in *P. falciparum*. Schematics of the genome organization within the nucleus for ring stage parasites, displaying the nucleolus constituted by rDNA transcription by Pol I, the transcription sites for the RNA Pol II and Pol III machinery, the telomeric and virulence gene families clusters in the nuclear periphery. Taken from [Scherf et al., 2017].

1.2.3 Transcription factors and DNA elements

Eukaryotic transcription is achieved by three RNA polymerases that transcribe different types of genes. As mentioned before, Pol I is responsible for rRNA synthesis. Pol II transcribes protein-coding genes into mRNA, but also synthesizes some micro RNAs (miRNAs) and other ncRNA. Finally, transcription by Pol III results in the formation of tRNA, 5S rRNA, and other small ncRNAs. RNA polymerases rely upon specific transcription factors (TFs) binding to DNA elements for proper recruitment 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 to which specialized TFs bind in order to enhance or repress transcription [Vannini and Cramer, 2012].

Protein-coding genes in *P. falciparum* are monocistronically transcribed by RNA Pol II and contain both UTRs and introns, similarly to other eukaryotes. Also the promoters resemble canonical eukaryotic promoters [Horrocks et al., 1998]. Core promoters with TATA box-like and Initiator-like sequences has been characterized for some genes [Ruvalcaba-Salazar et al., 2005, Calderwood et al., 2003]. However, the distances between the TSS to both the initial ATG and the regulatory element are variable [Horrocks et al., 2009]. Despite 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 [Lanzer et al., 1992b, Voss et al., 2003, López-Estraño et al., 2007, Olivieri et al., 2008, Osta et al., 2002]. Despite DNA elements with enhancer and repressive activities have been described, no insulator elements have been identified so far and neither a CTCF homolog.

There are currently 202 genes coding for TFs or proteins with DNA binding domains in *P. falciparum*, which represent 3.5% of the genome [Bischoff and Vaquero, 2010]. Amongst those, 73 specific TFs but also chromatin-associated proteins and components of the transcriptional machinery have been identified. Amongst the specific TFs, 27 belong to a highly conserved *Apicomplexa*-specific family of transcription factors [Balaji et al., 2005]. This family of TFs contains a modified version of the Apetala 2 (AP2)-integrase DNA binding domain found in a family of plant TFs [Riechmann and Meyerowitz, 1998] and hence termed ApiAP2. TFs from the ApiAP2 family are indeed central for *P. falciparum* transcription. TFs for which validation has been achieved are summarized in Table 1.2.

All ApiAP2 members contain at least one 60 aminoacid AP2 domain and their tran-

scriptional profiles correspond principally to four different stages of the intraerythrocytic life cycle [Balaji et al., 2005] but some of them are also transcribed in exoerythrocytic stages [Campbell et al., 2010]. Moreover, the majority of members exhibit unique DNA binding preferences that are conserved amongst apicomplexan orthologs [De Silva et al., 2008, Campbell et al., 2010]. The DNA binding specificity and expression throughout particular stages of the life cycle, strongly suggest the different ApiAP2 members to have key roles in stage-specific gene regulation

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

DNA binding protein	Gene ID	Class / Subclass	Biological process	Ref.
PfMyb1	PF3D7_1315800	Specific TF / HTH	Transcriptional regulation of genes involved in cell cycle regulation and progression	[Gissot et al., 2005]
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-var^{int}	PF3D7_1107800	Specific TF / ApiAP2	Binding to intron element of *var* genes in a complex with nuclear actin	[Zhang et al., 2011]
PfPREP	PF3D7_1011800	Specific TF / KH	Transcriptional regulation of intraerythrocytic 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 / HMGB	Non-specific DNA binding; regulates transcription of gametocyte genes and ookinete formation	[Briquet et al., 2006, Kumar et al., 2008]

Recently, two studies using transposase accessible chromatin sequencing (ATAC-seq),

identified novel regulatory regions in *P. falciparum* genome during intraerythrocytic stages [Toenhake et al., 2018, Ruiz et al., 2018]. Chromatin accessibility correlates with active transcription and with H3K9ac and H3K4me3 histone marks and can correspond to DNA elements such as TSS or enhancers [Ruiz et al., 2018]. The majority of ~4000 novel regulatory regions identified are found within 2 kb upstream of the transcribed gene [Toenhake et al., 2018]. Interestingly, motifs enriched in these regions are predictive of ApiAP2 TF binding [Toenhake et al., 2018, Ruiz et al., 2018].

1.2.4 Post-transcriptional and translational gene regulation

Regulation of gene expression occurs at many levels besides the previously described transcriptional control. Mechanisms regulating mRNA splicing and stability also play a key role together with translation and post-translational regulation. Many post-transcriptional regulation mechanisms rely on RNA-protein interactions that either target mRNA for degradation or avoid the access of the ribosome to its translation start codon [Day and Tuite, 1998]. *P. falciparum* genome encodes a large number of proteins coding for mRNA decay and translational rate control. However, post-translational regulation in this parasite has not been studied in depth. Different processes involved in gene regulation at the post-transcriptional level are briefly described below. Antisense ncRNAs participating in this layer of regulation are discussed with further detail in section 1.5.2.

1.2.4.1 Alternative Splicing

Once mRNA is transcribed, it must be processed to create mature RNA that is ready to be translated. This involves the removal of introns that do not code for protein. Alternative splicing is a modulator of gene expression by expanding the protein repertoire within the cell and by regulating the protein isoforms synthesized accordingly. Approximately half of *P. falciparum* genes contain introns and since over half of those have at least two introns, alternative splicing does play a role in gene regulation [Yeoh et al., 2019]. Several studies have focused on the alternative splicing of different genes in this parasite [Knapp et al., 1991, van Dooren et al., 2002, Muhia et al., 2003, Singh et al., 2004] and a genome-wide study detected alternative splicing events in 4.5% of the genes [Sorber et al., 2011]. Alternative splicing in *P. falciparum* is predicted to lead to the synthesis of truncated proteins, to the generation of multiple isoforms of the protein, and to further post-transcriptional

regulation. Besides, different serine/arginine-rich proteins that act as splicing regulators in other eukaryotes [Chen and Manley, 2009] have also been validated in *P. falciparum* [Dixit et al., 2010, Agarwal et al., 2011, Eshar et al., 2012] and there are more than 10 encoded in its genome [Sorber et al., 2011].

1.2.4.2 mRNA decay, stability and regulation of translation

Upon deadenylation mRNA transcripts are susceptible to degradation by the RNA exosome [Chen and Shyu, 2011]. The *P. falciparum* RNA exosome complex has recently been purified and is associated with degradation of diverse RNAs, but seems to play a particularly important role in degradation of cryptic and antisense RNA, highlighting the importance of post-transcriptional control in the regulation of gene expression [Droll et al., 2018]. Furthermore, another layer of post-transcriptional regulation is mediated by a chromatin-associated exoribonuclease (PfRNase II) that silences a subset of the *var* gene family by degradation of nascent cryptic RNA independently from the RNA exosome [Zhang et al., 2014].

Another mechanism of post-transcriptional regulation is mediated by the DNA/RNA-binding protein PfAlba1, which has recently been identified as a factor binding to mRNAs and maintaining mRNA stability [Vembar et al., 2015]. Translational delay occurs for over 30% of transcribed genes in *P. falciparum* [Vembar et al., 2016]. PfAlba1 binding to some invasion mRNAs during the trophozoite stage is linked to translation repression whereas their release from the protein in late stage parasites correlates with protein synthesis [Vembar et al., 2015]. This is the first described *P. falciparum* factor that regulates translation and protein expression timing by regulating mRNA association with the ribosome in a stage-dependent manner.

1.2.4.3 RNA epitranscriptome

Post-transcriptional modifications on mRNA and ncRNA influence RNA structure, expression and function [Schwartz, 2016]. In *P. falciparum*, a repertoire of chemical modifications of tRNA has been reported to be highly coordinated throughout the different stages [Ng et al., 2018]. Increased number of modifications in the trophozoite stage, consistent with the increased translation, are suggested to participate in the modulation of protein abundance by enhancing translational efficiency of codon-biased transcripts [Ng et al., 2018]. Additionally, the methylation modification N^6-methyladenosine (m^6A) has recently been described

as the dominant modification in the mRNA of the parasite and shown to be associated with mRNA outcome [Baumgarten et al., 2019]. The characterization of this modification together with its writer m^6A methyltransferase, PfMT-A70, has shown an inverse correlation between transcripts with m^6A and mRNA stability or translational efficiency [Baumgarten et al., 2019].

1.3 Antigenic variation in *P. falciparum*

Multiple pathogens undergo antigenic variation as an immune evasion mechanism. This strategy consists in coordinated changes of expression between variant surface antigens (VSA) that are exposed to the immune system, reducing immune clearance and leading to chronic infection [Deitsch et al., 2009]. Two main mechanisms enable pathogens to achieve antigenic variation: random and programmed variation. Random variation through DNA alterations such as recombination or errors in DNA replication and repair diversify the antigen repertoire. In contrast, programmed variation is characterized by the ability to switch between the expression of members of a gene family encoding proteins with similar functions. Nevertheless, in the case of programmed variation, the repertoire can also evolve by random changes despite an underlying mechanism regulating switching [Borst, 2003]. A summary of protozoan parasites using programmed antigenic variation and the involved gene families can be found 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	*vesα and vesβ*	~130-160	Variant expressed surface antigen
Giardia lamblia	*vsp*	~150	Variant surface protein
Plasmodium falciparum	*var*	~60	Erythrocyte membrane protein 1

The first evidences for antigenic variation in *Plasmodium* parasites came from the observation of relapses in parasite populations and infection for species not infecting humans [Cox, 1959,Brown and Brown, 1965,McLean et al., 1982,Handunnetti et al., 1987]. Subsequently, antigenic variation for *P. falciparum* was shown in squirrel monkeys [Hommel et al., 1983] and finally in humans (Fig. 1.14) [Miller et al., 1994]. The most important VSA involved in this mechanism in *P. falciparum* are PfEMP1 proteins (*P. falciparum* erythrocyte membrane protein 1), encoded by the *var* multigene family [Leech et al., 1984b]. This gene family is also found in all species from the *Laverania* subgenus of *Plasmodium* [Larremore et al., 2015, Otto et al., 2018b]. *Laverania* species infect African great apes and amongst these *P. falciparum* is the only one that infects humans. For non-*Laverania* species the ma-

jor multigene family identified associated with antigenic variation is the *pir* (*Plasmodium* interspersed repeat) present in *P. vivax* (*vir*), *P. knowlesi* (*kir*), *P. yoelii* (*yir*), *P. berghei* (*bir*), and *P. chabaudi* (*cir*) [Cunningham et al., 2010].

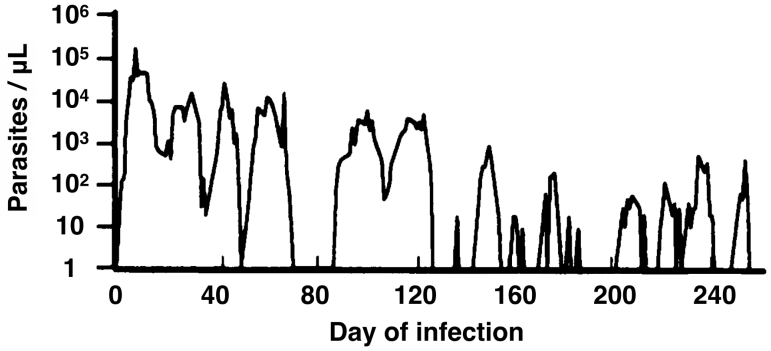


Figure 1.14: Waves of *P. falciparum* parasitemia. *P. falciparum* chronic infection in the blood after a single mosquito bite. Levels of parasitemia are shown for a period of 260 days. Taken from [Miller et al., 1994].

A particular mechanism of antigenic variation is mutually exclusive expression, in which only one member of the variant gene family is expressed at a given time, with the remaining members maintained in a repressed state. The number of distinct surface antigenic phenotypes is then at least as high as the number of members in the family, although it can be much higher than that if recombination occurs amongst them leading to unlimited diversity [Deitsch et al., 2009]. In *P. falciparum* the *var* gene family encoding for the VSA antigen PfEMP1 is expressed in a mutually exclusive manner and is the major player in immune evasion and pathogenesis [Scherf et al., 1998, Smith et al., 1995]. Despite advances in the understanding of regulation of this multigene family expression, major mechanistic questions concerning how a single member is activated while the other remains silenced still remain elusive. Gaining insight on the regulation of mutually exclusive expression of *var* genes was the main original goal of this thesis (see section 1.6).

Besides the *var* gene family, the genome of *P. falciparum* contains several other clonally variant multigene families encoding VSA which are displayed on the surface of iRBC and undergo antigenic variation. These VSA are a target of the host immune system but they show hypervariability in part of their sequences, increasing the repertoire of antigens and avoiding sterile immunity. A description of these gene families (summarized in Table 1.4) can be found in the following sections 1.3.1 and 1.3.2.

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

1.3.1 Mutually exclusive expression of the *var* gene family

Antigenic variation of *var* genes encoding for PfEMP1 is the most important immune evasion mechanism used by the fatal malaria parasite *P. falciparum* during its 48 hour intraeyrthrocytic asexual cycle. It prevents parasite clearance from the blood stream by switching between the display of PfEMP1 proteins. These proteins are assumed to be the major target of the host protective antibody response [Bull et al., 1998, Fried et al., 1998]. The periodic switching between the expression of different *var* members, ensures that the immune system remains naive to a parasite subpopulation, causing parasitemia waves, and ultimately leading to chronic infection. These antigens are also responsible for pathogenesis by mediating adhesion to the vascular endothelium and other tissues, and hence preventing spleen-dependent killing. Antigenic variation also gives rise to functionally distinct PfEMP1 that serve as ligands for different host receptors (Fig. 1.15) [Miller et al., 2002]. Thus, both immune evasion and functional diversity are mediated by antigenic variation and are key to pathogenesis [Scherf et al., 2008]. For instance, a single *var* member, *var2csa*, mediates adhesion to condroitin sulfate A (CSA) and is responsible for placental malaria [Duffy et al., 2006]. Other adhesion phenotypes, such as rosetting, binding to intracellular adhesion molecule 1 (ICAM-1) and endothelial protein C receptor (EPCR) are also associated to particular *var* types [Jensen et al., 2004, Wahlgren et al., 1992, Bernabeu and Smith, 2017].

Besides PfEMP1 expression on the surface of iRBC has broadly been studied, recently it has been shown that transcription also occurs during the mosquito stages and PfEMP1 are

also displayed on the surface of sporozoites [Zanghì et al., 2018, Gómez-Díaz et al., 2017]. PfEMP1 in this stage has been suggested to play a role in hepatocyte invasion [Zanghì et al., 2018]. The repurposing of PfEMP1 in the various transmission stages requires further characterization.

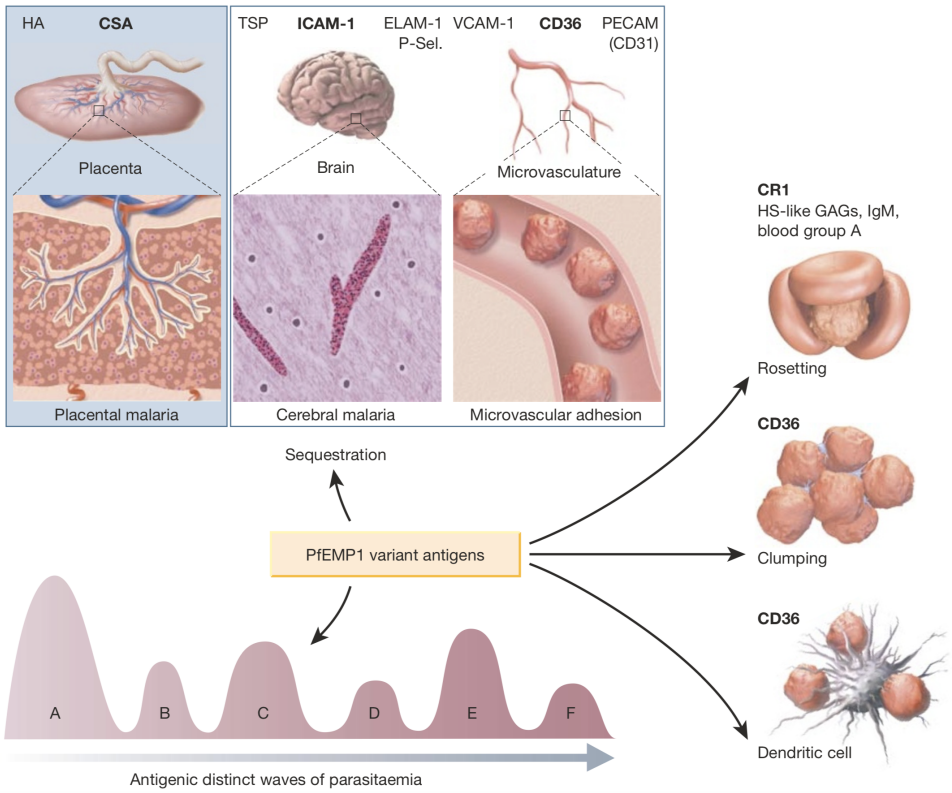


Figure 1.15: PfEMP1 variant antigen family is central to immune evasion and pathogenesis. PfEMP1 play a key role in host-parasite interactions. These proteins are expressed on the surface of mature iRBC, undergo clonal antigenic variation and can bind to many receptors of the host. Both sequestration by PfEMP1 to different tissues for evading spleen-dependent killing and antigenic variation for evading antibody-dependent killing contribute to virulence and pathogenesis in *P. falciparum* and are essential for parasite survival. Parasite sequestration in the brain and placenta contribute to the complications of cerebral and placental malaria, respectively. Binding to several receptors, rosetting with uninfected RBC, and clumping of iRBC through platelets are all associated with disease pathogenesis. The binding of parasite iRBC to dendritic cells downregulates the host immune response. Abbreviations: HA (hyaluronic acid), TSP (thrombospondin), ELAM-1 (endothelial/leukocyte adhesion molecule 1), P-Sel. (P-selectin), VCAM-1 (vascular cell adhesion molecule 1), PECAM (platelet endothelial cell adhesion molecule 1), CR1 (complement receptor 1), HS-like GAGs (heparin sulphate-like glycosaminoglycans), IgM (immunoglobulin M). Taken from [Miller et al., 2002].

As previously mentioned, *var* genes antigenic variation is mutually exclusive, and a single member of the multigene family is expressed at a given time while the others remain silenced. Epigenetics play a major role in this highly coordinated regulation and several layers of control are required to achieve activation of one member while simultaneous silencing of the rest. Details on the organization, structure and regulation of *var* genes are included in the following sections.

1.3.1.1 Organization of *var* genes within the genome

There is considerable variability on the repertoire of *var* genes amongst different *P. falciparum* strain isolates [Freitas-Junior et al., 2000,Kraemer et al., 2007,Otto et al., 2018a]. Most sequenced genomes have ~60 members but the number varies from 47 in the 7G8 isolate to 90 in the GN01 isolate [Otto et al., 2018a]. Despite this variability, all strains present the same patterns in organization and types of *var* genes. For simplicity, we focus the description on the 3D7 strain.

The genome of *P. falciparum* 3D7 strain contains 60 *var* genes. Amongst these, 36 *var* genes are subtelomeric and the other 24 are located in central chromosome clusters. There are one to three members in both subtelomeric regions of all 14 chromosomes, while only chromosomes 4, 7, 8 and 12 harbor central members distributed in head-to-tail tandem arrays; two clusters in chromosome 4 and a cluster in each of the other three chromosomes (Fig. 1.16) [Gardner et al., 2002]. The location of most *var* genes in subtelomeric regions enables frequent recombination events which contribute to a limitless diversity in *var* sequences [Freitas-Junior et al., 2000]. In fact, not only *var* but also other clonally variant multigene families encoding for VSA such as *rif*, *stevor* and *Pfmc-2TM* are also found in subtelomeric regions [Gardner et al., 2002].

The chromosomal position and orientation of *var* genes determines three different major types of 5' flanking regions, termed *ups*: *upsA*, *upsB* and *upsC*. *Var* genes are classified according to their *ups*. Subtelomeric genes transcribed towards the centromere are of *upsB* type; members transcribed towards the telomere are generally of *upsA* type; and finally, the genes in central chromosomal clusters are of the *upsC* type. However, some central members belong to an intermediate group termed *upsB/C* (Fig. 1.16) [Lavstsen et al., 2003, Kraemer et al., 2007]. Additionally, a particular subtelomeric member in chromosome 12, *var2csa*, transcribed towards the telomere belongs to its own group *upsE* [Lavstsen et al., 2003].

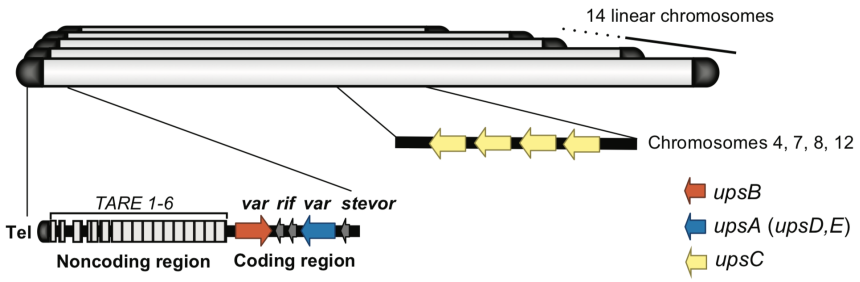


Figure 1.16: Genomic location and organization of *var* genes. Subtelomeric *var* genes are located adjacent to TAREs. Normally, the most proximal to the TAREs is a member from the *var upsB* type followed by a *var upsA* type member transcribed in the opposite direction, towards the telomere. Four chromosomes 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].

1.3.1.2 *Var* gene structure

The *var* gene structure is conserved amongst genetically different *P. falciparum* 3D7 strains with variable *var* repertoire [Kraemer et al., 2007]. *Var* genes contain two exons and one conserved intron of length 0.8-1.3 kb. Exon 1 is 3.9-9 kb long and codes for the variable extracellular domain of PfEMP1, whereas exon 2 is 1-1.3 kb long and codes for the transmembrane and intracellular domains, which are more conserved amongst the different members (Fig. 1.17) [Su et al., 1995].

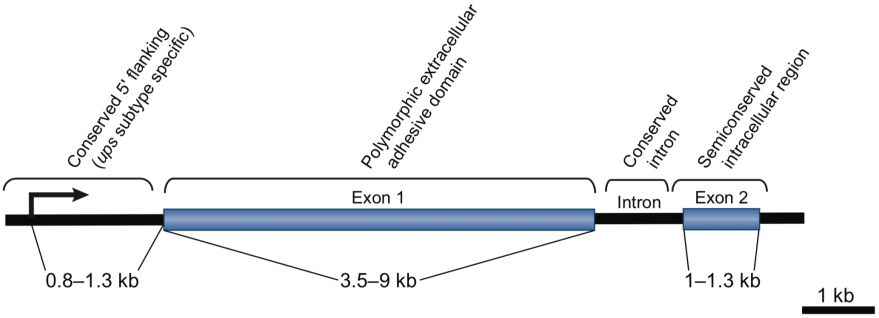


Figure 1.17: Structure and features of *var* genes. All *var* genes are composed of two exons and comprise one intron between them. Exon 1 codes for the polymorphic extracellular domain which contains variable numbers of Duffy-binding-like (DBL) adhesion domains and cysteine-rich interdomain regions (CIDR). Exon 2 encodes for the intracellular domain, containing the transmembrane domain and a highly conserved acidic terminal segment (ATS). Taken from [Scherf et al., 2008].

1.3.1.3 *Var* gene regulation

Var genes are transcribed by RNA Pol II and the expression is developmentally regulated throughout the intraerythrocytic cycle [Kyes et al., 2007, Schieck et al., 2007]. The transcriptional peak of the active *var* gene occurs around 12 hpi on ring stage parasites and expression is repressed during the later stages of the intraerythrocytic cycle (Fig. 1.18) [Kyes et al., 2000], when the encoded PfEMP1 proteins are displayed on the iRBC membrane [Gardner et al., 1996, Kriek et al., 2003]. However, the activated *var* gene remains in a poised state that enables reactivation during the next cycle [Lopez-Rubio et al., 2007]. Switching from one active *var* gene to another occurs at low frequency within a parasite population and the rate is variable depending on the member [Horrocks et al., 2004]. Central *var* genes have been shown to be more stable, with switch rates of ~0-0.3% per cycle, while subtelomeric members switch faster, at rates of at least ~1-2% per cycle [Frank et al., 2007].

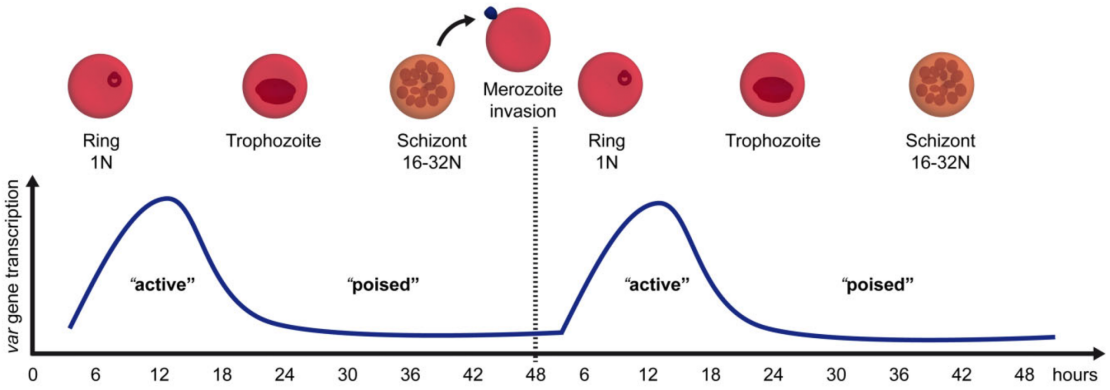


Figure 1.18: *Var* gene transcription throughout the intraerythrocytic cycle. A single *var* gene is transcribed in ring stages parasites and its transcription ceases before DNA replication. The active *var* remains in a poised state during trophozoite and schizont stages and is reactivated during the next cycle in the majority of the population, while some parasites switch to another member. Taken from [Guizetti and Scherf, 2013].

Default silencing of *var* genes requires distinct epigenetic mechanisms such as histone PTM, facultative heterochromatin and remodeling of the nuclear architecture. As mentioned in section 1.2.2.5 (Fig. 1.13), silenced *var* genes are tethered in repressive clusters in the nuclear periphery with the silencing histone modification H3K9me3 enriched in their promoter region [Freitas-Junior et al., 2005, Lopez-Rubio et al., 2007, Chookajorn et al., 2007, Lopez-Rubio et al., 2009]. The sirtuin (silent information regulator) HDAC PfSir2 is enriched in

telomeric and subtelomeric heterochromatic regions containing *var* genes [Freitas-Junior et al., 2005]. This HDAC could enable the establishment of H3K9me3 found in the promoter [Chookajorn et al., 2007] and exon 1 of repressed *var* genes (Fig. 1.19) [Lopez-Rubio et al., 2007]. In fact, disruption of *PfSir2A* caused the de-repression of several *var* genes, particularly of the *upsA* types, together with some *rif* genes located in the proximity [Duraisingh et al., 2005, Tonkin et al., 2009]. Additionally, a similar effect was observed when disrupting another HDAC gene from the same family, *PfSir2B*, that in this case affected mainly *upsB* type genes [Tonkin et al., 2009]. H3K9me3 modification promotes silencing heterochromatin formation by recruitment of PfHP1 [Pérez-Toledo et al., 2009]. Indeed, as previously mentioned in section 1.2.2.2, genome-wide distribution of PfHP1 highly correlates with H3K9me3 enrichment [Flueck et al., 2009]. However, it remains to be elucidated how are PfHP1 boundaries regulated. Another histone PTM enriched in silenced *var* genes is H3K36me3 and has been shown to be controlled by HKMT PfSETvs [Jiang et al., 2013]. Telomeres and thus subtelomeric *var* genes cluster at the nuclear periphery. Moreover, *var* genes located in chromosomal central regions, are also tethered to the nuclear periphery [Ralph et al., 2005b,Lopez-Rubio et al., 2009]. Indeed, this perinuclear localization seems to be essential for default silencing of *var* genes and, despite the mechanism enabling this positioning is not well understood, interaction of actin with the intron of central *var* genes seems to be required [Zhang et al., 2011].

The active *var* gene segregates away from the repressive clusters and locates to an elusive perinuclear expression site (Fig. 1.13) [Ralph et al., 2005b, Mok et al., 2008]. Transcription of *var* genes by RNA Pol II is controlled at the level of transcription initiation [Kyes et al., 2007, Schieck et al., 2007]. The histone marks H3K4me2/3 and H3K9ac are associated with the promoter of the active *var* gene [Lopez-Rubio et al., 2007]. Interestingly, the activating mark H3K4me2 remains associated with the poised *var* gene throughout the cycle (Fig. 1.19) [Lopez-Rubio et al., 2007], in concordance with the function of this mark in higher eukaryotes [Schneider et al., 2004]. Furthermore, the H3K4 HKMT PfSET10 colocalizes with the poised *var* gene once transcription has stopped and probably maintains it in this state during cell division [Volz et al., 2012]. Indeed, the active *var* gene is devoid of silencing mark H3K9me3 and PfHP1 [Lopez-Rubio et al., 2007, Pérez-Toledo et al., 2009]. Additionally, H4Kac and its potential writer PfMYST are also enriched in the promoter of the active *var* gene during transcription [Freitas-Junior et al., 2005, Miao et al., 2010]. The PfH2A.Z/PfH2B.Z double histone variants are incorporated to the active *var* gene only dur-

ing transcription and depleted later [Bártfai et al., 2010,Petter et al., 2011,Petter et al., 2013]. Conversely, PfH3.3 is enriched in the promoter region of both active and poised *var* gene but not in the silent ones [Fraschka et al., 2016].

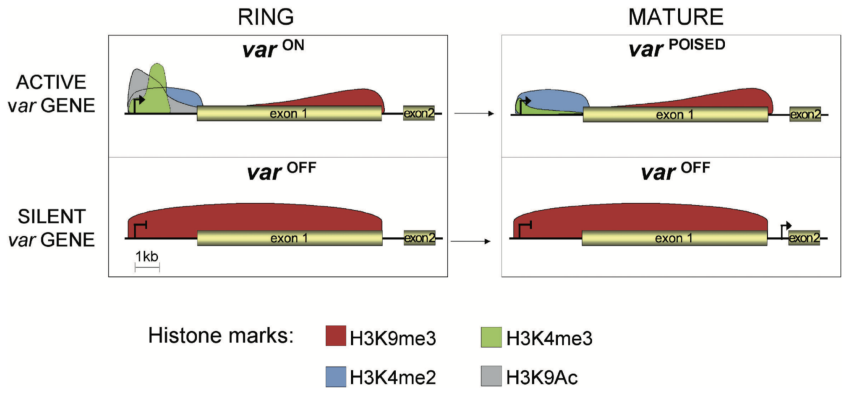


Figure 1.19: Histone H3 marks linked to *var* activation and silencing. Histone H3 modifications are shown for the active or poised *var* gene (var^{ON} or var^{POISED}) and for the other silenced members (var^{OFF}). Chromatin modifications for the intron and exon 2 have not been analyzed in depth due to the high sequence homology amongst most *var* genes. Taken from [Lopez-Rubio et al., 2007].

Perinuclear repositioning of the active *var* gene in a transcriptionally permissive site at the nuclear periphery has been 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] but remains ill defined as neither a mechanism for this localization nor factors associated with the expression site have been described so far. Furthermore, it has been suggested that the expression site could be a highly localized structure with this repositioning also happening within telomeric vicinity in episomally transfected lines [Duraisingh et al., 2005, Marty et al., 2006, Voss et al., 2006]. RNA FISH from an exceptional case in which two *var* genes were expressed at the same time, showed that simultaneous transcription occurred in the same expression site [Brolin et al., 2009]. Additionally, an activated episomal promoter of a *rif* gene family has also been reported to colocalize with the active *var* expression site [Howitt et al., 2009], suggesting that different clonally variant gene families could share a perinuclear expression site. Since in yeast, nuclear pores play a role in enhanced gene expression [Akhtar and Gasser, 2007], it was tested whether it could be the case for *var* genes in *P. falciparum* but no association between those and the *var* expression site was

identified [Guizetti et al., 2013].

Several studies have highlighted the importance of the promoter and intron elements to regulate *var* expression. These elements have been suggested to control silencing and activation at the level of transcription [Dzikowski et al., 2006, Voss et al., 2006]. The promoter has been shown crucial for the mutually exclusive expression counting mechanism [Voss et al., 2006]. Additionally, the ApiAP2 protein PfSIP2 binds to specific sequences in the promoters of *var* genes and seems to promote heterochromatin formation and *var* silencing [Flueck et al., 2010]. The intron has been reported to silence *var* genes by one-to-one interaction with the promoter [Deitsch et al., 2001, Frank et al., 2006, Swamy et al., 2011]. The *var* intron comprises a bidirectional promoter transcribing ncRNAs [Calderwood et al., 2003, Epp et al., 2009] and might be regulated by an ApiAP2 TF that putatively binds to a domain of the intron [Zhang et al., 2011]. Conversely, another study suggested that antisense ncRNA from the intron regulates *var* activation [Amit-Avraham et al., 2015]. However, all these studies relied on artificial systems and episomal expression. Recently, a novel marker-free deletion of the intron in its genomic location, has proved the *var* gene intron not essential for either activation or silencing in placental malaria associated *var* gene *var2csa*, but suggested it could potentially play a role in switching [Bryant et al., 2017].

As briefly mentioned in section 1.2.4.2, PfRNase II also plays a role in regulation of *var* expression. Chromatin-associated PfRNase II silences *var* genes of the *upsA* type by degradation of nascent cryptic *var* mRNA and also antisense intronic ncRNA [Zhang et al., 2014]. The described hallmarks for each state of *var* genes are summarized in Fig. 1.20.

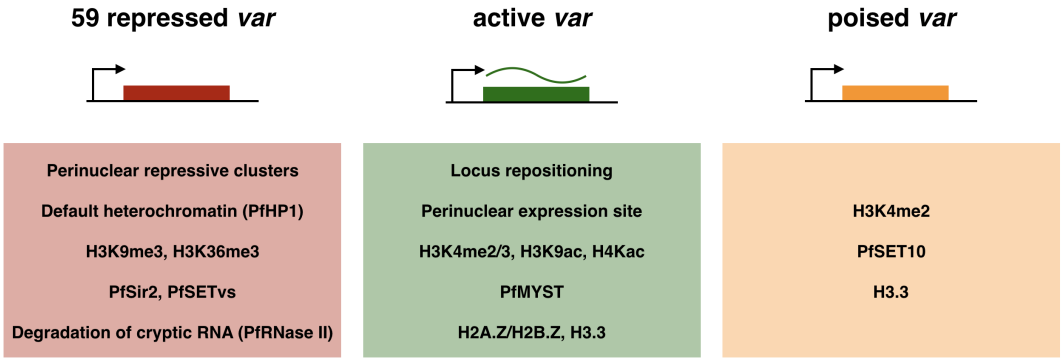


Figure 1.20: Hallmarks of silenced, active and poised *var* genes. Summary of mechanisms, modifications and factors associated with the different states of *var* genes.

Despite multiple factors known to be involved in default silencing of *var* genes, neither a specific regulator of mutually exclusive activation nor switching factors have been characterized yet. It is plausible that a limiting activation factor could drive single *var* transcription [Dzikowski et al., 2006]. In fact, for most *var* genes, an 8 base pair motif named mutually exclusive element (MEE) has been identified and seems to play a role in the *var* counting mechanism [Brancucci et al., 2012]. Activation of promoters lacking the MEE causes no competition with endogenous *var* genes [Brancucci et al., 2012], suggesting that could be the binding region for such limiting activation factor. It has also been hypothesized that an enhancer could drive mutually exclusive activation of *var* genes; however, no such element has been identified using genome-wide chromosome conformation capture (Hi-C) methods [Lemieux et al., 2013, Ay et al., 2014]. Thus, single *var* activation remains to be elucidated.

Lastly, *var* gene switching also remains elusive. Two studies have suggested that switching between different members is non-random [Recker et al., 2011, Enderes et al., 2011] but further investigation is required to confirm this. Similarly, little is known about the mechanisms driving switching and modulating switch rates. An interesting question is whether external factors play a role or the regulation is hard-wired.

1.3.2 Other clonally variant gene families associated with virulence

P. falciparum encodes for multiple clonally variant gene families regulated at the epigenetic level, several of which are associated with virulence. Nevertheless, clonal variation is not restricted to virulence and antigenic variation and affects also gene families involved in lipid metabolism, transcriptional regulation, protein folding amongst others. A common feature in the variegated transcription of all these gene families is the regulation by H3K9me3-linked heterochromatin [Rovira-Graells et al., 2012].

Similarly to PfEMP1 encoded by the *var* gene family, RIFIN and STEVOR proteins, encoded by clonally variant multigene families *rif* and *stevor* respectively, are VSA that undergo antigenic variation, have adhesion properties and contribute to virulence by binding to the microvasculature, non-infected erythrocytes, a mechanism known as rosetting, and immune cells [Kaviratne et al., 2002, Goel et al., 2015, Niang et al., 2014, Wahlgren et al., 2017, Saito et al., 2017].

The multigene families *rif* (repetitive interspersed families of polypeptides), *stevor* (sub-

telomeric variant open reading frame) and *Pfmc-2TM* (Maurer's cleft two transmembrane protein) are globally considered a superfamily of genes which code for proteins with two transmembrane domains (2TM). For all these three families the 2TM domains are found flanking an hypervariable region that codes for the exposed part of the antigen (Fig. 1.21) [Lavazec et al., 2006]. They also contain a PEXEL motif required for the export of the proteins [Hiller et al., 2004, Marti et al., 2004].

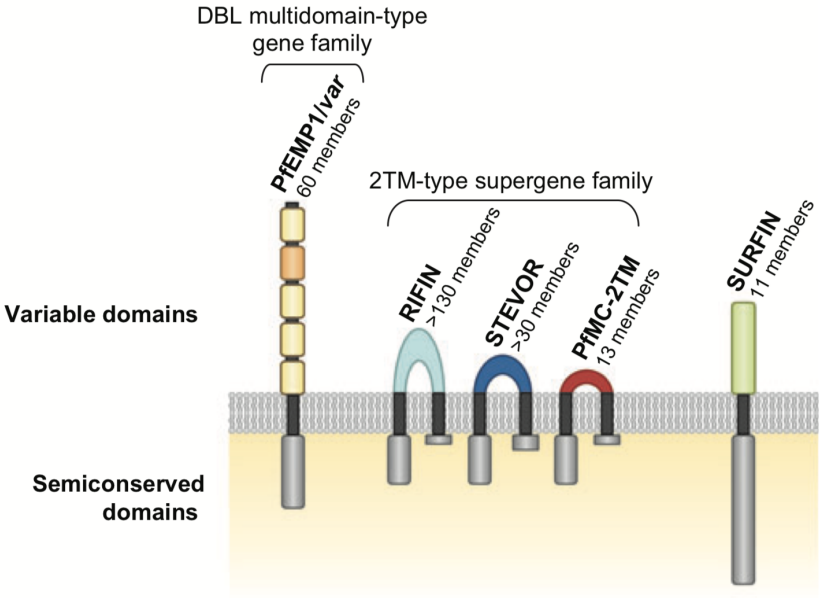


Figure 1.21: VSA displayed on the surface of *P. falciparum* iRBC. All molecules represented are trafficked via Marurer's clefts to the erythrocyte membrane. PfEMP1 are the first antigens exposed, at ~18 hpi, and the other are displayed later on in the intraerythrocytic cycle. These VSA are coded by clonally variant gene families. *surf* (surface-associated interspersed genes) are also found in subtelomeric regions like the other gene families, and code for SURFIN proteins that mediate merozoite invasion. Taken from [Scherf et al., 2008].

Similarly to *var* genes, these gene families contain a variable number of members amongst strain isolates [Otto et al., 2018a] and, interestingly, they are also conserved in all the *Laverania* species of *Plasmodium* [Otto et al., 2018b]. *P. falciparum* 3D7 strain contains approximately 150 *rif*, 30 *stevor* and 13 *Pfmc-2TM* genes. Most of the members are located in subtelomeric regions, and some within central chromosomal *var* gene clusters [Gardner et al., 2002]. A study suggested that all these families share a potential common regulator

factor since promoter titration for different genes results in downregulation for members of all the families [Howitt et al., 2009]. Additionally, transcriptionally active promoters for the different families were shown to colocalize in the same expression site, highlighting the similarities in regulation and the importance of nuclear architecture not only in regulation of *var* but also of other multigene families coding for VSA [Howitt et al., 2009].

1.3.2.1 *Rif* genes

Rif genes constitute the largest multigene family of *P. falciparum*. These genes are ~1 kb long, contain two exons [Joannin et al., 2008] and code for small proteins of 28-45 kDa [Kyes et al., 1999,Helmby et al., 1993,Fernandez et al., 1999]. Despite two transmembrane domains were initially predicted for RIFIN proteins, only one of the domains seems to be stably inserted into the membrane, while the other seems to be involved in dimerizing RIFINs [Goel et al., 2015]. They are subdivided in two major groups, A-RIFIN and B-RIFIN, according to an extra 25 aminoacid domain only found in A-RIFIN proteins [Joannin et al., 2008], and in the *rif* genes upstream sequences [Petter et al., 2008]. There are also functional differences between the proteins coded by the two groups. Only A-RIFINs, comprising 70% of the members, mediate rosetting (Fig. 1.22). Interestingly, A-RIFINs are known to interact with ABO blood group A antigen and glycophorin A (GYPA), which might explain why individuals from blood group A have worse outcome in severe malaria [Goel et al., 2015].

Importantly, RIFINs have recently been shown to participate in immune evasion by targeting two immune inhibitory receptors. The binding of a subset of RIFINs with leucocyte immunoglobulin-like receptor B1 (LILRB1) inhibits the activation of B cells and natural killer (NK) cells and is associated with severe malaria [Saito et al., 2017].

Clonally variant transcription of several *rif* genes occurs at around 18-24 hpi [Kyes et al., 2000, Cabral and Wunderlich, 2009, Kyes et al., 1999] and, as previously mentioned, their genomic organization and transcriptional regulation shares characteristics with those of *var* genes, despite not having been studied so extensively.

1.3.2.2 *Stevor* genes

STEVORs resemble RIFINs in their protein structure. In particular, they are similar to B-RIFINs as they lack the extra 25 aminoacid domain only found in A-RIFINs [Joannin et al., 2008]. However, similarly to A-RIFINs, STEVOR proteins also mediate rosetting (Fig. 1.22)

but by binding to glycophorin C (GYPC) instead of GYPA like 1-RIFINs [Goel et al., 2015]. In contrast with PfEMP1, some RIFINs and STEVORs are also found in the surface of merozoites, suggesting they might play a role in invasion [Petter et al., 2007, Khattab et al., 2008, Khattab and Meri, 2011].

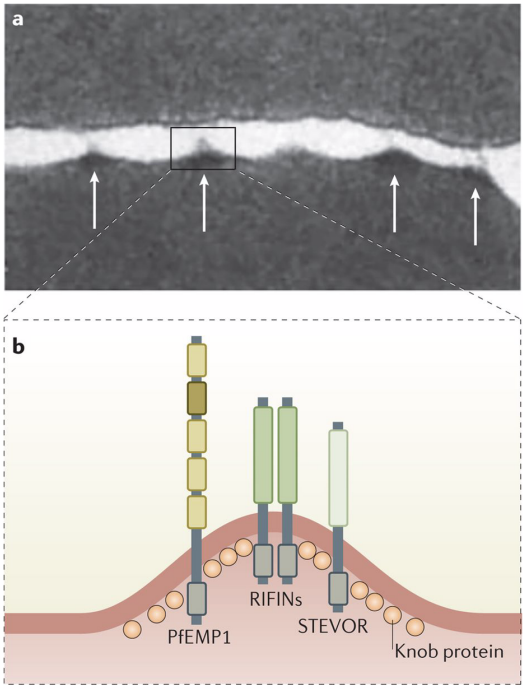


Figure 1.22: *P. falciparum* induced iRBC rosetting with uRBC. (a) Transmission electron micrograph showing knobs on a *P. falciparum* iRBC and binding to an uRBC. (b) Adhesins PfEMP1, RIFINs and STEVOR displayed on knobs at the surface of the iRBC mediate rosetting. Taken from [Wahlgren et al., 2017].

Stevor transcription occurs during the trophozoite stage and clonal variation and switching have been demonstrated [Lavazec et al., 2007]. Different patterns of transcription have been observed for this gene family. Initially, several genes were shown to be transcribed in single parasites [Kaviratne et al., 2002], however, a later study reported several clones with predominant transcription of a single member, suggesting an underlying mutually exclusive mechanism [Lavazec et al., 2007]. Additionally, switch rates for this gene family are very variable among different members [Lavazec et al., 2007].

1.3.2.3 *Pfmc-2TM* genes

Pfmc-2TM genes are transcribed during the trophozoite stage [Sam-Yellowe et al., 2004] and have been shown to be clonally variant and undergo switching [Lavazec et al., 2007]. Despite evidence showing that the encoded PfMC-2TM proteins are 20-25 kDa and displayed on the surface of iRBC exposing an hypervariable domain [Lavazec et al., 2006], the function of these VSA remains to be elucidated.

1.4 Mutually exclusive expression in other systems

Mutually exclusive expression mechanisms are not only characteristic of gene families like *var* that undergo antigenic variation in pathogens. In the following sections, we briefly describe some examples of other organisms that present mutually exclusive expression and their regulation in and outside the context of antigenic variation. Even though single gene activation has not been fully elucidated in any system studied, several epigenetic regulatory mechanisms such as histone PTMs, DNA methylation, chromosome looping, remodeling of nuclear architecture and ncRNAs, recur in different models. It is worth noticing that mutually exclusive expression can be also termed monoallelic gene expression or allelic exclusion, referring to single allele transcription in diploid organisms.

In *P. falciparum*, mutually exclusive expression has only been confirmed for the *var* gene family [Scherf et al., 1998] and for the *clag3* family, that contains two members and is involved in erythrocyte permeation pathways [Cortés et al., 2007].

1.4.1 Mutually exclusive expression in antigenic variation

As mentioned in section 1.3 many pathogens use antigenic variation as an immune evasion mechanism. In many cases, antigenic variation is mutually exclusive but the mechanisms underlying single gene expression vary among different organisms and gene families. In broad terms, mutually exclusive expression is achieved either by modifications of the underlying DNA sequence, by transcriptional regulation, or by post-transcriptional regulation. Nevertheless, in all cases, epigenetic regulation mechanisms seem to be involved.

1.4.1.1 *VSG* expression in *Trypanosoma brucei*

T. brucei VSG genes code for variant surface glycoproteins (VSGs). The *VSG* gene family comprises ~2500 members most of which are located in subtelomeric arrays [Berriman et al., 2005]. A subset of *VSG*s are located proximal to the telomeres, at the ends of polycistronic transcription units, named expression sites (ES). Strikingly, *VSG* genes are transcribed by RNA Pol I, representing an exception for this polymerase that normally only transcribes rDNA [Günzl et al., 2003]. Mutually exclusive expression of subtelomeric *VSG* genes in *T. brucei* occurs at an extranucleolar RNA Pol I transcription factory, known as the expression site body (ESB) [Navarro and Gull, 2001] which is depleted of nucleosomes [Figueiredo

and Cross, 2010, Stanne and Rudenko, 2010]. In fact, RNA Pol I transcription initiates at different sites but is then attenuated in the silent *VSG*s suggesting that the control does not occur only at the transcription initiation level but also during transcription elongation or RNA processing [Kassem et al., 2014].

VSG switching can be either transcriptional, activating an existing *VSG* gene, or recombinational, activating a novel chimeric *VSG* due to recombination. Despite the activation mechanism remains unclear, several factors have been identified involved with silencing, such as the chromatin remodeler ISWI [Hughes et al., 2007], the HMT responsible for H3K76me3, DOT1B [Figueiredo et al., 2008], a specific hypermodified DNA base J and H3.V [Schulz et al., 2016, Reynolds et al., 2016]. Recently, a *VSG* exclusion complex, VEX1-VEX2, has been identified to interact with the chromatin assembly factor CAF-1, and restrict expression to a single *VSG* and prevent others to activate [Glover et al., 2016, Faria et al., 2019]. Furthermore, it has also recently been reported that H3.V and H4.V regulate nuclear architecture and chromatin conformation playing a role in *VSG* expression. Depletion of these two histone variants increases ES clustering and DNA accessibility, enhancing recombinational switching [Müller et al., 2018].

1.4.2 Monoallelic expression in higher eukaryotes

Mutually exclusive expression in diploid higher eukaryotes is not linked to antigenic variation but to expression of genes from a single allele for necessity of the phenotypic outcome, in contrast to the more usual expression of both alleles at similar levels. Monoallelically expressed genes fall into three major classes: parent-of-origin imprinting, and two kinds of random monoallelic expression (RMAE), X-inactivated genes and autosomal RMAE. In the first deterministic case, always the same allele is chosen, whereas in the latter two stochastic cases, any of the alleles can be expressed [Chess, 2016].

1.4.2.1 Genomic imprinting

Genomic imprinting refers to the monoallelic expression of one of the two alleles in a parent-of-origin specific manner. Imprinted genes are expressed exclusively from the paternally or maternally inherited chromosome and are then functionally haploid [Chess, 2016]. This type of monoallelic expression is determined by epigenetic marks such as DNA methylation and histone PTMs placed during gender-specific gametogenesis. Thus, the fertilized egg

has different marks on the imprinted genes inherited maternally and paternally [Ferguson-Smith, 2011].

Imprinted genes are often found in clusters and amongst them one gene encodes a long non-coding RNA (lncRNA) that is essential for the regulation of allele-specific expression (Fig. 1.23). DNA methylation established during gametogenesis silences transcription of imprinted lncRNA genes and thereby activates imprinted coding genes within the cluster [Sleutels et al., 2002, Royo and Cavaillé, 2008]. An example of an imprinted gene cluster and its regulation in *cis* by lncRNA is the maternally expressed *Igf2r*/*Slc22a2*/*Slc22a3* gene cluster in mice and the cognate paternally expressed Air lncRNA encoded antisense of *Igf2r* (insulin-like growth factor 2 receptor gene). The Air promoter is located within an hypermethylated CpG island in the maternal allele and is thus repressed, allowing transcription of the nearby coding genes. Conversely, the paternally transcribed Air silences the paternal imprinted gene cluster, potentially by transcriptional interference [Sleutels et al., 2002].

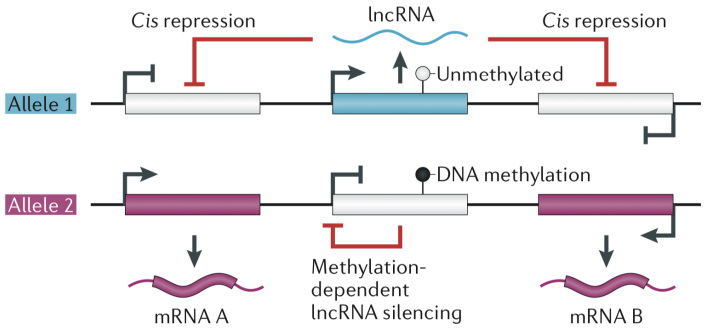


Figure 1.23: Genomic imprinting regulation by DNA methylation and lncRNAs. Allele-specific methylation silences expression of a lncRNA within the imprinted gene cluster, allowing coding genes of cluster to be expressed. In contrast, the other allele, lacking DNA methylation, exhibits expression of the lncRNA, repressing the coding genes of its cluster. Taken from [Quinn and Chang, 2015].

Similar mechanisms have been described for many imprinted lncRNA such as paternal Kcnq1ot1 at the maternal *Kcnq1*/*Cdkn1c*/*Ascl2* gene cluster [Pandey et al., 2008], paternal Nespas at the maternal *Gnas*/*Nesp* gene cluster [Williamson et al., 2011], maternal H19 at the paternal *Igf2* locus [Zhang and Tycko, 1992], and maternal Gtl2 at the paternal *Dlk1*/*Rtl1* gene cluster [Lin et al., 2003]. Additionally, CTCF binding to unmethylated alleles, which prevents interaction of enhancers with the lncRNA promoter, has also been shown to play a role in regulation of imprinted gene clusters [Bell and Felsenfeld, 2000].

1.4.2.2 X chromosome inactivation

Monoallelic expression is characteristic of sex chromosome dosage compensation, that inactivates one of the two X chromosome in females, equalizing the output of single X chromosome in males. X chromosome inactivation renders a transcriptionally silent heterochromatic X chromosome. In placental mammals, the choice of which X chromosome gets inactivated is random but the monoallelic expression choice is stably transmitted mitotically and remains inactivated in all the cells [Lyon, 1986]. Similarly to genomic imprinting, dosage compensation is regulated by epigenetic processes that depend on lncRNAs. The major player in this regulation is the lncRNA Xist (X-inactive specific transcript) that is transcribed in only one of the two X chromosomes and coats its chromosome in *cis*, leading to transcriptional silencing of most genes on the chromosome [Brockdorff et al., 1992, Clemson et al., 1996]. DNA methylation at CpG islands and histone hypoacetylation are epigenetic marks enriched in the inactive X chromosome. On the active X chromosome, the antisense lncRNA Tsix silences Xist in *cis*. A third lncRNA, Jpx is responsible for activation in *cis* and in *trans* of Xist in the inactive X chromosome. Several other ncRNAs have been identified to play a role in dosage compensation [Lee, 2009].

Interestingly, in marsupials, X inactivation is not random, and a similar process involving the lncRNA Rsx inactivates always the paternally inherited X chromosome, basically constituting a mechanism of chromosome-wide genomic imprinting [Grant et al., 2012].

1.4.2.3 Monoallelic expression of olfactory receptors

Autosomal allelic exclusion was first reported for immunoglobulins and T-cell receptors, in both cases by mechanisms involving DNA rearrangement [Hozumi and Tonegawa, 1976, Raulet et al., 1985]. However, many more genes have been reported to present monoallelic expression by different mechanisms. Interestingly, despite the functions of genes that undergo autosomal RMAE are very diverse, there is an overrepresentation of genes coding for cell surface molecules [Chess, 2016]. Amongst them, olfactory receptors (ORs) have been broadly studied [Chess et al., 1994].

The sense of smell allows the detection of an extensive range of odorants. This sensitivity is achieved by a highly regulated expression of the large gene family coding for ~1400 ORs in olfactory sensory neurons of mice [Wang et al., 1998]. ORs are expressed in a monogenic and monoallelic fashion, with only one out of the ~2800 alleles expressed in

each neuron [Chess et al., 1994]. Transcription of the chosen single OR is stabilized by a feedback loop preventing the activation of additional ORs [Serizawa et al., 2003, Shykind et al., 2004]. This feedback loop regulation depends on the detection of the newly translated OR [Dalton et al., 2013, Lyons et al., 2013]. This feedback loop is possible because constitutive heterochromatin causing OR silencing occurs during neuron differentiation, before the transcription onset of OR [Magklara et al., 2011]. Additionally, all silenced OR loci converge in few heterochromatic foci, while the single active OR dissociates of these foci for expression, highlighting the importance of spatial compartmentalization in monoallelic expression of ORs since disruption of this nuclear architecture causes multiple alleles to be expressed at the same time [Clowney et al., 2012, Armelin-Correa et al., 2014].

Despite single OR choice has not fully been elucidated yet, recent progress has been made in this direction. An enhancer H has been identified to associate with the active OR allele, whether from the same or different chromosomes [Lomvardas et al., 2006]. Notwithstanding, deletion of enhancer H only affects proximal OR loci [Khan et al., 2011], suggesting that other enhancer H-like elements could play a redundant role [Williams et al., 2010]. Indeed, circularized chromosome conformation capture (4C) and FISH analyses have shown that multiple enhancers converge with the single active OR locus and Hi-C analysis revealed a high number of interactions between these elements [Markenscoff-Papadimitriou et al., 2014]. Thus, intergenic OR enhancers have a critical role when acting in *cis*, which may help open up chromatin, and a redundant function when acting in *trans* together with other enhancers to activate the chosen OR (Fig. 1.24).

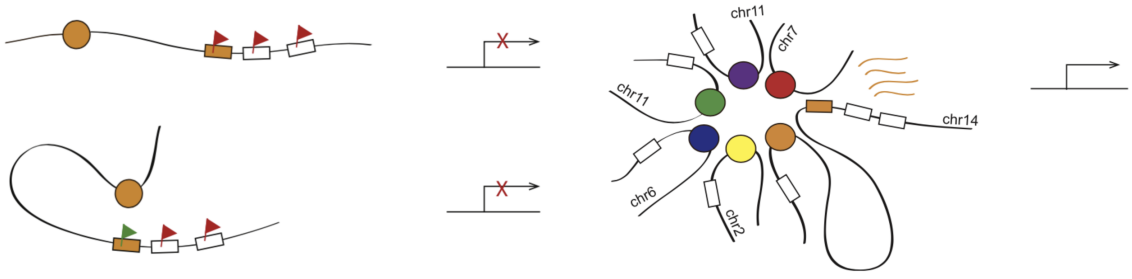


Figure 1.24: Essential and redundant function of OR enhancers. OR enhancers (circles) are essential to open chromatin for a proximal OR (box) in *cis*, but redundancy of multiple enhancers interacting in *trans* is required for OR transcriptional activation. Red flags indicate heterochromatin (H3K9me3) and green flag indicates derepressed chromatin landscape. Taken from [Markenscoff-Papadimitriou et al., 2014].

1.5 Regulatory non-coding RNA

Non-coding RNAs (ncRNAs) are defined as transcripts which do not encode for proteins and are thus not translated, in contraposition to messenger RNAs (mRNAs). ncRNA can be divided in two big classes: housekeeping or structural ncRNA and regulatory ncRNA. Housekeeping and structural ncRNAs comprise transfer RNA (tRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), and telomerase RNA component (TERC). 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 are further classified arbitrarily by their size: short ncRNAs, (< 50 bp), medium ncRNAs, (50-200 bp) and long ncRNAs (lncRNAs, > 200 bp) [Ponting et al., 2009, Ma et al., 2013].

The relatively new field studying regulatory ncRNAs and their functions is increasing exponentially and it is now well established that these ncRNAs play crucial roles in the regulation of gene expression via a wide variety of mechanisms. In fact, despite novel functions for regulatory ncRNAs are steadily emerging, the role of the vast majority of ncRNAs remains to be elucidated [Holoch and Moazed, 2015, Cech and Steitz, 2014, Quinn and Chang, 2015]. The functions of some regulatory ncRNAs are summarized in Table 1.5.

Table 1.5: Examples of regulatory ncRNAs.

	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

In this section, we focus in particular on medium and lncRNAs which constitute the

most diverse and prevalent type of regulatory ncRNAs, and are often defined according to their genomic location relative to coding genes and referred as a whole as lncRNAs (Fig. 1.25). In contrast, short regulatory ncRNAs such as siRNA, miRNA and piRNA have relatively well established functions (Table 1.5) and mechanisms of action, for example the RNA-interference (RNAi) pathway that mediates mRNA silencing [Fire et al., 1998].

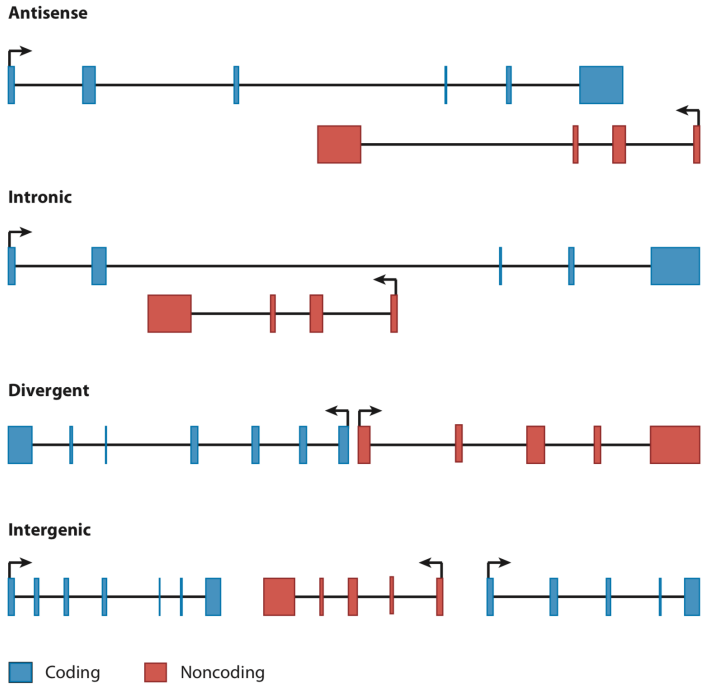


Figure 1.25: Genomic locations of lncRNAs. Antisense ncRNAs are transcribed on the opposite direction of coding genes and their initiation is inside or 3' of the coding genes, causing some overlap with them. Intronic ncRNAs initiate within the intron of a coding gene in either direction. Upstream divergent ncRNAs (uaRNAs) are transcribed within a few hundred base pairs of the promoter of a coding gene in the opposite direction. Intergenic lncRNAs (lincRNAs) are found between coding genes and further than 5 kb from these. Colored boxes represent exons. Taken from [Rinn and Chang, 2012].

1.5.1 Functions and mechanisms of action of ncRNA

RNA has a great biochemical versatility, allowing ncRNAs to have a broad diversity of functions and interaction partners to undergo their function. The ability to base pair with nucleic acids makes RNA unique in interacting with DNA and RNA targets via simple base pair-

ing, in contraposition to proteins that require much longer sequences to bind to nucleic acids with the same specificity. Additionally, RNA folding in three-dimensional structures allows for recognition and binding of many sorts of partners, ranging from small molecules to big proteins. Another key point in RNA versatility is its dynamism since it can be rapidly transcribed and degraded, giving an advantage over proteins for temporal and transient regulation. Furthermore, changes mediated by RNA can be inherited, as exemplified by RNA-templated modifications of the genome [Geisler and Coller, 2013].

The most broadly studied function of lncRNAs is regulation of transcription by chromatin modulation. However, an increasing number of functions are being elucidated within gene regulation but also on diverse biological processes such as allosteric regulation of proteins, organization of protein complexes or cell-cell signaling [Geisler and Coller, 2013, Quinn and Chang, 2015]. Here, we focus on the functions of ncRNAs as regulators of gene expression.

1.5.1.1 Transcriptional regulation by ncRNA

LncRNAs have crucial functions as *cis* and *trans* regulators of gene expression. They can play important functions in epigenetic regulation as described in section 1.4.2 for genomic imprinting and X inactivation. 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]. For example the lncRNA HOTAIR interacts with PRC2 (Polycomb repressive complex 2) and LSD1 (Lys-specific demethylase 1) which mediate the deposition of repressive H3K27me3 and removal of active H3K4me2 marks respectively [Tsai et al., 2010]. Many antisense lncRNA transcripts have been reported to repress mRNA transcription in different loci by transcriptional interference [Hongay et al., 2006, Latos et al., 2012]. Furthermore, lncRNAs transcription can regulate the binding of regulatory factors [Hirota et al., 2008, Lefevre et al., 2008, Bumgarner et al., 2012].

Another mechanism by which lncRNAs regulate transcription is by directly binding to RNA Pol II, like is the case of ncRNAs from Alu SINEs (short interspersed nuclear elements) that repress transcription of specific genes by binding to Pol II upon heat shock [Mariner et al., 2008]. Moreover, lncRNAs can also impair the assembly of the transcription PIC by 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 of some mechanisms of transcriptional regulation by ncRNA are found in Fig. 1.26.

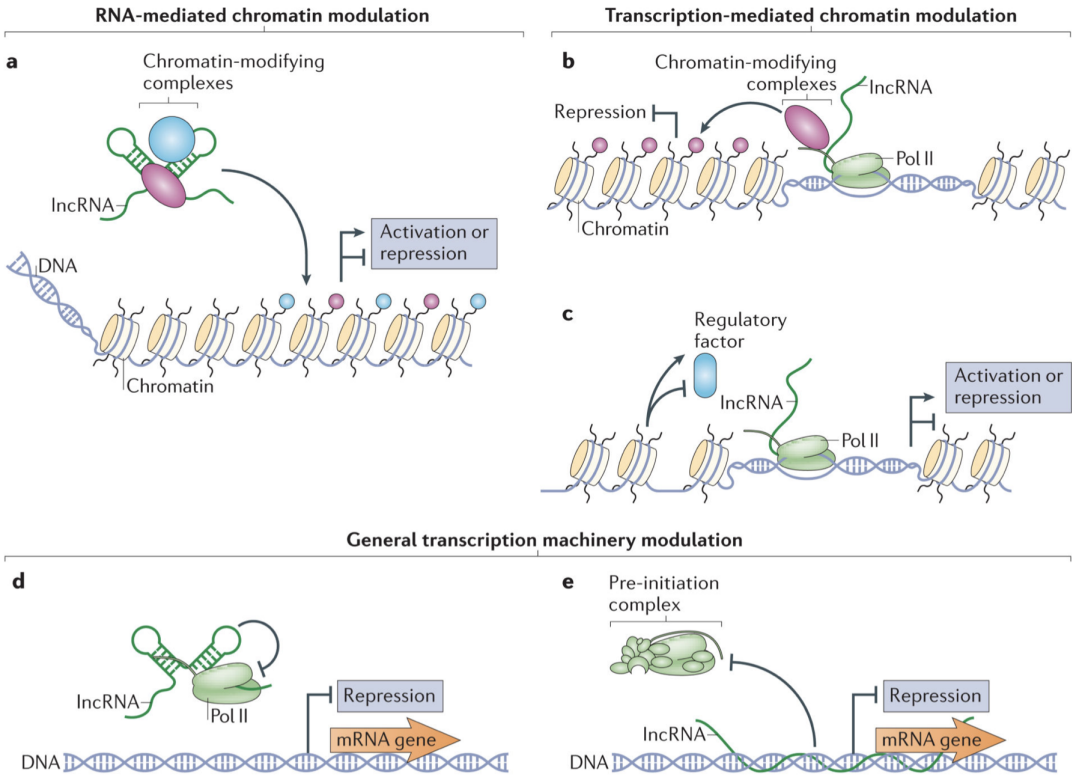


Figure 1.26: Examples of ncRNA-mediated transcriptional regulation. **(a)** LncRNAs can target chromatin-modifying complexes, or TFs, to specific loci. **(b)** Chromatin-modifying complexes can be bound to RNA Pol II and act while lncRNAs are being transcribed. **(c)** Transcription of lncRNAs can also 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].

Enhancers are DNA regulatory elements that normally regulate neighbouring genes in *cis*. Evidence has shown that, in fact, enhancers are transcriptionally active and produce enhancer RNAs (eRNAs), which can be bidirectionally transcribed by RNA Pol II. Enhancers and promoters interact physically and the eRNA expression tends to be correlated with that of the target gene mRNA [Kim et al., 2010, De Santa et al., 2010]. It has been reported that in some cases the eRNA transcript is required for the enhancer activity (Fig. 1.27) [Ørom et al., 2010, Wang et al., 2011], while in others it just seems to be a hallmark of the enhancer

being active [Li et al., 2016].

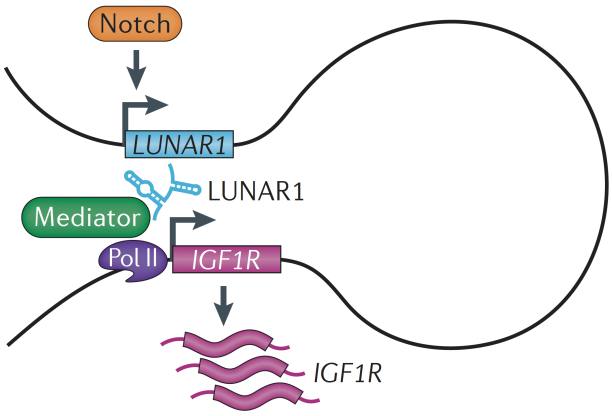


Figure 1.27: Enhancer RNA and chromosome looping. *Cis*-acting eRNAs, like LUNAR1, mediate chromosome looping between the enhancer and their target genes, in this case *IGF1R* (insulin-like growth factor 1 receptor), via protein complexes. Taken from [Quinn and Chang, 2015].

LncRNAs have also been associated with transcriptional activation and repression by mediating the organization of nuclear architecture. It remains unclear how nuclear subdomains are organized, but increasing evidence suggests that ncRNAs could play a role in the regulation or establishment of this compartmentalization [Yang et al., 2011, Yamazaki et al., 2018]. These ncRNAs can behave as scaffolds for subnuclear domains and have been reported to be implicated in the nucleation of interchromatin granules, histone locus bodies, nuclear stress bodies and paraspeckles [Carmo-Fonseca and Rino, 2011, Shevtsov and Dundr, 2011]. For instance, paraspeckles [Fox et al., 2002] have been reported to require the lncRNA NEAT1 for their assembly [Sunwoo et al., 2009, Sasaki et al., 2009] and ongoing NEAT1 transcription for their maintenance [Mao et al., 2011]. Recently, a study showed that NEAT1-dependent paraspeckle formation is achieved by phase separation upon NEAT1 interaction with NONO protein [Yamazaki et al., 2018]. Interestingly, several other studies have reported the implication of ncRNAs and even mRNAs in the regulation and formation of liquid-liquid phase separated granules both by interaction with RNA-binding proteins (RBPs) [Maharana et al., 2018, Langdon et al., 2018] and independently of RBPs [Jain and Vale, 2017], highlighting the importance of RNA as a driver of functional compartmentalization.

Notably, the broad range of functions and means deployed by lncRNAs known to date,

probably only represent a modest contribution of lncRNAs in gene regulation, with further mechanisms to be unveiled in the future.

Relationship between ncRNAs and the boundaries of heterochromatin

RNA Pol III transcription sites have been reported to function as boundary elements blocking chromatin spreading. In particular, genes coding for tRNA have a well established function as boundary elements, which requires the recruitment of TFIIIC, but not Pol III [Donze and Kamakaka, 2001, Scott et al., 2006, Ebersole et al., 2011, Raab et al., 2012]. Loci of some types of ncRNAs transcribed by RNA Pol III contain the A and B box sequences for the recruitment of TFs TFIIIC, TFIIIB and RNA Pol III [Orioli et al., 2012]. The boundary function only requires the B box sequence which is responsible for the recruitment of TFIIIC, but not Pol III, and is thus independent of transcription. TFIIIC plays an important role in genome organization and is responsible for clustering DNA regions at the nuclear periphery, resulting in the boundary function at tRNA loci [Noma et al., 2006]. SINEs, which are also transcribed into ncRNAs by Pol III, and additionally by Pol II, have also been shown to function as boundary elements when being transcribed [Lunyak et al., 2007, Ferrigno et al., 2001]. However, it remains unclear whether these ncRNAs are directly linked to the boundary function or whether it is linked to the transcription process.

Other mechanisms have been described by which ncRNAs, like BORDERLINE in fission yeast, can directly mediate the boundary function by directly binding to heterochromatin factors and blocking heterochromatin spreading [Cam et al., 2005]. Additionally, ncRNAs have also been associated with the regulation of CTCF function in chromatin insulation [Lefevre et al., 2008, Sun et al., 2013].

1.5.1.2 Post-transcriptional regulation by ncRNA

Natural antisense transcripts (NATs) can regulate splicing of their cognate mRNAs, however the mechanism of such regulation remains unclear [Beltran et al., 2008, Hastings et al., 1997, Munroe and Lazar, 1991, Faghihi and Wahlestedt, 2009]. Besides mRNA processing, lncRNAs are also involved in regulation of translational efficiency [Carrieri et al., 2012] and positive and negative regulation of mRNA stability [Faghihi et al., 2008, Gong and Maquat, 2011]. One mechanism mediated by lncRNA to stabilize target mRNA consists in competing for the same binding region of the mRNA with miRNAs involved in mRNA repres-

sion [Faghihi et al., 2008]. Additionally, lncRNAs can also act as miRNA sponges and directly compete with the mRNA for binding to the regulatory miRNA, inhibiting it and avoiding its interaction with mRNA targets [Salmena et al., 2011, Cesana et al., 2011, Wang et al., 2010]. Recently, a novel class of lncRNA with circular structure, named circular RNAs (circRNAs), have been described to act also as miRNA sponges with many miRNA binding sites [Hansen et al., 2013, Memczak et al., 2013].

1.5.1.3 Other functions of ncRNA

Many novel functions are continuously arising for ncRNAs. For example, lncRNAs, together with miRNAs and mRNAs, have been found in exosomes, suggesting that they can play a role as signalling molecules between cells. This cell to cell communication mediated by exosomal shuttle RNAs (exRNAs) could potentially serve as a signal to regulate gene expression in the arrival cell [Skog et al., 2008, Huang et al., 2013]. LncRNAs have also been shown to play important roles in the generation of genetic diversity by participating in V(D)J and class switch recombination events to generate antibody diversity [Cobb et al., 2006, Selsing, 2006].

1.5.2 ncRNA in *P. falciparum*

Since ncRNAs are increasingly recognized as important regulators of eukaryotic gene expression, as described in the previous sections, understanding the role of ncRNA in *P. falciparum* is of great interest, specially, their contribution to mutually exclusive expression of *var* genes. Furthermore, ncRNAs are particularly important for mutually exclusive expression in other systems (section 1.4.2). Thus, the study of regulatory ncRNAs in this parasite might be key not only to better understand how are gene regulation and epigenetic mechanisms regulated, but also to potentially decipher the elusive mechanism orchestrating single *var* choice and mutually exclusive expression.

Interestingly, *P. falciparum* genome sequencing unveiled a great share of predicted RBPs, suggesting an important potential for RNA-mediated regulation in the parasite [Gardner et al., 2002]. This parasite seems to lack the components of the RNAi pathway [Gardner et al., 2002, Baum et al., 2009] but many lncRNAs and NATs have been identified [Patankar et al., 2001, Gunasekera et al., 2004, Upadhyay et al., 2005, Chakrabarti et al., 2007, Mourier et al., 2008, Raabe et al., 2010, Broadbent et al., 2011, López-Barragán et al., 2011, Sierra-

Miranda et al., 2012, Siegel et al., 2014, Broadbent et al., 2015]. However, only structural but no regulatory ncRNAs have been functionally characterized in the parasite so far.

Centromeric ncRNAs were the first described chromatin-associated ncRNA in *P. falciparum*. These ncRNAs are transcribed from bidirectional promoters and associate with centromeres themselves [Li et al., 2008], but their function remains elusive. Additionally, telomere-associated ncRNAs have also been described in the parasite. In contraposition to TERC that plays a role in telomere elongation, many eukaryotes contain a type of subtelomeric lncRNA termed telomeric repeat-containing RNA (TERRA), which is constituted by stretches of G-rich repeats transcribed towards the telomere ends and seems to be involved in telomeric heterochromatin formation and maintenance, and in negatively affecting telomere length [Nergadze et al., 2009, Azzalin et al., 2007, Schoeftner and Blasco, 2008]. TERRA transcripts seem to colocalize in a particular foci at the nuclear periphery and form DNA:RNA hybrids with their DNA template at chromosome ends which can promote homologous recombination among telomeres and genome instability [Balk et al., 2013]. TERRA-like transcripts have been described in *P. falciparum*, transcribed from the TAREs (described in section 1.2.1) [Broadbent et al., 2011, Sierra-Miranda et al., 2012]. Similar to TERRA for other eukaryotes, ncRNAs from TAREs have been reported to cluster in perinuclear foci and be linked with heterochromatin [Sierra-Miranda et al., 2012], and potentially to maintain telomere ends since chromosome breakage and loss of subtelomeric regions leads to telomere elongation [Figueiredo et al., 2002]. The location in proximity of subtelomeric virulence genes, might be key to facultative heterochromatin and silencing of these clonally variant gene families.

P. falciparum NATs seem to be present in ~24% of all ORFs and be developmentally coordinated [Siegel et al., 2014]. For some NATs, transcriptional levels seem to associate with those of the downstream gene. However, most cases lack this relationship, suggesting NATs transcription from cryptic promoters [Siegel et al., 2014]. NATs, mRNA and protein levels seem to not be correlated in this parasite [López-Barragán et al., 2011, Siegel et al., 2014], and the function of these NATs remains uncharacterized. Recently, it has been shown that for one NAT presenting a particular multi-exonic structure, transcription is indeed highly correlated to that of its mRNA cognate coded by the gene *PfGDV1* (*P. falciparum* gametocyte development 1). As a consequence, it has been suggested that this antisense transcript could regulate *PfGDV1*, a key player in early sexual commitment [Broadbent et al., 2015]. Indeed, it has recently been shown that PfGDV1 targets heterochromatin and triggers HP1

eviction of *ap2g* to derepress it in response to the environmental changes, and this process is controlled by *PfGDV1* antisense RNA [Filarsky et al., 2018].

In concordance with the lack of RNAi, no *P. falciparum* miRNA have been detected [Xue et al., 2008]. Strikingly, human-derived miRNA signatures have been reported in particular mRNAs from the parasite, suggesting they are translocated into the parasite and incorporated to mRNAs to mediate ribosome loading blocking and consequently repress translation [LaMonte et al., 2012]. Interestingly, host miRNAs have been also detected in iRBC-derived exosomes, known to mediate cell-cell communication and induce sexual differentiation [Regev-Rudzki et al., 2013, Mantel et al., 2013]. A recent study validated several circRNAs and predicted over 1000 more in the *P. falciparum* genome [Broadbent et al., 2015]. Given their role as miRNA sponges discussed in section 1.5.1.2, it remains intriguing their potential role in the regulation of host-parasite interactions through host-derived miRNAs.

Finally, there is a great interest in the ncRNAs related to *var* genes that have the potential to regulate mutually exclusive expression. As discussed in section 1.3.1, intronic *var* ncRNAs from a bidirectional promoter (Fig. 1.28) have been studied but their function still remains unclear [Bryant et al., 2017, Calderwood et al., 2003, Ralph et al., 2005a, Epp et al., 2009, Amit-Avraham et al., 2015]. In the following section we briefly describe a novel family of ncRNA associated with these virulence genes.

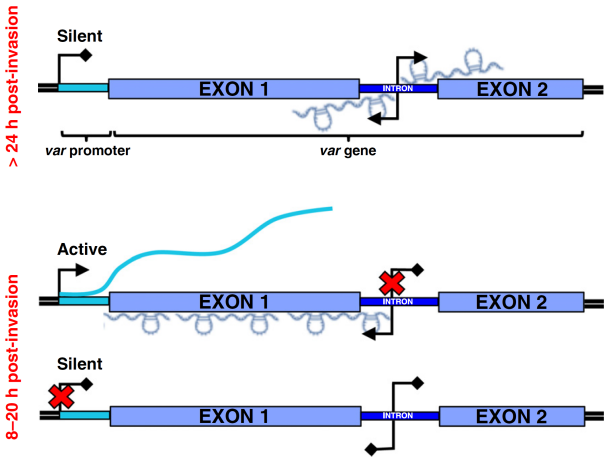


Figure 1.28: Transcriptional model for intronic ncRNAs of *var* genes. Intronic transcription from a bidirectional promoter in *var* genes is developmentally regulated. Late stage parasites exhibit sense and antisense intronic lncRNAs in silent var genes, while in ring stage parasite the active *var* gene presents ncRNA transcription only in the antisense direction. Taken from [Vembar et al., 2014].

1.5.2.1 RNA of unknown function 6 (RUF6)

Sequencing of the *P. falciparum* genome identified conserved sequences with high GC-content for the parasite, ~40%, in intergenic regions of central chromosomal *var* clusters [Hall et al., 2002, Hyman et al., 2002, Gardner et al., 2002]. This sequences code for a novel family of 15 highly homologous ncRNAs (89.9-100%) with loci interspersed between central *var* and *rif* genes, and annotated as RNA of unknown function 6 (*RUF6*) [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 suggest that they can be involved in the regulation of *var* expression (Fig. 1.28).

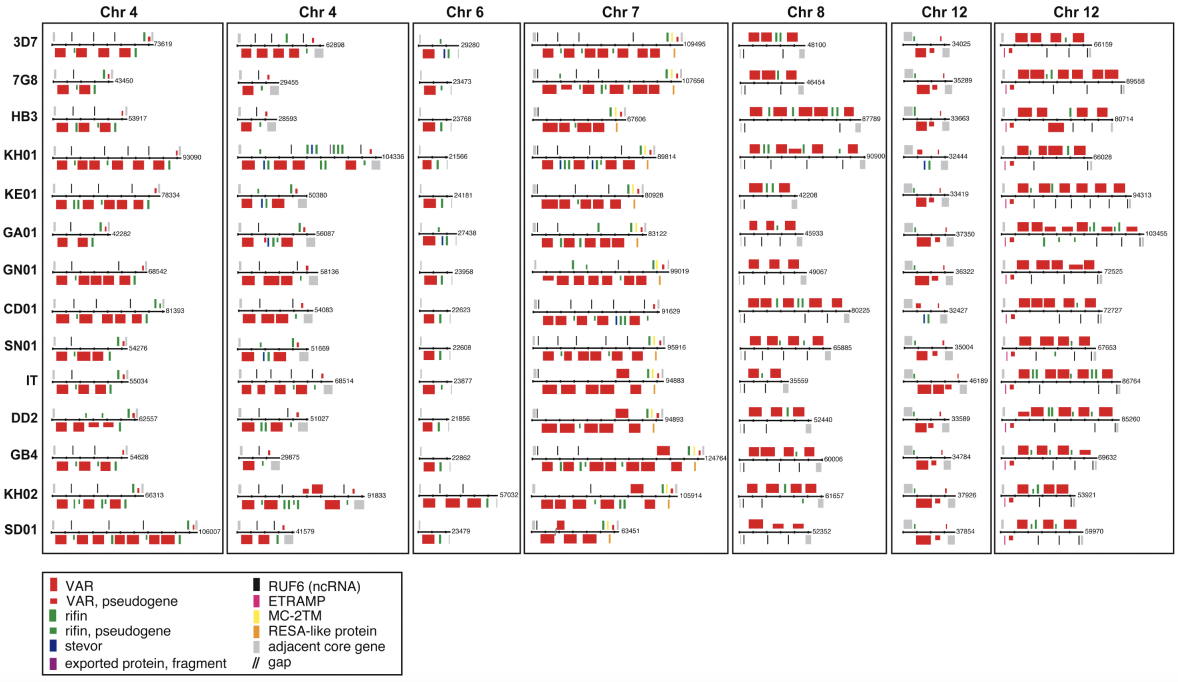


Figure 1.29: Central *var* genes in 13 *P. falciparum* strains. Organization of the central chromosomal *var* arrays in the genome of 13 sequenced *P. falciparum* strains. The number of *RUF6* genes per isolate are: 3D7, 15; 7G8, 10; HB3, 8; KH01, 16; KE01, 13; GA01, 10; GN01, 11; CD01, 16; SN01, 13; IT, 11; Dd2, 8; GB4, 12; KH02, 12; and SD01, 12. Colored boxes represent gene as indicated in the legend. Taken from [Otto et al., 2018a].

Transcripts from several *RUF6* loci have been detected in different transcriptomic studies and indicate the transcriptional peak for these ncRNAs occurs at late ring stages [Mourier et al., 2008, Otto et al., 2010, Broadbent et al., 2011, Siegel et al., 2014, Wei et al., 2014]. Ad-

ditionally, a recent study suggested *RUF6* transcription is clonally variant [Wei et al., 2015]. Interestingly, the exonuclease RNAse II responsible for the degradation of nascent cryptic *var* RNA transcripts, also regulates transcript levels for these RUF6 ncRNAs [Zhang et al., 2014].

As previously mentioned in section 1.3, all *Laverania* species have *var* genes and other virulence-related clonally variant gene families. Strikingly, all these *Laverania* species containing *var* genes also have the *RUF6* gene family [Otto et al., 2018b]. Despite there are differences in the repertoire of both *var* and *RUF6* genes amongst different species and also intraspecies variability, the general disposition is conserved (Fig. 1.28) [Chakrabarti et al., 2007, Otto et al., 2018a].

1.6 Scope of the thesis

The identification of the RNA of unknown function 6 (*RUF6*) gene family exclusively found in *Plasmodium* species that have *var* genes [Chakrabarti et al., 2007,Otto et al., 2018b], opens up avenues to gain insight into the regulation of *var* genes mutually exclusive expression. Strikingly, all *RUF6* members are interspersed between central *var* genes [Chakrabarti et al., 2007,Panneerselvam et al., 2011], pointing to a role in clonally variant gene expression.

This PhD project aims to elucidate the biological role of this GC-rich ncRNA family annotated as RUF6. We seek to determine whether this ncRNA family is involved in the regulation of *var* genes and to gain insight into the molecular mechanism by which this GC-rich ncRNA might regulate mutually exclusive expression of *var* genes. In addition, we also aim to determine whether this GC-rich ncRNA plays a role in regulating other clonally variant gene families associated with virulence.

In order to achieve the main goal of deciphering the function of the GC-rich ncRNA, the PhD has several concrete objectives, listed below:

- Overexpress different GC-rich ncRNA members and study the effect on the *var* gene family mutually exclusive expression. (Results section 2.1)

- Analyze the GC-rich ncRNA transcriptional profile in different parasite clones expressing distinct *var* gene members. (Results section 2.2)

- Develop methods to block GC-rich ncRNA transcription and analyze the transcriptome of the knockout or knockdown strains. (Results section 2.2)

- Elucidate the mechanism of action of the GC-rich ncRNA by identifying the binding partners. (Results sections 2.3 and 2.4)

Part 2

Results

Contents

This second part contains the main results of the thesis, linked to the study of RUF6. Each section is presented in the format of an independent manuscript including the summarized methods used and a brief discussion.

Section 2.1 comprises data that have already been published and is presented in the published format, and section 2.2 corresponds to results currently under review for publication. Finally, sections 2.3 and 2.4 describe preliminary results. Supplementary data linked to the main RUF6 results presented in this part can be found in the appendices A, B and C. Additional preliminary results can be found in Appendix D.

2.1 GC-rich non-coding RNA at the expression site affects *var* counting

The highlights of this section are:

- FISH analysis demonstrates that GC-rich ncRNA colocalizes *in trans* with the expression site of both central and subtelomeric *var* genes.

- Episomal overexpression of different GC-rich ncRNA members disrupts mutually exclusive expression of *var* genes activating concrete *var* subsets.

The results of this section have been published in Nucleic Acids Research (2016) under the title "Trans-acting GC-rich non-coding RNA at *var* expression site modulates gene counting in malaria parasites". My contribution to this work was generating constructs for the mutants. Additionally, I performed RNA sequencing (RNA-seq) analysis of the overexpression strains to validate that the GC-rich ncRNA element overexpressed was indeed the one with highest transcription (see Appendix A).

Trans-acting GC-rich non-coding RNA at *var* expression site modulates gene counting in malaria parasites

Julien Guizetti[1,2,3,*], Anna Barcons-Simon[1,2,3] and Artur Scherf[1,2,3,*]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
*Correspondence: julien.guizetti@pasteur.fr, artur.scherf@pasteur.fr

Published in Nucleic Acids Research. 2016 July 27
doi: 10.1093/nar/gkw664

Nucleic Acids Research Advance Access published July 27, 2016

doi: 10.1093/nar/gkw664

Trans-acting GC-rich non-coding RNA at *var* expression site modulates gene counting in malaria parasite

Julien Guizetti[1,2,3,*], Anna Barcons-Simon[1,2,3] and Artur Scherf[1,2,3,*]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France, [2]INSERM U1201, F-75724 Paris, France and [3]CNRS ERL9195, F-75724 Paris, France

Received November 27, 2015; Revised June 20, 2016; Accepted July 13, 2016

ABSTRACT

Monoallelic expression of the *var* multigene family enables immune evasion of the malaria parasite *Plasmodium falciparum* in its human host. At a given time only a single member of the 60-member *var* gene family is expressed at a discrete perinuclear region called the '*var* expression site'. However, the mechanism of *var* gene counting remains ill-defined. We hypothesize that activation factors associating specifically with the expression site play a key role in this process. Here, we investigate the role of a GC-rich non-coding RNA (ncRNA) gene family composed of 15 highly homologous members. GC-rich genes are positioned adjacent to *var* genes in chromosome-central gene clusters but are absent near subtelomeric *var* genes. Fluorescence *in situ* hybridization demonstrates that GC-rich ncRNA localizes to the perinuclear expression site of central and subtelomeric *var* genes in trans. Importantly, overexpression of distinct GC-rich ncRNA members disrupts the gene counting process at the single cell level and results in activation of a specific subset of *var* genes in distinct clones. We identify the first trans-acting factor targeted to the elusive perinuclear *var* expression site and open up new avenues to investigate ncRNA function in antigenic variation of malaria and other protozoan pathogens.

INTRODUCTION

Antigenic variation is an effective mechanism for pathogens to evade the immune response of the host organism. It is used by the human malaria parasite *Plasmodium falciparum*, who remains a major health threat and the cause of death for hundreds of thousands of people per year (1,2). During its 48h-long blood stage cycle the parasite invades and asexually replicates in red blood cells. It goes through three morphologically distinct developmental stages namely ring, trophozoite and schizont. Upon rupture of the infected red blood cell 16–32 newly formed daughter cells are released and start a new infectious cycle.

To promote adhesion to the vascular endothelium and prevent clearance by the spleen, the parasite remodels the infected red blood cells by exporting variant surface proteins (3). The virulence surface adhesion protein *P. falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) has been highly implicated in pathogenesis and is encoded by the well-studied *var* variant gene family. The *var* gene family is subject to monoallelic expression whereby only a single gene of the about 60-member gene family is transcribed while all others remain silenced (4). Switching between *var* genes allows the parasite to regularly expose surface proteins for which the human immune system is yet naïve. Transcription of the *var* gene peaks around 10–14 h post invasion (hpi) (5) and is repressed during later developmental stages, but the gene remains in a state 'poised' for reactivation in the next cycle (6,7).

Monoallelic expression in *P. falciparum* involves multiple layers of epigenetic regulation that mediate silencing, activation and poising of a single *var* gene while still allowing *var* gene switching rates optimized for efficient parasite infection and transmission (8–11). Post-translational modifications of histones and especially facultative heterochromatin regulated by various histone-modifying enzymes are required to establish a default silencing pathway for *var* genes (6,12–14). Heterochromatin Protein 1 silences *var* genes by binding to modified histones (15,16). Further, conserved genetic elements within the *var* gene play an important role in monoallelic expression. *Var* genes contain an upstream promoter sequence (ups), an exon 1 coding for the variant extracellular protein domain, a conserved intron and a highly conserved exon 2 containing the intracellular domain (Supplementary Figure S1A). A total of 36 *var* genes are subtelomeric and driven by subtype-specific pro-

*To whom correspondence should be addressed. Tel: +33 1 4061 3516; Email: julien.guizetti@pasteur.fr
Correspondence may also be addressed to Artur Scherf. Tel: +33 1 4568 8616; Email: artur.scherf@pasteur.fr
Present address: Julien Guizetti & Artur Scherf, Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, 25 Rue du Dr. Roux, 75724 Paris, France.

moters termed ups A or B (Supplementary Figure S1A). The 24 central *var* genes localize in 5 tandem arrays distributed over 4 chromosomes and are driven by ups C (or B/C) promoters (Figure 1A). The promoter contains motifs essential for monoallelic expression (8,17–19). The intron is thought to mediate *var* gene silencing by strict one-to-one pairing with the promoter and has bidirectional promoter activity (20–23).

In addition, spatial organization of *var* genes is tightly controlled and has been implied in *var* gene transcriptional control. All silent *var* genes are targeted to a few repressive clusters at the nuclear periphery except the active *var* gene, which localizes to a distinct expression site (14,24–27). Perinuclear localization is mediated by telomeric sequences as well as specific intronic motifs (25,28). Lastly, exonuclease activity is involved in controlling cryptic transcription of multiple *var* genes (29). Despite the tremendous advances in understanding regulation of monoallelic expression key questions remain unanswered. Default silencing and potential poising factors have been identified. The histone methyltransferase, PfSet10, associates with the poised *var* gene at trophozoite stage when transcription is halted for the rest of the blood stage cycle (7), but a factor specifically regulating the activation of *var* genes has yet to be identified. Hence, the perinuclear expression site remains undefined and no factors strictly associated with it have been found to date.

Non-coding RNAs (ncRNAs) are emerging as important regulators of eukaryotic gene expression and have been linked to monoallelic expression in other organisms (30). They frequently act as lynchpins for epigenetic regulators targeting non-specific enzymes to specific sites within the genome. Also, in *P. falciparum* ncRNAs have been associated with regulation of virulence genes (31,32). Various ncRNA transcripts have been detected by different methods (33–37). Telomeres produce long ncRNAs that localize to a distinct nuclear compartment that does not overlap with telomere clusters (38,39). Natural antisense transcripts are very frequent and associate with >24% of all open reading frames (36). Recently, transcription of TR2 ncRNA has been correlated with permeation pathway gene expression (40). *Var* genes produce 'sterile' transcripts from a bidirectional promoter within the intron (20,41). The function of the sense transcript produced from the intron of *var* genes remains elusive. The natural antisense RNA remains associated with chromatin and its transcription has been correlated with *var* gene activity (21,31,42). However, it is not yet clear whether this antisense RNA is cause or consequence of transcriptional activation.

The RNA of Unknown Function 6 (RUF-6) gene family encodes 15 ncRNAs with an unusually high GC-content for *P. falciparum* (>50% versus 20% genome-wide average) (43). Therefore, they have been referred to as GC-rich ncRNA or GC-rich elements (44–46). Their sequences are highly homologous (Supplementary Figure S1B) and interestingly, their gene loci are strictly associated with central *var* gene clusters where they are interspersed between *var* genes mostly in a tail-to-tail configuration (Figure 1A). Their transcripts have been detected in various transcriptomic studies, which indicate peak expression around late ring stage (34,36–38,47). Bioinformatical analysis shows conserved promoter elements, namely A- and B-box, within the GC-rich element sequence, suggesting transcription via RNA polymerase III (Supplementary Figure S1B) (48–50). A recent study shows that GC-rich ncRNA expression is clonally variant, although the members are not transcribed in a mutually exclusive fashion (51).

Here, we characterize GC-rich ncRNA using fluorescence *in situ* hybridization (FISH) and quantitative PCR. We find that even though GC-rich ncRNA genes can form multiple clusters within the nucleus, their transcripts localize to a single perinuclear locus in early ring stages. This locus colocalizes with the *var* gene expression site of central as well as subtelomeric *var* genes. Overexpression of distinct GC-rich ncRNAs causes activation of a specific subset of *var* genes and perturbs monoallelic expression at the single cell level. We uncover the first factor strictly associated with the active *var* gene expression site and provide functional characterization of a ncRNA gene in *P. falciparum*. These findings represent an important step toward better understanding of the mechanism of monoallelic expression in the malaria parasite.

MATERIALS AND METHODS

Parasite culture, synchronization and panning

P. falciparum blood stage parasites were cultivated as described previously (14). Parasite culture was synchronized by sorbitol lysis during ring stage, subsequent plasmagel enrichment in schizont stage, followed by another sorbitol treatment at 6 h post invasion. Synchronized parasites were harvested at 4% hematocrit and ∼3–5% parasitemia. Parasite development was monitored by Giemsa staining. To enrich for FCR3 parasites expressing var2csa panning on CSA receptor was carried out as previously described (52). In brief, cell culture flasks were coated with CSA receptor to specifically bind red blood cells infected with parasites expressing the PfEMP1 protein encoded by the *var2csa* gene.

RNA-FISH, DNA-FISH and immunofluorescence

Infected red blood cells were lysed with 0.015% Saponin in RPMI and the released parasites were fixed in suspension with 4% paraformaldehyde in phophate buffered saline (PBS) over night at 4°C. Parasites were then deposited on #1.5 cover glasses, permeabilized 15 min with 0.1% Triton-X-100 and subjected to DNA or RNA FISH as described previously in a detailed methods chapter (53). For sequential RNA-, DNA-FISH the transcript was labeled first with a GC-rich ncRNA probe. After another fixation step we carried out DNA-FISH using a probe labeling exclusively the non-transcribed 5′ region of all GC-rich genes. FISH probes were PCR amplified from genomic DNA using primers (Supplementary Table S1) and labeled by nick translation using FITC, DIG or Biotin High-Prime (Roche). RNA-FISH combined with immunofluorescence was carried out in an endogenously tagged PfSet10-HA parasite strain using anti-HA antibody (Roche, 3F10) at 1:200 in PBS/3%BSA as described previously (7,53,54).

RNA extraction and quantitative real-time PCR

RNA was extracted after Saponin lysis in 0.06% Saponin in PBS followed by one wash in PBS and resuspension in

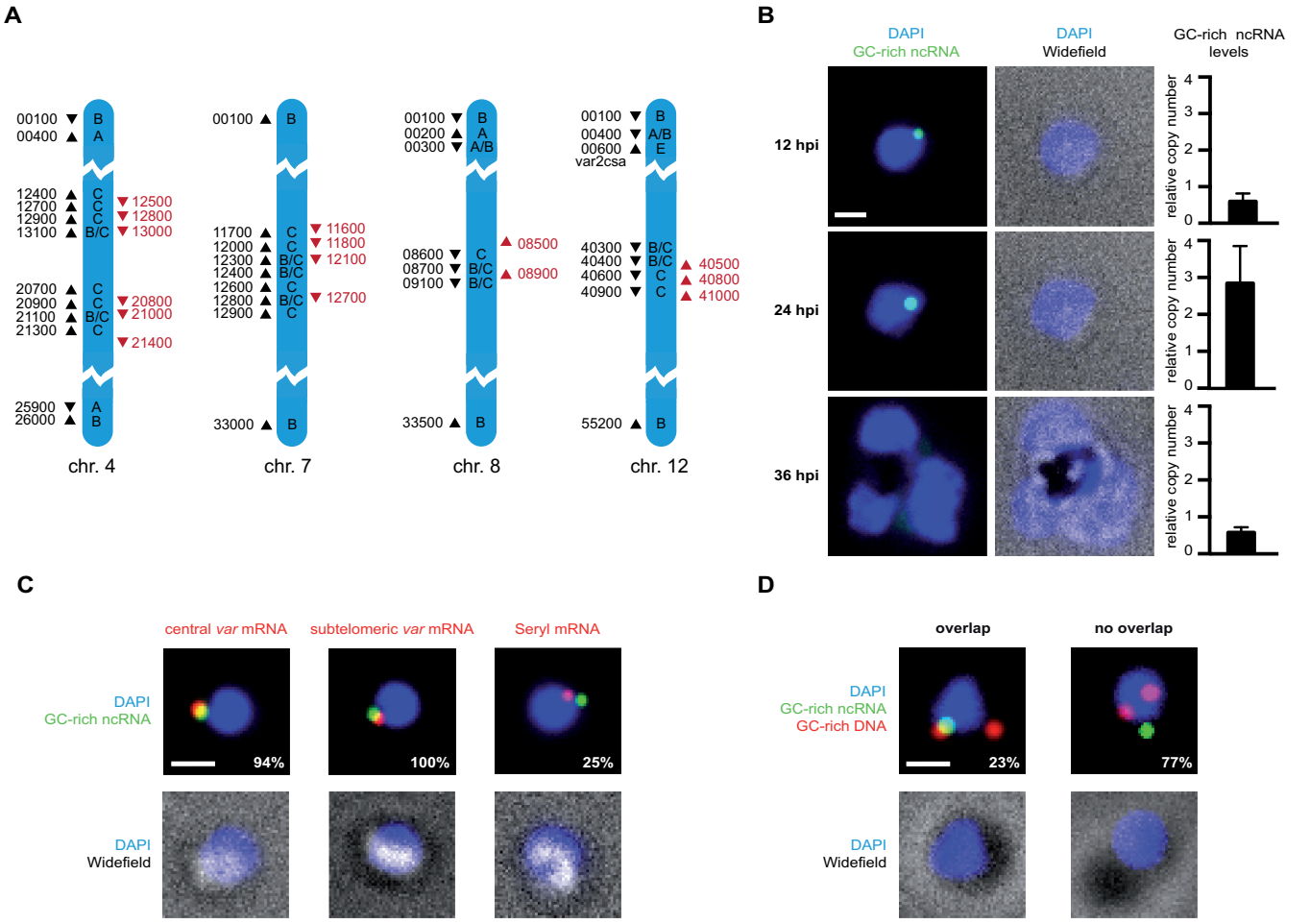


Figure 1. GC-rich ncRNA localizes to the perinuclear *var* gene expression site in trans. (**A**) Genomic organization of all GC-rich ncRNA elements (red arrowheads) and *var* genes (black arrowheads) on chromosomes 4, 7, 8 and 12. Only five last digits of gene IDs are displayed (PF3D7_chr#xxxxx). Direction of the arrowhead indicates orientation of gene. Drawing is not to scale and only parts chromosomes are shown. Upstream promoter sequence (ups) subtype for each *var* gene is labeled (A, A/B, B, B/C, C or E). (**B**) Fluorescent microscopy images of RNA-FISH labeling of parasites at different life cycle stages using a probe targeting all GC-rich ncRNA transcripts (green). All nuclei are stained with DAPI (blue) and overlayed with widefield image. All scale bars, 1 μm. On the right real-time PCR quantification of GC-rich ncRNA levels at the respective stages. (**C**) Colocalization analysis of GC-rich ncRNA and *var* gene expression site. Fluorescent microscopy images of ring stage parasites labeled in two colors with RNA-FISH probes targeting GC-rich ncRNAs (green) and mRNA (red) of a central (PF3D7_0412700) *var* (left), subtelomeric (*var2csa*) (middle) or Serine-tRNA ligase control gene (right). Percentages indicate frequency of colocalization. (**D**) Combination of DNA-FISH using a probe against GC-rich element gene loci (red) with RNA-FISH using probe targeting all GC-rich ncRNA transcript (green). Images of one representative cell with overlapping signal (left panels) and with non-overlapping signals (right panels) are shown. Percentages indicate occurrence of overlap or non-overlap within the imaged population.

Trizol. RNA was purified using the miRNeasy kit, which collects RNA fragments down to 20 bp in size (Qiagen). After on-column DNase treatment (Qiagen) reverse transcription of 200 ng of RNA was carried out using Superscript VILO kit (Life technologies) with random hexamer primers, to prevent loss of ncRNAs. cDNA levels were quantified in the CFX384 qPCR machine (Bio-Rad) using gene specific primers (Supplementary Table S2) and 2x Power SYBR Green master mix (Applied biosciences). Data were analyzed using the Bio-Rad CFX manager software after normalization to T-Serine ligase (PF3D7_0717700) transcription levels. GC-rich ncRNA was quantified by using two primer pairs covering small sequence variation in all members of the gene family. To assess total *var* gene levels expression values for all ups subtype specific primers were added. Quantitative real-time PCR experiments were carried out in triplicates. Limits of detection were calculated as mean of the blank plus three times the standard deviation.

Plasmid construction and transfection

To construct overexpression vectors synthetic gene fragments containing parts of U6 promoter either coding for a GC-rich ncRNA gene PF3D7_1241000, PF3D7_0808500 or a Luciferase control fragment of same size were ordered (Custom gene synthesis, Genscript) and cloned into pL6-egfp vector previously described in (55) using DraII and SwaI restriction site followed by Gibson assembly (Infusion, HD Clontech). pU6-Luc, pU6-GC12 and pU6-GC08 constructs were transfected as described previously (56) and maintained under WR99210 drug selection pressure.

Image acquisition and analysis

Images were captured using a Nikon Eclipse 80i microscope equipped with 100x NA 1.42 objective and CoolSnap HQ2 camera (Photometrics). Camera pixel size corresponds to 64 nm. NIS Elements 3.0 software (Nikon) was used for acquisition and Fiji (http://fiji.sc) for analysis. Numbers of FISH dots were quantified counting all cells positive for staining. Colocalization was scored in cell nuclei positive for both fluorescent signals when they overlapped while using the same contrast adjustments throughout all images. Scoring was performed by direct optical observation using Fiji (http://fiji.sc). For each analysis images from at least two independent replicas were combined.

RESULTS

GC-rich ncRNA localizes to the perinuclear *var* gene expression site in trans

GC-rich gene loci are adjacent to central *var* genes (Figure 1A), which have been reported to cluster at the nuclear periphery (14). Our DNA-FISH analysis confirms that GC-rich ncRNA genes form between 1–4 perinuclear clusters (n = 75, 48, 20 and 9, respectively) in a population of ring stage parasites prior to genome replication (Supplementary Figure S2). To characterize the localization pattern of GC-rich ncRNA transcripts we carried out RNA-FISH using a probe targeting all GC-rich ncRNAs. We analyzed fluorescent images of labeled cells at 12, 24 and 36 hpi (Figure 1B). Notably, at 12 hpi all labeled cells exhibited only a single perinuclear focus (n > 50). At 24 hpi most of the cells still show a single focus of increased fluorescence intensity. However, this signal was no longer particularly associated with the nuclear periphery. All detected signals were exclusively nuclear. In the trophozoite stage, at 36 hpi, we did not detect any cells with a distinct GC-rich ncRNA signal. Fluorescent signal intensities throughout the cell cycle correlate with the GC-rich ncRNA levels quantified by real time PCR carried out on RNA harvested at the same timepoints using primer pairs detecting all GC-rich ncRNA members (Figure 1B). Peak ncRNA levels occur at 24 hpi and drop again at 36 hpi. The lack of distinct GC-rich ncRNA signal at 36 hpi suggests that transcripts do not accumulate at a single site that can be detected by FISH. Next, we investigated the colocalization of GC-rich ncRNA with PfSet10, a histone methyltransferase that associates with the poised *var* gene during trophozoite stage (7). Due to the different temporal expression we could, however, not test for overlap between the PfSet10 and GC-rich ncRNA signal (Supplementary Figure S3). Importantly, the GC-rich ncRNA localization pattern that does not mimic the clustered distribution of the GC-rich gene loci in ring stage parasites, but suggests targeting to a specific perinuclear site.

Since the localization pattern of GC-rich ncRNA is reminiscent of the *var* gene expression site, we carried out dual color RNA-FISH of GC-rich ncRNA and *var* gene mRNA in ring stage parasites to test for colocalization. We used two different parasite lines that express either a central *var* gene (PF3D7_0412700) or the subtelomeric *var2csa* gene. As a negative control we used a T-serine ligase RNA (PF3D7_0717700, see Supplementary Figure S1A) probe. We found that GC-rich ncRNA signal overlapped with the central *var* mRNA in 30 of 32 cells, with *var2csa* mRNA in 19 out of 19 cells whereas T-Serine ligase mRNA showed colocalization in only 3 out of 16 cases (Figure 1C). While the central *var* gene could colocalize with GC-rich ncRNA due to spatial proximity of both genes on the chromosome, *var2csa* mRNA is expressed from a subtelomeric locus distant from any GC-rich element cluster (Supplementary Figure S1A). These data on GC-rich ncRNA localization demonstrate the first specific targeting of a molecular factor to the active *var* gene expression site.

Colocalization of *var* mRNA and GC-rich ncRNA could occur by two different mechanisms. The GC-rich ncRNA and active *var* gene loci could loop together via a long-range chromosomal interaction. Alternatively, GC-rich ncRNA could freely diffuse before being targeted to the *var* gene expression site. To test these hypotheses we combined DNA-FISH of the GC-rich gene loci with RNA-FISH of the GC-rich ncRNA transcript (Figure 1D). After FISH labeling GC-rich ncRNA in cells primarily expressing the subtelomeric *var2csa* gene, we labeled the GC-rich gene locus with a DNA-FISH probe exclusively targeting the 5′ region of all gene family members to avoid any recognition of the transcript. We found overlap between signals in only 7 out of 30 cells. These data suggest that GC-rich ncRNA can relocate to the *var* gene expression site independently of its gene locus.

Overexpression of GC-rich ncRNAs activates distinct subsets of *var* genes

Despite their high sequence similarity (Supplementary Figure S1B), *in silico* prediction using the RNAfold webserver indicates some structural variability within the gene family resulting in four major ‘classes’ of RNA secondary structures (Figure 2A). While the lower stem loop structure is highly preserved, the ‘head’ region displays much more flexible base pairing configurations. Conservation of specific promoter elements, the A- and B-box, suggests that RNA Polymerase III drives GC-rich ncRNA transcription (Supplementary Figure S1B). Hence, we designed a construct overexpressing GC-rich ncRNA under the highly active Pol III-dependent U6 promoter (Figure 2B). We inserted GC-rich ncRNA sequences of two distinct structural variant types (PF3D7_1241000, PF3D7_0808500) or a Luciferase control fragment of similar length and GC content into the previously described pL6-egfp vector backbone (55), generating pU6-GC12, pU6-GC08 and pU6-Luc, respectively. We transfected these constructs into a 3D7 parasite line and isolated one clone carrying pU6-Luc and two independent clones each carrying pU6-GC12 and pU6-GC08. Quantitative real time PCR analysis of cDNA derived from RNA harvested at 12 hpi demonstrated that pU6-GC12 and pU6-GC08 transfected clones have increased in GC-rich ncRNA levels (Figure 2C). Total *var* RNA levels, as quantified by a subset of qPCR primers covering the entire gene family, are also increased in GC-rich ncRNA overexpression clones.

Next, we analyzed the RNA levels for all individual *var* genes in the five clones also using T-Serine RNA ligase levels for normalization (Figure 2D). The pU6-Luc transfected clone expressed only a single *var* gene, displaying func-

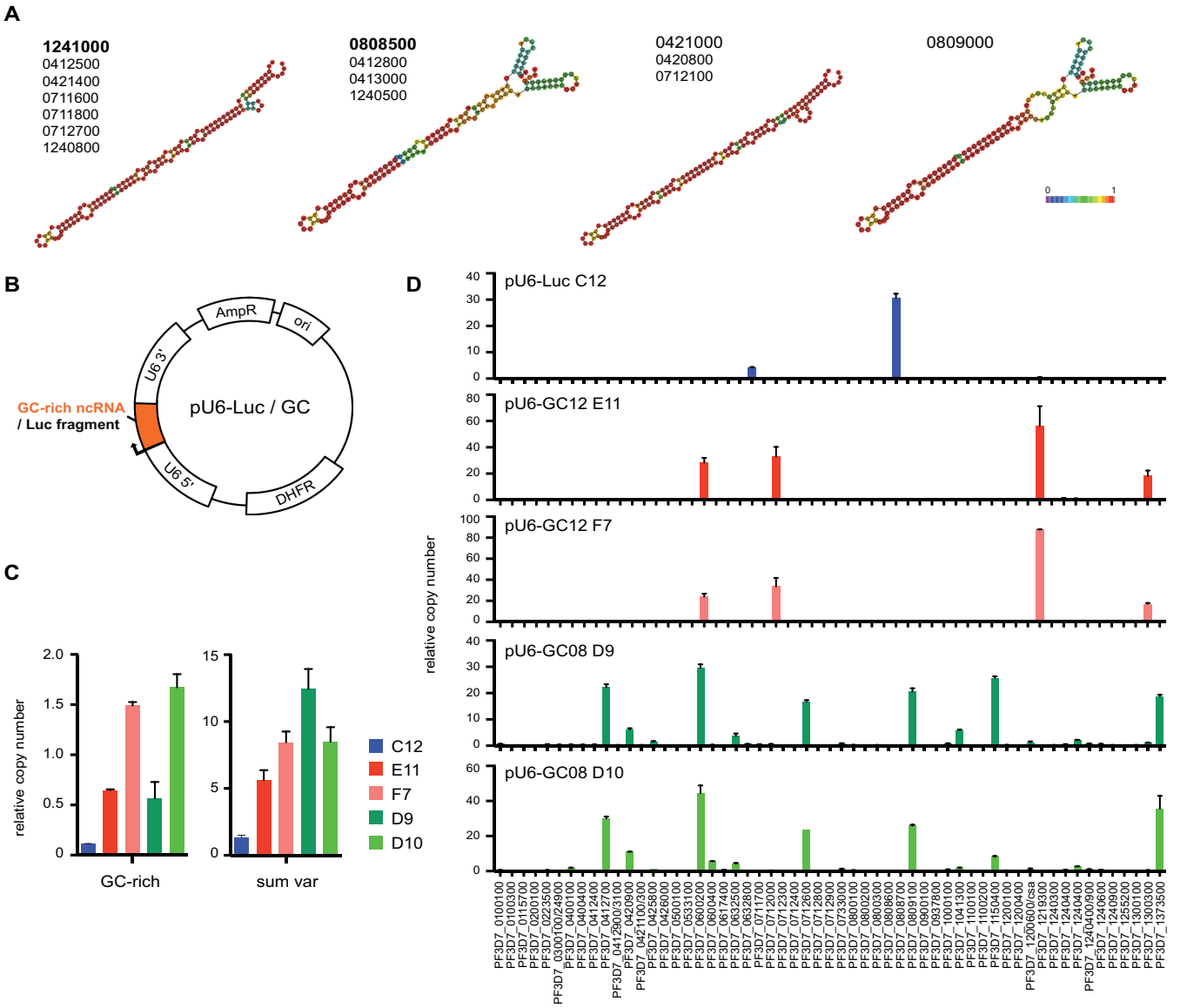


Figure 2. GC-rich ncRNA overexpression causes activation of distinct subsets of *var* genes. (**A**) Bioinformatic RNA structure prediction using RNAfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). Four representative structures covering diversity within entire GC-rich ncRNA gene family are shown. Members with similar structure are listed below. Colored scale represents relative base pairing probabilities. (**B**) Plasmid map of pU6-GC or pU6-Luc constructs used to overexpress GC-rich ncRNA (PF3D7_1241000 and PF3D7_0808500) or Luciferase control fragment, respectively. (**C**) Real time qPCR quantification of GC-rich ncRNA and total *var* mRNA levels in pU6-Luc transfected clone (C12, blue), two pU6-GC12 transfected clones (E11 and F7, red) and two pU6-GC08 clones (D9 and D10, green). (**D**) Expression profile of all *var* gene mRNAs as quantified by real time qPCR for pU6-Luc and pU6-GC clones. Relative copy numbers are normalized to T-Serine ligase (PF3D7_0717700) transcription levels. Limit of detection (LOD) for this data set is 0.0114.

tional monoallelic expression. Notably, both GC12 overexpressing clones primarily transcribed the same four *var* genes, while both GC08 overexpressing clones transcribed the same set of six *var* genes. Despite transfections being carried out in the same 3D7 parent strain, the expression profiles for pU6-Luc, pU6-GC12 and pU6-GC08 did not have any *var* genes in common except for PF3D7_0600200 (Figure 2D and Supplementary Figure S4). RNA levels of up-regulated *var* genes varied slightly within the same cloned parasite line, but the profile was virtually identical between two clones. Given, that we observe a similar subset of *var* gene expression in bulk culture before the cloning step, overexpression of distinct GC ncRNA could lead to the activation of that specific *var* gene subset (Supplementary Figure S4). The temporal regulation of *var* and GC-rich ncRNA levels was not altered in mutant parasite clones (Supplementary Figure S5A). Lastly, to interfere with GC-rich ncRNA function, we also attempted overexpression of GC-rich antisense RNA in the same vector, which did not cause a significant change in GC-rich ncRNA levels or affected *var* gene mRNA levels (Supplementary Figure S5B). Taken together, our data suggest that distinct GC-rich ncRNA members activate distinct subsets of *var* genes establishing a 'hardwired' expression profile.

GC-rich ncRNA overexpression perturbs monoallelic *var* gene expression at single cell level

Our quantitative PCR data suggest that GC-rich ncRNA overexpression interferes with monoallelic *var* gene expression. To test this in individual parasites we carried out multi-color RNA-FISH with gene specific probes targeting three of the highly expressed *var* mRNAs in the context of GC-rich ncRNA (PF3D7_1241000) overexpression (Figure 2D). Parasites carrying the pU6-Luc control construct show a single *var* gene (PF3D7_0808600) signal colocalizing with GC-rich ncRNA in 20 out of 21 cells (Figure 3A). When labeling GC-rich ncRNA overexpressing parasites with probes against a central *var* mRNA (PF3D7_0712000) and *var2csa* mRNA, we found that 31 out of 59 labeled cells were positive for both signals (Figure 3B). We found similar results when combining the central *var* mRNA (PF3D7_0712000) probe with another subtelomeric *var* mRNA (PF3D7_1300300) probe (double staining in 12 out of 19 cells). The lack of double staining in about half of cells could be explained by incomplete labeling of all cells by all probes. Nevertheless, our data show that overexpression of GC-rich ncRNA can disrupt monoallelic expression at single cell level, linking a trans-acting ncRNA to antigenic variation in malaria parasites.

DISCUSSION

In this work, we demonstrate the specific targeting of GC-rich ncRNA to the actively transcribed expression site of *var* genes. This is the first identification of a marker molecule for a nuclear structure that remains ill-defined despite being a hallmark for monoallelic expression in protozoan pathogens such as *P. falciparum* and African Trypanosomes (57). It has been postulated that, in the context of monoallelic expression, an as yet unknown limiting factor exists that allows for the activation of only a single *var* gene locus (10,18). GC-rich ncRNA is a potential candidate for such a limiting factor. Indeed, when we overexpress GC-rich ncRNA, we detect multiple different *var* transcripts in a single cell. Our study predicts that GC-rich ncRNA knockdown mutants could reduce *var* gene expression. However, due to the presence of 15 GC-rich ncRNA genes generating mutants by sequential knockout, despite the recent introduction of CRISPR technology to malaria parasites, represents a long-term approach (55,58). A recent study observed only a minor effect of GC-rich ncRNA overexpression on *var* gene transcription (51). Important technical differences used in their study such as the use of a Pol II instead of a Pol III promoter for the production of the GC-rich ncRNA could explain the observed phenotype. In addition, their study focused only on central *var* genes, omitting most members of the multigene family represented by the subtelomeric *var* genes from their transcriptional analysis (51).

Previous studies in *P. falciparum* described the transcription and localization of various ncRNA species such as subtelomeric RNA and natural antisense RNA, but their mechanisms or biological roles are still unclear (32). In this study, we present experimental data that posits GC-rich ncRNA as a trans-acting factor that contributes to the control of *var* gene activation at the perinuclear expression site. Because GC-rich ncRNA elements are only in close proximity to central *var* genes (Supplementary Figure S1), a cis-regulatory function has been suggested for these elements. Furthermore, their positioning in intergenic spaces and recent experimental evidence based on GC-rich element associated silencing of an adjacent GFP gene on episomal constructs has led to speculation that they might act as insulator elements that prevent spreading of epigenetic marks (51). While this hypothesis remains possible, our data strongly support a general trans-acting function linked to monoallelic expression of all *var* gene members (Figure 3C). The mechanism by which GC-rich ncRNA is targeted to the expression site remains unclear. Bioinformatic prediction, however, suggests a relatively strong secondary RNA structure (Figure 2A), which could mediate specific RNA–RNA and RNA–protein interactions (34,43). Since cryptic var gene transcription has recently been demonstrated (29) GC-rich ncRNA could, hypothetically, bind to nascent var mRNA to target transcriptional activation factors to the expression site. Transcriptional activation would generate more var RNA recruiting more GC-rich ncRNA thereby generating the positive feedback loop that drives singular gene choice. A similar mechanism has recently been described in the Epstein-Barr virus (59). Here, a viral ncRNA, EBER2, recruits a host transcription factor to the viral DNA locus via base-pairing interactions with nascent transcript and thereby modulates transcription. Hence, it is of utmost importance to identify interacting proteins to obtain insight into the underlying molecular mechanism by which this ncRNA regulates monoallelic expression. It is noteworthy that GC-rich ncRNA levels peak only at 24 hpi when *var* gene transcription is already repressed. GC-rich ncRNA function at this stage remains elusive, but despite its distinct colocalization with *var* genes, it is possible that it is linked to the expression of other clonally variant genes such as rifins as well (60,61). RNA-seq analysis of parasites that overexpress GC-rich ncRNA may reveal regulation of other target genes or gene families. Interestingly, *Plasmodium reichenowi*, the only other species carrying a GC-rich ncRNA gene family with almost identical sequences and structures, is also the only other sequenced malaria species encoding the *var* and *rifin* gene families (plasmodb.org).

The most striking result of this study is that overexpression of a single member of the GC-rich gene family disrupts monoallelic gene counting. Interestingly, five out of nine of the preferentially activated *var* genes are subtelomeric, emphasizing that GC-rich ncRNA function is not restricted to central *var* genes. Further, in distinct overexpression clones, the same subset of *var* genes is up-regulated, indicating a 'hardwired' activation program. A similar observation was made when the histone deacetylase PfSir2a was inactivated, resulting in the de-repression of a specific subset of *var* genes (62). Considering that GC-rich ncRNAs are clonally variant it is tempting to speculate that individual GC-rich elements could have a preference for specific subsets of *var* genes and thus, decide on which silent *var* gene may be activated at the next switch in expression. Given that switching is not understood in malaria parasites at the molecular level, our work points to a promising new research avenue that may lead to a better insight into a process that is key to antigenic variation.

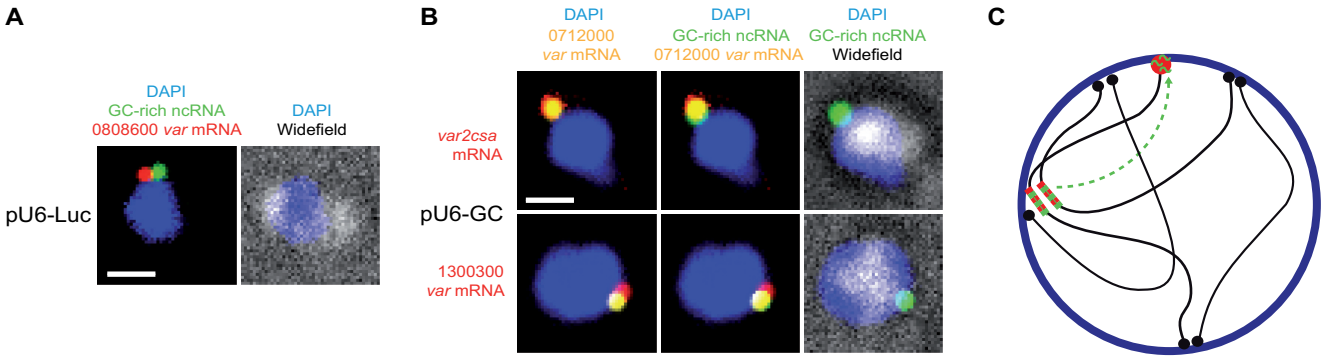


Figure 3. GC-rich ncRNA overexpression disrupts monoallelic expression. (**A**) Fluorescent microscopy images of multicolor RNA-FISH labeling. pU6-Luc transfected parasites were labeled with probe against GC-rich ncRNA (green) and *var* mRNA PFD3D7_0808600 (red). (**B**) pU6-GC transfected parasites were triple labeled with probe against GC-rich ncRNA (green), *var* mRNA PFD3D7_0712000 (yellow) and *var2csa* mRNA (red, middle panels) or *var* mRNA PFD3D7_1300300 (red, lower panels). Nuclei are stained with DAPI (blue) and overlayed with widefield image. Scale bars, 1 μm. (**C**) Schematic working model shows relocalization of GC-rich ncRNA (green) from central *var* gene clusters (red-green stripes) to the distinct perinuclear *var* expression site (red). Telomeres (black) cluster at the nuclear membrane (blue).

Many years after the discovery of monoallelic *var* gene expression, we are only now beginning to understand the elusive expression site. These findings expand our knowledge about ncRNA genes as novel players in monoallelic expression and will open avenues to deepen our understanding of antigenic variation in the malaria parasite and other protozoan pathogens.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Allon Weiner, Shruthi Vembar and Jessica Bryant for critical reading of the manuscript, Sebastian Baumgarten for help with data analysis and Jose-Juan Lopez-Rubio for providing reagents.

FUNDING

European Research Council Advanced [PlasmoEscape 250320, PlasmoSilencing 670301]; Agence Nationale de la Recherche [ANR-11-LABEX-0024-01 ParaFrap, ANR-13-ISV3-0003-01 NSFC MalVir]; Human Frontier Science Program fellowship [LT000620/2012-L to J.G.].
Conflict of interest statement. None declared.

REFERENCES

1. Snow,R.W., Guerra,C.A., Noor,A.M., Myint,H.Y. and Hay,S.I. (2005) The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature*, **434**, 214–217.
2. Murray,C.J., Rosenfeld,L.C., Lim,S.S., Andrews,K.G., Foreman,K.J., Haring,D., Fullman,N., Naghavi,M., Lozano,R. and Lopez,A.D. (2012) Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet*, **379**, 413–431.
3. Maier,A.G., Cooke,B.M., Cowman,A.F. and Tilley,L. (2009) Malaria parasite proteins that remodel the host erythrocyte. *Nat. Rev. Microbiol.*, **7**, 341–354.
4. 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**, 5418–5426.
5. Schieck,E., Pfahler,J.M., Sanchez,C.P. and Lanzer,M. (2007) Nuclear run-on analysis of var gene expression in Plasmodium falciparum. *Mol. Biochem. Parasitol.*, **153**, 207–212.
6. 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. Microbiol.*, **66**, 1296–1305.
7. Volz,J.C., Bártfai,R., Petter,M., Langer,C., Josling,G.A., Tsuboi,T., Schwach,F., Baum,J., Rayner,J.C., Stunnenberg,H.G. *et al.* (2012) PfSET10, a Plasmodium falciparum methyltransferase, maintains the active var gene in a poised state during parasite division. *Cell Host Microbe*, **11**, 7–18.
8. Dzikowski,R., Li,F., Amulic,B., Eisberg,A., Frank,M., Patel,S., Wellems,T.E. and Deitsch,K.W. (2007) Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO Rep.*, **8**, 959–965.
9. Scherf,A., Lopez-Rubio,J.J. and Riviere,L. (2008) Antigenic variation in Plasmodium falciparum. *Annu. Rev. Microbiol.*, **62**, 445–470.
10. Guizetti,J. and Scherf,A. (2013) Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum. *Cell Microbiol.*, **15**, 718–726.
11. Chookajorn,T., Ponsuwanna,P. and Cui,L. (2008) Mutually exclusive var gene expression in the malaria parasite: multiple layers of regulation. *Trends Parasitol.*, **24**, 455–461.
12. 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**, 899–902.
13. Coleman,B.I., Skillman,K.M., Jiang,R.H.Y., Childs,L.M., Altenhofen,L.M., Ganter,M., Leung,Y., Goldowitz,I., Kafsack,B.F.C., Marti,M. *et al.* (2014) A Plasmodium falciparum Histone Deacetylase Regulates Antigenic Variation and Gametocyte Conversion. *Cell Host Microbe*, **16**, 177–186.
14. Lopez-Rubio,J.J., Mancio-Silva,L. and Scherf,A. (2009) Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe*, **5**, 179–190.
15. Perez-Toledo,K., Rojas-Meza,A.P., Mancio-Silva,L., Hernandez-Cuevas,N.A., Delgadillo,D.M., Vargas,M., Martinez-Calvillo,S., Scherf,A. and Hernandez-Rivas,R. (2009) Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res*, **37**, 2596–2606.
16. Brancucci,N.M., Bertschi,N.L., Zhu,L., Niederwieser,I., Chin,W.H., Wampfler,R., Freymond,C., Rottmann,M., Felger,I., Bozdech,Z. *et al.* (2014) Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe*, **16**, 165–176.
17. Brancucci,N.M., Witmer,K., Schmid,C.D., Flueck,C. and 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.

18. Dzikowski,R., Frank,M. and Deitsch,K. (2006) Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathog.*, **2**, e22.
19. Voss,T.S., Healer,J., Marty,A.J., Duffy,M.F., Thompson,J.K., Beeson,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**, 1004–1008.
20. Calderwood,M.S., Gannoun-Zaki,L., Wellems,T.E. and 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.
21. 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**, 116–127.
22. Frank,M., Dzikowski,R., Costantini,D., Amulic,B., Berdougo,E. and 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.
23. Gannoun-Zaki,L., Jost,A., Mu,J., Deitsch,K.W. and Wellems,T.E. (2005) A silenced Plasmodium falciparum var promoter can be activated in vivo through spontaneous deletion of a silencing element in the intron. *Eukaryot. Cell*, **4**, 490–492.
24. Duraisingh,M.T., Voss,T.S., Marty,A.J., Duffy,M.F., Good,R.T., Thompson,J.K., Freitas-Junior,L.H., Scherf,A., Crabb,B.S. and Cowman,A.F. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. *Cell*, **121**, 13–24.
25. Figueiredo,L.M., Freitas-Junior,L.H., Bottius,E., Olivo-Marin,J.C. and Scherf,A. (2002) A central role for Plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J.*, **21**, 815–824.
26. Freitas-Junior,L.H., Bottius,E., Pirrit,L.A., Deitsch,K.W., Scheidig,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. falciparum. *Nature*, **407**, 1018–1022.
27. Ralph,S.A., Scheidig-Benatar,C. and 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.
28. 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**, 451–463.
29. Zhang,Q., Siegel,T.N., Martins,R.M., Wang,F., Cao,J., Gao,Q., Cheng,X., Jiang,L., Hon,C.C., Scheidig-Benatar,C. *et al.* (2014) Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. *Nature*, **513**, 431–435.
30. Yang,P.K. and Kuroda,M.I. (2007) Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell*, **128**, 777–786.
31. Amit-Avraham,I., Pozner,G., Eshar,S., Fastman,Y., Kolevzon,N., Yavin,E. and 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.
32. 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.
33. Li,F., Sonbuchner,L., Kyes,S.A., Epp,C. and Deitsch,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**, 5692–5698.
34. Mourier,T., Carret,C., Kyes,S., Christodoulou,Z., Gardner,P.P., Jeffares,D.C., Pinches,R., Barrell,B., Berriman,M., Griffiths-Jones,S. *et al.* (2008) Genome-wide discovery and verification of novel structured RNAs in Plasmodium falciparum. *Genome Res.*, **18**, 281–292.
35. 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. *et al.* (2010) A global view of the nonprotein-coding transcriptome in Plasmodium falciparum. *Nucleic Acids Res.*, **38**, 608–617.
36. Siegel,T.N., Hon,C.C., Zhang,Q., Lopez-Rubio,J.J., Scheidig-Benatar,C., Martins,R.M., Sismeiro,O., Coppee,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**, 150–165.
37. Wei,C., Xiao,T., Zhang,P., Wang,Z., Chen,X., Zhang,L., Yao,M., Chen,R. and Wang,H. (2014) Deep profiling of the novel intermediate-size noncoding RNAs in intraerythrocytic Plasmodium falciparum. *PLoS One*, **9**, e92946.
38. 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. *et al.* (2011) A global transcriptional analysis of Plasmodium falciparum malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol.*, **12**, R56.
39. Sierra-Miranda,M., Delgadillo,D.M., Mancio-Silva,L., Vargas,M., Villegas-Sepulveda,N., Martinez-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**, 36–47.
40. Rovira-Graells,N., Crowley,V.M., Bancells,C., Mira-Martinez,S., Ribas de Pouplana,L. and Cortes,A. (2015) Deciphering the principles that govern mutually exclusive expression of Plasmodium falciparum clag3 genes. *Nucleic Acids Res.*, **43**, 8343–8357.
41. Kyes,S.A., Christodoulou,Z., Raza,A., Horrocks,P., Pinches,R., Rowe,J.A. and Newbold,C.I. (2003) A well-conserved Plasmodium falciparum var gene shows an unusual stage-specific transcript pattern. *Mol. Microbiol.*, **48**, 1339–1348.
42. Jiang,L., Mu,J., Zhang,Q., Ni,T., Srinivasan,P., Rayavara,K., Yang,W., Turner,L., Lavstsen,T., Theander,T.G. *et al.* (2013) PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. *Nature*, **499**, 223–227.
43. Chakrabarti,K., Pearson,M., Grate,L., Sterne-Weiler,T., Deans,J., Donohue,J.P. and Ares,M.J. (2007) Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA*, **13**, 1923–1939.
44. Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S. *et al.* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, **419**, 498–511.
45. Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D., Mungall,K., Bowman,S., Atkin,R., Baker,S. *et al.* (2002) Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13. *Nature*, **419**, 527–531.
46. 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**, 149–158.
47. Otto,T.D., Wilinski,D., Assefa,S., Keane,T.M., Sarry,L.R., Bohme,U., Lemieux,J., Barrell,B., Pain,A., Berriman,M. *et al.* (2010) New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol. Microbiol.*, **76**, 12–24.
48. Dieci,G., Conti,A., Pagano,A. and Carnevali,D. (2013) Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. *Biochim. Biophys. Acta*, **1829**, 296–305.
49. Marck,C., Kachouri-Lafond,R., Lafontaine,I., Westhof,E., Dujon,B. and 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.
50. Schramm,L. and Hernandez,N. (2002) Recruitment of RNA polymerase III to its target promoters. *Genes Dev*, **16**, 2593–2620.
51. Wei,G., Zhao,Y., Zhang,Q. and 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.
52. Pouvelle,B., Meyer,P., Robert,C., Bardel,L. and Gysin,J. (1997) Chondroitin-4-sulfate impairs in vitro and in vivo cytoadherence of Plasmodium falciparum infected erythrocytes. *Mol. Med.*, **3**, 508–518.
53. Mancio-Silva,L. and 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.
54. Guizetti,J., Martins,R.M., Guadagnini,S., Claes,A. and Scherf,A. (2013) Nuclear pores and perinuclear expression sites of var and

ribosomal DNA genes correspond to physically distinct regions in plasmodium falciparum. *Eukaryot. Cell*, **12**, 697–702.
55. Ghorbal,M., Gorman,M., Macpherson,C.R., Martins,R.M., Scherf,A. and Lopez-Rubio,J.J. (2014) Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat. Biotechnol.*, **32**, 819–821.
56. Hasenkamp,S., Merrick,C.J. and Horrocks,P. (2013) A quantitative analysis of Plasmodium falciparum transfection using DNA-loaded erythrocytes. *Mol. Biochem. Parasitol.*, **187**, 117–120.
57. Borst,P. (2002) Antigenic variation and allelic exclusion. *Cell*, **109**, 5–8.
58. MacPherson,C.R. and Scherf,A. (2015) Flexible guide-RNA design for CRISPR applications using Protospacer Workbench. *Nat. Biotechnol.*, **33**, 805–806.
59. Lee,N., Moss,W.N., Yario,T.A. and Steitz,J.A. (2015) EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. *Cell*, **160**, 607–618.
60. Goel,S., Palmkvist,M., Moll,K., Joannin,N., Lara,P., Akhouri,R.R., Moradi,N., Öjemalm,K., Westman,M., Angeletti,D. *et al.* (2015) RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. *Nat. Med.*, **21**, 314–317.
61. Kyes,S.A., Rowe,J.A., Kriek,N. and Newbold,C.I. (1999) Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. *Proc. Natl. Acad. Sci. U.S.A*, **96**, 9333–9338.
62. Tonkin,C.J., Carret,C.K., Duraisingh,M.T., Voss,T.S., Ralph,S.A., Hommel,M., Duffy,M.F., Silva,L.M., Scherf,A., Ivens,A. *et al.* (2009) Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum. *PLoS Biol.*, **7**, e84.

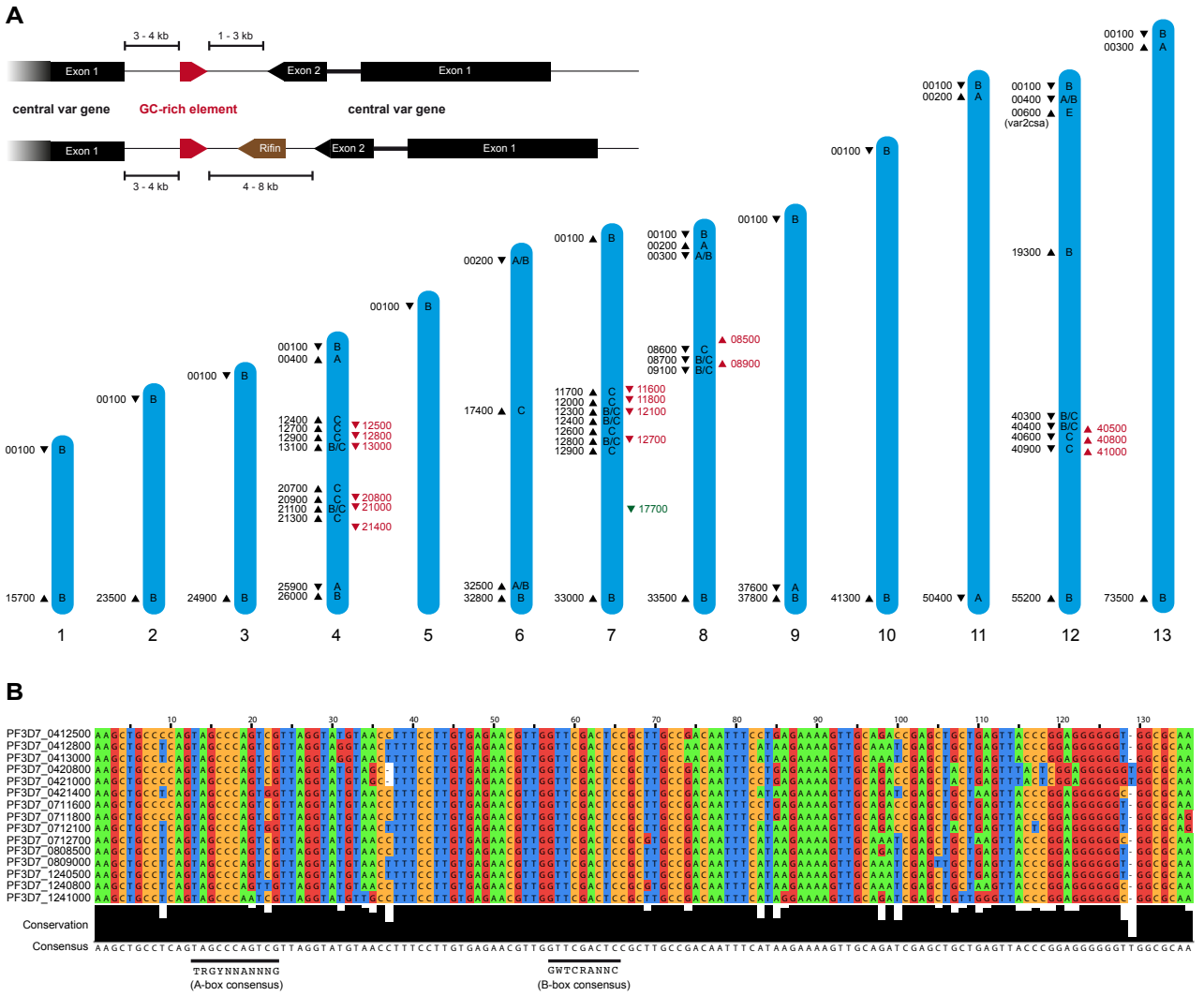
SUPPLEMENTARY DATA

Trans-acting GC-rich non-coding RNA at *var* expression site modulates gene counting in malaria parasites

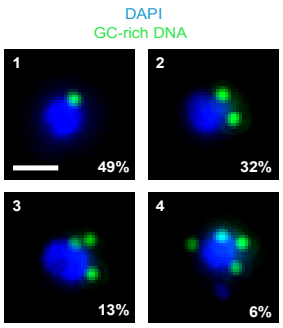
Julien Guizetti[1,2,3,*], Anna Barcons-Simon[1,2,3] and Artur Scherf[1,2,3,*]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[*]Correspondence: julien.guizetti@pasteur.fr, artur.scherf@pasteur.fr

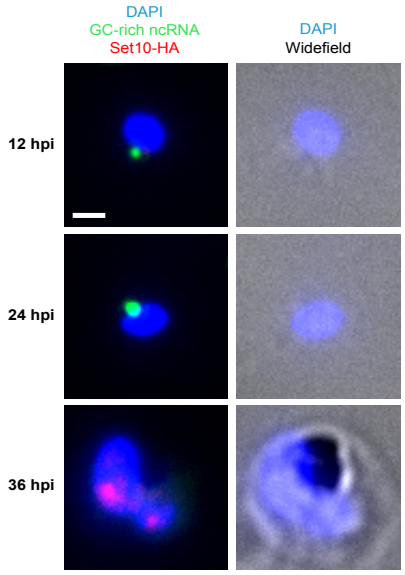
Published in Nucleic Acids Research. 2016 July 27
doi: 10.1093/nar/gkw664



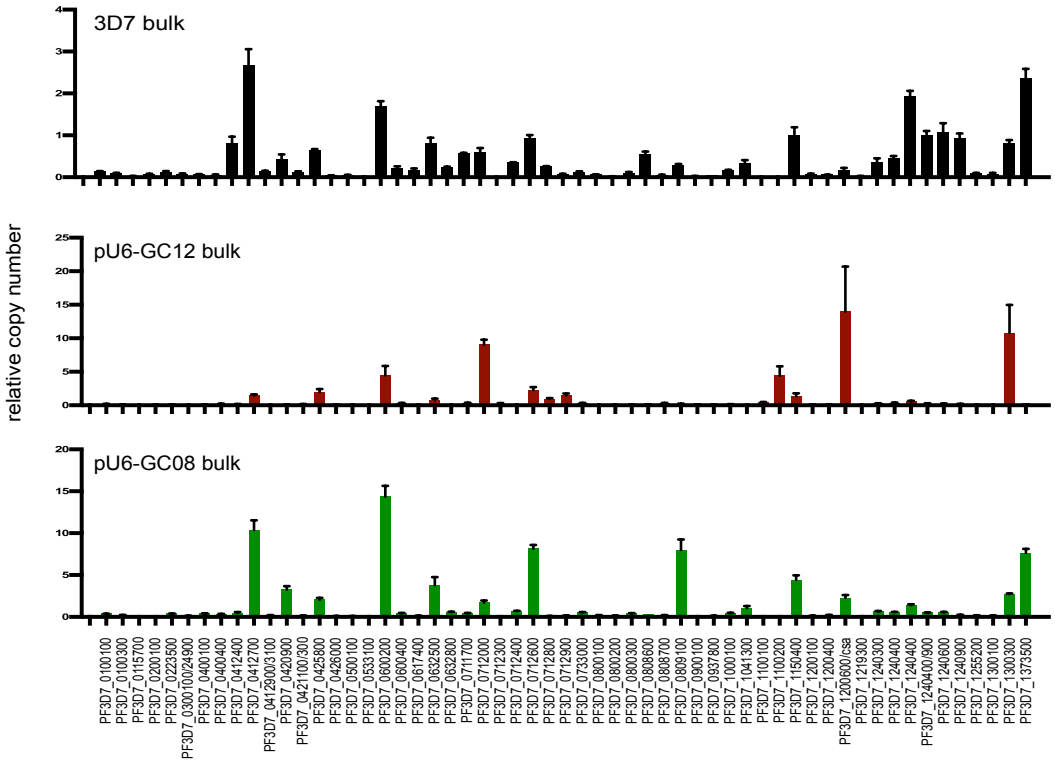
Supplementary Figure 1. Genomic organization and sequence alignment of GC-rich ncRNA gene family. (**A**) Genomic organization of GC-rich elements and shows proximity to central var genes. Top panel displays two occuring genomic arrangements for GC-rich elements. In most cases they are in a tail to tail conformation with the closest var gene (9/15) while sometimes (5/15) a rifin gene is intercalated between them. Schematics below show 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). (**B**) Multiple sequence alignement of highly conserved GC-rich non-coding 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).



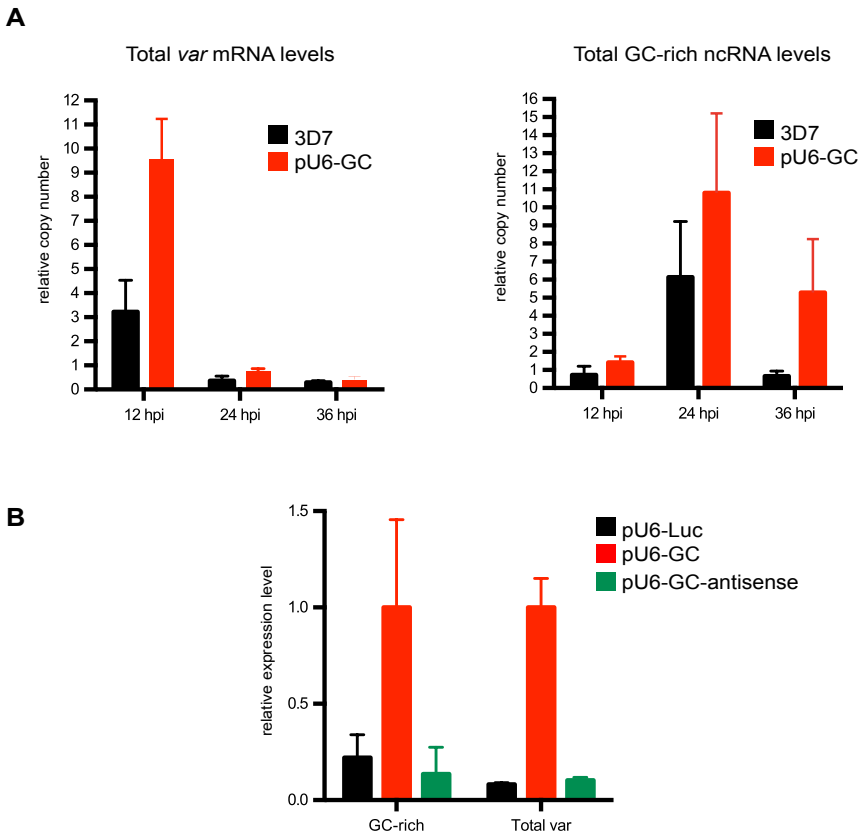
Supplementary Figure 2. Perinuclear clustering of GC-rich ncRNA gene loci. Fluorescent microscopy images of DNA-FISH labeling of four different ring stage parasites using a probe targeting all GC-rich elements (green). Nucleus is stained with DAPI (blue). Scale bar, 1 μm. Percentages indicate occurrence of cells with 1,2,3, or 4 foci within the imaged population.



Supplementary Figure 3. Differential temporal expression of PfSet10 and GC-rich ncRNA. Fluorescence images show combination of RNA-FISH using probe against all GC-rich ncRNA trasncripts (red) with immunolabeling of endogeneously tagged PfSet10-HA using anti-HA antibody (red) for three stages of the blood stage cycle. Nucleus is stained with DAPI (blue) and overlayed with widefield on the right. Scale bar, 1 μm. We used identic acquisition parametres and contrast settings for all images.



Supplementary Figure 4. *Var* gene mRNA profile in GC-rich ncRNA overexpressing bulk cultures and control strains. Individual *var* gene mRNA levels are quantified by real time PCR as relative copy numbers after normalization to T-Serine Ligase levels. Error bars shown standard deviation within triplicates. Limit of detection (LOD) for this data set is 0.0114 (relative copy number).



Supplementary Figure 5. Temporal regulation of *var* and GC-rich ncRNA gene expression is not affected by overexpression of GC-rich ncRNA or GC-rich ncRNA antisense. (**A**) Total var gene mRNA levels are quantified by real time PCR using ups subtype specific primers covering all genes at three timepoints of the blood stage development. *Var* mRNA levels of 3D7 wildtype (black) and GC-rich ncRNA overexpressing (orange) parasites peak at 12 hpi. GC-rich ncRNA levels are also quantified for both parasite lines and peak at 24 hpi. Limit of detection (LOD) for this data set is 0.0114 (relative copy number). (**B**) Total *var* gene mRNA and GC-rich ncRNA levels are quantified by real time PCR in three different parasite lines at 12 hpi. Control line pU6-Luc expresses luciferase fragment (black), pU6-GC line overexpresses GC1241000 ncRNA (red), and pU6-GC-antisense GC1241000 antisense transcript (green). No significant difference between pU6-Luc and pU6-GC-antisense can be detected. Graph shows relative expression levels. Limit of detection (LOD) for this graph is 0.0019 (relative expression level).

Supplementary Table 1. Primer pairs for FISH probes		
Target	Forward primer	Reverse primer
Seryl (0717700)	AGGAGGGAATCCTGACAAGA	AAGAAGACAACATAAGAATC
FBA (1444800)	ATTAGCAACCACCGCCCAAA	CCTGCATTTTCACCACCTGC
GC-rich-ncRNAs	CTGCCTCAGTAGCCCAGTCG	CGGGTAACTCAGCAGCTCGAT
var2csa (1200600)	AGCTGATCCTAGTGAAGTG	TGAAGTATCTTGTTCAGCAG
upsA-subtype var	ATTCCATACATCCGATATAGG	CCGAAATCACCTGTTGACCTC
central var (0412700)	CTA GTG GAT GAG GAT TGG GTG	CTA GTG GAT GAG GAT TGG GTG
GC-rich 5'UTR F1	GAATTTGGTCTAGTTTGTCTA	GTCGAACCAACGTTCTCA
GC-rich 5'UTR F2	GAATTTGGTCTGGTTTGTCTAT	GTCGAACCAACGTTCTCA
upsA var (1300300)	AAGTCACAACAGTGCAAGGG	ACAATCGTAACCCTCACGAC
upsC var (0712000)	CGCCTACAAGTGATGGTGGT	TACTTGCATCAGTTCCGCGA
upsA var (0600200)	GCAACGTGCCGAAGATTCTAA	TGCATTTATTCGAACACGGAGT
upsC var (0808600)	GGTTGGTGGTAGTCCACAGG	TCCACGACATTTTTGTATCGCA

Supplementary Table 2. Real time PCR primer pairs

Target	Forward primer	Reverse primer
Seryl (0717700)	AAGTAGCAGGTCATCGTGGTT	TTCGGCACATTCTTCCATAA
FBA(1444800)	TGTACCACCAGCCTTACCAG	TTCCTTGCCATGTGTTCAAT
GC-rich-ncRNA-A	AAGCTGCCTCAGTAGCCCA	AAAAATTGCGCCACCCCC
GC-rich-ncRNA-B	AAGCTGCCCCAGTAGCCCA	AAAAATTGCGCCGCCCCC
Var genes		
PF3D7_1240400/900	AAAGCCACTAGCGAGGGTAA	TGTTTTTGCCCACTCCTGTA
PF3D7_1240600	CATCCATTACGCAGGATACG	AAATAGGGTGGGCGTAACAC
PF3D7_1300300	CACAGGTATGGGAAGCAATG	CCATACAGCCGTGACTGTTC
PF3D7_0412700	TAAAAGACGCCAACAGATGC	TCATCGTCTTCGTCTTCGTC
PF3D7_0412900/3100	ACTTTCTGGTGGGGAATCAG	TTCACCGCCACTTCATTCAG
PF3D7_0425800	AAACACGTTGAATGGCGATA	GACGCCGAGGAGGTAAATAG
PF3D7_0500100	GAAGCTGGTGGTACTGACGA	TATTTTCCCACCAGGAGGAG
PF3D7_0632500	ATGTGTGCGAGATGAAG	TGCCTTCTAGGTGGCATACA
PF3D7_0421100/300	ACCAAGTGGTGACAAAGCAG	GGGTGGCACACAAACACTAC
PF3D7_0617400	ATTTGTCGCACATGAAGGAA	AACTTCGTGCCAATGCTGTA
PF3D7_0412400	ACCGCCCCATCTAGTGATAG	CACTTGGTGATGTGGTGTCA
PF3D7_0420900	AGAGGGTTATGGGAATGCAG	GCATTCTTTGGCAATTCCTT
PF3D7_0533100	AAGAAAGTGCCACAACATGC	GTTCGTACGCCTGTCGTTTA
PF3D7_0937800	CACACGTGGACCTCAAGAAC	AAAACCGATGCCAATACTCC
PF3D7_0600200	TGGAAAGAACATGGACCTGA	TTCCTCGAGGGAAGAATCAC
PF3D7_1200600	TGGTGATGGTACTGCTGGAT	TTTATTTTCGGCAGCATTTG
PF3D7_0712800	ACGTGGTGGAGACGTAAACA	CCTTTGTTGTTGCCACTTTG
PF3D7_0712400	GCGACGCTCAAAAACATTTA	TCATCCAACGCAATCTTTGT
PF3D7_1200400	TCGATTATGTGCCGCAGTAT	TTCCCGTACAATCGTATCCA
PF3D7_0600200	TGGAAAGAACATGGACCTGA	TTCCTCGAGGGAAGAATCAC
PF3D7_0100300	TCATTATGGGAAGCACGATT	TGATTTCTACCATCGCAAGG
PF3D7_0712000	GTTGAGTCTGCGGCAATAGA	CTGGGGTTTGTTCAACACTG
PF3D7_0712600	CGTGGTAGTGAAGCACCATC	CCCACCTTCTTGTGGTTTCT
PF3D7_0711700	CAATTTTTCCGACGCTTGTA	CACATATAGCGCCGTCCTTA
PF3D7_0712900	CACACATGTCCACCACAAGA	ACCCTTCTGTGGTGTCTTCC
PF3D7_0808600	CCTAAAAAGGACGCAGAAGG	CCAGCAACACTACCACCAGT
PF3D7_0712300	GGTGGAGGTAGTCCACAGGA	CAGCTATTTCCCCACCAGAA
PF3D7_0809100	TGCAAGGGTGCTAATGGTAA	CCTGCATTTTGACATTCGTC
PF3D7_0808700	TTTGTCCGGAAGACGATACA	ATCTGGGGCAGAATTACCAC
PF3D7_1240300	AGCAAAATCCGAAGCAGAAT	CCCACAGATCTTTTCCTCGT
PF3D7_0800200	GGTGTCAAGGCAGCTAATGA	TATGTCCTGCGCTATTTTGC
PF3D7_0400400	ATATGGGAAGGGATGCTCTG	TGAACCATCGAAGGAATTGA
PF3D7_1100200	GACGGCTACCACAGAGACAA	CGTCATCATCGTCTTCGTTT
PF3D7_0600400	CGTAAAACATGGTGGGATGA	GGCCCATTCAGTTAACCATC
PF3D7_1150400	TGCTGAAGACCAAATTGAGC	TTGTTGTGGTGGTTGTTGTG
PF3D7_0632800	GACAAATACGGCGACTACGA	TGTTTCACCCCATTCTTCAA
PF3D7_0733000	TGACGACGATAAATGGGAAA	TTCTTTTGGAGCAGGGAGTT
PF3D7_0800100	GTCGTGGAAAAACGAAAGGT	TATCTATCCAGGGCCCAAAG
PF3D7_1000100	GACGAGGAGTCGGAAAAGAC	TGGACAGGCTTGTTTGAGAG
PF3D7_1041300	GTGCACCAAAAGAAGCTCAA	ACAAAACTCCTCTGCCCATT
PF3D7_1100100	GAGGCTTATGGGAAACCAGA	AGGCAGTCTTTGGCATCTTT
PF3D7_1300100	ACAAAGGAACGTCCATCTCC	GCCAATACTCCACATGATCG
PF3D7_1373500	CGGAATTAGTTGCCTTCACA	CATTGGCCACCAAGTGTATC
PF3D7_0100100	TGCGCTGATAACTCACAACA	AGGGGTTCATCGTCATCTTC
PF3D7_0115700	AACCCCCAATACCATTACGA	TTCCCCACTCATGTAACCAA
PF3D7_0200100	ATGTGCGCTACAAGAAGCTG	TTGATCTCCCCATTCAGTCA
PF3D7_0223500	CAATTTTGGGTGTGGAATCA	CACTGGCCACCAAGTGTATC

PF3D7_0300100/24900	CAATCTGCGGCAATAGAGAC	CCACTGTTGAGGGGTTTTCT
PF3D7_0400100	GACGACGATGAAGACGAAGA	AGATCTCCGCATTTCCAATC
PF3D7_0426000	TGACGACTCCTCAGACGAAG	CTCCACTGACGGATCTGTTG
PF3D7_0900100	TGCAAACCACCAGAAGAAAG	GTTCTCCGTGTTGTCCTCCT
PF3D7_1200100	CGGAGGAGGAAAAACAAGAG	TGCCGTATTTGAGACCACAT
PF3D7_1219300	GACGCCTGCACTCTCAAATA	TTGGAGAGCACCACCATTTA
PF3D7_1255200	GGCACGAAGTTTTGCAGATA	TTTGTGCGTCTTTCTTCGTC
PF3D7_0800300	TTTGGGATGACACCAAGAAA	GTCGCTTGATGAAGGAGTCA
PF3D7_1240400	TGCCGACAAGCCAAATACCG	CAAAAGTCTTCTGCCCATTCCTC
PF3D7_1240400	ACCGCCTTTGAAGAGGACGATG	TTGGTGCTGGTGTTTGTGCTGGAC
PF3D7_1240900	CAAAATGGTAGTGATGGTGGTCG	CCCCCTGCTTTATTATCTTTCGTC

Var ups subtypes

upsA-1	TTGGGRAATBTGTTAGTTAYRGCAA	CTGCAAAACTKCGWGCAAG
upsA-2	AACCCATCTGTRRATGATATACCTATGGA	GTTCCAASGATCCATTRGATGTATTA
upsB-1	CATCCGCCATGCAAGTATAA	CGTGCACGATTTCGATTTTTTG
upsB-2	ATCAAGGTAATTTCATACATATGTGATA	GTCCGTGCACGATTTCGATTTT
upsBC-1	AATGATCGGTGTAACCACTATC	GACAAAACTTTCACCCAATAGA
upsBC-2	CATCTGTTGCAAATTTATTCCAAATAC	TCAGTAGTATCAGACATAAATGCATA
upsC-1	CACATCGATTACATTTTAGCGTTT	TGTGGTAATATCATGTAATGG
upsC-2	CATTGTTAACATAGTCTACCATTA	GTAGCGACAACCACGRYATCATGG

2.2 Down-regulation of GC-rich ncRNA by CRISPRi inhibits *var* expression

The highlights of this section are:

- Characterization of the GC-rich ncRNA as a clonally variant gene family with mutually exclusive profile coordinated with adjacent active *var* gene.

- Development of a new CRISPRi method to down-regulate an entire gene family in *P. falciparum*.

- Down-regulation of all GC-rich ncRNA members by CRISPRi leads to repression of *var* expression.

- GC-rich ncRNA might play a role regulating other clonally variant multigene families coding for VSA.

The results of this section are currently under review in mBio under the title "CRISPR interference of a clonally variant GC-rich non-coding RNA family leads to general repression of *var* genes in *Plasmodium falciparum*". Supplementary tables with NGS data have not been included in this thesis due to space reasons. My contribution to this work included: conceptualization, experimental design, constructs and strains generation, performing CRISPRi, RT-qPCR, RNA-seq and ChIP-seq experiments, analyzing the data, and writing the manuscript.

Additional results linked to the CRISPRi method and to the down-regulation phenotype can be found in Appendix B.

CRISPR interference of a clonally variant GC-rich non-coding RNA family leads to general repression of *var* genes in *Plasmodium falciparum*

Anna Barcons-Simon[1,2,3,4], Carlos Cordon-Obras[1,2,3], Julien Guizetti[1,2,3,*], Jessica M. Bryant[1,2,3] and Artur Scherf[1,2,3,#]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[*]Present address: Centre for Infectious Diseases, Parasitology Heidelberg University Hospital, 69120 Heidelberg, Germany
[#]Correspondence: artur.scherf@pasteur.fr

Manuscript under review

ABSTRACT

The human malaria parasite *Plasmodium falciparum* uses mutually exclusive expression of the PfEMP1-encoding *var* gene family to evade the host immune system. Despite progress in the molecular understanding of the default silencing mechanism, the activation mechanism of the uniquely expressed *var* member remains elusive. A GC-rich ncRNA gene family has co-evolved with *Plasmodium* species that express *var* genes. Here, we show that this ncRNA family is transcribed in a clonally variant manner with predominant transcription of a single member when the ncRNA is located adjacent to and upstream of an active *var* gene. We developed a specific CRISPR interference (CRISPRi) strategy that allowed for transcriptional repression of all GC-rich members. Lack of GC-rich ncRNA transcription led to down-regulation of the entire *var* gene family in ring stage parasites. Strikingly, in mature blood stage parasites, the GC-rich ncRNA CRISPRi affected transcription patterns of other clonally variant gene families, including the down-regulation of all *Pfmc-2TM* members. We

provide novel evidence for the necessity of GC-rich ncRNA transcription in *var* gene activation and discovered a molecular link between transcriptional control of various clonally variant multigene families involved in parasite virulence. This work opens new avenues for elucidating the molecular processes that control immune evasion and pathogenesis in *P. falciparum*.

IMPORTANCE

Plasmodium falciparum is the deadliest malaria parasite species, accounting for the vast majority of disease cases and deaths. Virulence of this parasite is reliant upon the mutually exclusive expression of cytoadherence proteins encoded by the 60-member *var* gene family. Antigenic variation of this multigene family serves as an immune evasion mechanism ultimately leading to chronic infection and pathogenesis. Understanding the regulation mechanism of antigenic variation is key to developing new therapeutic and control strategies. Our study uncovers a novel layer in the epigenetic regulation of transcription of this family of virulence genes by means of a multigene-targeting CRISPR interference approach.

Keywords: ncRNA, *Plasmodium falciparum*, antigenic variation, mutually exclusive expression, epigenetics, *var* genes, virulence

INTRODUCTION

The protozoan parasite *Plasmodium falciparum* is responsible for the deadliest form of human malaria, causing hundreds of thousands of deaths every year (1) and, akin to other protozoan parasites (2), relies on mutually exclusive expression of virulence gene families to survive within the host (2–6). Among them, the 60-member *var* gene family encodes PfEMP1, an important variant surface adhesion molecule (7). *Var* genes are located in subtelomeric regions and central chromosomal clusters, and their transcription is epigenetically controlled (4). Transcription of a single *var* gene peaks around 12 hours post invasion (hpi) (8) and is then silenced, but poised during the later stages of the 48-hour intraerythrocytic cycle (9, 10). All other *var* genes remain transcriptionally silenced throughout the cycle and are tethered in repressive heterochromatic clusters enriched in H3K9me3 and H3K36me3 at the nuclear periphery (9, 11, 12). In contrast, the active *var* gene is euchromatic, enriched in H3K4me3 and H3K9ac, and localizes to a distinct perinuclear expression site (9,

11, 13) (schematic shown in Fig. 1A). Since the first description of this expression site (13), the mechanism of activation has remained elusive. Long non-coding RNAs (lncRNA) transcribed from the *var* intron have been implicated in a *cis*-activation process (14, 15); however, the necessity of the intron-derived lncRNA was questioned by a recent study showing that an intron-less *var* gene could be activated and silenced (16).

The initial characterization of a gene family encoding 15 GC-rich ncRNAs (annotated as *RUF6*) located adjacent to central *var* genes suggested a role in regulation of *var* genes (17, 18). Fluorescence *in situ* hybridization (FISH) revealed that these ncRNAs co-localized to the *var* gene expression site, and episomal overexpression of distinct GC-rich ncRNAs resulted in de-regulation of mutually exclusive *var* gene expression (17). However, their mechanism of action remains unknown. In this study, we use CRISPR interference (CRISPRi) to down-regulate the entire GC-rich gene family and provide evidence for the necessity of the GC-rich ncRNA in mutually exclusive *var* gene activation. Our results demonstrate a clear link between the transcription of both gene families, along with other clonally variant gene families involved in malaria virulence.

RESULTS

GC-rich ncRNA shows predominant transcription of a single member when adjacent to an active *var* gene

Recently, RNA FISH was used to demonstrate the physical association of GC-rich ncRNAs with the expression site of central and subtelomeric *var* genes (17). Given the restricted genomic location of the GC-rich genes adjacent to some, but not all, central *var* genes (Fig. 1A), we investigated whether transcription of GC-rich and *var* multigene families are coordinated. To this end, we generated an array of *P. falciparum* 3D7 wild type (WT) clones and performed paired-end RNA sequencing (RNA-seq) analysis at 12 hpi to determine the transcriptional profile of the highly homologous GC-rich ncRNA and *var* gene families. In three clones, a single member of the 15 GC-rich genes was predominantly transcribed (Fig. 1B and Fig. S1A). In these same clones, we observed mutually exclusive transcription of the central *var* gene adjacent to the upstream region of the active GC-rich gene (Fig. 1B and Fig. S1A). Notably, when the active central *var* gene lacks an adjacent GC-rich gene, ncRNA transcripts from several loci were detected, but at much lower levels compared to the former situation (Fig. 1C). Since subtelomeric *var* genes are prone to switch faster in

culture, we were only able to isolate a clone with a dominant subtelomeric *var* gene that also expressed a second central *var* gene. In this clone, we observed low-level transcription from several GC-rich ncRNA genes in addition to the dominant GC-rich ncRNA adjacent to the transcribed central *var* (Fig. S1B). Additionally, we performed receptor panning with chondroitin sulfate A (CSA) to enrich for parasites expressing subtelomeric *var2csa*. In this parasite panned line, we observed predominant transcription of *var2csa* and low levels of transcripts from several GC-rich genes (Fig. S1B). These data suggest that the GC-rich gene family is expressed in a clonally variant manner related to the chromosomal location of the active *var* gene. GC-rich genes are being predominantly transcribed from a single locus when adjacent to and upstream of an active central *var* gene.

CRISPRi of the entire GC-rich gene family leads to transcriptional down-regulation of the *var* gene family

To determine the role of GC-rich ncRNA in *var* gene expression regulation, we aimed to down-regulate the transcription of all GC-rich ncRNA genes. An attempted simultaneous knockout of all 15 highly homologous members was unsuccessful, likely due to the widespread distribution of the GC-rich genes or the diversity of their up- and down-stream regions required for such an approach. Thus, we developed a CRISPR interference (CRISPRi) (19) strategy for multigene families to block transcription via binding of dead Cas9 (dCas9), a mutated Cas9 protein lacking endonuclease activity. We designed an sgRNA targeting a homologous region common to all GC-rich gene members (Fig. 2A). We transfected the 3D7 G7 parasite strain (WT clone expressing the central *var* gene PF3D7_0412700) with the pUF-dCas9-GFP-3HA plasmid and either pL8-gRNA-GC-tc or a control plasmid with a scrambled guide RNA (pL8-gRNA-control). Chromatin immunoprecipitation (ChIP) of dCas9 followed by quantitative PCR with universal GC-rich gene primers showed a strong enrichment of dCas9 at GC-rich ncRNA loci in two independent CRISPRi clones when compared to two scrambled control clones (Fig. S2A). ChIP followed by massively parallel DNA sequencing (ChIP-seq) analysis of dCas9 showed enrichment at all 15 GC-rich gene loci in both CRISPRi clones, but not in the scrambled control (12 hpi in Fig. 2B and 24 hpi in Fig. S2B).

To determine the transcriptional effect of GC-rich gene CRISPRi, we performed reverse transcription quantitative PCR (RT-qPCR) with universal GC-rich ncRNA primers for two clones each of CRISPRi and scrambled control lines. The housekeeping gene fructose-

bisphosphate aldolase (PF3D7_1444800) was used for normalization. CRISPRi clones showed significantly reduced levels of GC-rich ncRNA compared to the control line clones at 12 and 24 hpi (Fig. 2C and Fig. 4A, respectively).

The global transcriptional effects of GC-rich gene CRISPRi were analysed by RNA-seq. Strikingly, CRISPRi lines exhibited a global down-regulation of *var* genes in two independent clones, suggesting a role for the GC-rich ncRNA in the activation of *var* gene transcription (Fig. 3A). Conversely, scrambled control clones showed transcription of a single dominant *var* gene (Fig. 3A), similar to wild type clones (Fig. 1B), suggesting that the expression of dCas9 alone does not affect *var* gene transcription. Additionally, a rescue experiment was conducted by removing the drug pressure required to maintain the plasmids and using negative drug selection to ensure plasmid removal from the CRISPRi lines (Fig. S5). Rescue control clones recovered *var* mutually exclusive transcription (Fig. 3A). Differential gene expression analysis of CRISPRi and scrambled control clones with a false discovery rate (FDR) cutoff of 0.01 returned 125 genes, 115 (92%) of which were down-regulated in the CRISPRi lines in three independent replicates (Fig. 3B and Fig. 3C, Fig. S3 and Table S1). Among these down-regulated genes were 13 GC-rich genes, 23 *var* genes (including the active *var* in the scrambled control clones, PF3D7_0712000), and several *rif* genes (Table S1). Differentially expressed genes other than GC-rich genes were validated for not presenting off target dCas9 binding. Altogether, our results suggest that GC-rich ncRNA transcription is essential for mutually exclusive expression of *var* genes.

Given that the GC-rich gene family is located within several central chromosome regions that are silenced by facultative heterochromatin enriched in Heterochromatin Protein 1 (PfHP1), we hypothesized that HP1 occupancy would determine the variegated transcription profile for the GC-rich genes. Indeed, ChIP-seq using anti-HP1 antibodies revealed that, as with *var* genes, all GC-rich genes are enriched for HP1 except the single active GC-rich gene, adjacent to the active *var* gene (Fig. S6). Notably, in the CRISPRi clones, all GC-rich genes were enriched in HP1 (Fig. S6).

GC-rich ncRNA CRISPRi affects other multigene families encoding variant surface antigens

GC-rich ncRNA is highly transcribed at the same later time point (~24 hpi) as *rif* and *stevor* genes (20, 21). Interestingly, several virulence gene families with a transcriptional peak later

in the blood stage cycle than *var* genes showed significant transcriptional down-regulation upon GC-rich gene CRISPRi, even at 12 hpi (Fig. 3B and Table S1). To investigate a potential role for GC-rich ncRNA in the transcriptional control of these gene families, we performed RNA-seq and differential expression analysis in the CRISPRi lines at 24 hpi. The total number of significantly differentially expressed genes (FDR cutoff of 0.01) in CRISPRi compared to scrambled control clones at 24 hpi was 77, of which the majority (77%) were down-regulated (Fig. 4B, Fig. S4 and Table S1) in three independent replicates. Besides GC-rich genes, most significantly down-regulated genes belonged to multigene families encoding variant surface antigens with 2 transmembrane (2TM) domains (22): *rif*, *Pfmc-2TM* and *stevor* (Fig. 4B). These three multigene families exhibited a transcription down-regulation of most members in the CRISPRi compared to the control lines (Fig. 4C). In the case of *Pfmc-2TM* gene family, the global transcription level was significantly lower in the CRISPRi compared to the control lines whereas total levels of *stevor* and *rif* gene families were not significantly lower. Altogether, these data strongly suggest that the GC-rich ncRNA is an important *trans*-activating element shared by at least *Pfmc-2TM* and *var* gene families.

DISCUSSION

The perinuclear compartment that is key to mutually exclusive expression of a single *var* gene in *P. falciparum* remains poorly understood. In a previous study, we identified the GC-rich gene family as the first *trans*-acting ncRNA localizing to this expression site (17). Here, we show that this ncRNA is essential for the transcriptional activation of a single *var* gene, and we provide evidence that this function of the GC-rich element is shared with other clonally variant gene families.

By performing RNA-seq analysis on freshly cloned parasite lines that each transcribed a single *var* gene, we showed that GC-rich ncRNAs are transcribed in a clonally variant manner (Fig. 1B). We observed two profiles of GC-rich gene transcription depending on the relative chromosomal location of the GC-rich genes and active *var* gene. In cases where there is one GC-rich gene predominantly transcribed (5-10 fold higher than other members), it is always found adjacent to the 5' region of an active central *var* gene. The ncRNA transcription profile for a clone with an active central *var* gene or subtelomeric *var* lacking an adjacent member of the GC-rich gene family at its 5' upstream region showed multiple ncRNA transcripts, but at much lower levels than in the former case. It is tempting to spec-

ulate that high levels of GC-rich ncRNA transcription adjacent to a central *var* gene may stabilize the expression site of a central *var* gene over subtelomeric *var* genes. A previous study reported variable switch rates depending on the chromosomal location of *var* genes, with central *var* genes more stably expressed and less prone to switching compared to the subtelomeric ones (23). Our data suggest that varying levels of ncRNA at the *var* gene expression site modulate the switch rate of individual *var* genes. Furthermore, transcription of GC-rich genes may open the local chromatin structure and enhance the accessibility for the transcription machinery to the adjacent *var* gene. This hypothesis finds support in a recent study showing increased chromatin accessibility of GC-rich genes when they are adjacent to the active *var* and/or *rif* gene (24).

Until recently, it was not possible to inactivate an entire multigene family dispersed over many chromosomes in *P. falciparum*. We adapted the CRISPRi technique for simultaneous knockdown of the entire GC-rich multigene family by targeting a conserved region that includes part of the Pol III B-box present in all 15 members. All members of this GC-rich family have unique DNA motifs (internal A- and B-boxes) (17, 18) only found at polymerase III-transcribed tRNA genes and short interspersed nuclear elements (SINE) in other organisms (25), suggesting that transcription of this multigene family is mediated by polymerase III (Pol III). Upon down-regulation of GC-rich ncRNA transcription, *var* gene expression was abolished, revealing an unprecedented regulatory interaction between Pol III- and Pol II-transcribed clonally variant genes. Interestingly, the GC-rich gene family is conserved throughout all *Laverania* subgenera of *Plasmodium* along with *var* genes and other clonally variant gene families involved in immune evasion, such as *rif* and *stevor* (26–28). Since GC-rich genes are transcribed at their highest levels at approximately 24 hpi, the question arises whether GC-rich ncRNA might also play a direct or indirect role in regulating clonally variant virulence gene families expressed at later stages of the asexual blood cycle, such as *rif* (20, 29). A previous study suggested that an activation factor may be common to multiple clonally variant families (30). Our work suggests that GC-rich ncRNA could be indeed such factor regulating different clonally variant gene families.

Although the precise molecular mechanism of ncRNA action remains to be investigated using techniques such as chromatin isolation by RNA purification (ChIRP), we postulate that the ncRNA associates with *var* gene control regions and acts as an activator. Indeed, a recent study showed that Pol III-transcribed SINEs act as enhancers of Pol II genes activation in response to depolarization of neurons (31). However, the lack of sequence homology

between GC-rich genes and *var* loci suggests the need of additional protein factors for such a physical interaction. Alternatively, GC-rich ncRNA could interact with nascent *var* mRNA stabilizing it for transcription. It is also possible that GC-rich ncRNA could participate in ncRNA-mediated HP1 eviction from heterochromatic *var* genes as previously described in fission yeast (32). Whichever hypothesis is correct, an essential next step would be to use this ncRNA as a molecular tool to pull down interacting partners from the *var* expression site, elucidating the molecular mechanism of *var* gene activation.

In conclusion, we developed a novel CRISPRi system that allows for the simultaneous down-regulation of the entire GC-rich multigene family. In doing so, we establish the GC-rich ncRNA as an epigenetic regulatory element that plays a role in the activation of *var* genes transcription and several other clonally variant gene families. We also provide a first glimpse into the molecular process that controls switch rates of *var* genes, which is currently a black box in the field of *var* gene transcription. The identification of a *trans*-activating factor of the expression site opens novel experimental approaches to identify regulatory proteins needed for mutually exclusive *var* expression.

METHODS

Parasite culture and synchronization

Blood stage *P. falciparum* parasites were cultured as previously described (11). Parasites were synchronized with a 6 h time window by sorbitol lysis during ring stage, followed by plasmagel enrichment in schizont stage and another sorbitol treatment 6 h after corresponding to 3±3 hpi. Synchronized parasites were harvested a 3.3% haematocrit and ~2-4% parasitemia. Parasite development was monitored by Giemsa staining.

Receptor panning

Plastic cell culture dishes were coated with 1mg/mL CSA receptor (Sigma, C9819) in PBS overnight at 4°C. The dishes were blocked with 1% BSA (Sigma, A4503) in PBS for 1 h at 37°C. Mature stages iRBC were isolated plasmagel enrichment and resuspended in 8 mL of binding medium (RPMI 1640 with 25mM HEPES pH 7.2). Culture dish was washed with binding medium and iRBC were added and allowed to adhere for 1 h with gentle agitation. The dish was washed until no unbound cells were observed by microscopy and bound cells were recovered with culture medium.

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 Qiazol. Total RNA was extracted using the miRNeasy mini kit and performing on-column DNase treatment (Qiagen). Reverse transcription 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 (17). Transcript levels were quantified by normalizing the starting quantity mean to the one of a housekeeping gene (fructose-bisphosphate aldolase, PF3D7_1444800). Starting quantity means of three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA.

Chromatin immunoprecipitation and next generation sequencing

ChIP was performed as previously described (11) with some modification using ring or trophozoite stage parasites (12 or 24 hpi). Sonicated chromatin (500 ng DNA content) was combined with either 0.5 μg of anti-HP1 (Genscript) or 1 μg anti-HA (AbcamAb9110) polyclonal rabbit antibodies. After overnight incubation, 25 μL of Dynabeads Protein G (Invitrogen) were added to each sample and additional two hour incubation was conducted. Subsequent washing, cross-link reversion and DNA extraction were carried out as described before (11). Sequencing libraries were produced with the immunoprecipitated DNA using the MicroPlex Library Preparation Kit v2 (Diagenode) with the KAPA HiFi 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. A minimum of two biological replicates were analyzed for each clone and time point.

RNA library preparation and sequencing

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 (Illumina). Fastq files were obtained by demultiplexing the data using bcl2fastq (Illumina) prior to downstream analysis. A minimum of three biological replicates were analyzed for each clone for the knockdown experiment and a minimum of two biological replicates for each WT clone (Table S2).

Plasmid construction, transfection and plasmid removal

A sequence coding for a control gRNA (targeting Luciferase) flanked with an autocleavable ribozyme (custom gene synthesis, GeneScript) was ligated with backbone pL6-eGFP vector (35) using DraII and SwaI restriction sites and T4 DNA ligation to obtain the pL8-gRNA-control plasmid. A sequence coding for a gRNA targeting all GC-rich ncRNA loci flanked with an autocleavable ribozyme (gBlock, Intregrated DNA Technologies) was cloned into the pL8-gRNAcontrol using HpaI and PmlI restriction sites followed by Gibson assembly (In-Fusion HD cloning kit, Clontech) to generate the pL8-gRNA-GC-tc plasmid. Constructs were transformed into *Escherichia coli* XL10-Gold Ultracompetent Cells (Agilent Technologies) and isolated with NucleoBond Xtra Maxiprep (Macherey-Nagel). pUF-dCas9-GFP-HA was constructed by primer mutagenesis of pUF1-Cas9 and addition of a GFP tag, three HA tags and an inducible riboswitch. pUF-dCas9-GFP-HA together with either pL8-gRNA-control or pL8-gRNA-GC-tc were transfected in ring stage parasites as previously described (36) and maintained under WR99210 and DSM1 drug selection pressure. Parasite clones were obtained by limiting dilution. Rescue lines were obtained by drug removal from the CRISPRi lines and negative selection treatment over 21 days using 5-fluorocytosine (Ancotil) to ensure pL8-gRNA-GC-tc removal. Negative selection was further maintained during parasite cloning.

ChIP-seq and RNA-seq data analysis

Fastq files were subjected to quality control using the FastQC software (37). Sequencing reads were mapped to the *P. falciparum* genome (38) (PlasmoDB, v.9) using Burrows-Wheeler Alignment tool (BWA-MEM) with default settings (39). PCR duplicates were removed without further quality score filtering since 4 GC-rich ncRNA fall in low mappability regions. ChIP-seq data were normalized over input and likelihood ratio calculation and peak calling was performed using the MACS2 (40) software with default parameters and a false discovery rate (FDR) cutoff of 0.05. RNA-seq data gene counts were calculated us-

ing bedtools (41). Differential gene expression analysis was performed with the R package edgeR (42) with an FDR threshold of 0.01. RPKM (reads per kilobase per million of mapped reads) normalization of gene counts for gene length and sequencing depth was calculated using the R:limma package (43). Data were visualized using Integrative Genomics Viewer (44).

Data availability

Sequencing data from this study are available in Genebank repository under accession number PRJNA498234 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA498234).

ACKNOWLEDGEMENTS

We thank Gretchen M. Diffendall for proofreading the manuscript. This work was supported by a European Research Council Advanced Grant (PlasmoSilencing 670301) and the French Parasitology consortium ParaFrap (ANR-11-LABX0024) awarded to A.S. and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 741447 to C.C.O.

AUTHOR CONTRIBUTIONS

A.B.S., J.G. and A.S. conceived and designed the experiments. A.B.S. performed dCas9 knockdown, RNA-seq experiments. A.B.S and C.C.O performed ChIP experiments and isolation of clones expressing distinct var genes. A.B.S. analysed RNA-seq and ChIP-seq data. J.G. and J.M.B. constructed dCas9 plasmid. A.B.S. and A.S. wrote the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. WHO. 2018. World Malaria Report 2018.

2. Duraisingh, MT, Horn, D. 2016. Epigenetic Regulation of Virulence Gene Expression in Parasitic Protozoa. Cell Host Microbe 19:629–640.

3. Su, XZ, Heatwole, VM, Wertheimer, SP, Guinet, F, Herrfeldt, JA, Peterson, DS, Ravetch, JA, Wellems, TE. 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. Cell 82:89–100.
4. 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.
5. Chen, Q, Fernandez, V, Sundström, A, Schlichtherle, M, Datta, S, Hagblom, P, Wahlgren, M. 1998. Developmental selection of var gene expression in Plasmodium falciparum. Nature 394:392–395.
6. Guizetti, J, Scherf, A. 2013. Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum. Cell Microbiol 15:718–726.
7. Miller, LH, Baruch, DI, Marsh, K, Doumbo, OK. 2002. The pathogenic basis of malaria. Nature 415:673–679.
8. Schieck, E, Pfahler, JM, Sanchez, CP, Lanzer, M. 2007. Nuclear run-on analysis of var gene expression in Plasmodium falciparum. Mol Biochem Parasitol 153:207–212.
9. Lopez-Rubio, JJ, Gontijo, AM, Nunes, MC, Issar, N, Hernandez Rivas, R, Scherf, A. 2007. 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. Mol Microbiol 66:1296–1305.
10. Volz, JC, Bártfai, R, Petter, M, Langer, C, Josling, GA, Tsuboi, T, Schwach, F, Baum, J, Rayner, JC, Stunnenberg, HG, Duffy, MF, Cowman, AF. 2012. PfSET10, a Plasmodium falciparum methyltransferase, maintains the active var gene in a poised state during parasite division. Cell Host Microbe 11:7–18.
11. Lopez-Rubio, JJ, Mancio-Silva, L, Scherf, A. 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5:179–190.
12. Jiang, L, Mu, J, Zhang, Q, Ni, T, Srinivasan, P, Rayavara, K, Yang, W, Turner, L, Lavstsen, T, Theander, TG, Peng, W, Wei, G, Jing, Q, Wakabayashi, Y, Bansal, A, Luo, Y, Ribeiro, JM, Scherf, A, Aravind, L, Zhu, J, Zhao, K, Miller, LH. 2013. PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. Nature 499:223–227.
13. Ralph, SA, 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.
14. 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:E982–91.
15. Epp, C, Li, F, Howitt, CA, Chookajorn, T, Deitsch, KW. 2009. Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite Plasmodium falciparum. RNA 15:116–127.
16. Bryant, JM, Regnault, C, Scheidig-Benatar, C, Baumgarten, S, Guizetti, J, Scherf, A. 2017. Genome Editing Reveals That the Intron Is Not Essential for var2csa gene activation or silencing in Plasmodium falciparum. MBio 8
17. 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.
18. 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.
19. Qi, LS, Larson, MH, Gilbert, LA, Doudna, JA, Weissman, JS, Arkin, AP, Lim, WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183.
20. Kyes, S, Pinches, R, Newbold, C. 2000. A simple RNA analysis method shows var and rif multigene family expression patterns in Plasmodium falciparum. Mol Biochem Parasitol 105:311–315.
21. M. Kaviratne, SMK, W. Jarra, and P. R. Preiser. 2002. Small Variant STEVOR Antigen Is Uniquely Located within Maurer's Clefts in Plasmodium falciparum-Infected Red Blood Cells. Eukaryotic Cell Vol. 1, No. 6
22. Scherf, A, Lopez-Rubio, JJ, Riviere, L. 2008. Antigenic variation in Plasmodium falciparum. Annu Rev Microbiol 62:445–470.
23. Frank, M, Dzikowski, R, Amulic, B, Deitsch, K. 2007. Variable switching rates of malaria virulence genes are associated with chromosomal position. Mol Microbiol 64:1486–1498.
24. Ruiz, JL, Tena, JJ, Bancells, C, Cortés, A, Gómez-Skarmeta, JL, Gómez-Díaz, E. 2018. Characterization of the accessible genome in the human malaria parasite Plasmodium falciparum. Nucleic Acids Res.
25. Arimbasseri, AG, Maraia, RJ. 2016. RNA Polymerase III Advances: Structural and tRNA

Functional Views. Trends Biochem Sci 41:546–559. 26. PasmoDB.
27. Larremore, DB, Sundararaman, SA, Liu, W, Proto, WR, Clauset, A, Loy, DE, Speede, S, Plenderleith, LJ, Sharp, PM, Hahn, BH, Rayner, JC, Buckee, CO. 2015. Ape parasite origins of human malaria virulence genes. Nat Commun 6:8368.
28. Otto, TD, Gilabert, A, Crellen, T, Böhme, U, Arnathau, C, Sanders, M, Oyola, SO, Okouga, AP, Boundenga, L, Willaume, E, Ngoubangoye, B, Moukodoum, ND, 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.
29. Goel, S, Palmkvist, M, Moll, K, Joannin, N, Lara, P, Akhouri, RR, Moradi, N, Öjemalm, K, Westman, M, Angeletti, D, Kjellin, H, Lehtiö, J, Blixt, O, Ideström, L, Gahmberg, CG, Storry, JR, Hult, AK, Olsson, ML, von Heijne, G, Nilsson, I, Wahlgren, M. 2015. RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. Nat Med 21:314–317.
30. Howitt, CA, Wilinski, D, Llinás, M, Templeton, TJ, Dzikowski, R, Deitsch, KW. 2009. Clonally variant gene families in Plasmodium falciparum share a common activation factor. Mol Microbiol 73:1171–1185.
31. Policarpi, C, Crepaldi, L, Brookes, E, Nitarska, J, French, SM, Coatti, A, Riccio, A. 2017. Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. Cell Rep 21:2879–2894.
32. Keller, C, Adaixo, R, Stunnenberg, R, Woolcock, KJ, Hiller, S, Bühler, M. 2012. HP1(Swi6) mediates the recognition and destruction of heterochromatic RNA transcripts. Mol Cell 47:215–227.
33. Fastman, Y, Noble, R, Recker, M, Dzikowski, R. 2012. Erasing the epigenetic memory and beginning to switch–the onset of antigenic switching of var genes in Plasmodium falciparum. PLoS One 7:e34168.
34. 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.
35. Ghorbal, M, Gorman, M, Macpherson, CR, Martins, RM, Scherf, A, Lopez-Rubio, JJ. 2014. Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. Nat Biotechnol 32:819–821.
36. Hasenkamp, S, Merrick, CJ, Horrocks, P. 2013. A quantitative analysis of Plasmodium falciparum transfection using DNA-loaded erythrocytes. Mol Biochem Parasitol 187:117–120.
37. Wingett, SW, Andrews, S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res 7:1338.

38. Gardner, MJ, Hall, N, Fung, E, White, O, Berriman, M, Hyman, RW, Carlton, JM, Pain, A, Nelson, KE, Bowman, S, Paulsen, IT, James, K, Eisen, JA, Rutherford, K, Salzberg, SL, Craig, A, Kyes, S, Chan, MS, Nene, V, Shallom, SJ, Suh, B, Peterson, J, Angiuoli, S, Pertea, M, Allen, J, Selengut, J, Haft, D, Mather, MW, Vaidya, AB, Martin, DM, Fairlamb, AH, Fraunholz, MJ, Roos, DS, Ralph, SA, McFadden, GI, Cummings, LM, Subramanian, GM, Mungall, C, Venter, JC, Carucci, DJ, Hoffman, SL, Newbold, C, Davis, RW, Fraser, CM, Barrell, B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511.
39. Li, H, Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
40. Zhang, Y, Liu, T, Meyer, CA, Eeckhoute, J, Johnson, DS, Bernstein, BE, Nusbaum, C, Myers, RM, Brown, M, Li, W, Liu, XS. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137.
41. Quinlan, AR, Hall, IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. 42. Robinson, MD, McCarthy, DJ, Smyth, GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.
43. Ritchie, ME, Phipson, B, Wu, D, Hu, Y, Law, CW, Shi, W, Smyth, GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47.
44. Robinson, JT, Thorvaldsdóttir, H, Winckler, W, Guttman, M, Lander, ES, Getz, G, Mesirov, JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26.

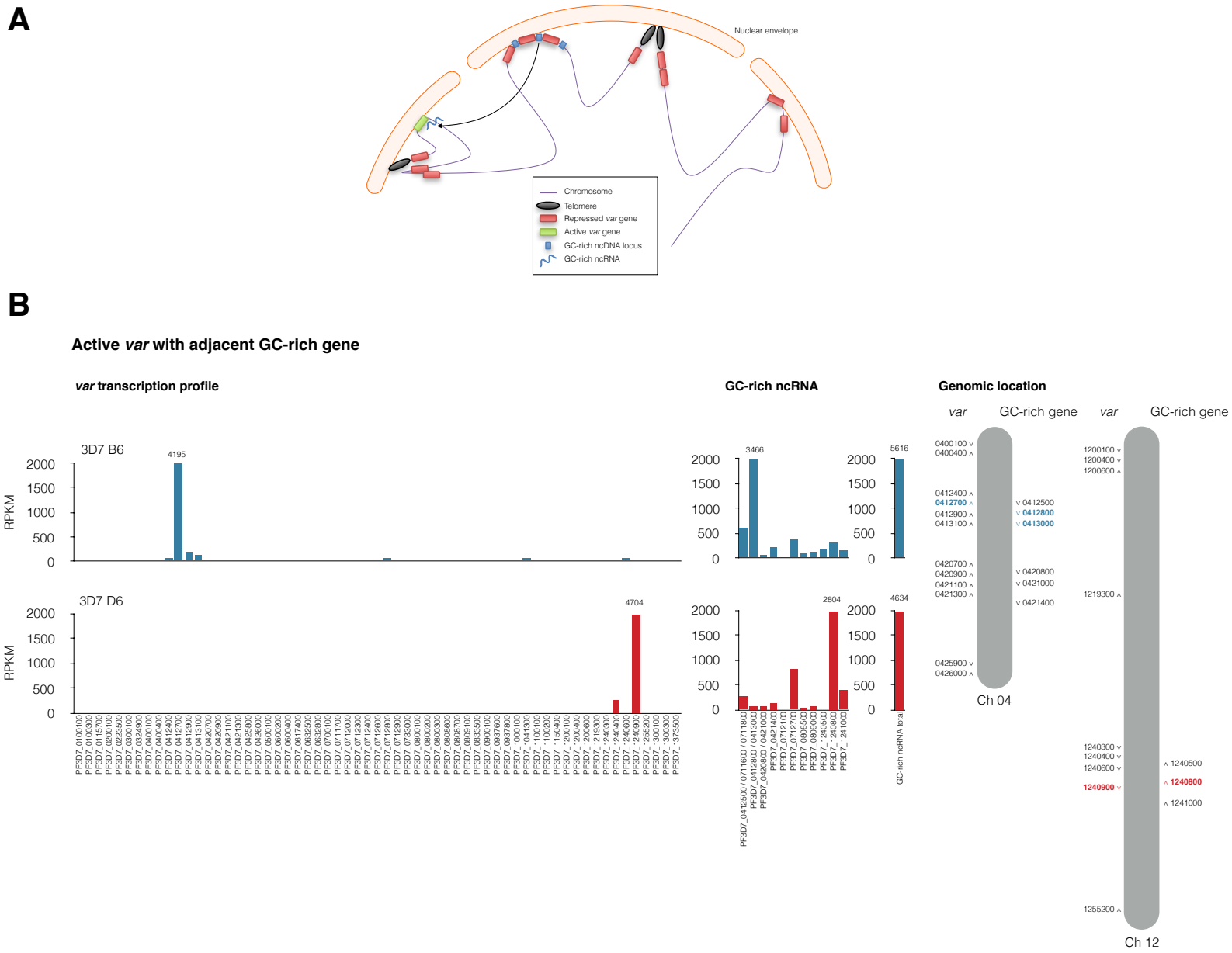


Figure 1. GC-rich ncRNA and *var* clonal variation. (Continued next page).

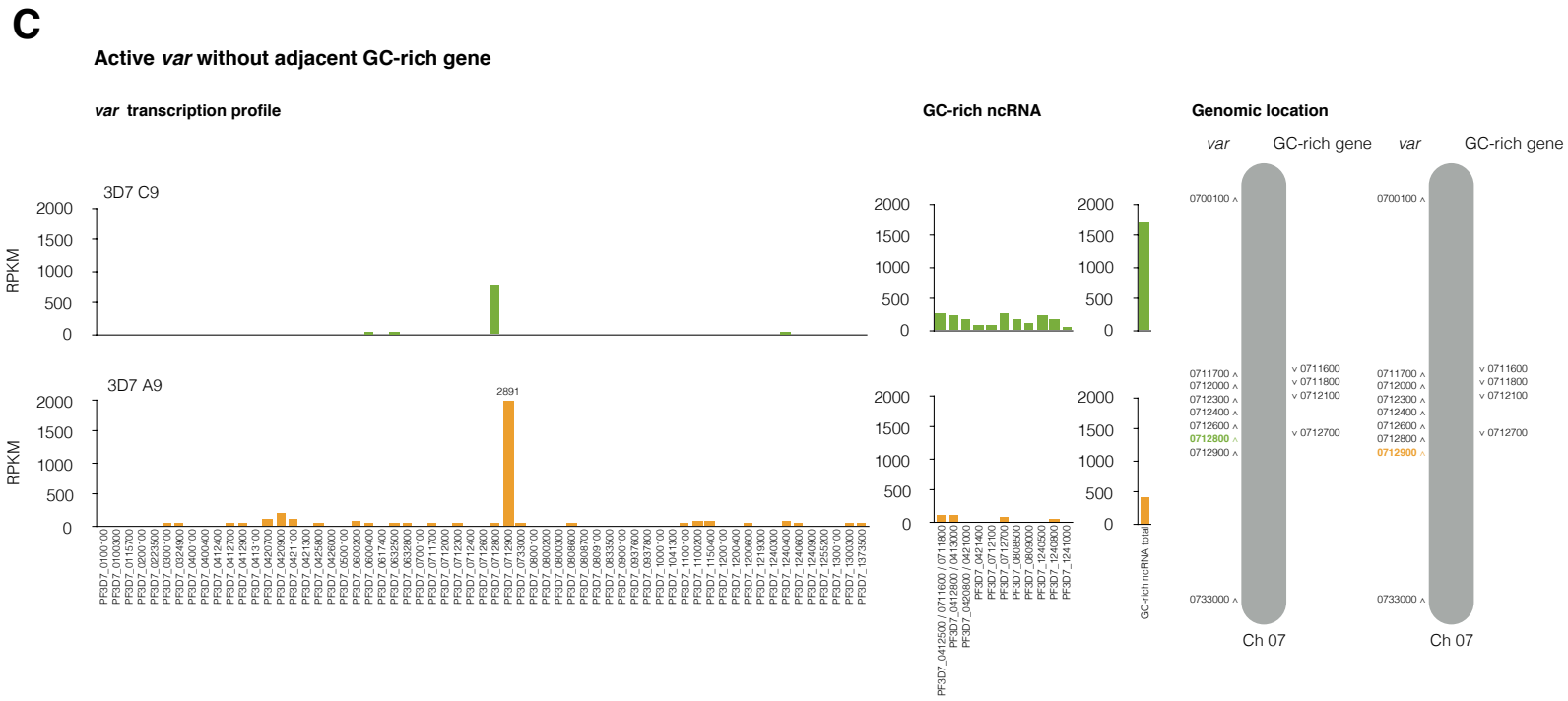
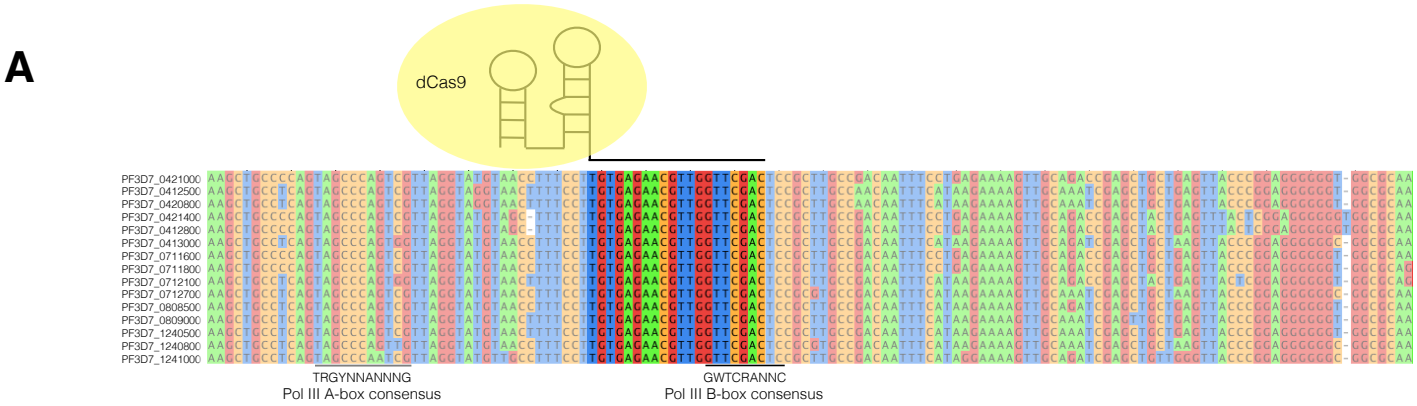
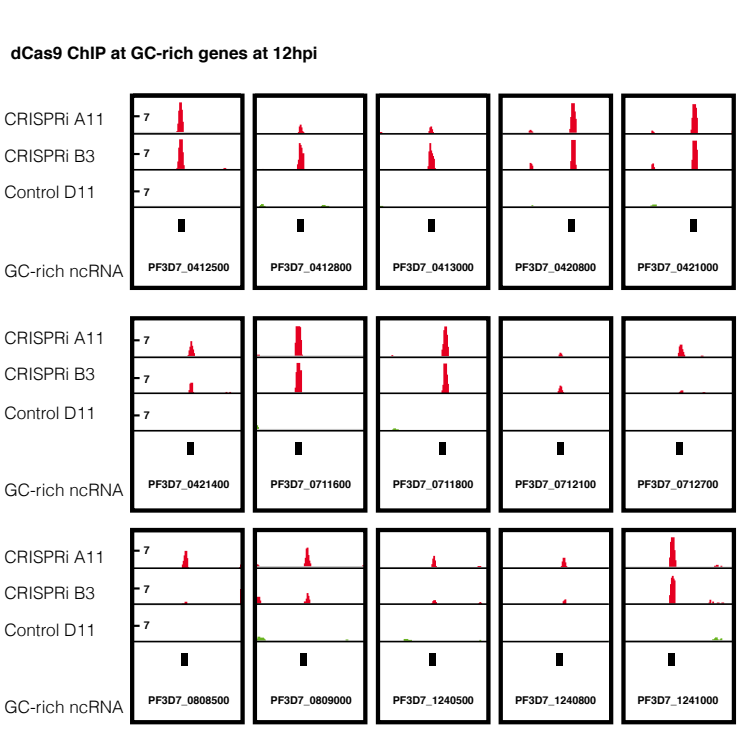


Figure 1. GC-rich ncRNA and *var* clonal variation. (A) Schematic model showing repressed *var* genes perinuclear clustering and active *var* gene relocation to active expression site colocalizing with GC-rich ncRNA. **(B)** Transcriptional profile of *var* genes and GC-rich ncRNA at 12 hpi in WT clones assessed by RNA sequencing. Chromosome schematics (modified from (33)) highlighting active *var* genes and GC-rich genes in different clones. Clones 3D7 B6 (blue) and 3D7 D6 (red) have predominant transcription of the GC-rich ncRNA adjacent to their active *var* locus. **(C)** Transcriptional profile of *var* genes and GC-rich ncRNA at 12 hpi in clones 3D7 C9 (green) and 3D7 A9 (yellow) that do not have a GC-rich gene upstream from their respective active central *var* locus and have several GC-rich ncRNA transcripts at lower levels. Mean of RPKM of two independent experiments at 12 hpi. Further clones are shown in Fig. S1.


A
dCas9
PF3D7_0421000
PF3D7_0412500
PF3D7_0420800
PF3D7_0421400
PF3D7_0412800
PF3D7_0413000
PF3D7_0711600
PF3D7_0711800
PF3D7_0712100
PF3D7_0712700
PF3D7_0808500
PF3D7_0809000
PF3D7_1240500
PF3D7_1240800
PF3D7_1241000
TRGYNNANNNG
Pol III A-box consensus
GWTCRANNC
Pol III B-box consensus


B
dCas9 ChIP at GC-rich genes at 12hpi
CRISPRi A11
CRISPRi B3
Control D11
GC-rich ncRNA
PF3D7_0412500
PF3D7_0412800
PF3D7_0413000
PF3D7_0420800
PF3D7_0421000
PF3D7_0421400
PF3D7_0711600
PF3D7_0711800
PF3D7_0712100
PF3D7_0712700
PF3D7_0808500
PF3D7_0809000
PF3D7_1240500
PF3D7_1240800
PF3D7_1241000

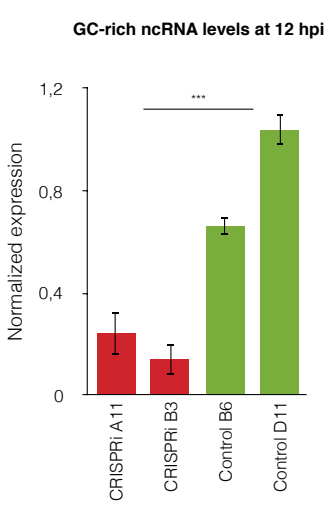

C
GC-rich ncRNA levels at 12 hpi
Normalized expression
1,2
0,8
0,4
0
***
CRISPRi A11
CRISPRi B3
Control B6
Control D11

Figure 2. GC-rich ncRNA knockdown by CRISPR interference. (A) Multiple sequence alignment of the 15 GC-rich ncRNA members showing the dCas9 target region (not shaded) where the sgRNA binds on the DNA coding strand of all GC-rich genes. Below, the black lines show the position of polymerase III A- and B-box consensus motifs (34). **(B)** ChIP sequencing data shows enrichment of dCas9 in the 15 GC-rich gene loci for the CRISPRi line. Logarithmic scale of likelihood ratio of fold enrichment over input for dCas9 computed with MACS2 is represented in red for the CRISPRi clones and in green for the scrambled control clone D11. Data range for each track is 0 to 14. Data are representative of two independent experiments at 12 hpi. **(C)** GC-rich ncRNA levels at 12 hpi as quantified by RT-qPCR for two GC-rich gene CRISPRi clones (B3 and A11) and two scrambled control clones (Control D11 and B6). Expression is normalized to fructose-bisphosphate aldolase (PF3D7_1444800) transcript levels. Mean±SEM of three independent experiments are shown. Statistical significance was determined by two-tailed Student's t-test (*** $p<0.001$).

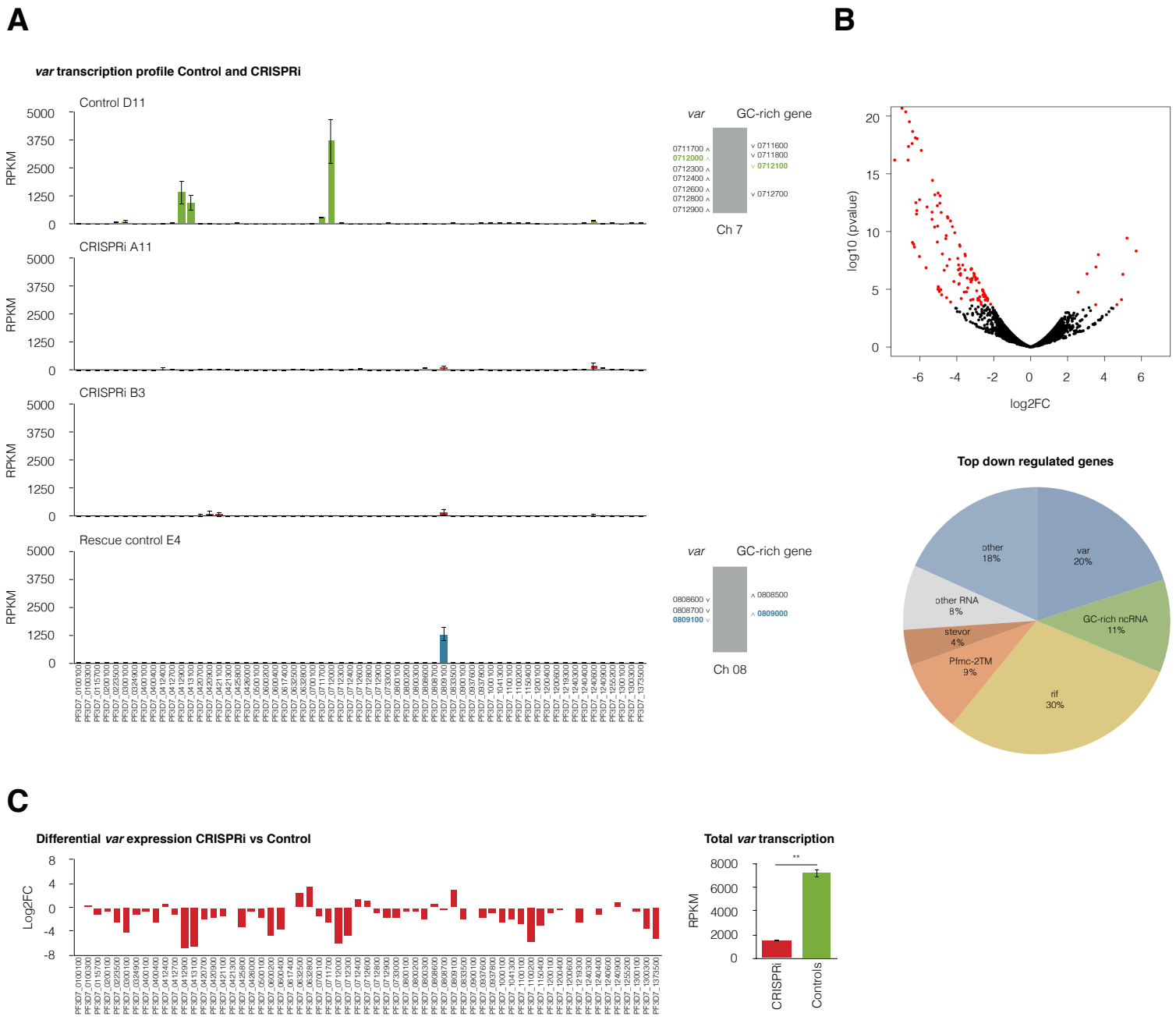

A
var transcription profile Control and CRISPRi
Control D11
CRISPRi A11
CRISPRi B3
Rescue control E4
RPKM
var
GC-rich gene
Ch 7
Ch 08
B
log10 (pvalue)
log2FC
Top down regulated genes
var 20%
GC-rich ncRNA 11%
rif 30%
Pfmc-2TM 9%
stevor 4%
other RNA 8%
other 18%
C
Differential var expression CRISPRi vs Control
Log2FC
Total var transcription
RPKM
CRISPRi
Controls

Figure 3. GC-rich ncRNA knockdown leads to the down-regulation of *var* gene expression. (A) Transcriptional *var* genes profile at 12 hpi assed by RNA sequencing for the control gRNA clone D11, two GC-rich ncRNA knockdown clones (CRISPRi B3 and A11) and the rescue control clone E4. Two control gRNA clones and two rescue control clones were analyzed but only one representative example for each control is shown. Central chromosome cluster schematics schematics highlighting active *var* and GC-rich genes in control clones. Mean±SEM of three independent experiments are shown. **(B)** Top panel: Volcano plot for differential expression between CRISPRi clones and scrambled control line. Differentially expressed genes with a 0.01 FDR cutoff are represented by dots highlighted in red. Total number of significantly differentially expressed genes is 125; 92% of which are down-regulated and 8% up-regulated. Lower panel: families of top down-regulated genes in CRISPRi clones compared to scrambled control clones, significantly differentially expressed with a 0.01 FDR cutoff in three independent replicates. **(C)** Left panel: Differential expression of *var* genes for CRISPRi clones compared to scrambled control line shows down-regulation of the entire *var* gene family. Log of fold enrichment (log2FC) for three replicates of two CRISPRi clones compared to two scrambled control clones is represented. Right panel: Mean±SEM transcriptional levels of *var* gene family as assessed by RNA-seq for three replicates of two CRISPRi and two scrambled control clones. Statistical significance was determined by two-tailed Student's t-test (** $p<0.01$).

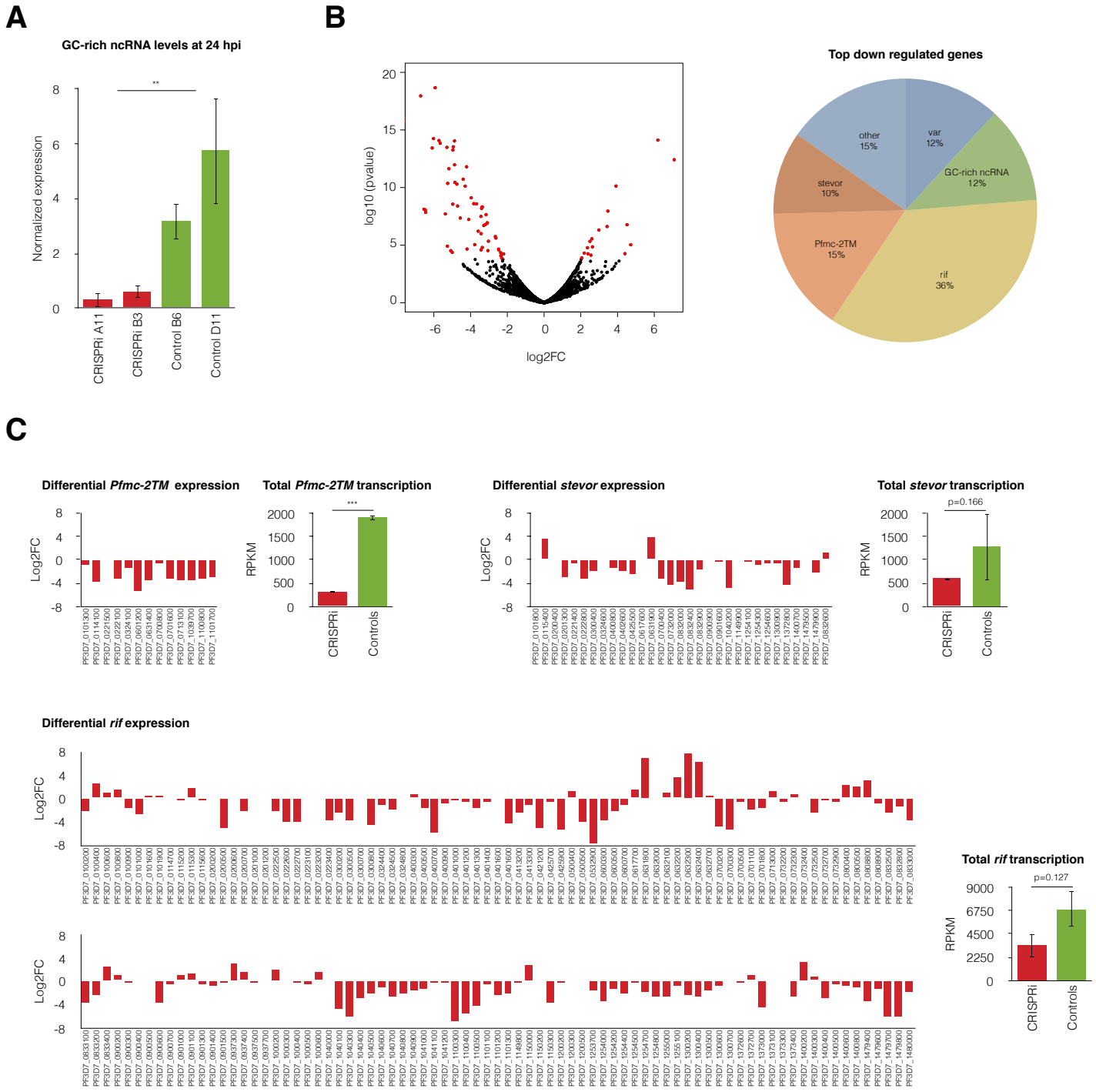

A
GC-rich ncRNA levels at 24 hpi
Normalized expression
CRISPRi A11
CRISPRi B3
Control B6
Control D11
B
log10 (pvalue)
log2FC
Top down regulated genes
other 15%
var 12%
GC-rich ncRNA 12%
stevor 10%
Pfmc-2TM 15%
rif 36%
C
Differential Pfmc-2TM expression
Total Pfmc-2TM transcription
Differential stevor expression
Total stevor transcription
p=0.166
Differential rif expression
Total rif transcription
p=0.127
Log2FC
RPKM
CRISPRi
Controls

Figure 4. GC-rich ncRNA knockdown lines exhibit down-regulation of 2TM-type supergene family. (A) GC-rich ncRNA levels at 24 hpi as quantified by RT-qPCR for two CRISPRi clones (B3 and A11) and two control gRNA clones (Control D11 and B6). Transcription is normalized to house-keeping gene fructose-bisphosphate aldolase (PF3D7_1444800) levels. Mean±SEM of three independent experiments are shown. Statistical significance was determined by two-tailed Student's t-test (** p<0.01). **(B)** Left panel: Volcano plot for differential expression between CRISPRi clones and scrambled control line. Differentially expressed genes with a 0.01 FDR cutoff are represented by dots highlighted in red. Total number of significantly differentially expressed genes is 77; 77% of which are down-regulated and 23% up-regulated. Right panel: families of top down-regulated genes in CRISPRi clones compared to scrambled control clones, significantly differentially expressed with a 0.01 FDR cutoff in three independent replicates. **(C)** Left panels: Differential expression of 2TM multigene families for CRISPRi clones compared to scrambled control line shows general down-regulation of most *Pfmc-2TM*, *stevor* and *rif* gene family members. Log of fold enrichment (log2FC) for three replicates of two CRISPRi clones compared to two scrambled control clones is represented. Right panels: Mean±SEM transcriptional levels of entire gene families as assessed by RNA-seq for three replicates of two CRISPRi and two scrambled control clones. Statistical significance was determined by two-tailed Student's t-test (*** p<0.001, ** p<0.01).

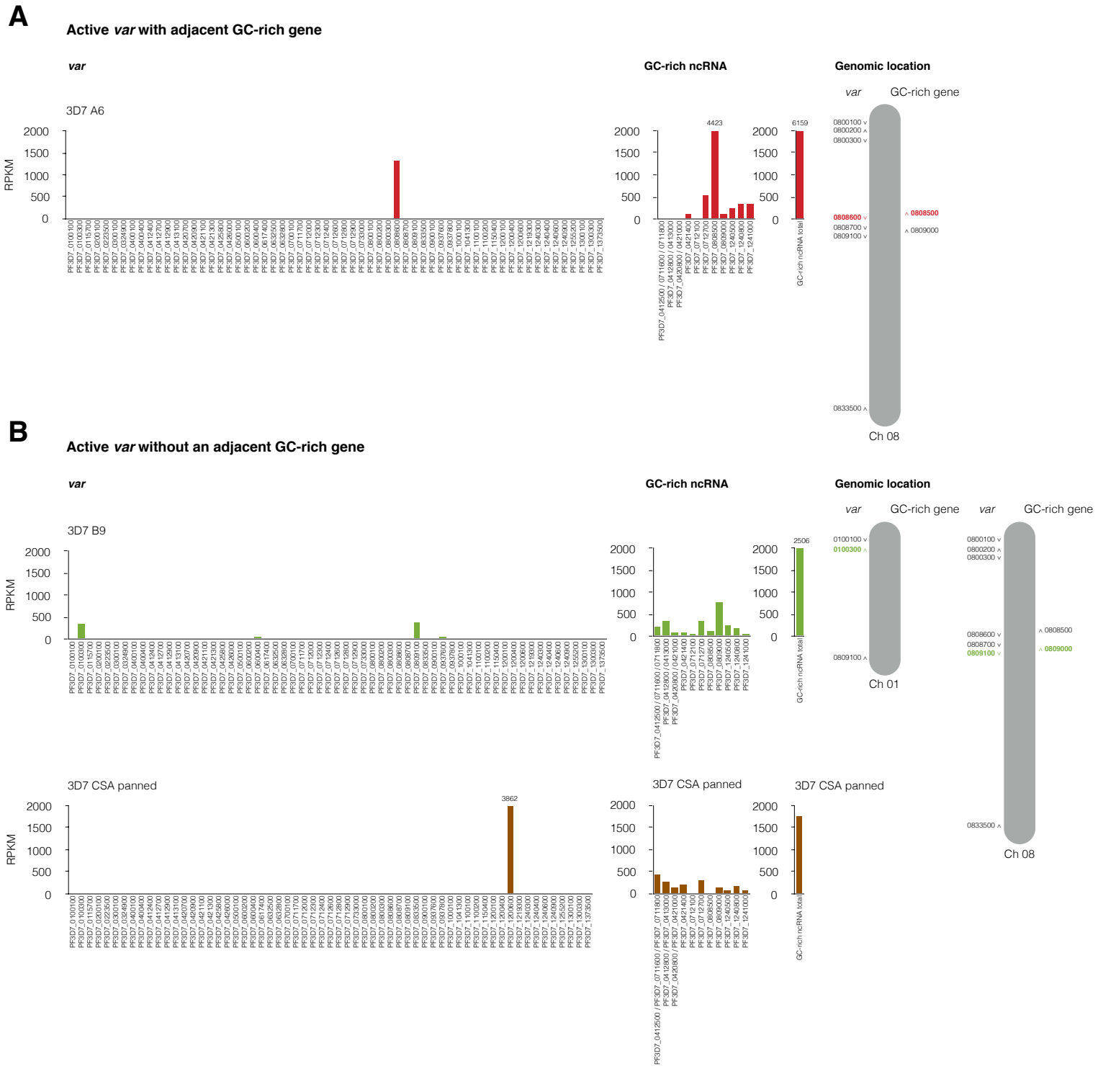
SUPPLEMENTARY DATA

CRISPR interference of a clonally variant GC-rich non-coding RNA family leads to general repression of *var* genes in *Plasmodium falciparum*

Anna Barcons-Simon[1,2,3,4], Carlos Cordon-Obras[1,2,3], Julien Guizetti[1,2,3,*], Jessica M. Bryant[1,2,3] and Artur Scherf[1,2,3,#]

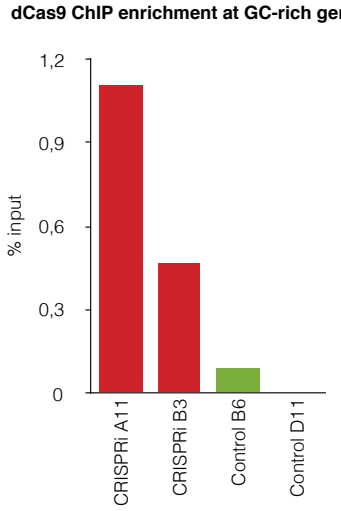
[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[*]Present address: Centre for Infectious Diseases, Parasitology Heidelberg University Hospital, 69120 Heidelberg, Germany
[#]Correspondence: artur.scherf@pasteur.fr

Manuscript under review

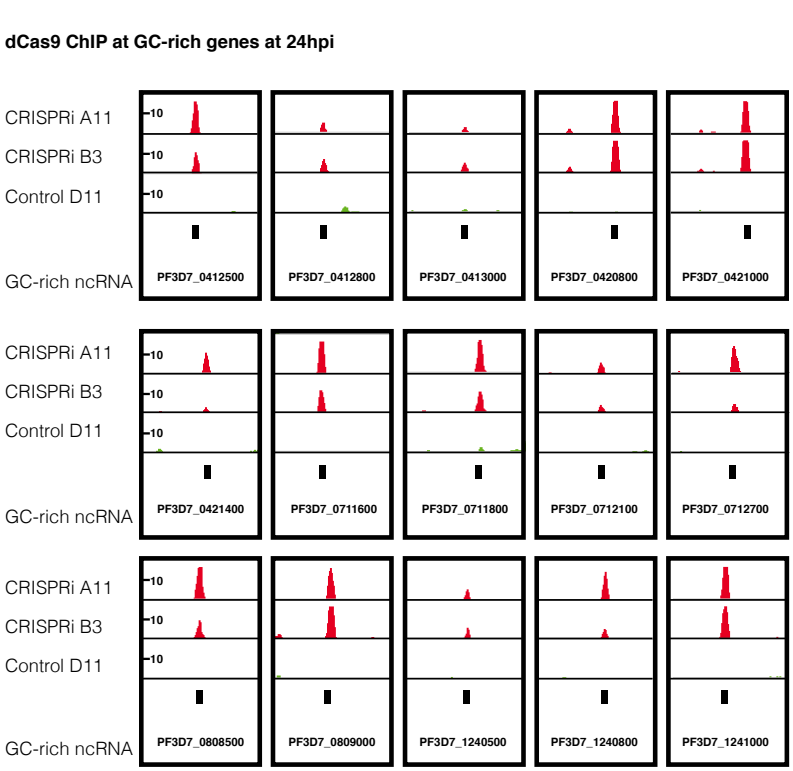

A
Active var with adjacent GC-rich gene
var
GC-rich ncRNA
Genomic location
3D7 A6
RPKM
4423
6159
var
GC-rich gene
0808500
0809000
Ch 08
B
Active var without an adjacent GC-rich gene
3D7 B9
2506
Ch 01
3D7 CSA panned
3862
Ch 08

Supplementary Figure 1. GC-rich ncRNA and *var* clonal variation. Transcriptional profile of *var* genes and GC-rich ncRNA at 12 hpi in WT clones assed by RNA sequencing. Chromosome schematics (modified from (33)) highlighting active *var* genes and GC-rich genes in different clones. **(A)** Clone 3D7 A6 (red) has predominant transcription from the GC-rich gene (PF3D7_0808500) adjacent to their active var locus (PF3D7_0808600). **(B)** Clone 3D7 B9 (green) has two dominant active *var* genes, one central and one subtelomeric and several GC-rich ncRNA transcripts with dominance of the one adjacent to the active central *var*. Clone 3D7 CSA panned (brown) has predominant transcription from subtelomeric *var2csa* and several GC-rich ncRNA transcripts. Mean of RPKM of two independent experiments at 12 hpi.

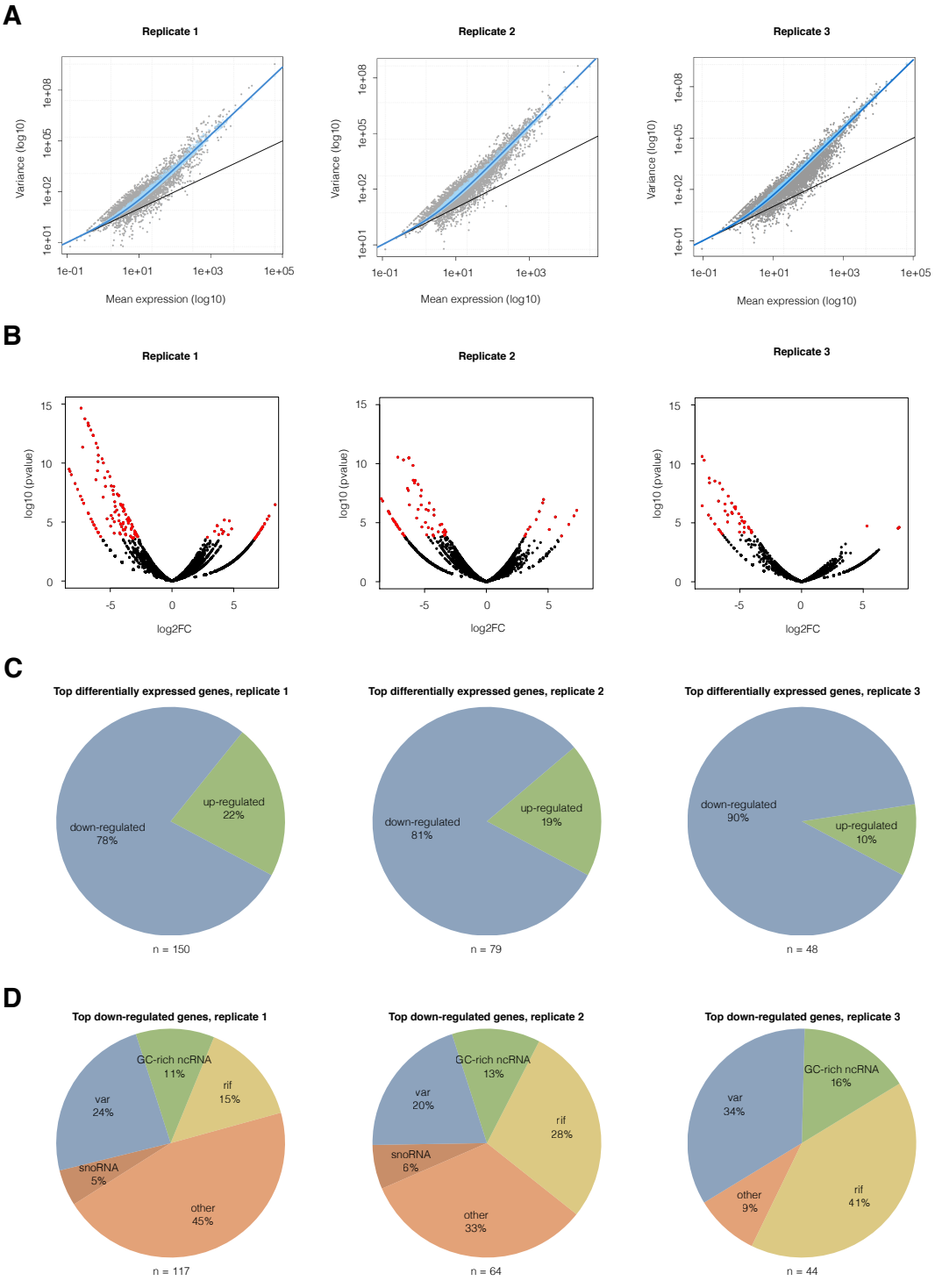
A


dCas9 ChIP enrichment at GC-rich genes at 12hpi
% input
1,2
0,9
0,6
0,3
0
CRISPRi A11
CRISPRi B3
Control B6
Control D11

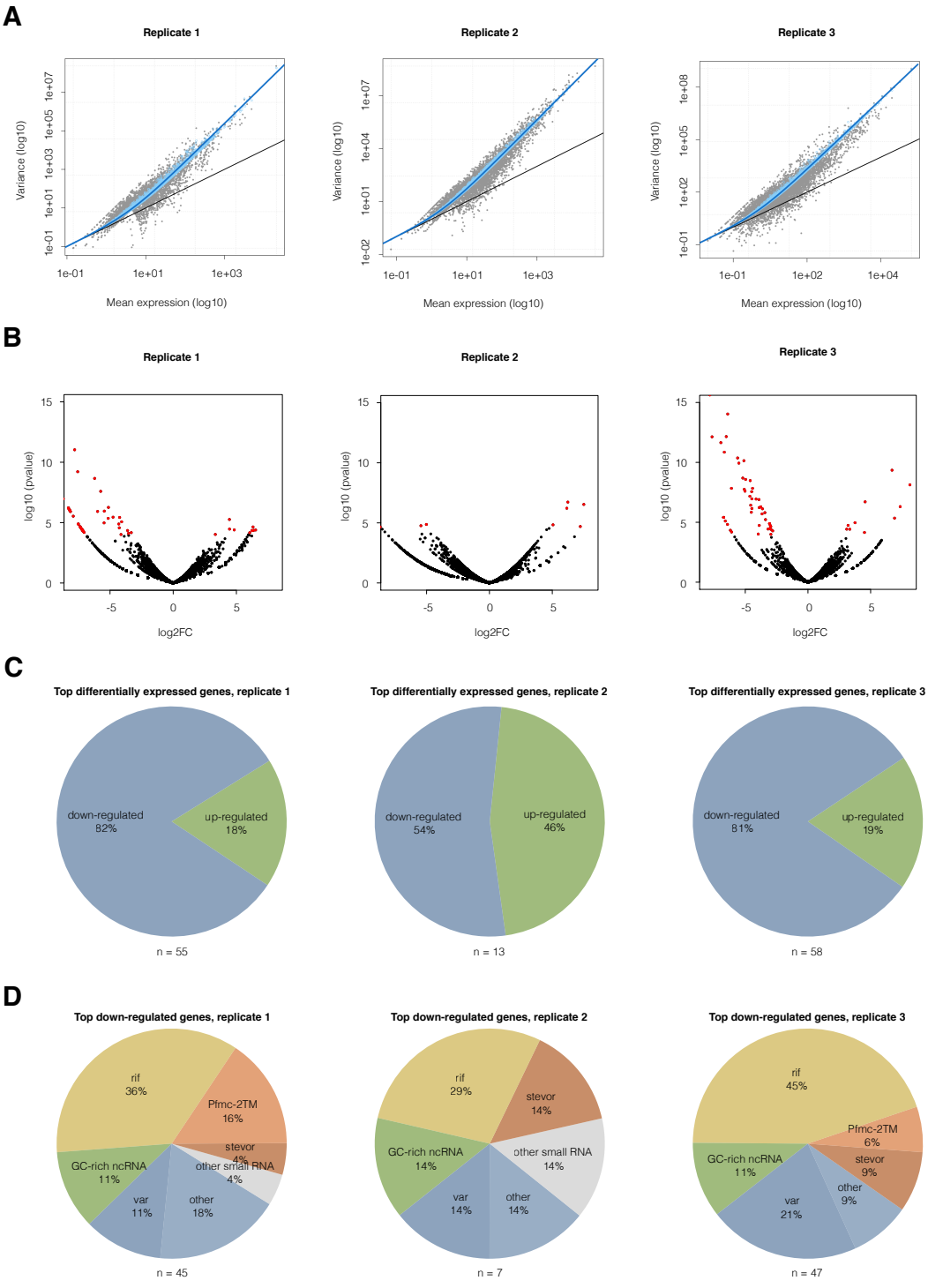
B


dCas9 ChIP at GC-rich genes at 24hpi
CRISPRi A11
CRISPRi B3
Control D11
GC-rich ncRNA
PF3D7_0412500
PF3D7_0412800
PF3D7_0413000
PF3D7_0420800
PF3D7_0421000
PF3D7_0421400
PF3D7_0711600
PF3D7_0711800
PF3D7_0712100
PF3D7_0712700
PF3D7_0808500
PF3D7_0809000
PF3D7_1240500
PF3D7_1240800
PF3D7_1241000

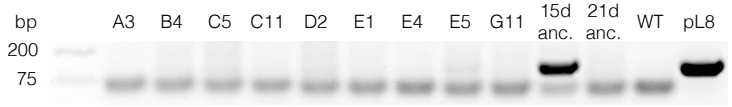
Supplementary Figure 2. GC-rich ncRNA knockdown by CRISPR interference. (A) dCas9 enrichment at GC-rich genes calculated as percent of input quantified by qPCR for two GC-rich ncRNA knockdown clones (CRISPRi B3 and A11) and two control clones with dCas9 and a control gRNA (Control D11 and B6) at 12hpi. **(B)** ChIP sequencing data shows enrichment of dCas9 in the 15 GC-rich gene loci for the CRISPRi line. Logarithmic scale of likelihood ratio of fold enrichment over input for dCas9 computed with MACS2 is represented in red for the CRISPRi clones and in green for the control gRNA clone D11. Data range for each track is 0 to 20. Data are representative of two independent experiments at 24 hpi.


A
Replicate 1
Replicate 2
Replicate 3
Variance (log10)
Mean expression (log10)
B
Replicate 1
Replicate 2
Replicate 3
log10 (pvalue)
log2FC
C
Top differentially expressed genes, replicate 1
down-regulated 78%
up-regulated 22%
n = 150
Top differentially expressed genes, replicate 2
down-regulated 81%
up-regulated 19%
n = 79
Top differentially expressed genes, replicate 3
down-regulated 90%
up-regulated 10%
n = 48
D
Top down-regulated genes, replicate 1
GC-rich ncRNA 11%
rif 15%
var 24%
snoRNA 5%
other 45%
n = 117
Top down-regulated genes, replicate 2
GC-rich ncRNA 13%
var 20%
rif 28%
snoRNA 6%
other 33%
n = 64
Top down-regulated genes, replicate 3
GC-rich ncRNA 16%
var 34%
other 9%
rif 41%
n = 44

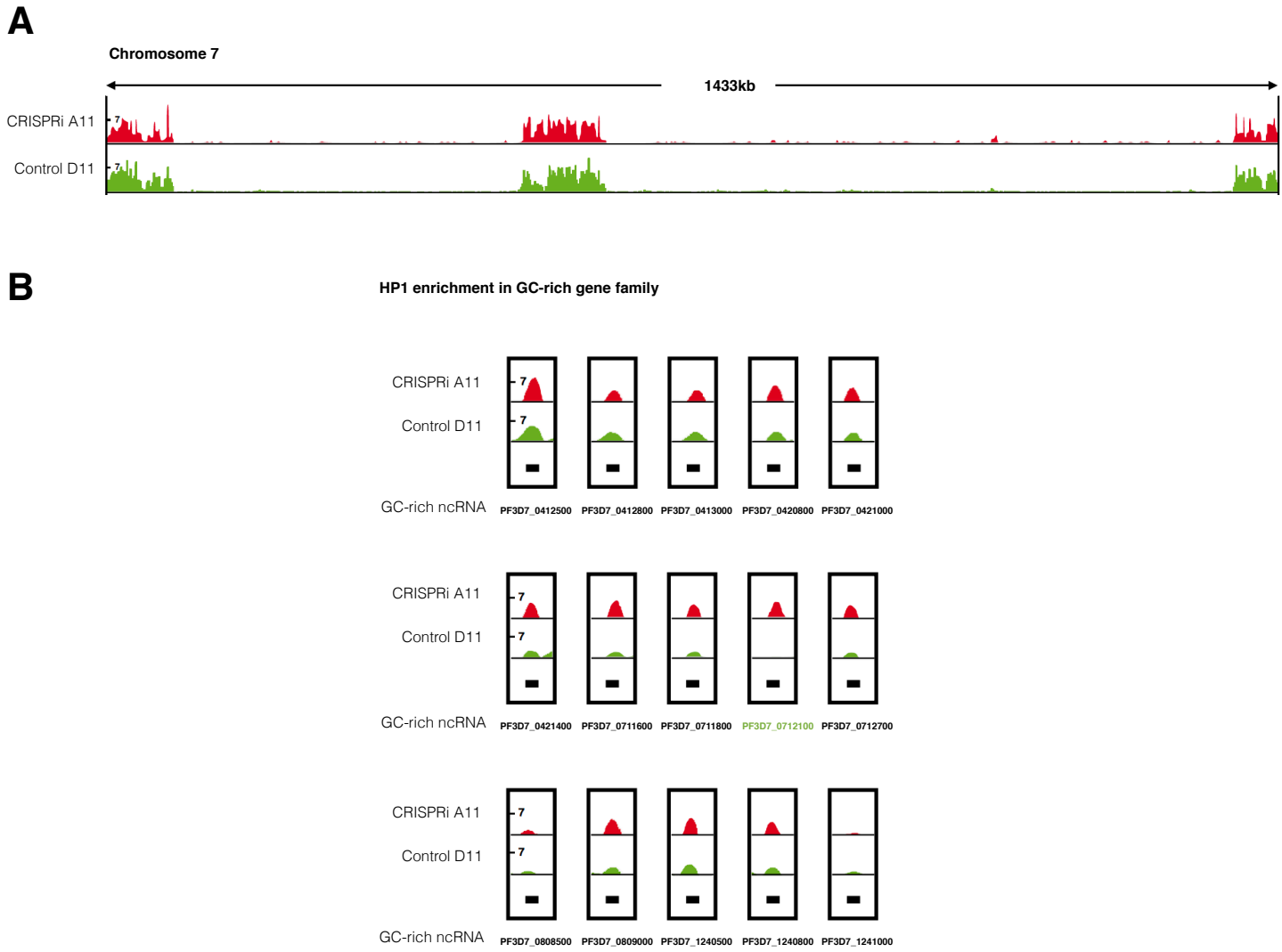
Supplementary Figure 3. Differential gene expression in GC-rich ncRNA knockdown at 12 hpi. (A) Mean-variance scatterplot for differential expression between CRISPRi clones (A11 and B3) and scrambled control clones (D11 and B6) for three independent replicates at 12 hpi. **(B)** Volcano plot for differential expression between CRISPRi clones (A11 and B3) and scrambled control clones (D11 and B6) for three independent replicates at 12 hpi. Differentially expressed genes with a 0.01 FDR cutoff are represented by dots highlighted in red. **(C)** Plot showing percentage of up- and down-regulated genes amongst the differentially expressed genes with a 0.01 FDR cutoff in CRISPRi clones (A11 and B3) compared to scrambled control clones (D11 and B6) for three independent replicates at 12 hpi. **(D)** Families of top down-regulated genes in CRISPRi clones (A11 and B3) compared to scrambled control clones (D11 and B6), significantly differentially expressed with a 0.01 FDR cutoff in three independent replicates at 12 hpi.


A
Replicate 1
Replicate 2
Replicate 3
Variance (log10)
Mean expression (log10)
B
Replicate 1
Replicate 2
Replicate 3
log10 (pvalue)
log2FC
C
Top differentially expressed genes, replicate 1
down-regulated 82%
up-regulated 18%
n = 55
Top differentially expressed genes, replicate 2
down-regulated 54%
up-regulated 46%
n = 13
Top differentially expressed genes, replicate 3
down-regulated 81%
up-regulated 19%
n = 58
D
Top down-regulated genes, replicate 1
rif 36%
Pfmc-2TM 16%
stevor 4%
other small RNA 4%
GC-rich ncRNA 11%
var 11%
other 18%
n = 45
Top down-regulated genes, replicate 2
rif 29%
stevor 14%
other small RNA 14%
GC-rich ncRNA 14%
var 14%
other 14%
n = 7
Top down-regulated genes, replicate 3
rif 45%
Pfmc-2TM 6%
stevor 9%
other 9%
GC-rich ncRNA 11%
var 21%
n = 47

Supplementary Figure 4. Differential gene expression in GC-rich ncRNA knockdown at 24 hpi. (A) Mean-variance scatterplot for differential expression between CRISPRi clones (A11 and B3) and scrambled control clones (D11 and B6) for three independent replicates at 24 hpi. **(B)** Volcano plot for differential expression between CRISPRi clones (A11 and B3) and scrambled control clones (D11 and B6) for three independent replicates at 24 hpi. Differentially expressed genes with a 0.01 FDR cutoff are represented by dots highlighted in red. **(C)** Plot showing percentage of up- and down-regulated genes amongst the differentially expressed genes with a 0.01 FDR cutoff in CRISPRi clones (A11 and B3) compared to scrambled control clones (D11 and B6) for three independent replicates at 24 hpi. **(D)** Families of top down-regulated genes in CRISPRi clones (A11 and B3) compared to scrambled control clones (D11 and B6), significantly differentially expressed with a 0.01 FDR cutoff in three independent replicates at 24 hpi.



Supplementary Figure 5. CRISPRi rescue phenotype experiment. PCR results with oligos for the DHFR cassette from pL8 plasmids on parasite gDNA. Plasmid was successfully removed in all rescue control clones (first 9 lanes). Plasmid removal was obtained by removing drug pressure and using negative selection with 5-fluorocytosine (anc.) over 21 days. Subsequently, clones were obtained by limiting dilution. pL8 plasmid and gDNA from a WT strain are used as controls.



Supplementary Figure 6. HP1 profile for GC-rich genes. (A) ChIP sequencing data shows similar enrichment profile of HP1 for GC-rich ncRNA knockdown (CRISPRi A11) and control gRNA line (Control D11). Coverage of fold enrichment over input for dCas9 computed with MACS2 is represented in red for the CRISPRi clone, in green for the control gRNA clone D11 and in blue for the WT. Data are representative of two independent experiments at 12 hpi. **(B)** GC-rich ncRNA members show variegated HP1 enrichment corresponding to transcriptomic data. In Control D11 only the active GC-rich gene PF3D7_0712100 (highlighted in green), adjacent to the active *var* (PF3D7_0712000, see Fig. 3A) is not enriched by HP1 while all down-regulated GC-rich genes in CRISPRi A11 clone present HP1 enrichment.

2.3 Mechanistic insight on the GC-rich ncRNA

The highlights of this section are:

- Development of methods for studying RNA interaction in *P. falciparum*: ChIRP, RNA EMSA and RNA pulldowns.

- Identification of candidate proteins interacting with GC-rich ncRNA.

The results of this section are preliminary data summarized in a draft under the title "Exploring the interactome of a ncRNA gene family that activates expression of a gene family linked to severe malaria". My contribution to this work included: conceptualization, experimental design, performing ChIRP, RNA EMSAs and RNA pulldowns experiments, analyzing the data, and writing the manuscript. Supplementary ChIRP results can be found in subsection 2.3.1.

Exploring the interactome of a ncRNA gene family that activates expression of a gene family linked to severe malaria

Anna Barcons-Simon[1,2,3,4], Gretchen M. Diffendall[1,2,3,4], Florent Dingli[5], Damarys Loew[5] and Artur Scherf[1,2,3,#]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[5]Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, F-75248 Paris, France
[#]Correspondence: artur.scherf@pasteur.fr

Preliminary data

ABSTRACT

The role of ncRNAs in the human malaria parasite *P. falciparum* remains ill defined. Recently, the knockdown of a GC-rich ncRNA gene family has been associated with the expression of a major pathogenesis factor encoded by the *var* gene family. The interactome of this ncRNA (annotated as RNA of unknown function 6 RUF6) with chromatin remains unknown. Here, we developed a robust Chromatin Isolation by RNA Purification (ChIRP) protocol for *P. falciparum* to explore the RUF6 interacting partners in their native context. This method was used to pull down proteins that interact with RUF6. We identified several candidate proteins that are linked to gene transcription. Our studies using ChIRP point to RUF6 as a potential chromatin assembly factor involved in gene transcription of clonally variant virulence genes.

Keywords: ncRNA, *RUF6*, *Plasmodium falciparum*, ChIRP, RNA pulldown

INTRODUCTION

Non-coding RNAs (ncRNAs) are increasingly recognized as key players in eukaryotic gene regulation by a broad variety of mechanisms (1, 2). In the protozoan parasite *Plasmodium falciparum*, responsible for the most severe form of human malaria (3), many potential regulatory ncRNAs have been detected by transcriptomic analyses (4–7) but they remain mostly uncharacterized. Recently, a gene family coding for 15 highly homologous ncRNAs, annotated as RNA of unknown function 6 (*RUF6*) (8), has been linked to the regulation of mutually exclusive transcription of the *var* multigene family (9, 10) encoding variant surface adhesion molecules associated with disease pathogenesis, virulence and chronic infection (11). In fact, ncRNAs have also been shown to be crucial for the regulation of mutually exclusive expression in other systems (12, 13).

RUF6 members contain the A- and B-box required for RNA polymerase III (Pol III) transcription and have a GC-content higher than 50% compared to 20% genome-wide. 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 (14) but not in other species causing malaria. All *RUF6* genes are interspersed between *var* genes in central chromosomal clusters (8). Additionally, these virulence gene families also contain members in subtelomeric regions (15), which are devoid of *RUF6* loci. Transcriptional profiling of *RUF6* and *var* genes demonstrated an association between the two gene families. In all the cases studied, parasites with an active *var* gene harboring a *RUF6* gene in their upstream region, also transcribe this proximal ncRNA predominantly over the other members (10). Furthermore, it has been reported that all *RUF6* loci show transcriptional activity but most RUF6 members are degraded *in situ* by a plasmodial RNase II resulting in clonal variation of these ncRNAs at the steady state RNA level (16). Importantly, these ncRNAs have been associated with the perinuclear *var* expression site and with transcriptional activation of *var* and other virulence gene families (9, 10, 17). However, the molecular mechanism of RUF6-dependent gene activation remains unknown. In particular, proteins that bind directly or form a complex with this ncRNA may lead to new insight into the process of *var* gene activation.

In this study, we shed light on the mode of action of this ncRNA by exploring its chromatin interaction partners. To this end, we adopted a recently reported method to identify ncRNA interacting partners to the cellular context of malaria parasites. This method called Chromatin Isolation by RNA Purification (ChIRP) (18) allowed the identification of several

candidate chromatin factors linked to gene activation that co-precipitate with RUF6. This work points to this ncRNA as a potential chromatin assembly factor involved in gene transcription of virulence genes.

RESULTS

Identification of RUF6 protein binding partners *in vitro*

Proteins are often associated with ncRNAs and their functionality (1). Furthermore, the predicted secondary structures (9) of the RUF6 family are highly stable and conserved within the different members suggesting potential protein binding regions that may contribute to *var* gene activation. Additionally, the lack of high sequence homology between *var* and *RUF6* genes, suggests that if the ncRNA can bind to *var* loci, such interaction should be mediated by protein partners. To test this hypothesis, we conducted electrophoretic mobility shift assays (EMSAs) using a biotinylated probe for a RUF6 member (PF3D7_1241000). A specific gel shift was observed when incubating the probe with nuclear extract of asynchronous parasites and was reproducible among four independent experiments. The specific shift was verified by competition with unlabeled probes. Namely, in presence of the unlabeled probe (200-800x), the target biotinylated probe migrated similarly as the free probe. Conversely, in presence of unspecific competitor, the shift of the probe did not have a significant change (Fig. 1).

To identify the RUF6 binding proteins, we performed streptavidin-biotin affinity purification. A biotinylated RUF6 probe was used in parallel with an antisense probe as a control. Bound proteins were eluted with RNase A and analyzed by SDS-PAGE followed by silver stain prior to mass spectrometry. Eluted proteins were identified after trypsin digestion by LC-MS/MS using Orbitrap-based detection of HCD. In total 177 unique peptides from 45 *P. falciparum* proteins were identified and 12 proteins showed enrichment in the RUF6 probe compared to the control in two independent experiments (Table S1). Other 33 candidate proteins were identified with at least 2 unique peptides in only one experiment. Interestingly, the candidate protein with highest peptide enrichment which also had no peptides identified in the control pulldown corresponds to a transcription factor (PRE-binding protein, PF3D7_1011800) activating gene expression in the intraerythrocytic stage and containing four K-homology RNA-binding domains (19).

However, due to the limitations of the *in vitro* interaction and the lack of a high number

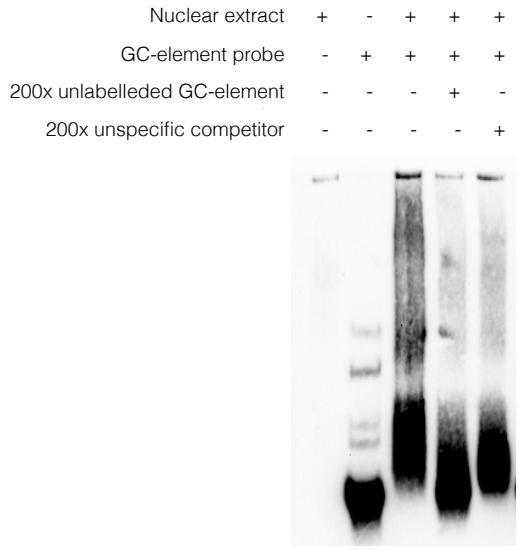


Figure 1. RNA electrophoretic mobility shift assays for GC-rich RUF6. Representative example of a native EMSA gel showing specific shift for the RUF6 probe in presence of nuclear extract. Lane 1: nuclear extract only; lane 2: biotinylated probe only; lane 3: nuclear extract and biotinylated probe; lane 4: unlabeled probe, nuclear extract and biotinylated probe; lane 5: unlabeled unspecific competitor, nuclear extract and biotinylated probe. tRNA is used to minimize unspecific binding in all lanes.

of promising candidates, we used ChIRP followed by mass spectrometry (ChIRP-MS) in order to explore the protein interactome in the native cellular context of asexual parasites prior to establish a list of candidate genes.

Optimization of ChIRP for RUF6 in *P. falciparum* blood stages

In order to elucidate the interactome of RUF6, we used ChIRP on the WT 3D7 clone B6 which expresses the *RUF6* gene PF3D7_0412800 and the adjacent *var* gene PF3D7_0412700 (10). We designed two sets of antisense probes (odd and even) hybridizing the active ncRNA, targeting the regions less prone to base pair on a loop according to the predicted secondary structure (Fig. 2A, Table S2). Additionally, we used scrambled probes of similar size and GC-content as negative controls (Table S2). In a pilot experiment with chromatin harvested at 24 hpi, we first tested different crosslinking conditions and analyzed the retrieval of target and control RNAs compared to the input (Fig. 2B). For all the tested conditions, the two sets of probes retrieved specifically RUF6 and not the control RNA. Crosslinking with 1%

formaldehyde yielded the highest retrieval (50-80%) but has been suggested suboptimal for RNA crosslinking (20). Subsequent experiments were performed with 1% formaldehyde and 1% glutaraldehyde, which yielded the second highest retrieval (25-40%), and 3% formaldehyde for ChIRP-MS experiments, as previously suggested (21).

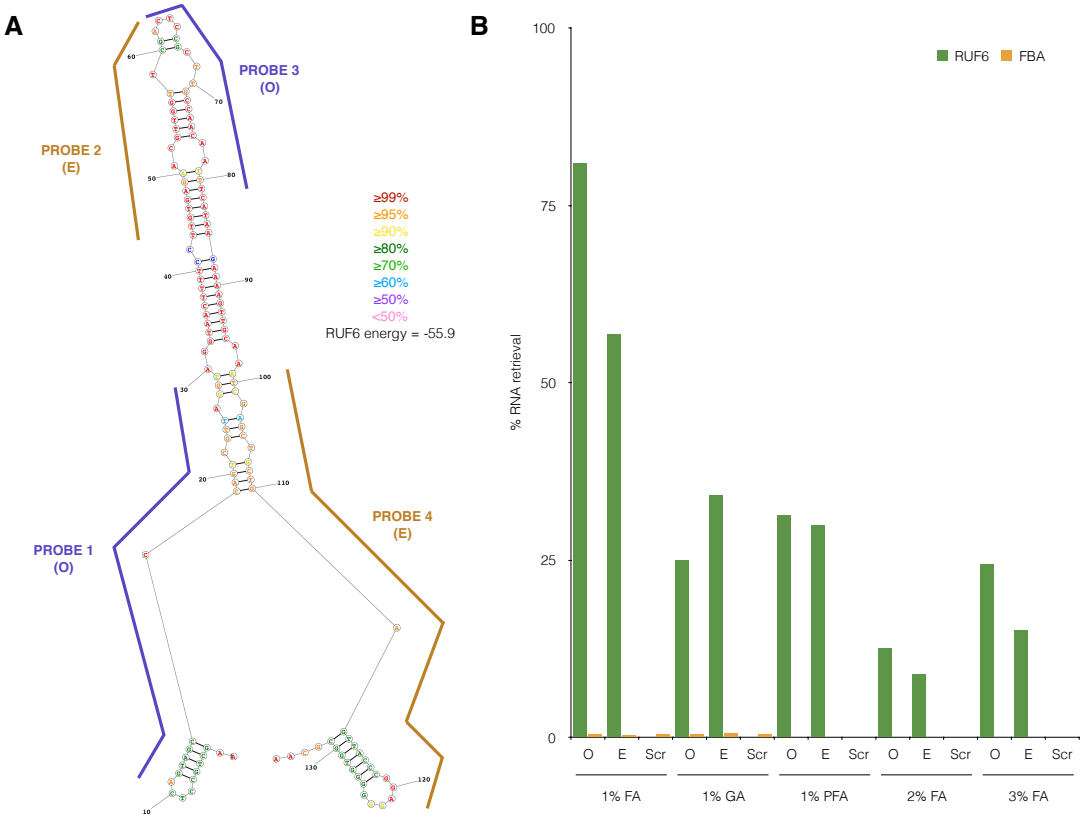


Figure 2. ChIRP of RUF6. (A) Lowest free energy secondary structure of the RUF6 coded by PF3D7_0412800, as predicted by RNAstructure bioinformatics webserver. The four oligonucleotide probes used for ChIRP are shown along their binding regions of the ncRNA structure. Odd probe (O) set is colored in purple and even probe (E) set is colored in brown. Base coloring in the structure indicates probability of pairing. **(B)** Percentage of RNA retrieval in ChIRP compared to input samples for different crosslinking conditions using odd (O), even (E) and scrambled (Scr) sets of probes. Transcript levels were assessed by RT-qPCR and fructose-bisphosphate aldolase (FBA, PF3D7_1444800) levels were used as negative control. FA, formaldehyde; GA, glutaraldehyde; PFA, paraformaldehyde.

Identification of RUF6-binding proteins by ChIRP-MS

We performed ChIRP followed by MS at 18 hpi to identify proteins that interact with RUF6 *in vivo*. We used odd, even and scrambled sets of probes, with an additional control sample that underwent RNAse A treatment in order to discard further non RUF6-specific interactions (Fig. 3A). Using this method, we obtained sixfold increase of the number of unique peptides compared to the in vitro pulldowns: 977 compared to 177.

We did 4 experimental replicates and performed statistical analysis for enrichment in samples with odd and even probes compared to those with scrambled probes and RNAse A treatment. In order to restrict the list of candidate proteins, we excluded those appearing in only one or two replicates and those presenting with high enrichment compared to only one of the two controls. Finally, we obtained a list of 13 candidates and we selected 6 to begin with the further validation according to their RNA-binding properties (Fig. 3B, Table 1). We selected uncharacterized putative RNA-binding proteins, a conserved protein of unknown function and transcription and chromatin associated proteins such a subunit of RNA polymerase II (Pol II) and an helicase. Validation for these 6 candidates is currently ongoing.

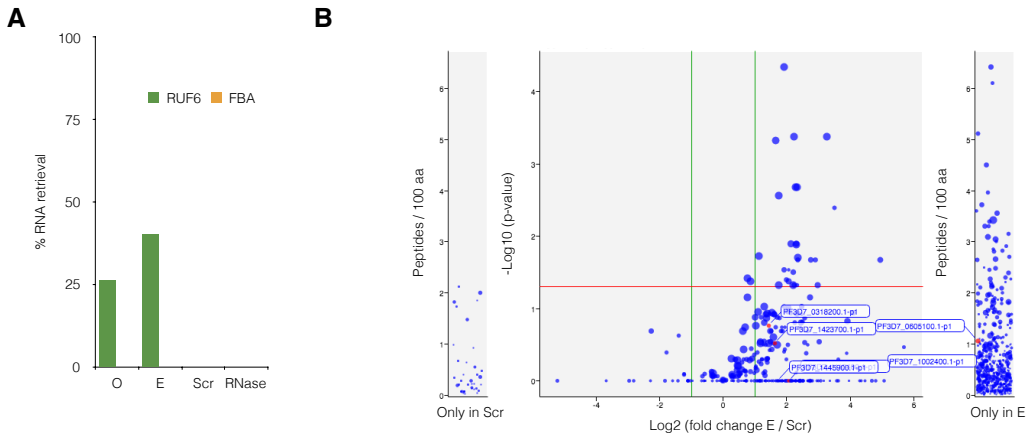


Figure 3. RUF6 ChIRP-MS. (A) Percentage of RNA retrieval in ChIRP-MS compared to input samples using odd (O), even (E) and scrambled (Scr) sets of probes, and in RNase A treated sample. Transcript levels were assessed by RT-qPCR and fructose-bisphosphate aldolase (FBA, PF3D7_1444800) levels were used as negative control. Only one out of four replicates is shown as an example. **(B)** Volcano plot of enrichment for all replicates in samples with even probes (E) compared to those with scrambled probes (Scr). Highlighted protein IDs belong to the final candidate list (see Table 2).

Table 2. Putative RUF6 interacting candidates identified by ChIRP-MS. Protein candidates from ChIRP identified by LC-MS/MS with at least three peptides in two independent replicates and enriched over the control. Highlighted proteins are currently being functionally validated. Number of peptides for samples with odd (O), even (E) and scrambled (Scr) sets of probes and RNase A treated sample (R) are shown for the 4 replicates.

Protein ID	**Annotation**	**MW (kDa)**	**S 1**	**S 2**	**S 3**	**S 4**	**R 1**	**R 2**	**R 3**	**R 4**	**O 1**	**O 2**	**O 3**	**O 4**	**E 1**	**E 2**	**E 3**	**E 4**
PF3D7_1237700	conserved protein, unknown function	23.7	3	-	-	-	5	-	-	-	-	-	3	-	5	4	4	-
PF3D7_1468700	eukaryotic initiation factor 4A	45.3	-	-	3	-	7	-	-	-	-	-	3	-	12	4	-	-
PF3D7_1002400	**transformer-2 protein homolog beta, putative**	30.1	-	-	4	-	5	-	-	-	-	-	-	-	7	5	6	-
PF3D7_1002400	**transformer-2 protein homolog beta, putative**	29.8	-	-	4	-	4	-	-	-	-	-	-	-	6	5	5	-
PF3D7_1202900	high mobility group protein B1	11.3	-	-	3	-	4	-	-	-	-	-	-	-	7	3	4	-
PF3D7_1110400	**RNA-binding protein, putative**	173.9	-	-	-	-	13	-	-	-	-	-	-	-	21	4	6	3
PF3D7_1423700	**conserved Plasmodium protein, unknown function**	183.6	-	-	-	-	6	-	-	-	-	-	-	-	7	3	3	-
PF3D7_0605100	**RNA-binding protein, putative**	85.6	-	-	-	-	4	-	-	-	-	-	-	-	3	3	3	-
PF3D7_1445900	**ATP-dependent RNA helicase DDX5, putative**	60	-	-	-	-	-	-	-	-	-	-	-	-	5	4	-	4
PF3D7_0318200	**DNA-directed RNA polymerase II subunit RPB1**	278.7	-	-	-	-	-	-	-	-	-	-	-	-	5	-	3	3
PF3D7_1023900	chromodomain-helicase-DNA-binding protein 1 homolog, putative	381.3	-	-	-	-	-	-	-	-	-	-	-	-	3	-	3	-
PF3D7_1459000	ATP-dependent RNA helicase DBP5	84,3	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	5

DISCUSSION

Despite the association of ncRNA with malaria parasite virulence and pathogenesis, the biology of ncRNAs remains largely unexplored in *P. falciparum*. For this reason, we established first a robust ChIRP protocol, to explore the interactome of ncRNAs of this parasite in their natural chromatin context. Once validated, this new tool was used to identify the protein interactome of RUF6, a ncRNA reported to be involved in activation of the *var* virulence gene family (9, 10). The lack of *var* mRNA retrieved with ChIRP, suggests the mechanism of action of RUF6 may not involve direct interaction with *var* transcripts during transcription (data not shown).

We used four biological replicates with a parallel RNAse treatment control to identify protein candidates that either bind directly or are associated with RUF6. Our ChIRP-MS allowed for the identification of potential interaction proteins. Those hits were compared with RUF6 pulldowns obtained from nuclear extracts. Both techniques revealed strikingly different candidates, demonstrating that the "native" ChIRP method may detect interactions that cannot be reconstituted once the nucleus is disrupted. We decided to give priority to candidates obtained by the ChIRP method. Among those candidates are genes that encode proteins implicated in gene transcription and candidates with unknown function. We hypothesize that some proteins bind specifically to RUF6 and others form complexes via protein-protein interactions. Interestingly, *RUF6* genes are predicted to be transcribed by RNA Pol III but one of the candidates is a subunit of RNA Pol II, which transcribes *var* genes. In fact, it has recently been reported that SINEs transcribed by Pol III can act as enhancers of Pol II genes in neurons (22). Further validation will confirm whether RUF6 could play a similar role in activation of *var* genes. We are currently analyzing these candidate proteins to gain insight in the underlying molecular mechanism of RUF6-dependent activation of clonally variant virulence gene families. Several key questions linked to the expression of *var* genes, such as mutually exclusive expression and switching remain elusive and exploring the function of this ncRNA may open avenues to understand these important processes in malaria biology.

METHODS

Parasite culture and synchronization

Blood stage *P. falciparum* parasites were cultured as previously described (11).

Parasites were synchronized with a 6 h time window by sorbitol lysis during ring stage, followed by plasmagel enrichment in schizont stage and another sorbitol treatment 6 h after corresponding to 3±3 hpi. Synchronized parasites were harvested a 3.3% haematocrit and ~2-4% parasitemia. Parasite development was monitored by Giemsa staining.

Chromatin isolation by RNA purification (ChIRP)

ChIRP was performed as previously described (18, 20) with the following modifications. $8x10^8$ parasites were harvested and subjected to saponin lysis prior to 1% glutaraldehyde crosslinking or after 1% or 2% formaldehyde or 1% paraformaldehyde crosslinking. An extra lysis step was performed prior to sonication with lysis buffer containing 1% NP40. DNA sonication for different times at 4°C (30 s on/off intervals, high power) with a Bioruptor (Diagenode) was assessed on an agarose gel. Optimal sonication times varied depending on the crosslinking conditions: 12 min for 1% formaldehyde or paraformaldehyde, 17 min for 2% formaldehyde, 30 min for glutaraldehyde and 22 min for 3% formaldehyde.

ChIRP-MS was performed as previously described (21) with the following modifications. $4x10^{10}$ parasites 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 SDS-PAGE followed by silver stain prior to mass spectrometry.

Mass spectrometry analysis

On-beads digestion was performed for one hour with 0.6 μg of trypsine/LysC (Promega). Samples were then loaded onto home-made C18 StageTips for desalting. Peptides were eluted using 40/60 MeCN/H2O + 0.1% formic acid and vacuum concentrated to dryness.

The online chromatography was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled online 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 PepMapTM 100, Thermo Scientific) with buffer A (2/98 MeCN/H_2O in 0.1% formic acid) at a flow rate of 2.5 μL/min over four minutes. Separation was then performed on a 50 cm x 75 μm C18 column (nanoViper Acclaim PepMapTM RSLC, 2 μm, 10 nm, 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 375−1,500 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 × 105. Ions with charge state from 2+ to 6+ were selected for screening. Normalized collision energy (NCE) was set at 27 with a dynamic exclusion of 40s.

For identification, the data were searched against the *P. falciparum* FASTA database using Sequest HF through proteome discoverer (version 2.2). Enzyme specificity was set to trypsin and a maximum of two-missed cleavage sites were allowed. Oxidized methionine 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 (version 3.6) (24). FDR calculation used Percolator 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 (25). 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 ($n\bar{4}$). 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 with a control threshold set to 0.05. Proteins with at least 1.5 fold enrichment, P value < 0.05, and at least three total peptides in all replicates were considered significantly enriched in sample comparisons.

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

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 (9). 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.

Nuclear extract preparation

Asynchronous parasite cultures were lysed in 0.15% saponin in PBS followed by three washes in PBS. Parasites were resuspended in cold low salt lysis buffer (20 mM Tris-HCl, 10 mM NaCl, 1.5 mM $MgCl_2$, 2 mM DTT, 1% Triton and protease inhibitors) and incubated for 1 h at 4°C under rotation. Nuclei were pelleted at 17000 g for 20 min at 4°C and the supernatant corresponding to the cytoplasmic fraction was removed and stored at -80°C. Nuclei was extracted with high salt lysis buffer (20 mM Tris-HCl, 600 mM NaCl, 1.5 mM $MgCl_2$, 2 mM DTT, 1% Triton and protease inhibitors) for 30min at 4°C under rotation. After short vortexing, the sample was sonicated for 3 min (30 s on/off intervals, high power) at 4°C with a Bioruptor (Diagenode). Cell debris was pelleted by at 17000 g for 15 min at 4°C. The supernatant containing the nuclear fraction was stored at -80°C.

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 μg 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 μL of binding reaction containing 5 μg of nuclear extract, 2 μg of tRNA, 10, 20 or 40 fmol of biotinylated probe and when indicated 8 pmol of unlabelled probe were incubated in REMSA buffer (10mM HEPES, 20mM KCl, 1mM $MgCl_2$, 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).

Affinity purification of RNA binding proteins

RNA-binding proteins were purified by streptavidin-biotin affinity. Nuclear extract corresponding to 4x109 parasites was concentrated and subsequently precleared with Dynabeads MyOne Streptavidin T1 (Thermo Fisher Scientific) for 30min at 4°C. Same RNA probes used for EMSAs were relaxed and refolded by incubation to 64°C and gradual cooling to 4°C prior to binding. RNA pulldown was performed based on a recently published protocol (26). Precleared nuclear extract was diluted in binding and wash buffer (20 mM Tris-HCl, 300 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT, 2 mM ribonucleoside vanadyl complex and protease inhibitors) containing RNAsin Ribonuclease Inhibitor (Promega) and tRNA as unspecific competitor. After addition of 500 pmol of 5' biotinylated RNA (GC-rich ncRNA or control) the binding reaction was incubated for 1 h at 4°C under rotation. To isolate RNA-protein complexes, Dynabeads MyOne Streptavidin T1 (Thermo Fisher Scientific) were added to the binding reaction before further incubation of 1 h at 4°C under rotation. Bead-RNA-protein complexes were immobilized on a magnetic rack and washed 5 times with binding and wash buffer containing RNAsin Ribonuclease Inhibitor (Promega), changing the tube every two washes, and once with elution buffer (20 mM Tris-HCl, 30 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT). Proteins were eluted after 15 min at 4°C incubation with 100 μL of elution buffer containing 3μg of RNase A. A 10 μL aliquot was used for SDS-PAGE followed by silver staining and the remaining 90 μL of the final samples were processed for mass spectrometry analysis.

ACKNOWLEDGEMENTS

We thank Valentin Sabatet from assistance with ChIRP-MS data analysis. This work was supported by a European Research Council Advanced Grant (PlasmoSilencing 670301) and the French Parasitology consortium ParaFrap (ANR-11-LABX0024) awarded to A.S and "Région Ile-de-France" and Fondation pour la Recherche Médicale grants to D.L.

AUTHOR CONTRIBUTIONS

A.B.S. and A.S. conceived and designed the experiments. A.B.S. performed ChIRP development, EMSAs and RBP pulldowns. G.M.D and A.B.S performed ChIRP-MS experiments. F. D. and D. L. performed mass spectrometry analyses. A.B.S. wrote the manuscript.

REFERENCES

1. Quinn, JJ, Chang, HY. 2015. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17:47–62.
2. Cech, TR, Steitz, JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157:77–94.
3. WHO. 2018. World Malaria Report 2018.
4. Patankar, S, Munasinghe, A, Shoaibi, A, Cummings, LM, Wirth, DF. 2001. Serial analysis of gene expression in Plasmodium falciparum reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. Mol Biol Cell 12:3114–3125.
5. Broadbent, KM, Park, D, Wolf, AR, Van Tyne, D, Sims, JS, Ribacke, U, Volkman, S, Duraisingh, M, Wirth, D, Sabeti, PC, Rinn, JL. 2011. A global transcriptional analysis of Plasmodium falciparum malaria reveals a novel family of telomere-associated lncRNAs. Genome Biol 12:R56.
6. Siegel, TN, Hon, CC, Zhang, Q, Lopez-Rubio, JJ, Scheidig-Benatar, C, Martins, RM, Sismeiro, O, Coppée, JY, Scherf, A. 2014. Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum. BMC Genomics 15:150.
7. Broadbent, KM, Broadbent, JC, Ribacke, U, Wirth, D, Rinn, JL, Sabeti, PC. 2015. Strand-specific RNA sequencing in Plasmodium falciparum malaria identifies developmentally regulated long non-coding RNA and circular RNA. BMC Genomics 16:454.
8. Chakrabarti, K, Pearson, M, Grate, L, Sterne-Weiler, T, Deans, J, Donohue, JP, Ares, M. 2007. Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. RNA 13:1923–1939.
9. 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.
10. Barcons-Simon, A, Cordon-Obras, C, Guizetti, J, Bryant, JM, Scherf, A. 2019. CRISPR interference of a clonally variant GC-rich non-coding RNA family leads to general repression of var genes in Plasmodium falci- parum. Manuscript under review
11. Miller, LH, Baruch, DI, Marsh, K, Doumbo, OK. 2002. The pathogenic basis of malaria. Nature 415:673–679.

12. Zhang, Y, Tycko, B. 1992. Monoallelic expression of the human H19 gene. Nat Genet 1:40–44.
13. Brockdorff, N, Ashworth, A, Kay, GF, McCabe, VM, Norris, DP, Cooper, PJ, Swift, S, Rastan, S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71:515–526.
14. Otto, TD, Gilabert, A, Crellen, T, Böhme, U, Arnathau, C, Sanders, M, Oyola, SO, Okouga, AP, Boundenga, L, Willaume, E, Ngoubangoye, B, Moukodoum, ND, 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.
15. 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.
16. Zhang, Q, Siegel, TN, Martins, RM, Wang, F, Cao, J, Gao, Q, Cheng, X, Jiang, L, Hon, CC, Scheidig-Benatar, C, Sakamoto, H, Turner, L, Jensen, AT, Claes, A, Guizetti, J, Malmquist, NA, Scherf, A. 2014. Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. Nature 513:431–435.
17. 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.
18. 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.
19. Komaki-Yasuda, K, Okuwaki, M, Nagata, K, Kawazu, S, Kano, S. 2013. Identification of a novel and unique transcription factor in the intraerythrocytic stage of Plasmodium falciparum. PLoS One 8:e74701.
20. Chu, C, Quinn, J, Chang, HY. 2012. Chromatin isolation by RNA purification (ChIRP). J Vis Exp
21. Chu, C, Zhang, QC, da Rocha, ST, Flynn, RA, Bharadwaj, M, Calabrese, JM, Magnuson, T, Heard, E, Chang, HY. 2015. Systematic discovery of Xist RNA binding proteins. Cell 161:404–416.
22. Policarpi, C, Crepaldi, L, Brookes, E, Nitarska, J, French, SM, Coatti, A, Riccio, A. 2017. Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. Cell Rep 21:2879–2894.

23. Lopez-Rubio, JJ, Mancio-Silva, L, Scherf, A. 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5:179–190.
24. 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.
25. Valot, B, Langella, O, Nano, E, Zivy, M. 2011. MassChroQ: a versatile tool for mass spectrometry quantification. Proteomics 11:3572–3577.
26. Marín-Béjar, O, Huarte, M. 2015. RNA pulldown protocol for in vitro detection and identification of RNA-associated proteins. Methods Mol Biol 1206:87–95.

SUPPLEMENTARY DATA

Exploring the interactome of a ncRNA gene family that activates expression of a gene family linked to severe malaria

Anna Barcons-Simon[1,2,3,4], Gretchen M. Diffendall[1,2,3,4], Florent Dingli[5], Damarys Loew[5] and Artur Scherf[1,2,3,#]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[5]Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, F-75248 Paris, France
[#]Correspondence: artur.scherf@pasteur.fr

Preliminary data

Table S1. Putative RUF6 interacting candidates identified *in vitro*. Protein candidates from *in vitro* protein affinity purification identified by LC-MS/MS with at least two peptides in two independent experiments and enriched over the control. Number of peptides is shown.

Protein ID	Annotation	MW (kDa)	RUF6 (exp1)	Control (exp1)	RUF6 (exp2)	Control (exp1)
PF3D7_0818900	Heat shock protein 70	73.9	14	6	3	2
PF3D7_1011800	PRE-binding protein	131.6	6	-	10	-
PF3D7_1357000	Elongation factor 1-alpha	49.0	7	3	7	-
PF3D7_0818200	14-3-3 protein	30.2	6	5	7	-
PF3D7_0500800	Mature parasite-infected erythrocyte surface antigen	168.3	9	3	2	-
PF3D7_1006200	DNA/RNA-binding protein Alba 3	12.0	2	2	7	4
PF3D7_0617200	conserved Plasmodium protein, unknown function	53.0	6	2	2	-
PF3D7_0202000	Knob-associated histidine-rich protein	69.4	3	-	3	2
PF3D7_1410400	Rhoptry-associated protein 1	90.1	3	1	2	1
PF3D7_0929400	RhopH2	162.6	2	-	3	-
PF3D7_1471100	Exp-2 protein	33.0	2	-	2	-
PF3D7_0905400	RhopH3	104.9	2	-	2	-

Table S2. ChIRP probe sequences. Biotinylated (3'-TEG Biotin) DNA oligonucleotides used for 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

2.3.1 Insight on the RUF6 interactome by ChIRP

The highlights of this section are:

- Development of ChIRP-seq method for studying ncRNA interaction with the genome in *P. falciparum*.

The results of this section are supplementary ChIRP results not included in the previous draft and presented as preliminary data under the title "Insight on the DNA and RNA interaction partners of a GC-rich ncRNA family in *P. falciparum*". My contribution to this work included: conceptualization, experimental design, performing ChIRP, analyzing the data and writing.

Insight on the DNA and RNA interaction partners of a GC-rich ncRNA family in *P. falciparum*

Anna Barcons-Simon[1,2,3,4], and Artur Scherf[1,2,3,#]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[#]Correspondence: artur.scherf@pasteur.fr

Preliminary data

INTRODUCTION

The mechanisms by which ncRNAs mediate chromatin regulation are very diverse still often involve targeting to a specific genomic locus. These interactions can be either directly with the DNA sequence, through canonical Watson-Crick base pairing or non-canonical complexes such as triple helices, or mediated by proteins or nascent transcripts (1). In the human malaria parasite *Plasmodium falciparum*, the role of a the GC-rich ncRNA family in mutually exclusive activation of the *var* virulence genes has been inferred from descriptive observations and indirect effects of ncRNA perturbation (2, 3). However, the molecular mechanism of action of this ncRNA, annotated as RNA of unknown function 6 (RUF6), remains unknown. Here, we developed ChIRP-seq (4), to study the potential interactions of this ncRNA with the parasite genome with the aim of gaining insight into RUF6 mode of action. Genome-wide analysis revealed RUF6 enrichment in several *var* loci but also in many other genes, and no clear distinct interaction was observed specifically for the active *var* gene. Additionally, ChIRP-RNA-seq was performed in order to investigate whether RUF6 might interact with mRNA or other ncRNAs but no RNA interaction partners could be identified.

RESULTS

Optimization of ChIRP-seq for RUF6 in *P. falciparum* blood stages

In order to elucidate the interactions of RUF6 with the genome, we used ChIRP on the WT 3D7 clone B6 which expresses the RUF6 gene PF3D7_0412800 and the adjacent *var* gene PF3D7_0412700 (3). We performed the experiments in parallel for synchronized parasites at 14 hours post invasion (hpi) and 24 hpi, timepoints with *var* and *RUF6* transcriptional peaks respectively. We used two sets of antisense probes (odd and even) hybridizing the active ncRNA and negative control scrambled probes, as described in the previous section. Additionally, we used sense probes as a second negative control (Table S1). For all the tested crosslinking conditions described in the previous section, 1% formaldehyde gave the highest retrieval (50-80%) but has been suggested suboptimal for RNA crosslinking (5). Thus, we tested both 1% formaldehyde and 1% glutaraldehyde, which gave the second highest retrieval (25-40%), for performing ChIRP-seq. DNA eluted from ChIRP was sequenced and similar enrichment over input profiles were observed in samples crosslinked with 1% formaldehyde and 1% glutaraldehyde. Specific peaks in samples with the odd and even probe sets not present in the scrambled control sample were similar with both crosslinking conditions. However, samples crosslinked with 1% glutaraldehyde displayed significantly less background of unspecific enrichment also observed in the control (Fig. S1). Thus, 1% glutaraldehyde crosslinked samples have been used for the analysis.

Study of RUF6 genome interactions by ChIRP-seq

Sequencing analysis of the DNA eluted by ChIRP showed specific enrichment in all 15 *RUF6* loci not present in the scrambled control but observed at lower levels with sense probes that do not retrieve the ncRNA (Fig. 1A and Fig. 1C). The high sequence similarity amongst all *RUF6* members (>90%) is likely to cause the probes designed for the predominantly transcribed ncRNA (Fig. 1A and Fig. 1B) to bind transcripts from all members, and additionally to also bind all *RUF6* loci independently of the transcript, since enrichment is also observed when using sense probes (Fig. 1C).

ChIRP on chromatin harvested at 12 and 24 hpi showed similar results but levels were much higher at 24 hpi, so we proceeded with the analyses at this time point. Genome-wide analysis of fold enrichment detected 3405 genes with peaks in the even probes sample, 2257

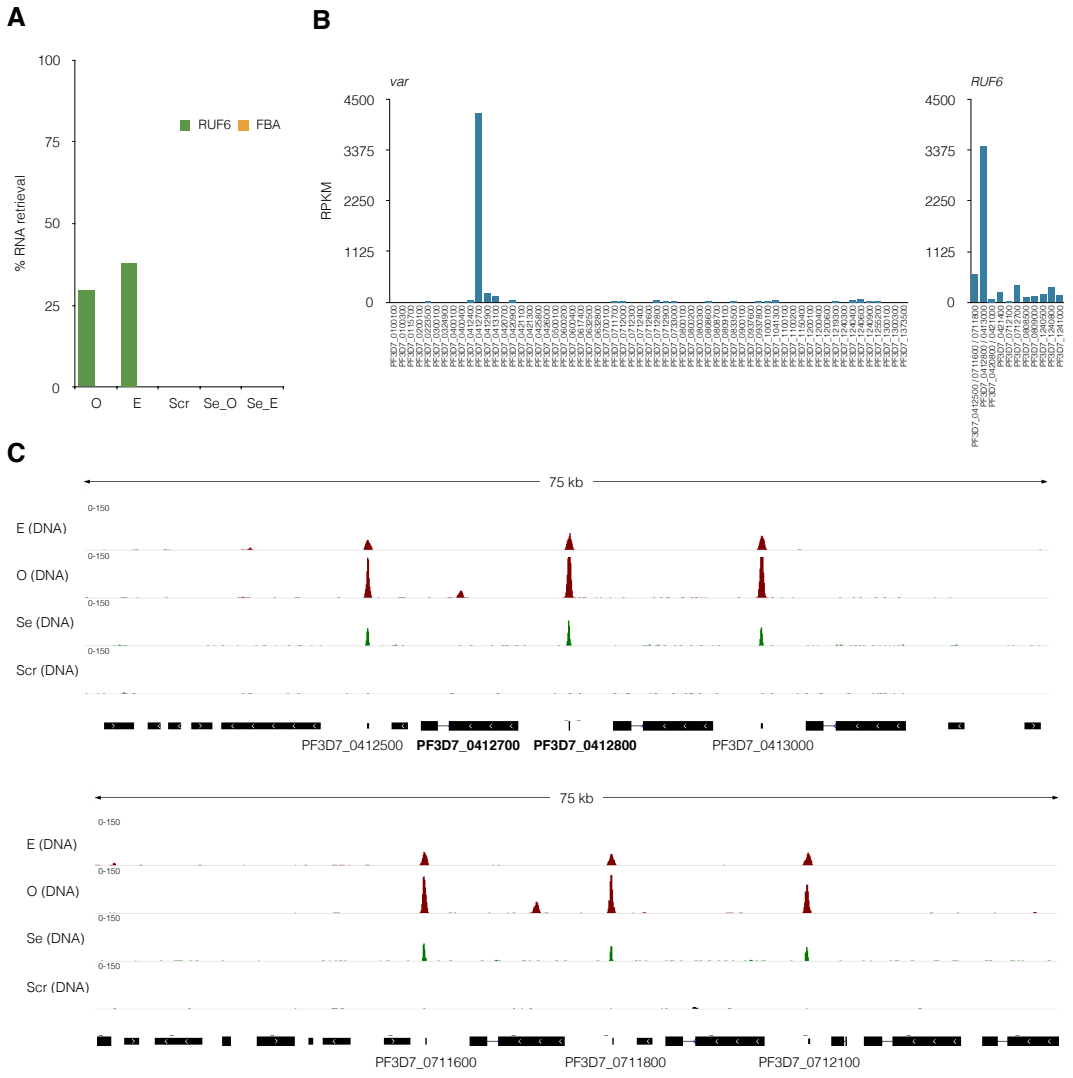


Figure 1. ChIRP-seq of GC-rich RUF6. (A) Percentage of RNA retrieval in ChIRP compared to input samples using odd (O), even (E), scrambled (Scr), sense odd (Se_O) and sense even (Se_E) sets of probes. Transcript levels were assessed by RT-qPCR and fructose-bisphosphate aldolase (FBA, PF3D7_1444800) levels were used as negative control. **(B)** Transcriptional profile of *var* genes and *RUF6* at 12 hpi in 3D7 B6 clone assessed by RNA-seq with predominant transcription of the *RUF6* (PF3D7_0412800) adjacent to the active *var* locus (PF3D7_0412700). **(C)** IGV tracks for RUF6 ChIRP-seq of 3D7 B6 clone at 24 hpi. Fold enrichment compared to input is shown for samples pulled with odd (O), even (E), sense (Se) and scrambled (Scr) probes. Two regions with *RUF6* genes in chromosomes 4 and 7 are shown. Upper tracks contain the active *var* and active *RUF6* genes (in bold). Lower tracks contain *RUF6* member with second highest transcriptional levels (PF3D7_0711800) that is located upstream of a *var* gene. *RUF6* genes are labeled.

in the odd probes sample and 1916 in the scrambled probes sample with FDR < 0.05. 482 genes had significant enrichment only in the odd and even probes samples but not in the control sample with scrambled probes. Amongst those, the ones with highest peak score were all members of the *RUF6* family. This list contains 5 *var* genes, 2 of which present the highest scores after *RUF6* loci, but not the active member for the studied clone. Gene ontology analysis for molecular function of the genes with significant RUF6 enrichment, other than their own loci, revealed that many of them code for proteins with transferase activity or cation binding properties (Fig. 2). Analysis of the gene ontology enrichment for biological process revealed that many are involved in carbon fixation, mRNA catabolism, cellular communication, cellular response to stimulus and actin filaments regulation (Fig. S2). In-depth analysis of the genes with the strongest RUF6 enrichment peaks is currently ongoing.

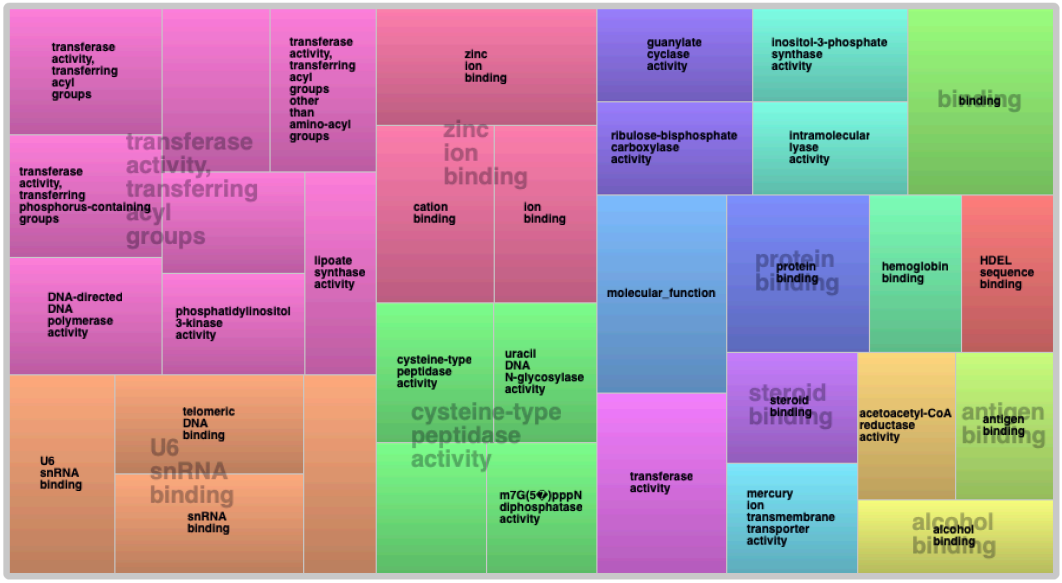


Figure 2. ChIRP-seq gene ontology enrichment. Tree map created with Revigo (revigo.irb.hr) for gene ontology terms related to molecular function of genes with ChIRP-seq peaks in samples with odd and even probes but not in control with scrambled probes.

Interestingly, the exon 1 of the active *var* gene (PF3D7_0412700), located adjacent to the predominantly transcribed *RUF6* gene (PF3D7_0412800), showed specific enrichment

with the odd probes set but not with the controls or even sets of probes (Fig. 1C). We infer that the sample with the even probe set did not display such enrichment due to lower total sequencing depth, or alternatively, because the even probes targeted regions required for such interaction. Furthermore, the silent *var* gene adjacent to the *RUF6* gene that has the second highest transcriptional levels (PF3D7_0711800) (Fig. 1B), also displayed similar enrichment in exon 1, suggesting that transcripts from ncRNAs adjacent and upstream of a *var* gene could interact with that gene (Fig. 1B). Notwithstanding, it would be required to reproduce the data in another WT clone that has different active *var* and *RUF6* genes and also test whether the same interaction can be observed when using the even set of probes in another experiment.

Analysis of ChIRP eluted RNA

Since ChIRP-seq results did not give a clear clue to a regulatory mechanism for mutually exclusive *var* gene activation, we analyzed the ChIRP eluted RNA to elucidate whether RUF6 could interact with *var* transcripts and participate in the regulation of nascent transcription or mRNA stability.

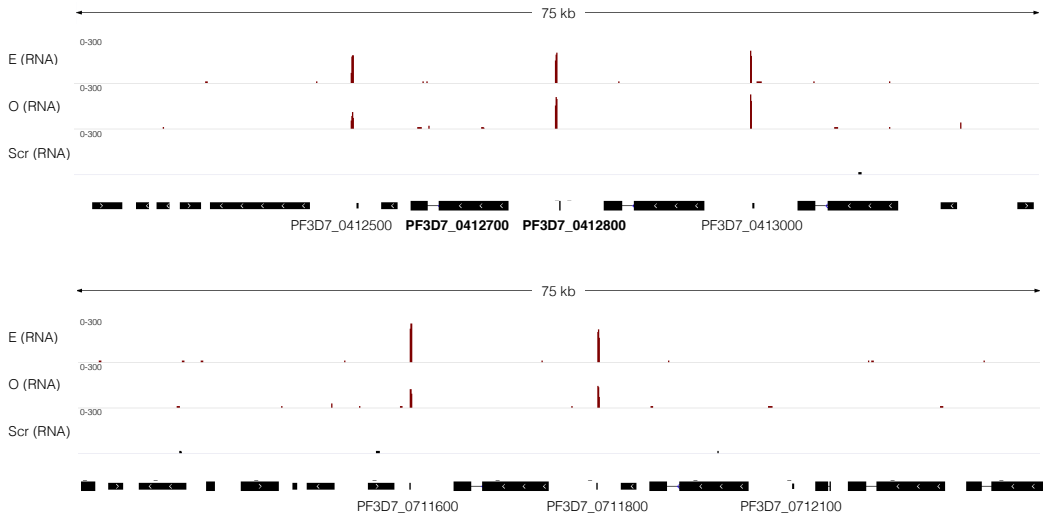


Figure 3. ChIRP-RNA-seq of GC-rich RUF6. IGV tracks for RUF6 ChIRP-RNA-seq of 3D7 B6 24 hpi chromatin. Fold enrichment compared to input is shown for samples pulled with odd (O), even (E), and scrambled (Scr) probes. Regions with *RUF6* genes in chromosomes 4 and 7 are shown. *RUF6* genes are labeled. Targeted active *RUF6* (PF3D7_0412800) and active *var* (PF3D7_0412700) genes are highlighted (in bold).

Sequencing of ChIRP eluted RNA, treated with DNase I, pulled the targeted RUF6 but also most other members of the highly homologous ncRNA family (Fig. 3), as expected from the ChIRP-seq results (Fig. 1C). No other significant enriched RNA was observed suggesting albeit the probes are not specific only for predominantly transcribed RUF6, there is no off-target effect and no RNAs other than RUF6 are nonspecifically pulled.

DISCUSSION

The role of ncRNAs remains largely unexplored in *P. falciparum*, here we optimized the ChIRP-seq protocol for this parasite, which is an important tool for their study. We analyzed the interactome of RUF6, a ncRNA reported to be involved in activation of the *var* virulence gene family (2, 3). The lack of *var* mRNA retrieved with ChIRP eluted RNA, suggests the mechanism of action of RUF6 does not involve interaction with *var* transcripts during transcription or post-transcriptionally. We could not find a clear association of RUF6 with the active *var* gene reproducible using different sets of probes. Despite detecting high enrichment in several *var* genes using both sets of probes, we could only detect interaction with the active one using one set of probes. Additionally, RUF6 enrichment was not specific for *var* genes and was present in many other genes with diverse functions. It is possible that these ncRNAs mark *var* genes with the potential of being expressed or switched to, while also regulating other genes.

We hypothesize that RUF6 might require protein partners in order to perform its function which could mediate the interactions with the genome or *var* transcripts and thus pose a limitation for using ChIP-seq to detect them.

METHODS

Parasite culture and synchronization

Blood stage *P. falciparum* parasites were cultured as previously described (11). Parasites were synchronized with a 6 h time window by sorbitol lysis during ring stage, followed by plasmagel enrichment in schizont stage and another sorbitol treatment 6 h after corresponding to 3±3 hpi. Synchronized parasites were harvested a 3.3% haematocrit and ~2-4% parasitemia. Parasite development was monitored by Giemsa staining.

Chromatin iisolation by RNA purification and next generation sequencing

ChIRP was performed as previously described (18, 20) with the following modifications. $8x10^8$ parasites were harvested and subjected to saponin lysis prior to 1% glutaraldehyde crosslinking or after 1% or 2% formaldehyde or 1% paraformaldehyde crosslinking. An extra lysis step was performed prior to sonication with lysis buffer containing 1% NP40. DNA sonication for different times at 4°C (30 s on/off intervals, high power) with a Bioruptor (Diagenode) was assessed on an agarose gel. Optimal sonication times varied depending on the crosslinking conditions: 12 min for 1% formaldehyde or paraformaldehyde, 17 min for 2% formaldehyde, 30 min for glutaraldehyde and 22 min for 3% formaldehyde.

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

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 (9). 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.

ChIRP-DNA and ChIRP-RNA library preparation and sequencing

Sequencing libraries were produced with the ChIRP eluted DNA using the MicroPlex Library Preparation Kit v2 (Diagenode) with the KAPA HiFi polymerase (Kapa Biosystems) for the PCR amplification. For each ChIRP sample a control DNA corresponding to the ChIRP input was processed in parallel.

RNA eluted from ChIRP was subjected 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. For each ChIRP sample a control RNA corresponding to the ChIRP 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. A minimum of two biological replicates were analysed for each time point.

ChIRP-seq and ChIRP-RNA-seq data analysis

Fastq files were subjected to quality control using the FastQC software (7). Sequencing reads were mapped to the P. falciparum genome (8) (PlasmoDB, v.9) using Burrows-Wheeler Alignment tool (BWA-MEM) with default settings (9). PCR duplicates were removed without further quality score filtering since 4 GC-rich ncRNA fall in low mappability regions. ChIRP-seq data were normalized over input and fold enrichment calculation and peak calling was performed using the MACS2 (10) software with default parameters and a false discovery rate (FDR) cutoff of 0.05. Data were visualized using Integrative Genomics Viewer (11).

REFERENCES

1. Koziol, MJ, Rinn, JL. 2010. RNA traffic control of chromatin complexes. Curr Opin Genet Dev 20:142–148.

2. 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.

3. Barcons-Simon, A, Cordon-Obras, C, Guizetti, J, Bryant, JM, Scherf, A. 2019. CRISPR interference of a clonally variant GC-rich non-coding RNA family leads to general repression of var genes in Plasmodium falci- parum. Manuscript under review

4. 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.

5. Chu, C, Quinn, J, Chang, HY. 2012. Chromatin isolation by RNA purification (ChIRP). J Vis Exp

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

7. Wingett, SW, Andrews, S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res 7:1338.

8. Gardner, MJ, Hall, N, Fung, E, White, O, Berriman, M, Hyman, RW, Carlton, JM, Pain, A, Nelson, KE, Bowman, S, Paulsen, IT, James, K, Eisen, JA, Rutherford, K, Salzberg, SL, Craig, A, Kyes, S, Chan, MS, Nene, V, Shallom, SJ, Suh, B, Peterson, J, Angiuoli, S, Pertea, M, Allen, J, Selengut, J, Haft, D, Mather, MW, Vaidya, AB, Martin, DM, Fairlamb, AH, Fraunholz, MJ,

Roos, DS, Ralph, SA, McFadden, GI, Cummings, LM, Subramanian, GM, Mungall, C, Venter, JC, Carucci, DJ, Hoffman, SL, Newbold, C, Davis, RW, Fraser, CM, Barrell, B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511.
9. Li, H, Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
10. Zhang, Y, Liu, T, Meyer, CA, Eeckhoute, J, Johnson, DS, Bernstein, BE, Nusbaum, C, Myers, RM, Brown, M, Li, W, Liu, XS. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137.
11. Robinson, JT, Thorvaldsdóttir, H, Winckler, W, Guttman, M, Lander, ES, Getz, G, Mesirov, JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26.

SUPPLEMENTARY DATA

Insight on the DNA and RNA interaction partners of a GC-rich ncRNA family in *P. falciparum*

Anna Barcons-Simon[1,2,3,4], and Artur Scherf[1,2,3,#]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[#]Correspondence: artur.scherf@pasteur.fr

Preliminary data

Table S1. ChIRP probe sequences. Biotinylated (3'-TEG Biotin) DNA oligonucleotides used for 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
Sense RUF6 1	AAGCTGCCYCAGTAGCCCAG
Sense RUF6 2	TBGTTAGGTAKGTARCYTTT
Sense RUF6 3	CCTTGTGAGAACGTTGGTTC
Sense RUF6 4	CCGCKTGCCRACAATTTCMT
Sense RUF6 5	GAAAAGTTGCARAYCGAGCT
Sense RUF6 6	CYCGGAGGGGGGYGGCGCAA

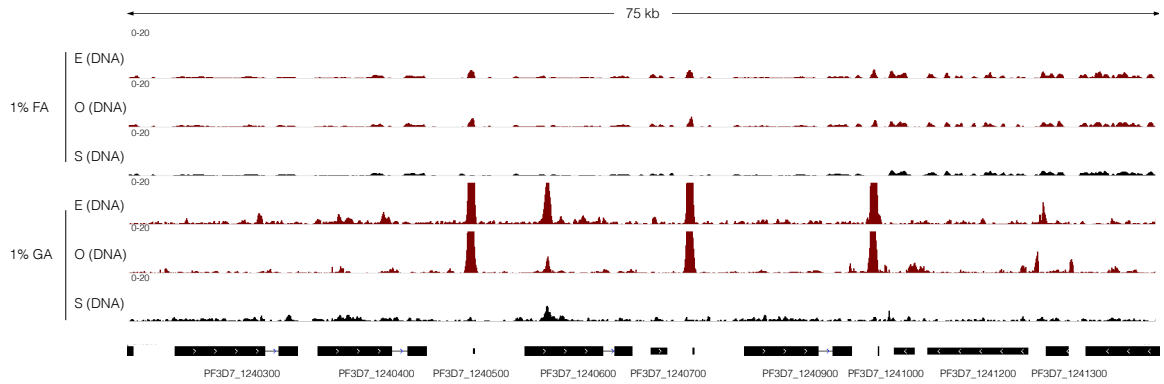


Figure S1. Formaldehyde produces more background than glutaraldehyde crosslinking. IGV tracks for RUF6 ChIRP-seq of 3D7 B6 24 hpi chromatin crosslinked with 1% formaldehyde (FA) and 1% glutaraldehyde (GA). Fold enrichment compared to input is shown for samples pulled with odd (O), even (E) and scrambled (S) probes. An example region of chromosome 12 containing several *RUF6* genes is shown.

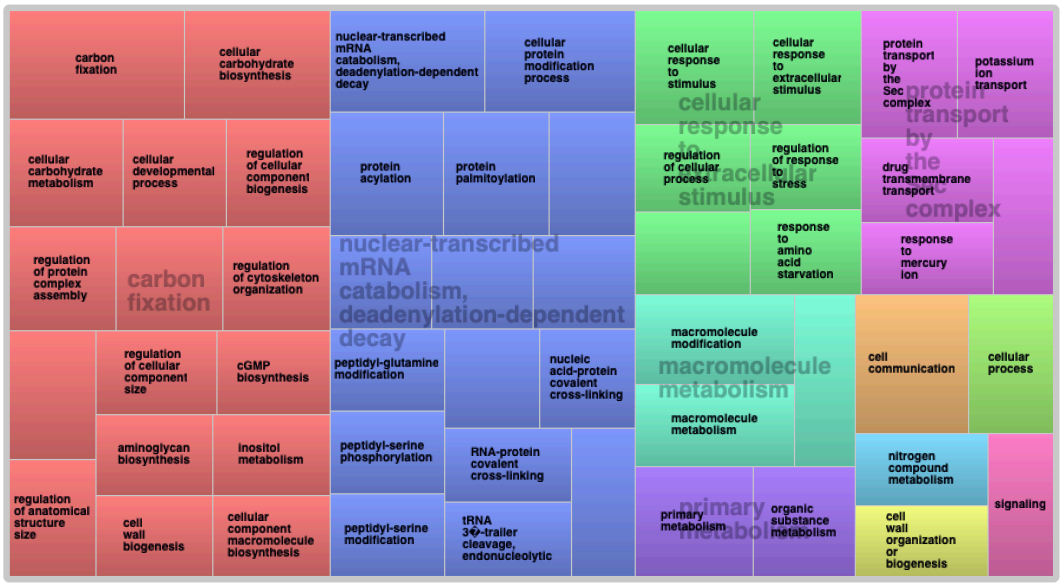


Figure S2. ChIRP-seq gene ontology enrichment. Tree map created with Revigo (revigo.irb.hr) for gene ontology terms related to biological process of genes with ChIRP-seq peaks in samples with odd and even probes but not in control with scrambled probes.

2.4 Caracterization of the GC-rich DNA element

The highlights of this section are:

- GC-rich ncRNA is transcribed by RNA Pol III.
- Development of Capture-C technology in *P. falciparum* to study genome organization.
- The GC-rich DNA elements cluster and present long-range chromosomal interactions.

The results of this section are preliminary data summarized in a draft under the title "High resolution Capture-C analysis reveals long-range interactions between dispersed Pol III transcribed GC-rich ncRNA genes in *P. falciparum*". My contribution to this work included: conceptualization, experimental design, performing Pol III inhibition assays and Capture-C experiments, analyzing the data, and writing the manuscript.

High resolution Capture-C analysis reveals long-range interactions between dispersed Pol III transcribed GC-rich ncRNA genes in *P. falciparum*

Anna Barcons-Simon[1,2,3,4,#], Carlos Cordon-Obras[1,2,3,#] and Artur Scherf[1,2,3,*]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[#]These authors contributed equally

[*]Correspondence: artur.scherf@pasteur.fr

Preliminary data

ABSTRACT

Singular expression of gene family members determines phenotypic outcomes at the single cell level in a wide variety of organisms. In the human malaria parasite *Plasmodium falciparum*, the underlying mechanisms leading to mutually exclusive expression of the major virulence gene family, encoded by the 60-member *var* gene family, is coordinated by multiple epigenetic layers involving silencing by perinuclear heterochromatin and active locus repositioning. The singular activation process, however, remains vague. Recently, a GC-rich ncRNA encoded by a gene family of 15 members, has been identified as an activation factor of *var* genes. Here, we show that this ncRNA is transcribed by RNA Pol III using a Pol III specific inhibitor. Furthermore, we used the high resolution Capture-C technique to demonstrate long-range interactions between the GC-rich gene members. Given their location interspersed between *var* gene clusters, the GC-rich genes may contribute to their spatial organization into foci at the nuclear periphery. Our work might establish a link between Pol III and nuclear organization of Pol II transcribed genes and the process of immune

evasion and virulence of malaria parasites.

Keywords: RUF6, *Plasmodium falciparum*, Capture-C, oligonucleotide capture technology, RNA polymerase III, ncRNA

INTRODUCTION

The protozoan parasite *Plasmodium falciparum* undergoes antigenic variation of surface molecules in order to escape the immune system and survive within the host (1, 2). The major player in immune evasion and antigenic variation are the PfEMP1 proteins, encoded in a mutually exclusive manner by the *var* multigene family (3). *Var* genes are located in subtelomeric regions and central chromosomal clusters (4) and their expression is controlled by different layers of epigenetic regulation (1). Silenced *var* genes are tethered together in heterochromatic repressive clusters at the nuclear periphery and the single euchromatic active *var* member relocates to a distinct perinuclear expression site (5–7). Recently, a GC-rich ncRNA family, annotated as RNA of unknown function (RUF6), with location interspersed between central *var* genes has been associated with *var* gene activation (8, 9). Interestingly, these ncRNAs contain the canonical A- and B- boxes for recognition by the RNA polymerase III (Pol III) machinery, while neighboring *var* genes are transcribed by RNA pol II. Annotated putative RNA Pol III subunits in *P. falciparum* share 25-55% identity with the human homologues and 30-53% identity with *Saccharomyces cerevisiae* homologues, while these two model organisms share 30-70% identity between their subunits as compared by BLAST.

RNA Pol III-transcribed loci of class 2 contain A- and B-boxes in their sequences allowing the direct binding of the RNA Pol III initiation factor TFIIIC and have been associated with genome organization (10, 11). TFIIIC binding to class 2 genes such as tDNA and SINE elements, which can occur independently of the RNA Pol III machinery assembly, has been reported to mediate clustering at the nuclear periphery and have a role as heterochromatin boundary in yeast, mice and human (12, 13). Additionally, TFIIIC binding to SINE genes has been also reported to mediate their relocation to transcriptional factories in response to external stimuli and thereby mediating the activation of nearby poised genes (14). Since nuclear organization is critical for mutually exclusive expression, we hypothesized that *RUF6* loci could participate in the spatial organization of *var* genes and thus contribute to their regulation. Here, we used high resolution Capture-C to study potential long-range genome

interactions of *RUF6* loci. This method combines chromosome conformation capture (3C), oligonucleotide capture technology (OCT), and high-throughput sequencing, to analyze the interactome of these genes with the rest of the genome. We show that *RUF6* genes present long-range interchromosomal interactions between them. Our data suggest a role of these Pol III transcribed ncRNA genes in clustering of central *var* genes.

RESULTS

RUF6 belong to class 2 of RNA Pol III transcribed genes

We performed RNA Pol III inhibition assays with synchronized wild type parasites to test whether *RUF6* genes are transcribed by this polymerase as predicted by the A- and B- boxes found in their sequences. 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, not observed in the transcript levels of control housekeeping gene, ubiquitin-conjugating enzyme (PF3D7_0812600) transcribed by RNA Pol II (Fig. 1). Our data strongly suggest that *RUF6* genes are transcribed by RNA Pol III, corroborating DNA sequence predictions.

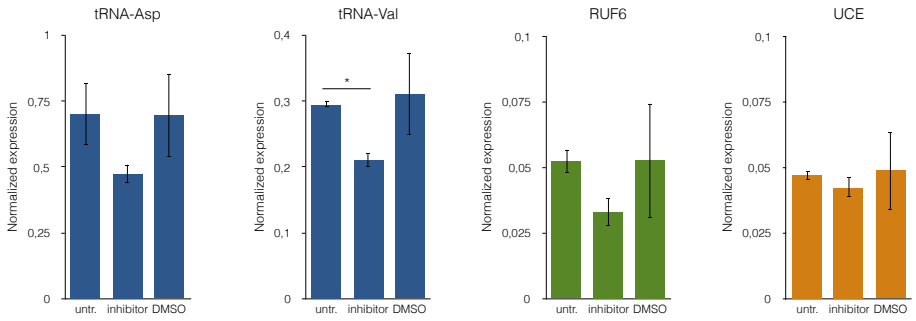


Figure 1. RNA polymerase III inhibition assay. Transcript levels as quantified by RT-qPCR in synchronized wild type parasites at 24 hpi for tRNA-Asp, tRNA-Val, RUF6 and housekeeping control gene coding for ubiquitin-conjugating enzyme, in untreated cultures and in the presence of RNA Pol III inhibitor 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).

Optimization of Capture-C for *P. falciparum* blood stages

In order to study whether *RUF6* genes can be involved in 3D organization of the genome, similarly to other RNA Pol III-transcribed genes in other systems (11–14), we adapted the high resolution Capture-C method established recently (15, 16) to asexual *P. falciparum* blood stage parasites (Fig. 2).

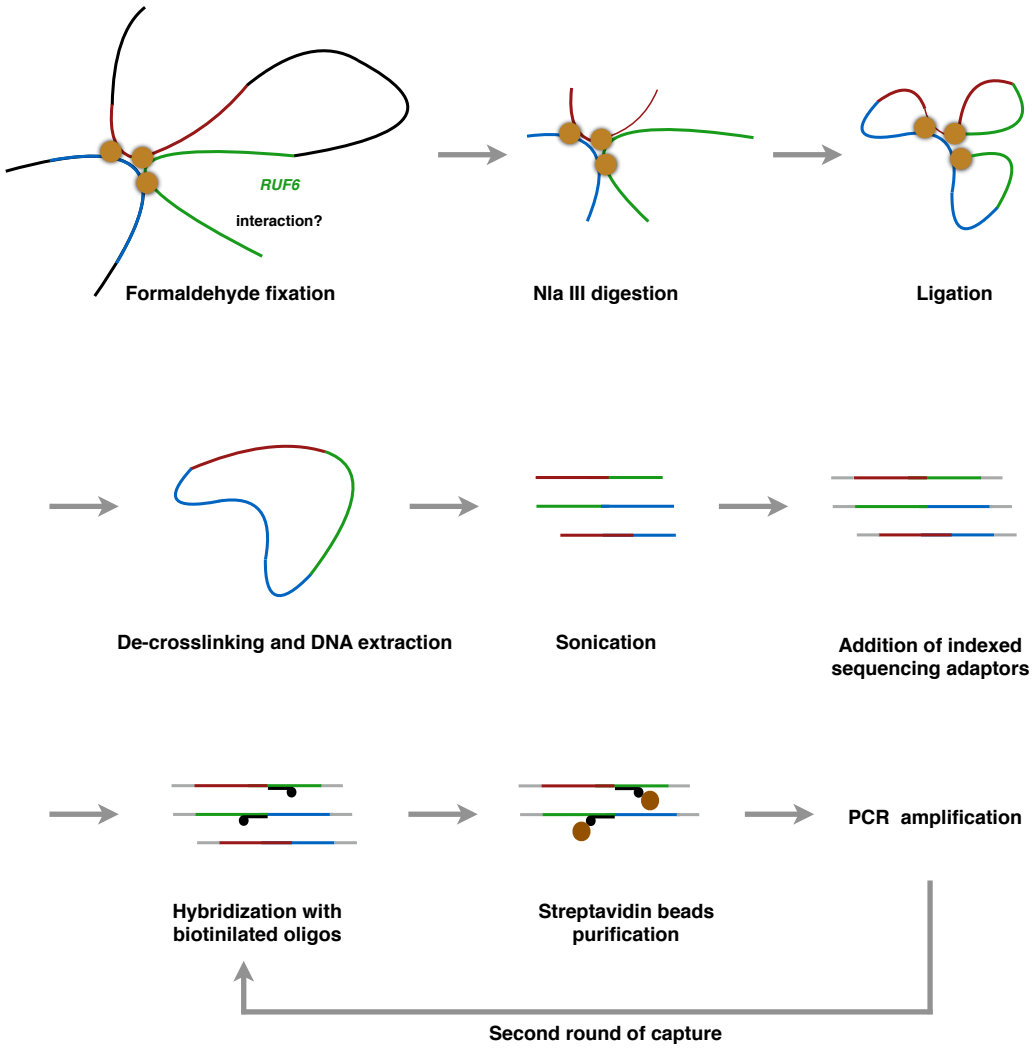


Figure 2. Capture-C method overview. Schematics of Capture-C method for detecting chromosomal interactions of *RUF6* capture points.

We designed specific probes to target three different *RUF6* genes as "viewpoints" (Table S1) to optimize the resolution with NlaIII digested 3C libraries. Since *RUF6* genes se-

quence similarity is extremely high (>90%) we targeted the most diverse regions outside the gene body that would allow to differentiate the *RUF6* genes used in this study. The viewpoints of interest were the following *RUF6* genes: PF3D7_1240800 (chromosome 12, with a restriction fragment of 1142 bp), PF3D7_1241000 (chromosome 12, 972 bp fragment) and PF3D7_0809000 (chromosome 8, 861 bp fragment). We used three different WT clones, NF54 A11, 3D7 D6 and 3D7 B6 (9). The first two clones, A11 and D6, presented predominant transcription of one member of the *RUF6* gene family used as viewpoint, PF3D7_1240800 (9).

We tested different conditions of parasite lysis, crosslinking and sonication to optimize the generation of 3C library preparation and determined that 2% formaldehyde crosslinking prior to parasite lysis gave the highest DNA yields. We assessed the 3C library generation by agarose gels (Fig. 3A) before proceeding with enrichment of capture points by hybridization with biotinylated oligos. Enrichment of viewpoints in Capture-C compared to 3C libraries was confirmed by qPCR (Fig. 3B). All targeted regions presented a very significant enrichment after oligonucleotide capture (50- to 600-fold) and the retrieval of a control gene was negligible.

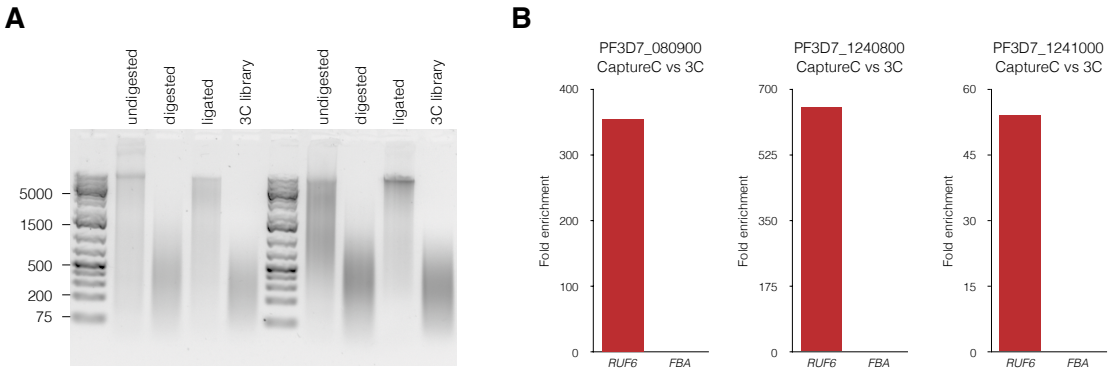


Figure 3. Capture-C for *RUF6*. **(A)** Examples of assessment of 3C library generation on a 1% agarose gel. The different lanes correspond to purified DNA before digestion, after NlaIII digestion, after ligation and after sonication. **(B)** Fold enrichment of viewpoints in Capture-C compared to 3C libraries as assessed by qPCR using the housekeeing gene fructose-bisphosphate aldolase (PF3D7_1444800) as a non targeted control. Example enrichment for the 3D7 B6 clone is shown.

Capture-C of *RUF6* viewpoints reveals long-range chromosomal interactions amongst different members of the gene family

High-throughput sequencing analysis of Capture-C libraries revealed a highly conserved profile of intra and interchromosomal interactions for each viewpoint in the three independent wild type clones used (Table S2), suggesting a high reproducibility of the method. As expected for this technique, most enrichment of viewpoints concentrated in proximal regions (Fig. 4) since sequences from capture points interact with the surrounding regions of the genome in a distance-dependent manner (15, 16). Strikingly, the strongest interchromosomal interactions observed for *RUF6*, corresponded to other members of the same gene family (Fig. 4). This points to physical interactions of *RUF6* genes and would suggest that they could mediate central *var* clustering by zipper-like interchromosomal binding. We did not observe *RUF6* interaction differences between strains in which the viewpoint was transcriptionally active or silent. Interestingly, the strongest intrachromosomal interaction observed outside a *RUF6* viewpoint central cluster, corresponded to a gene coding for an ApiAP2 transcription factor (PF3D7_0802100). In-depth analysis of obtained Capture-C data and validation of the detected inter and intrachromosomal interactions by PCR on 3C libraries are currently ongoing.

DISCUSSION

Transcription of the *RUF6* ncRNA gene family has been associated with the active state of *var* and other virulence gene families, by mechanisms that remain elusive. Here, we provide evidence for a novel transcription-independent function of these elements based on long-range DNA interactions.

Genome-wide 3C (Hi-C), has previously been used in this parasite (17, 18) to explore the genome organization and predict regulatory regions. This method confirmed clustering of central and subtelomeric *var* genes by chromosomal looping established earlier using single cell FISH analysis (7, 5). However, since these studies focused in all-vs-all loci interactions, despite producing a comprehensive global picture of the parasite genome organization, they lack the resolution required to study particular loci of interest. Thus, to analyze *RUF6* interactions we opted for the Capture-C method, which uses OCT to pull loci of interest from 3C libraries combined with multiplexed high-throughput sequencing. The observed interplay between dispersed *RUF6* genes points to a role as spatial organizers of central *var* gene clus-

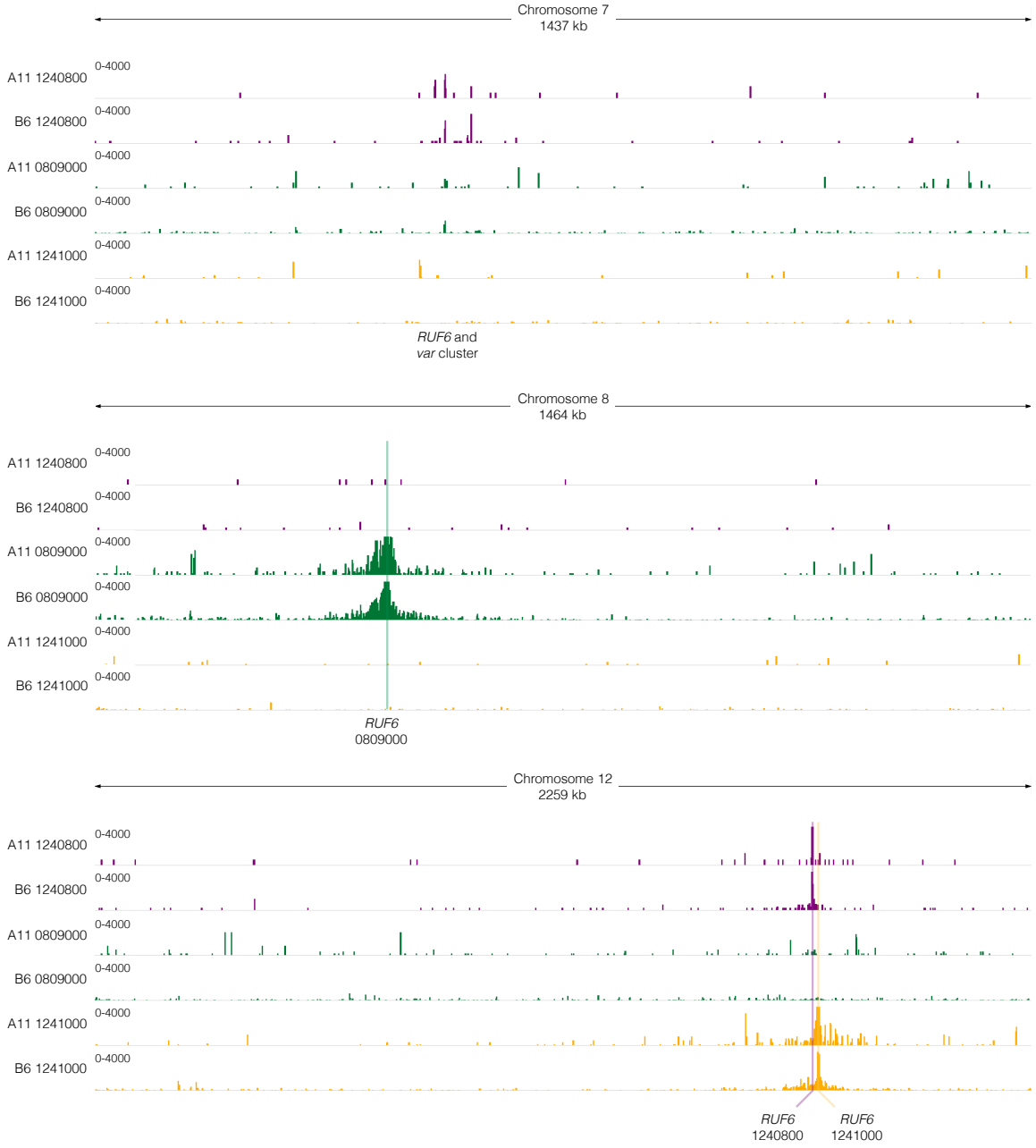


Figure 4. *RUF6* genes display long-range interactions among each other. Capture-C normalized interactions displayed on IGV for the A11 and B6 wild type clones and the three *RUF6* captured points (PF3D7_0809000, PF3D7_1240800 and PF3D7_1241000). Full view of chromosomes 7, 8 and 12 is shown.

ters and could represent a novel layer of *var* gene regulation. We suggest that similarly to telomeric sequences mediating clustering of subtelomeric *var* genes at the nuclear periphery, *RUF6* genes could play a comparable function and target central members to the nuclear periphery. The A- and B-box binding protein TFIIIC may be the best candidate to promote the long-range interactions similarly to what has been observed in other organisms (10–13). FISH DNA analysis may reveal if episomes containing a *RUF6* gene are able to target the episome to the perinuclear *var* foci. Ongoing studies will explore TFIIIC binding to *RUF6* genes.

It is tempting to hypothesize that TFIIIC binding independently of RNA Pol III transcription could mediate the observed long-range interactions between different members and by these means facilitate heterochromatization and perinuclear clustering of central *var* loci, as previously reported in other systems (12, 13). Interestingly, we observed derepression of a *var* gene when knocking out the *RUF6* member located immediately downstream (Guizetti, unpublished data) supporting this hypothesis.

In summary, we show that the Capture-C method applied to *P. falciparum* is a new high resolution and high throughput tool to study specific biological questions linked to parasite virulence. This method applies to any gene or loci of interest such as *var* promoters and represents a largely improved tool compared to the previously used Hi-C methods (17, 18). It may help to uncover insulator, enhancer and other regulatory regions of gene expression in *P. falciparum*.

METHODS

Parasite culture and synchronization

Blood stage *P. falciparum* parasites were cultured as previously described (5). Parasites were synchronized with a 6h time window by sorbitol lysis during ring stage, followed by plasmagel enrichment in schizont stage and another sorbitol treatment 6 h after corresponding to 3±3 hpi. Synchronized parasites were harvested a 3.3% haematocrit and ~2-4% parasitaemia. Parasite development was monitored by Giemsa staining.

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.

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 Qiazol. Total RNA was extracted using the miRNeasy mini kit and performing on-column DNase treatment (Qiagen). Reverse transcription 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 (17). Transcript levels were quantified by normalizing the starting quantity mean to the one of a housekeeping gene (fructose-bisphosphate aldolase, PF3D7_1444800). Starting quantity means of three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA.

Capture-C and sequencing

Capture-C was performed as previously described (15, 16, 19) with the following modifications. Synchronized parasite cultures at 14 and 24 hpi were crosslinked with 2% formaldehyde for 10 min at RT with shaking and subsequently quenched with 125 mM cold glycine. Parasites were lysed by two rounds of 0.15% saponin/PBS treatment. 3C libraries were generated with digestion enzyme NlaIII, sonicated at 4°C for 3 min (30 s on/off intervals, high power) with Bioruptor (Diagenode) and assessed on an agarose gel. Sequencing adaptors were added with the MicroPlex Library Preparation Kit v2 (Diagenode) with the KAPA HiFi polymerase (Kapa Biosystems) for the PCR amplification and resulting libraries were assessed on a DNA High Sensitivity Bioanalyzer chip (Agilent Technologies). Enrichment by oligonucleotide capture of *RUF6* viewpoints was performed twice in order to increase the fold enrichment of the Capture-C library over 3C library, as previously described (16, 19). Final Capture-C libraries were assessed on a DNA High Sensitivity Bioanalyzer chip (Agilent Technologies). 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.

Capture-C data analysis

Fastq files were subjected to quality control using the FastQC software (20). Sequencing reads were mapped to the *P. falciparum* genome (4) (PlasmoDB, v.29) using Bowtie. PCR duplicates were removed. The rest of the analysis was performed as described in the pipeline of CapC-MAP (21). Data were visualized using Integrative Genomics Viewer (22).

ACKNOWLEDGEMENTS

This work was supported by a European Research Council Advanced Grant (PlasmoSilencing 670301) and the French Parasitology consortium ParaFrap (ANR-11-LABX0024) awarded to A.S. and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 741447 to C.C.O.

AUTHOR CONTRIBUTIONS

A.B.S., C.C.O. and A.S. conceived and designed the study. A.B.S and C.C.O optimized and performed Capture-C experiments and analysis. A.B.S. performed RT-qPCR experiments and wrote the manuscript.

REFERENCES

1. 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.
2. Chen, Q, Fernandez, V, Sundström, A, Schlichtherle, M, Datta, S, Hagblom, P, Wahlgren, M. 1998. Developmental selection of var gene expression in Plasmodium falciparum. Nature 394:392–395.
3. Miller, LH, Baruch, DI, Marsh, K, Doumbo, OK. 2002. The pathogenic basis of malaria. Nature 415:673–679.
4. Gardner, MJ, Hall, N, Fung, E, White, O, Berriman, M, Hyman, RW, Carlton, JM, Pain, A, Nelson, KE, Bowman, S, Paulsen, IT, James, K, Eisen, JA, Rutherford, K, Salzberg, SL, Craig, A, Kyes, S, Chan, MS, Nene, V, Shallom, SJ, Suh, B, Peterson, J, Angiuoli, S, Pertea, M, Allen, J, Selengut, J, Haft, D, Mather, MW, Vaidya, AB, Martin, DM, Fairlamb, AH, Fraunholz, MJ,

Roos, DS, Ralph, SA, McFadden, GI, Cummings, LM, Subramanian, GM, Mungall, C, Venter, JC, Carucci, DJ, Hoffman, SL, Newbold, C, Davis, RW, Fraser, CM, Barrell, B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511.

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

6. Jiang, L, Mu, J, Zhang, Q, Ni, T, Srinivasan, P, Rayavara, K, Yang, W, Turner, L, Lavstsen, T, Theander, TG, Peng, W, Wei, G, Jing, Q, Wakabayashi, Y, Bansal, A, Luo, Y, Ribeiro, JM, Scherf, A, Aravind, L, Zhu, J, Zhao, K, Miller, LH. 2013. PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. Nature 499:223–227.

7. Ralph, SA, 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.

8. 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.

9. Barcons-Simon, A, Cordon-Obras, C, Guizetti, J, Bryant, JM, Scherf, A. 2019. CRISPR interference of a clonally variant GC-rich non-coding RNA family leads to general repression of var genes in Plasmodium falci- parum. Manuscript under review

10. Canella, D, Praz, V, Reina, JH, Cousin, P, Hernandez, N. 2010. Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. Genome Res 20:710–721.

11. Noma, K, Cam, HP, Maraia, RJ, Grewal, SI. 2006. A role for TFIIIC transcription factor complex in genome organization. Cell 125:859–872.

12. Kirkland, JG, Raab, JR, Kamakaka, RT. 2013. TFIIIC bound DNA elements in nuclear organization and insulation. Biochim Biophys Acta 1829:418–424.

13. Raab, JR, Chiu, J, Zhu, J, Katzman, S, Kurukuti, S, Wade, PA, Haussler, D, Kamakaka, RT. 2012. Human tRNA genes function as chromatin insulators. EMBO J 31:330–350.

14. Crepaldi, L, Policarpi, C, Coatti, A, Sherlock, WT, Jongbloets, BC, Down, TA, Riccio, A. 2013. Binding of TFIIIC to sine elements controls the relocation of activity-dependent neuronal genes to transcription factories. PLoS Genet 9:e1003699.

15. Hughes, JR, Roberts, N, McGowan, S, Hay, D, Giannoulatou, E, Lynch, M, De Gobbi, M, Taylor, S, Gibbons, R, Higgs, DR. 2014. Analysis of hundreds of cis-regulatory landscapes at

high resolution in a single, high-throughput experiment. Nat Genet 46:205–212.
16. Davies, JO, Telenius, JM, McGowan, SJ, Roberts, NA, Taylor, S, Higgs, DR, Hughes, JR. 2016. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat Methods 13:74–80.
17. Lemieux, JE, Kyes, SA, Otto, TD, Feller, AI, Eastman, RT, Pinches, RA, Berriman, M, Su, XZ, Newbold, CI. 2013. Genome-wide profiling of chromosome interactions in Plasmodium falciparum characterizes nuclear architecture and reconfigurations associated with antigenic variation. Mol Microbiol 90:519–537.
18. Ay, F, Bunnik, EM, Varoquaux, N, Bol, SM, Prudhomme, J, Vert, JP, Noble, WS, Le Roch, KG. 2014. Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. Genome Res 24:974–988.
19. Oudelaar, AM, Downes, DJ, Davies, JOJ, Hughes, JR. 2017. Low-input Capture-C: A Chromosome Conformation Capture Assay to Analyze Chromatin Architecture in Small Numbers of Cells. Bio Protoc 7
20. Wingett, SW, Andrews, S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res 7:1338.
21. Buckle, A, Gilbert, N, Marenduzzo, D, Brackley, CA. 2019. capC-MAP: software for analysis of Capture-C data. Bioinformatics
22. Robinson, JT, Thorvaldsdóttir, H, Winckler, W, Guttman, M, Lander, ES, Getz, G, Mesirov, JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26.

SUPPLEMENTARY DATA

High resolution Capture-C analysis reveals long-range interactions between dispersed Pol III transcribed GC-rich ncRNA genes in *P. falciparum*

Anna Barcons-Simon[1,2,3,4,#], Carlos Cordon-Obras[1,2,3,#] and Artur Scherf[1,2,3,*]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[#]These authors contributed equally

[*]Correspondence: artur.scherf@pasteur.fr

Preliminary data

Table S1. Capture-C probe sequences. 5' biotinylated oligonucleotides used for Capture-C.

Probe name	Sequence
0809000_CapC_1	CC ATT GTA TGT ACA TAT ATA AAA TAT ATC ATA ATT GAA ATA TAA ATA TTT TGA AAA TAA AAG TAC AAA TTT TAT ATT TAA ATT TAA TGA AAA AAA ATA TAT GTT ATA TAT TTA TAT ATA TAT ATA TAT ATA TTC ATG
0809000_CapC_2	CA TGA ATT ATA GCA CAA ATG CAA TAT ATA TAT AAG TGT ATA TAA TTA TTA TTA GTT ACG AAT CAG CAG AAC AAA ATT TTG AAA ATA ATA ATT TTG TCA TAT GGA TAA ATA TAT AAT ATA ATA ATG TGT G
1240800_CapC_1	GC ACA TAA TAT ATA TAT ATA TAT ATA TAT TTA CCA CTG TAT GTA CAT ATA TAA AAT ATA TCA TAA TTG AAA TAT AAA TAT TTT GAA TAT AAA AGT ACG AAT TTT ATA TTT AAA TTT AAT GAA AAG AAA AAT ATA TGT TAT ATA TTT ATA TAT ATA TAT ATA TAT ATT CAT G
1240800_CapC_2	CA TGT TAT ATG AAT ACT TTC CTT TTT TAT ATT TTA TTT TAA TTA TCA TTA TTA ATA TTC TTC CTC TCA TCA TAA TAT AAT ATA ATA TAA TTA TTT GAT AAA ATC AAT TCA TTT TGC TAA ATA CTG TCA AAC G
1241000_CapC_1	CA TAA AAG GGT AAA ATA TAT TTT TTT CTT TTT TTT TTT TGT ATG CAC GTT TTT TCC ATT TTG AAT ATA TAA CTA TTT ATA TTC CAA AAA AGA ATA ACA TAA TAT ATT TAT TTA TAG AAT ATA ATT TCA TG
1241000_CapC_2	CA TGA ATA TGC ATA TCA CAT TAT GTT GGA ACA TAA ATT ATA TAA ATA TAT AAT TAT TTA ATA AAA GAA TTA AAA CAA AAA AAA AAA AAT ATA ATA GAC ATA CAT AAT AAA TGA ATA AAC CAA TTA TC

Table S2. Summary of valid interactions for the different *RUF6* viewpoints used. Percentages of intra and interchromosomal interactions are shown for each wild type clone and viewpoint used.

Clone	ID of *RUF6* viewpoint	Intrachrom. interactions (%)	Interchrom. interactions (%)	Total valid interactions
NF 54 A11	0809000	31%	69%	2470
	1240800	83%	17%	775
	1241000	64%	36%	2549
3D7 D6	0809000	32%	68%	205
	1240800	94%	6%	70
	1241000	58%	42%	3650
3D7 B6	0809000	32%	68%	7147
	1240800	79%	21%	1629
	1241000	53%	47%	74010

Part 3

General discussion and outlook

Contents

This third part of the thesis summarizes and contextualizes the main results presented. It also aims to discuss the concepts that remain yet unclear and to serve as a basis for future research perspectives.

3.1 Characterization of the GC-rich ncRNA

Default heterochromatin silencing of *var* genes has been broadly studied and many factors involved with it have been identified. Notwithstanding, activation and switching of this gene family and their mutually exclusive counting are subject of many open questions and remain ambigous. Our study aimed to shed light to this topic with the hypothesis that a GC-rich ncRNA family, annotated as RNA of unknown function 6 (RUF6), could be involved in *var* gene activation and contribute to *var* switch rates. Importantly, as described in section 1.5.2.1, *RUF6* gene loci are found interspersed between *var* genes of central chromosomal clusters and their sequences are highly homologous [Gardner et al., 2002].

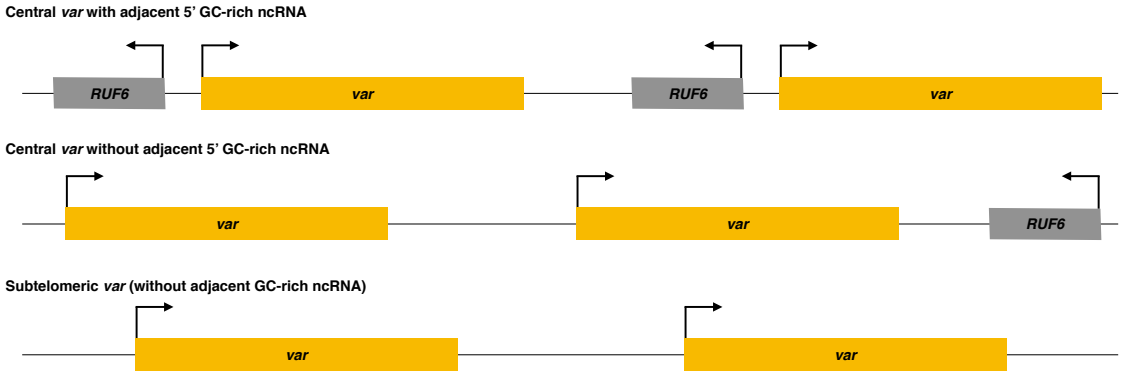


Figure 3.1: Scheme of GC-rich and *var* genes relative chromosomal location. There are three different situations for the chromosomal location of *var* genes relative to that of GC-rich *RUF6* genes. Only central *var* genes harbor GC-rich genes within their clusters and when they are adjacent to the 5' upstream sequence of an active *var*, that ncRNA presents the predominant transcriptional levels.

We analyzed the steady state transcript levels of the 15 GC-rich ncRNA members. Our data indicate a clonally variant expression mode with two different profiles according to their chromosomal location relative to *var* genes (Fig. 3.1 and section 2.2). When a single GC-rich gene is predominantly transcribed, it is always found adjacent to the 5' region of an active central *var* gene. Conversely, clonal populations with an active subtelomeric or central *var* devoid of a GC-rich gene in this configuration, show ncRNA transcription from multiple loci at lower levels. We infer that high transcriptional levels of a given GC-rich ncRNA could stabilize the expression site for its adjacent active *var* gene and thus reduce the switch rate when compared to subtelomeric members. Indeed, it has previously been

reported that switch rates are linked to *var* chromosomal location and that central members are less prone to switch than subtelomeric ones [Frank et al., 2007], concurring with this hypothesis. Active transcription of the adjacent ncRNA could help open local chromatin and enhance accessibility of the transcriptional machinery to the *var* gene. This rationale is supported by a recent study reporting increased accessibility at GC-rich genes adjacent to actively transcribed *var* and *rif* genes [Ruiz et al., 2018]. Our work may have identified the first molecular component associated to *var* gene switching and may be explored further to study this mechanism that is key to antigenic variation in *P. falciparum*. For example, we speculate that insertion of a GC-rich gene adjacent to a subtelomeric *var* might significantly reduce its switch rate.

3.2 GC-rich ncRNA is associated with *var* gene activation and expression site

Several studies have described perinuclear repositioning of the active *var* gene to a transcriptionally permissive site [Ralph et al., 2005b, Freitas-Junior et al., 2005, Voss et al., 2006], but neither the mechanism for this localization nor factors associated with it had previously been described. We have identified the first factor targeted in *trans* to the active *var* expression site of both central and subtelomeric *var* genes (section 2.1) [Guizetti et al., 2016], a nuclear structure that remains elusive (Fig. 3.2). It has been previously suggested that a limiting factor might allow the activation of a single *var* gene locus [Howitt et al., 2009]. In fact, GC-rich ncRNA could potentially be such factor as episomal overexpression disrupts *var* mutually exclusive expression at single cell level (section 2.1) [Guizetti et al., 2016]. Furthermore, we 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 (section 2.2). Additionally, another study has also suggested that transcripts from this ncRNA are associated with *var* activation [Wei et al., 2015].

Moreover, it has been postulated that *P. falciparum* clonally variant gene families could share a common activation factor because promoter titration of *var*, *rif*, *stevor* and *Pfmc-2TM* genes results in down-regulation also of other families [Howitt et al., 2009]. It is noteworthy, that the GC-rich ncRNA family is conserved throughout all species from the *Laverania* subgenus of *Plasmodium*, along with these clonally variant gene families coding for VSA

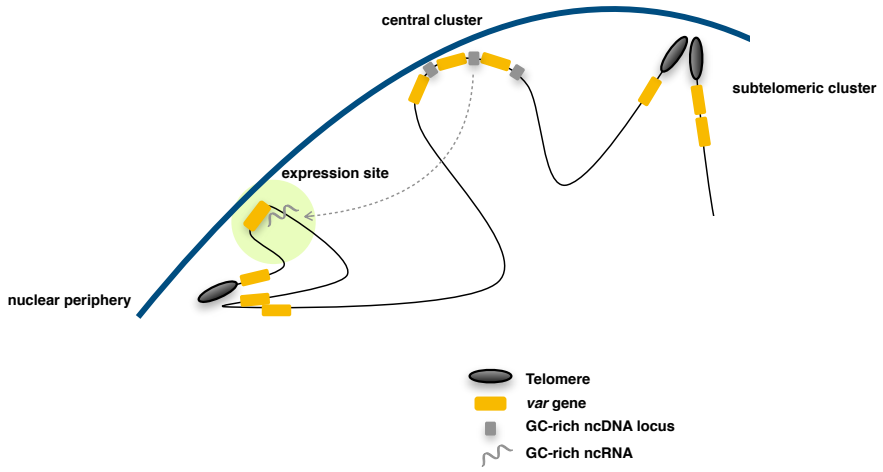


Figure 3.2: Scheme of GC-rich ncRNA and *var* expression site colocalization in *trans*. *Var* genes cluster at the nuclear periphery and the active member relocates to a transcriptionally permissive site colocalizing with GC-rich ncRNA.

associated with virulence [Otto et al., 2018b]. Our work suggests that the GC-rich ncRNA could be the postulated common activator factor since CRISPRi of the GC-rich ncRNA family leads to transcriptional down-regulation also of these other clonally variant gene families coding for surface antigens (section 2.2). Additionally, episomal overexpression of GC-rich ncRNA also perturbs the transcriptional profile of *var* and *stevor* gene families in a similar manner (Appendix A).

The work presented in this thesis strongly associates these ncRNAs with *var* activation (sections 2.2 and 2.1) but we are just starting to have a glimpse into their mechanism of action (section 2.3) which is the current focus of our work. In order to gain in-depth insight, we developed a ChIRP protocol to identify chromatin interacting partners of RUF6. First, we identified proteins potentially associated with RUF6 in the chromatin context. Several of the identified candidates are involved in gene transcription, indicating that this ncRNA might assemble these factors at the *var* expression site in order to promote transcription. Further validation of these candidates is currently ongoing.

We also report interaction of actively transcribed GC-rich ncRNA with exon 1 of adjacent active *var* gene and some other members, including the one adjacent to the second pre-

dominantly transcribed ncRNA, although these data need further validation as come from a single clone and were detected with only one set of ChIRP probes. Additionally, we observe interaction of this ncRNA with many other genes, including some other subtelomeric *var* members. However, in the parasite population, albeit clonal, we can not distinguish what is happening at single cell level. In fact, we aim to understand whether at single cell level predominant transcription of GC-rich genes is in fact mutually exclusive (see Appendix C). It is possible that transcription of these ncRNAs marks *var* genes with the potential of being transcribed but only the one with highest transcription is able to stabilize the active *var* locus. Interestingly, both *var* and GC-rich genes are controlled at the level of nascent transcription [Zhang et al., 2014], suggesting that albeit several genes can display transcription initiation, a feedback loop could end up stabilizing a single locus.

Further studies should shed light in the molecular mechanisms involving these ncRNAs and their association with the *var* expression site and activation. Validation of interacting proteins could indicate whether they target chromatin modifying enzymes or TFs to the active *var* locus. Alternatively, RUF6 could play a role as architectural RNAs (arcRNAs) and thus serve as scaffolds for nuclear bodies [Chujo et al., 2016] associated to *var* gene control or for the *var* expression site itself.

3.3 *RUF6* genes as spatial organizers

Spatial organization of the genome plays a critical role in gene regulation, and in particular of mutually exclusive gene expression by separating silent from active loci [Yang and Kuroda, 2007]. Indeed, nuclear organization of *var* genes seems to be a key layer of regulation for their expression [Ralph et al., 2005b, Lopez-Rubio et al., 2007, Chookajorn et al., 2007, Lopez-Rubio et al., 2009].

RUF6 genes contain the canonical A- and B-boxes for RNA Pol III recognition in their sequence. Furthermore, we report a decrease of *RUF6* transcription when using an RNA Pol III inhibitor, similar to that of tRNA genes, suggesting transcription is indeed mediated by this polymerase (section 2.4). Thus, we assume that *RUF6* loci can harbor the RNA Pol III initiation factor TFIIIC, which can also associate with chromatin to promote changes in the nuclear organization independently of RNA Pol III assembly [Kirkland et al., 2013].

As reviewed in section 1.5.1.1, TFIIIC binding to genes transcribed by RNA Pol III is associated with genome organization, clustering at the nuclear periphery and heterochro-

matin boundary function. In particular, it is the case for RNA Pol III transcribed genes of class 2, which are those containing A- and B-boxes in their sequences allowing the direct binding of TFIIIC [Canella et al., 2010]. Insulation mediated by TFIIIC bound DNA elements has been described in many eukaryotes so far, including yeast, mice and human and has been reported to occur both by barrier activity and by enhancer-blocking activity in transcription independent and dependent manners [Gdula et al., 1996, Noma et al., 2006, Lunyak et al., 2007, Raab et al., 2012]. In fact, TFIIIC binding to SINE genes has been also reported to mediate their relocation to transcriptional factories in response to external stimuli and thereby mediating the activation of nearby poised genes [Crepaldi et al., 2013]. Despite evidence for a relationship between GC-rich ncRNA transcription and the active state of *var* genes, the question arises whether the *RUF6* DNA sequence recruits TFIIIC to mediate interactions with other loci, repositioning or insulation, similarly to tRNA or SINE genes in other systems.

Interestingly, our 3D analysis revealed that *RUF6* genes present long-range interactions between them (section 2.4), supporting this hypothesis. We are currently validating these interactions and aim to analyze whether they interact with other regions of the genome. It is possible that these elements facilitate the default silencing of proximal *var* genes, and that active transcription of a member into ncRNA, facilitates the escape from this silencing configuration allowing the adjacent *var* to be ready for activation and enhancing its transcription. Indeed, a *cis*-regulatory function for these genes has previously been suggested based on GC-rich gene associated silencing of an adjacent *gfp* gene on episomal plasmids, leading to speculations that they might act as insulator factors preventing spread of heterochromatin [Wei et al., 2015]. Additionally, we observed the derepression of a *var* gene when deleting specifically the *RUF6* located immediately downstream (Guizetti, unpublished data) and are currently analyzing whether the deletion of another member has a similar effect. Further investigation will be required to confirm the function of *RUF6* in spatial organization and insulation, and to characterize its involvement in the regulation of neighboring *var* genes.

3.4 Hypothetical models

Taken all our data together, we can hypothesize several models that would require further validation for the mechanism of action of the GC-rich genes.

The observation of long-range interactions between several *RUF6* genes and their transcription by RNA Pol III, strongly suggests they participate in nuclear organization by the binding of TFIIIC. We can hypothesize *RUF6* genes mediate perinuclear heterochromatinized clustering of central *var* genes through TFIIIC anchoring, and ongoing transcription of the ncRNA allows activation of *var* genes (Fig. 3.3). One possibility could be that high transcriptional levels of the ncRNA could facilitate the stabilization of the expression site by RNA phase separation mechanisms. Further studies should formally confirm TFIIIC binding to *RUF6* genes by ChIP. Additionally, in order to test whether these elements mediate relocation or anchoring to the nuclear periphery, it would be useful to perform DNA FISH with an episome containing a *RUF6* gene and analyze its nuclear positioning.

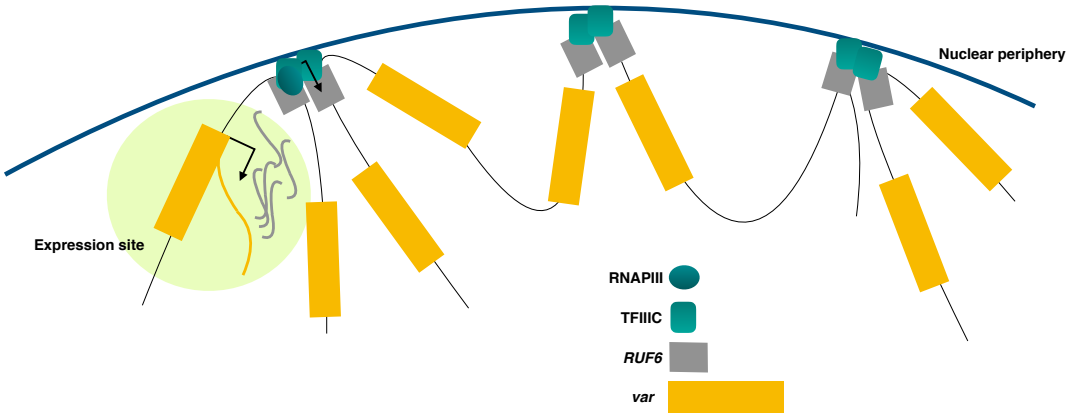


Figure 3.3: Hypothetical model of 3D chromosomal organization by GC-rich DNA elements. TFIIIC binding to *RUF6* genes B-box could mediate the observed clustering and anchoring to the nuclear periphery. The spatiotemporal events leading to transcriptional activation and relocation of a single *var* member remain unknown. The *trans*-acting ncRNA might associate with the *var* once in the expression site or initiate the selection process of a singular *var* member followed by the relocation.

It remains to be elucidated which mechanism triggers the activation of GC-rich genes and whether their activation occurs through signaling in response to external stimuli, which is a tempting hypothesis since RNA Pol III machinery has been reported to mediate transcription of genes in response to nutrient sensing and external stimuli [Dieci et al., 2007,Policarpi et al., 2017].

One of the protein candidates predicted to interact with GC-rich ncRNA is a subunit of RNA Pol II, polymerase that transcribes *var* but not *RUF6* genes, suggesting the ncRNA could

be a potential molecular link between RNA Pol III and Pol II transcriptional machineries and thus activate *var* transcription (Fig. 3.4). Further validation of RNA Pol II binding to the GC-rich ncRNA is the scope of current study.

Figure 3.4: Hypothetical enhancer model of GC-rich ncRNA interaction with RNA Pol II. Transcripts from GC-rich ncRNA could enhance adjacent *var* transcription by binding to RNA Pol II at *var* promoter and mediating the relocation to the transcription site. This could lead to more stable transcription of central *var* and less switching to other members when compared to subtelomeric members

3.5 Conclusion and further directions

Our findings broaden the number of players in antigenic variation by identifying a new factor involved in the activation process and potentially participating in the switch rate modulation of clonally variant virulence genes. Furthermore, this study highlights the importance of ncRNA as regulators of gene expression in *P. falciparum*, and provides new tools for their study which have not been previously used in malaria parasites such down-regulation of an entire gene family and identification of ncRNA interactions with chromatin.

We strongly associate transcription of GC-rich ncRNA with the active state of *var* genes but can only start to get a glimpse on the molecular mechanism of action of these ncRNAs. Our current data still miss the molecular details on how ncRNA transcripts contribute to *var* gene activation. Several candidate proteins have been identified to potentially interact with these ncRNAs and might be new leads to achieve mechanistic insight into mutually exclusive expression. Furthermore, we highlight the potential of the GC-rich loci as regulatory DNA

elements, which in coordination with ncRNA transcription could orchestrate a mechanism involving several layers of regulation.

We are currently analyzing in-depth the obtained ChIRP and Capture-C data while also validating protein candidates interacting with this ncRNA, which can shed light on the molecular mechanism involved in activation of *var* and other clonally variant gene families coding for VSA.

Appendix A

Additional results of GC-rich ncRNA overexpression strains

RNA-seq analysis of overexpression strains showed that the GC-rich gene in the episome was indeed overexpressed in the parasites (Fig. A.1).

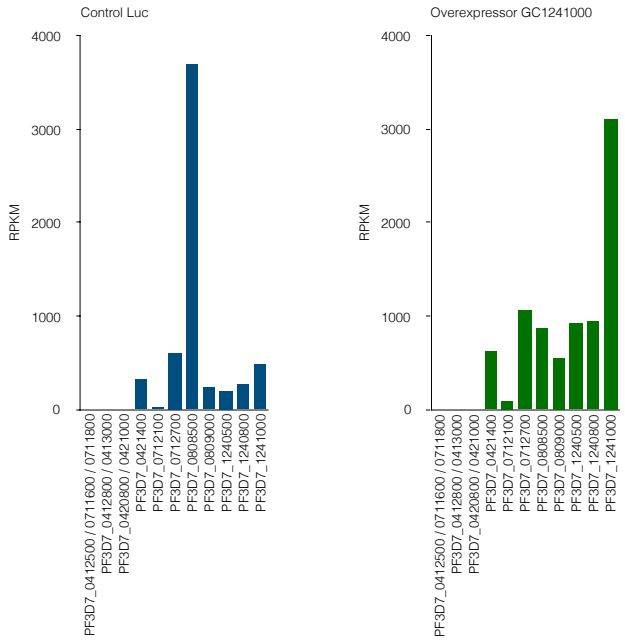


Figure A.1: GC-rich ncRNA transcription in overexpression strain. Transcriptional profile of GC-rich ncRNA at 24 hpi in GC1241000 overexpressor and Luc control clones assessed by RNA-seq. Example replicate of clones GC1241000 F7 and Luc H2 is shown.

RNA-seq analysis of overexpression strains confirmed the *var* profile phenotype observed by RT-qPCR (Fig. A.2).

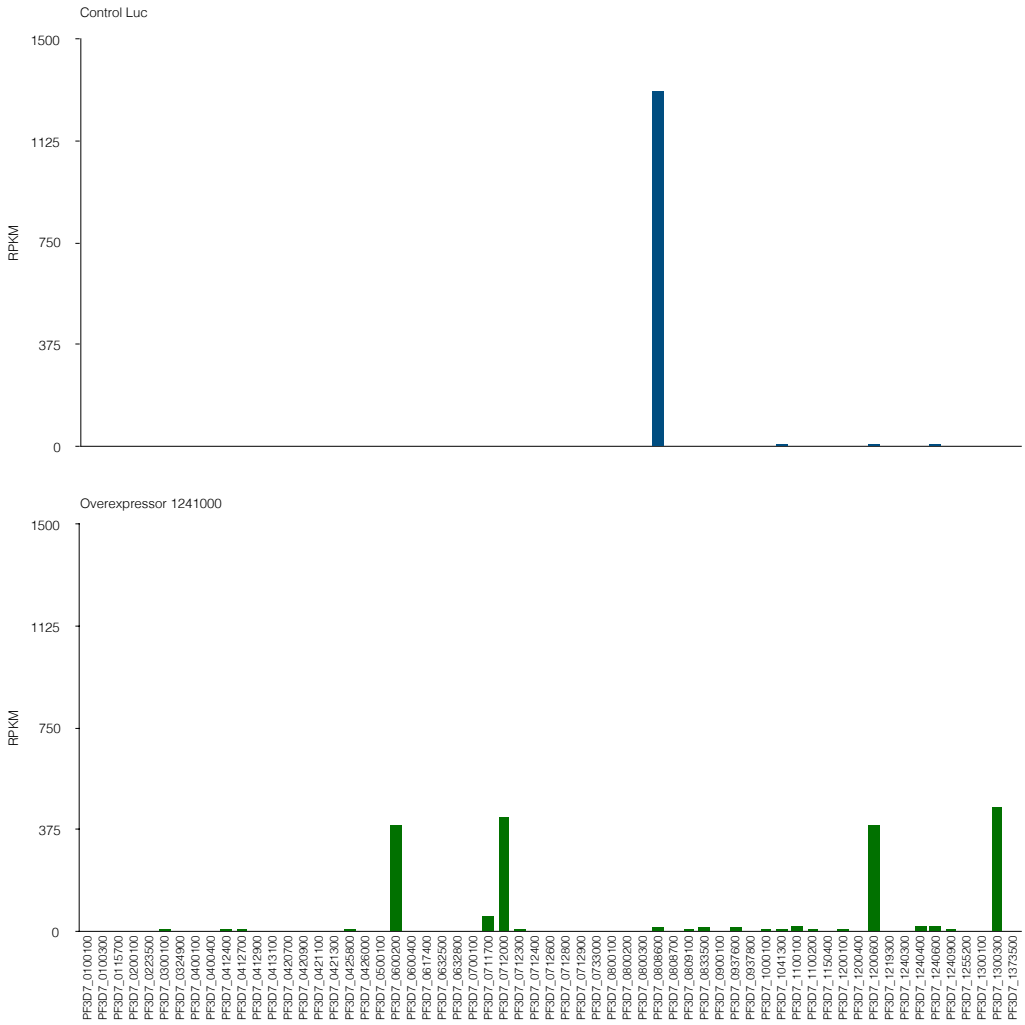


Figure A.2: *Var* transcription in overexpression strain. Transcriptional profile of *var* genes at 12 hpi in GC1241000 overexpressor and Luc control clones assessed by RNA-seq. Example replicate of clones GC1241000 F7 and Luc A6 is shown.

Additionally, the genome-wide transcriptome revealed that GC-rich ncRNA overexpression clones present a subset of active *stevor* genes compared to the clones with the control plasmid that show major transcription of a single member, similarly to the observed phenotype for *var* genes (Fig. A.3).

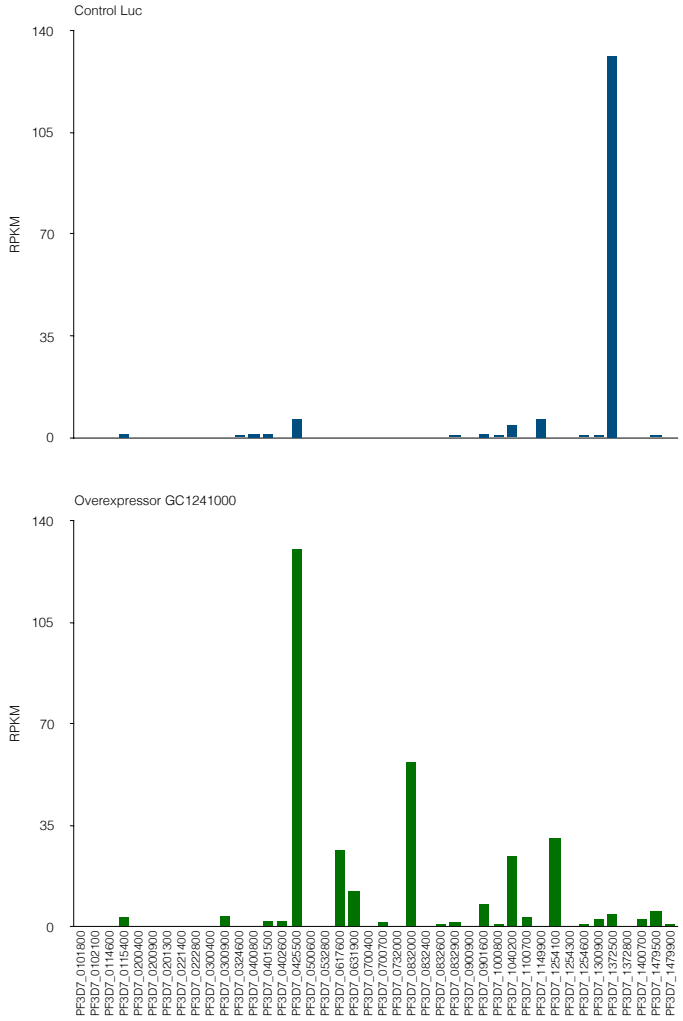


Figure A.3: *Stevor* transcription in overexpression strain. Transcriptional profile of *stevor* genes at 24 hpi in GC1241000 overexpressor and Luc control clones assessed by RNA-seq. Example replicate of clones GC1241000 F7 and Luc H2 is shown.

Appendix B

Additional results of CRISPRi

CRISPRi strains show lower levels of PfEMP1 on the surface of iRBC (Fig. B.1). Western blot of membrane fraction for PfEMP1 was performed as previously described [Bryant et al., 2017]. PfEMP1 proteins were detected with guinea pig anti-ATS [Nacer et al., 2015] primary antibody and goat anti-guinea pig horseradish peroxidase (HRP) (Abacam; ab6908). Total protein was detected by Ponceau staining.

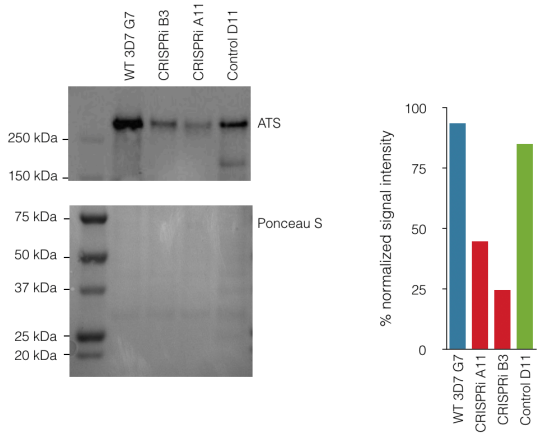


Figure B.1: *Var* down-regulation in CRISPRi strains is conserved at the protein level. Western blot of PfEMP1 in WT, CRISPRi clones and CRISPRi control with anti-ATS antibody.

The dCas9 of the CRISPRi strains is tagged with HA and GFP. Additionally, it also contains an inducible riboswitch which degrades the transcript in presence of glucosamine. We confirmed the nuclear localization of the episome (Fig. B.2) and assessed the levels of dCas9 protein in presence of 1.25mM glucosamine (Fig. B.3). Despite observing a decrease in the dCas9 protein levels the band was still visible indicating presence of dCas9 molecules. An increase in the GC-rich ncRNA expression was observed in presence of glucosamine but the repression of *var* transcription was sustained, suggesting the remaining dCas9 could still mediate the down-regulation phenotype (data not shown).

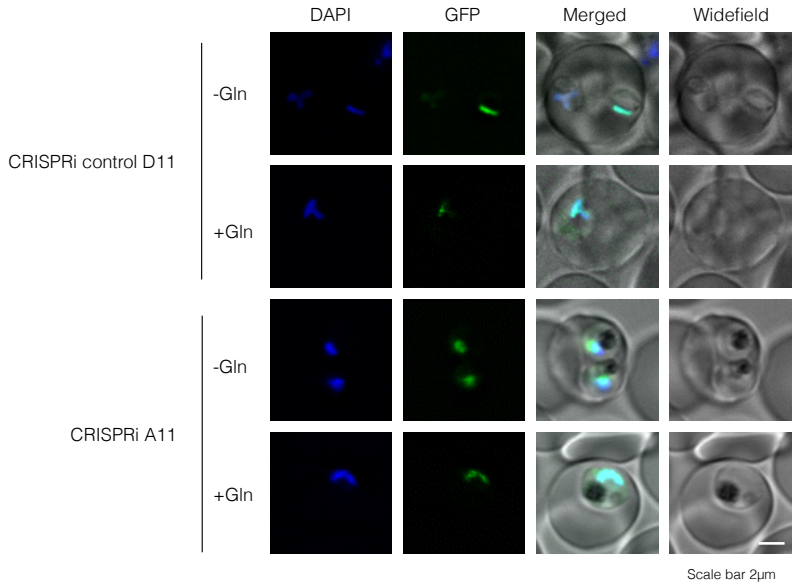


Figure B.2: dCas9 episome is targeted to the parasite nucleus. Localization of pUF-dCas9-3HA-GFP in mutant strains. One CRISPRi clone (A11) and a control clones (D11) are shown.

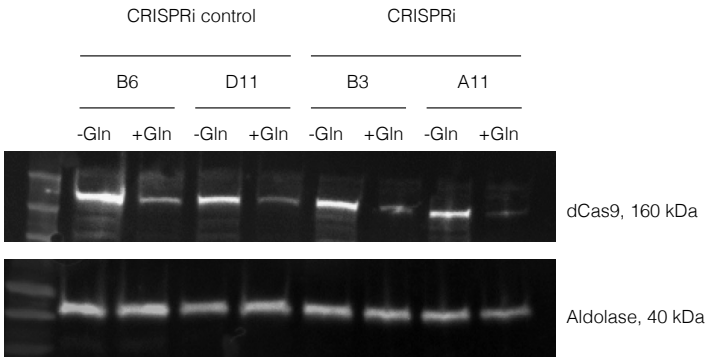


Figure B.3: Inducible riboswitch lowers the levels of dCas9. Western blot showing dCas9 expression in presence of 1.25mM of glucosamine CRISPRi control clones using an anti-HA antibody. Aldolase is used as a loading control.

Appendix C

Single cell RNA sequencing

In order to investigate the transcriptome of GC-rich ncRNA overexpression strains at single cell level, single cell RNA-seq of wild type and GC1241000 overexpressor parasites was performed with the Chromium technology (10X Genomics). Synchronized 18 hpi rings and 30 hpi trophozoites were enriched by plasmion. Single cell data was analyzed using the scShinyHub. The estimated number of cells and run statistics are summarized in Table C.1 and Fig. C.1.

Table C.1: Single cell RNA-seq stats.

Stats	**Rings (3D7)**	**Rings (GC12)**	**Trophozoites (3D7)**	**Trophozoites (GC12)**
Number of cells	692	2954	7406	3032
Number of reads	73029679	86078475	215783712	365509863
Reads mapped to genome	65.2%	66.4%	87.5%	82.7%
Mean reads per cell	105534	29139	29136	120550
Median genes per cell	123	107	386	463

Overexpression of different *var* genes was confirmed at single cell level in the overexpression but not wild type strain (Fig. C.2). However, only transcripts from a GC-rich ncRNA member could be detected likely due to the polyA enrichment step undergone using this technology. We aim to use the obtained data to study clonal variation of different gene families at the single cell level.

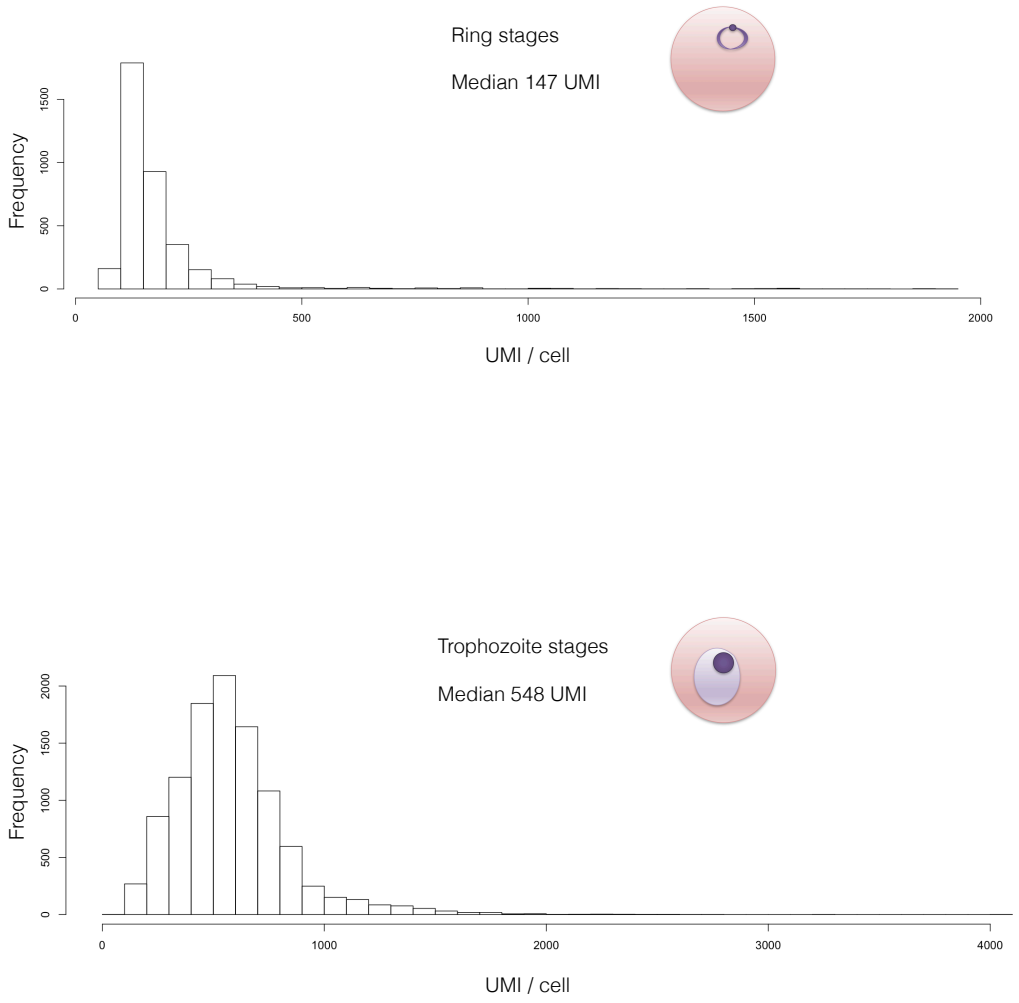


Figure C.1: UMI histogram of single cell RNA-seq samples. Histogram for the frequency of detected unique molecular identifiers (UMI) per cell in single cell RNA-seq samples from ring and trophozoite stages.

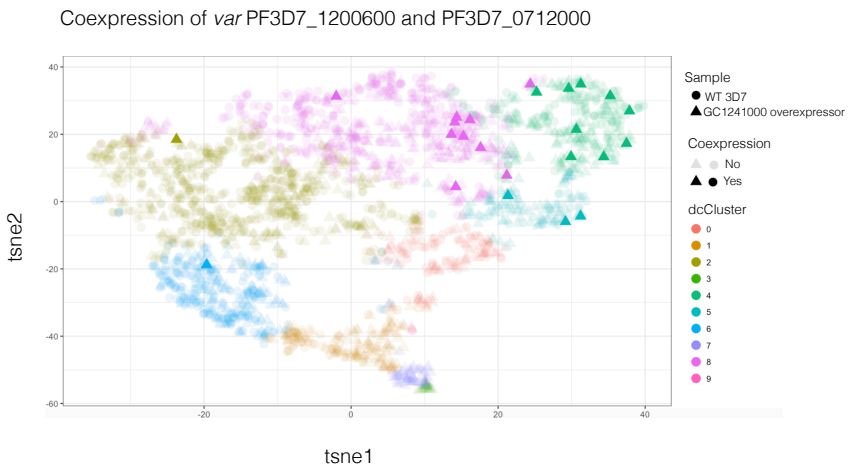


Figure C.2: *Var* coexpression in GC1241000 overexpression strain. Coexpression of two active *var* genes assessed by single cell RNA-seq. 2 dimension t-sne plot is shown.

Appendix D

Identification of the first boundary element in *Plasmodium falciparum*

The highlights of this appendix are:

- Identification of the first heterochromatin boundary element in *Plasmodium falciparum*.

- Master regulator of sexual commitment *ap2-g* harbors a boundary element in the 3' end of the ORF.

The results presented in this appendix are preliminary data summarized in a draft under the title "Uncovering the first heterochromatin boundary element in the locus encoding the master regulator for sexual commitment in *Plasmodium falciparum*". My contribution to this work was performing RT-qPCR experiments and preparing ChIP-seq libraries.

Uncovering the first heterochromatin boundary element in the locus encoding the master regulator for sexual commitment in *Plasmodium falciparum*

Carlos Cordon-Obras[1,2,3], Anna Barcons-Simon[1,2,3,4], Christine Scheidig-Benatar[1,2,3], Aurelie Claës[1,2,3] and Artur Scherf[1,2,3,*]

[1]Unité de Biologie des Interactions Hôte-Parasite, Institut Pasteur, Paris, 75724, France
[2]INSERM U1201, F-75724 Paris, France
[3]CNRS ERL9195, F-75724 Paris, France
[4]Sorbonne Université, Ecole doctorale Complexité du Vivant ED515, F-75005 Paris, France
[*]Correspondence: artur.scherf@pasteur.fr

Preliminary data

ABSTRACT

Heterochromatin is essential in all eukaryotic systems to maintain genome integrity, long-term gene repression and to help chromosome segregation during mitosis. However, heterochromatin regions must be isolated to avoid its spreading over actively transcribed loci. Such function is accomplished by chromatin boundaries, DNA elements that block heterochromatin self-propagation by interaction with other partners. In *Plasmodium falciparum*, facultative heterochromatin is important to regulate parasite virulence, antigenic variation and transmission. The underlying molecular mechanisms that confine these repressive regions remain unknown. To address this topic, we studied a singular heterochromatin island of 12 kb found within an euchromatic region controlling the variagated expression of a transcription factor (AP2G) that is a master regulator of sexual commitment in this pathogen. In this gene, Heterochromatin Protein 1 (PfHP1) coverage is restricted to the exon and 5' upstream region of *ap2-g*. Upon replacement of the 3' region of this gene by *gfp*, we found a shift in the heterochromatin boundary resulting in PfHP1 spreading into the intergenic region downstream. EMSA analysis detected a specific protein complex binding to the cen-

tral part of a 746 bp fragment deleted in the AP2G-GFP line. We are currently identifying the precise binding sites and interacting protein complexes. In addition, we are exploring the boundary effect of the region by re-inserting it in the *ap2-g* locus. Our studies will provide novel insight into the topic of heterochromatin boundaries in *P. falciparum* and likely uncover promising therapeutic pathways that would target heterochromatin-linked prasite virulence and transmission.

INTRODUCTION

The apicomplexan parasite *Plasmodium falciparum* is the causative agent of the deadliest and more prevalent form of malaria in humans, infecting more than 200 million people and causing almost half a million of deaths per year, especially in Africa (1). During its complex life cycle, the parasite undergoes major morphological and transcriptional changes in the arthropod vector, mosquitoes from the genus *Anopheles*, and in the human host (2–4). To cope with developmental changes of transcription, *P. falciparum* has only a limited set of transcription factors (TFs), mainly restricted to the 27-member *Apicomplexa*-specific ApiAP2 family (5). However, multiple epigenetic control layers at the transcriptional and post-transcriptional level complement the regulation of gene expression during its life cycle (6–8).

In *P. falciparum*, some typical eukaryotic features associated to transcription at the chromatin level are well conserved, as the depletion of nucleosomes within promoters and a higher nucleosome occupancy in genic regions (9, 10). Also, the histone variant H2A.Z is widely spread over intergenic regions, putatively rendering them more relaxed and accessible to the transcription machinery, as it weakens nucleosome stability (11). Additionally, most of the post-translational modifications (PTMs) of the N-terminal histone tails associated with transcription regulation in other organisms have been also reported in *P. falciparum* (6). The repressive marks H3K36me3 and H4K20me3, and classical activation marks H3K9ac and H3K4me3 showed a broad distribution across the parasite genome, whereas the repressive modification H3K9me3 appears to be restricted to subtelomeric regions and to some specific internal clusters of chromosomes 4, 6, 7, 8 and 12 (12, 13). This particular histone PTM shows a strong bias towards genes involved in virulence, antigenic variation, stage-specific development and host-parasite interaction, most of which belong to clonally variant gene (CVG) families (12). Among them, the best studied are the *var* genes, which

encode the PfEMP1 surface antigens, one of the major virulence factors involved in cytoadhesion and pathogenesis (14). H3K9me3 is the landing pad for Heterochromatin Protein 1 (PfHP1) (15), which is essential for heterochromatin formation. This chromatin state, characterized for dense nucleosome packing and refractoriness to transcription, plays a pivotal role in *P. falciparum* gene regulation.

Around 400 genes are enriched with PfHP1 in *P. falciparum*. The vast majority of them belong to clusters whose boundaries mark synteny breakpoints or species-specific indels, and almost all of them are part of CVG families encoding exported proteins to the infected RBC surface (16). A notable exception to the aforementioned is a member of the ApiAP2 gene family. In this case, only a single gene out of 27 members is associated with PfHP1. This TF, named AP2G, has been shown to be a master regulator for sexual commitment (17). The *ap2-g* gene remains repressed in most parasites during the intraerythrocytic developmental cycle (IDC) and is only activated in a small subset of cells to initiate gametocytogenesis (2). A number of features make the *ap2-g* gene an appealing model for studying heterochromatin biology, which is underexplored in this pathogen despite its crucial role in transmission and virulence gene control. Firstly, it can be genetically manipulated since it is not essential for asexual development, and secondly, the heterochromatin island of *ap2-g* is embedded into euchromatin with well defined boundaries, facilitating the search for DNA control elements that limit the spreading of PfHP1. Such boundary elements have been suggested (12, 16), given the sharp PfHP1 profiles observed, but never found.

Here, we engineered a transgenic *P. falciparum* line unable to produce gametocytes by replacing the *ap2-g* AP2 domain at the C-terminus by GFP. We observe that upon such replacement, the marked PfHP1 boundary at the 3' end of the ORF is shifted ~2 kb downstream and infer that this depleted region contains a barrier element in order to keep heterochromatin confined to *ap2-g* and avoid spreading into adjacent genes. We characterize and provide the first description of this kind of elements found in *P. falciparum* so far.

RESULTS & DISCUSSION

Replacement of AP2 domain from *ap2-g* makes *P. falciparum* unable to differentiate and respond to gametocyte induction

Transfection of NF54 A11 clone with pL6-AP2G-GFP and pUF-Cas9 plasmids resulted in the replacement of the last 3' 765 bp of the *ap2-g* ORF by GFP protein (Fig. 1). That includes a

part of the AP2 domain and mimics the natural mutation of F12 clone, known to be unable to accomplish gametocytogenesis (17, 27).

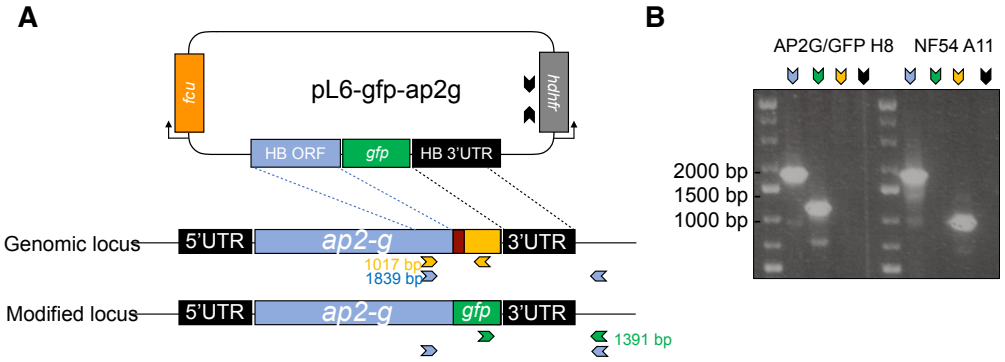


Figure 1. Development and validation of AP2G-GFP line. (A) Depiction of the pL6-gfp-ap2g plasmid co-transfected with pUF-Cas9 to replace the last 765 bp of *ap2-g* ORF by *gfp* gene. Arrows mark the primers used for further validation. Red square marks the AP2 domain, included in the replaced region (yellow). HB: homology box. **(B)** PCR validation of correct integration for AP2G-GFP H8 clone and parental line.

Consistently with this, upon induction, two clones of the AP2G-GFP line yielded no gametocytes, whereas the parental line reached ~8% conversion rate (Fig. 2A and 2B). GFP was visible on live imaged cells at very low rate (<0.1%), and we never obtained higher ratios of expressers even after gametocyte induction (Fig. 2C). Furthermore, transcription of *ap2-g* varies in AP2G-GFP clones, it peaks in late ring stages rather than in early rings and schizonts as occurs in the parental line (Fig. 2D). These data are in accordance with previous reports, suggesting that AP2-G expression is regulated by an auto feedback loop, and after initial expression of the protein, its binding to a putative regulatory element upstream of its own gene is required to trigger gametocytogenesis (2, 17). Partial depletion of DNA-binding domain AP2 abolishes the capacity of AP2-G to continue the feedback and keeps its expression at basal levels.

The 3' end of *ap2-g* ORF contains a boundary element

To evaluate the impact on the heterochromatin profile of replacing the terminal part of *ap2-g*, we performed ChIP-seq with anti-HP1 on tightly synchronised 36 hpi parasites from both

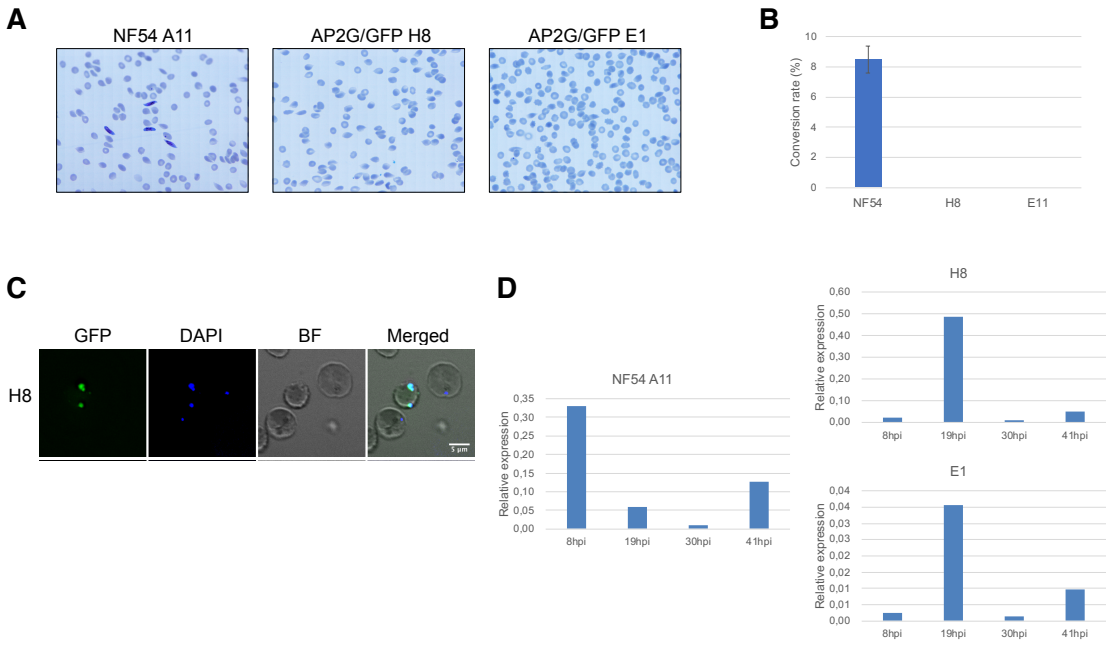


Figure 2. AP2G-GFP line validation. (A) Giemsa staining of induced gametocytes of parental line NF54 A11 and two clones of AP2G-GFP line. **(B)** Gametocyte conversion rate of lines in A. Mean ±SEM is shown. **(C)** Live imaging of AP2G-GFP H8 clone. Scale bar at 5 μm. **(D)** Expression levels of AP2G at different time points measured by RT-qPCR. Transcript levels are shown relative to ubiquitin-conjugating enzyme (PF3D7_0812600).

parental and two clones of the AP2G-GFP lines. In NF54 A11, PfHP1 spreads over the complete ORF, sharply dropping after the stop codon. Accumulation extends ~3 kb upstream the start codon, progressively falling after a first drop and increases again at ~1.7-1.9 kb from the start codon. AP2G-GFP clones exhibit a similar pattern upstream the *ap2-g* ORF, but the PfHP1 accumulation extends after the stop codon ~2 kb beyond the limit shown in the parental line (Fig. 3). This boundary shift strongly suggests the presence of DNA elements acting as a barrier for heterochromatin spreading in the 3' end of *ap2-g*. In other organisms such elements have been described to operate by multiple, often synergistic, mechanisms (28–30). Most of the means by which this is accomplished imply altering the chromatin environment required for PfHP1 spreading. Complexes containing any histone modifying enzyme regularly need specific DNA sequences to recognize the site where they act, but such elements have never been described in *Plasmodium*. The only suspect are the ITSs

(interstitial telomeric repeat sequences) from *P. knowlesi*, which locate adjacent to most heterochromatin clusters, although their role as border elements has never been formally demonstrated (18). Nucleosome depleted regions (NDR) or with high nucleosome turnover are also refractory to heterochromatin spreading, as they negate the histone marks required by PfHP1. High AT-rich content in intergenic regions of *P. falciparum*, which are prone to lack nucleosomes (11), are therefore natural obstacles for heterochromatin spreading. Interestingly, ATAC-seq data (31) revealed an NDR where PfHP1 accumulation stops in the AP2G-GFP clones, pointing to a gradual PfHP1 decrease due to this fact. However, the immediate vicinity downstream the *ap2-g* ORF does not show nucleosome depletion, and therefore, the sharp drop of PfHP1 in the parental line cannot be attributed to this mechanism (Fig. 3).

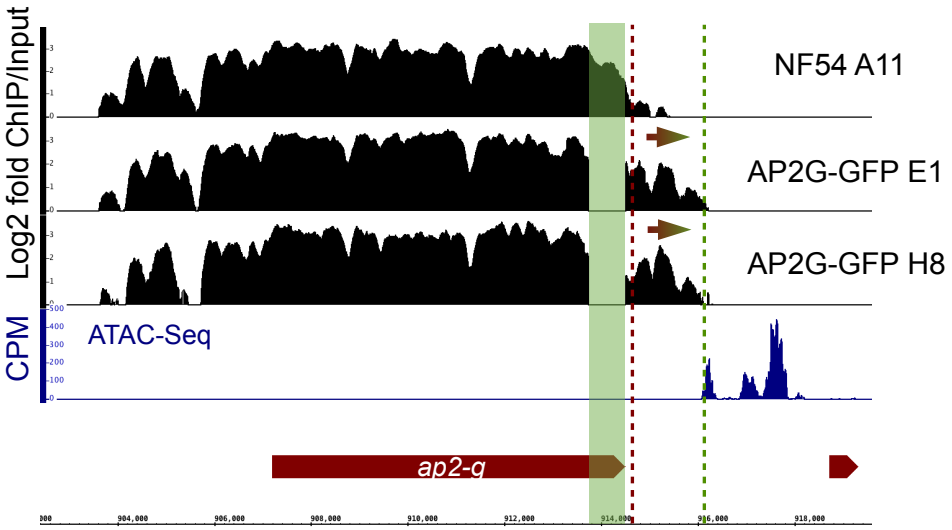


Figure 3. Heterochromatin boundary shift in *ap2-g* locus. ChIP-seq with anti-HP1 antibodies expressed as log2 fold enrichment over input. ATAC-seq data from (31) are shown in dark blue as normalised tags (counts per million reads -CPM-). Region replaced in AP2G-GFP lines is shaded in green. Boundaries at 3' end are marked with red and green lines for parental and AP2G-GFP lines, respectively.

Currently, we are assessing whether the re-insertion of the boundary region within the *ap2-g* locus can induce a change in PfHP1 boundary, and if so, whether this barrier works

in an orientation dependent manner.

A protein complex specifically binds to the boundary element

In order to elucidate the players operating in the identified boundary, we performed a screening of the replaced region searching proteins able to bind there in a sequence-specific manner. To do so, we synthesised biotinylated DNA fragments *in vitro* and used them as probes for EMSA. A specific band shift was observed in the fragment 2.2 (183 bp), indicating a protein complex interacting with this region (Fig. 4). We continue narrowing down the minimum required fragment keeping the shifted band and we plan to identify the complex bound by mass spectrometry. Until then, we can only speculate about the proteins involved in the boundary function. As PfHP1 needs H3K9me3 modification to propagate, histone acetyl transferases (HAT) or demethylases (HDM) replacing or removing such mark effectively block the heterochromatin spreading. Some bromodomain-containing proteins, which bind to acetylated histones, are able to prevent methylation of H3K9 by occupying such residue and making it inaccessible to further modifications. *P. falciparum* genome contains at least seven HATs (six from GNAT family, including PfGCN5, and one MYST family) and four HDMs (two lysine-specific histone demethylases -LSD1 and 2-, and two JmjC domain-containing proteins -JmjC1 and 2-). There are also seven predicted bromodomain-containing proteins, including again PfGCN5 (PF3D7_0823300). This protein is a homologue of yeast yGCN5 (32) and known to acetylate H3K9 in *P. falciparum* (33). In yeast, it is part of the SAGA (Spt-Ada-Gcn5 acetyltransferase) and SLIK (SAGA-like) complexes, involved in heterochromatin boundary function, and *gcn5* disruption completely abolishes such ability (34). There are some evidences pointing to a conserved function in *P. falciparum* and yeast. A homologue for another component of the SAGA complex was identified, PfADA2, and in yeast two-hybrid experiments, PfGCN5 and PfADA2 were shown to interact *in vitro*. Additionally, complementation assays showed that PfGCN5 can partially restore the phenotype of yGCN5 deletion (32). We hypothesize that the combination of histone acetyl transferase activity and the presence of bromodomain in PfGCN5 allows such protein to establish effective boundaries for heterochromatin spreading, likely interacting with other components, avoiding methylation of H3K9me3.

Aside from this promising candidate, the targets of other HATs and HDMs are described in less detail. PfJmjC1 (PF3D7_0809900) and LSD1 (PF3D7_1211600) preferentially demethy-

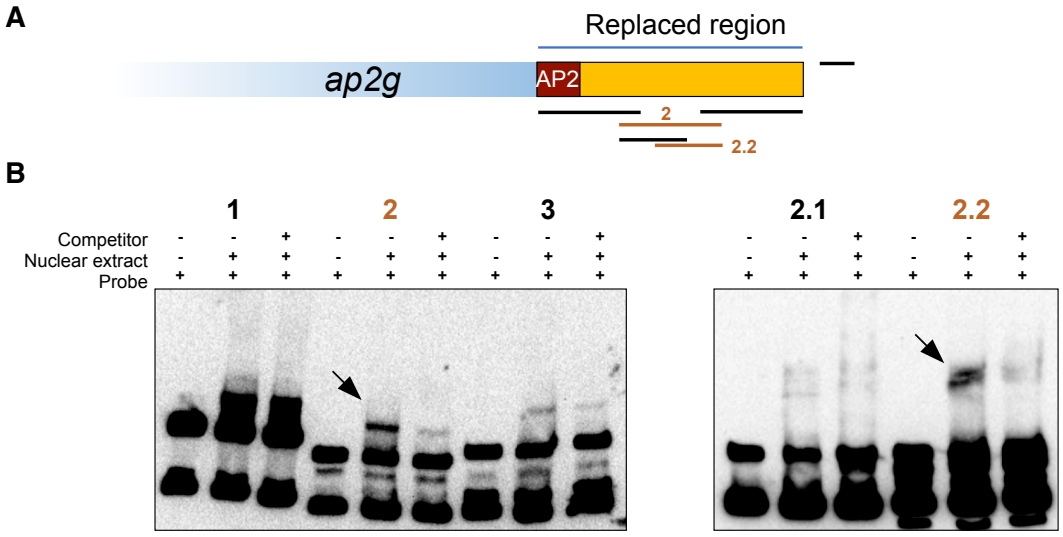


Figure 4. Search of protein complex binding to the 3' end *ap2-g* locus. (A) Depiction of the pL6-gfp-ap2g plasmid co-transfected with pUF-Cas9 to replace the last 765 bp of *ap2-g* ORF by *gfp* gene. Arrows mark the primers used for further validation. Red square marks the AP2 domain, included in the replaced region (yellow). HB: homology box. **(B)** PCR validation of correct integration for AP2G-GFP H8 clone and parental line.

late H3K36 and H3K9 (35, 36), and bromodomain proteins PfBDP1 (PF3D7_1033700) and PfBDP2 (Pf3D7_1212900) are part of a complex recognizing acetylated H3 (37). All them could be part of a potential complex involved in heterochromatin boundary formation, although further research will be required to confirm or rule out their involvement in this process.

In other organisms, as yeast or *Drosophila*, heterochromatin spreading can be blocked by physical means via loop formation and isolation of the region in the nuclear periphery (38, 39). Re-positioning of heterochromatin areas in such localisation have been described in *P. falciparum*, including *ap2-g* locus (12), and one 25-bp element found in *var* genes intron has been identified to bind an actin containing complex and to be sufficient to direct a plasmid to the nuclear periphery (40). However, there is no evidence of such element in *ap2-g* or other loci and it seems to be more involved in heterochromatin assembly than in the boundary function, as normally both exons of the *var* gene, surrounding the intron, maintain high levels of PfHP1 enrichment.

Lastly, transcription itself has been shown to perform a blocking function in some model organisms by means of the transcription machinery components or direct competition of

nascent RNA with HP1 binding to H3K9me3. It is especially relevant the case of tRNA genes and RNA polymerase III (RNAP III) transcription, which are essential for the boundary function in silent mating type locus and rDNA of budding yeast and pericentromeric heterochromatin in fission yeast. In mice, transcription of short interspaced nuclear elements (SINEs) is also required to establish the boundary which allows growth hormone to be expressed in the later stages of development (29, 41). tRNA genes are found close to some PfHP1-enriched areas of *P. falciparum*, especially in telomeres (12), although their role in boundary function is unclear and has been never researched.

To understand the heterochromatin biology of *P. falciparum* is crucial to comprehend the mechanisms of host-parasite interaction, as most genes encoding proteins involved in such processes are regulated via heterochromatin repression. Our knowledge about this topic suffers fundamental gaps, as we have very limited, and mostly speculative, information about heterochromatin establishment, spreading, maintenance and confinement in this parasite. Given the role of boundary elements in parasite virulence and transmission, in-depth insight into this essential process may reveal new targets for drug intervention strategies against this deadly pathogen.

MATERIALS & METHODS

Parasites culture and synchronisation

Blood stage NF54 A11 clone P. falciparum parasites were cultured in human O+ red blood cells in RPMI-1640 medium supplemented with Albumax II (10% v/v), hypoxanthine and gentamycin in 3% CO2 and 5% O2 at 37°C. Parasites were synchronized with a 6 h time window by sorbitol lysis during ring stage, followed by plasmagel enrichment in schizont stage and another sorbitol treatment 6 h after. Synchronized parasites were harvested at % hematocrit and ~2-4% parasitemia. Parasite development was monitored by Giemsa staining.

Plasmids, transfections and transgenic lines

Plasmids used for replacement of *ap2-g* 3' end and insertions of boundary elements derived from previously described pL6-GFP (19). Homology boxes, obtained through sequential PCRs were inserted into the pL6 vector digested with AflII and SpeI sites. The guide for Cas9 was obtained by direct hybridization of the oligos and inserted in in the same plasmid by Gibson assembly using the BtgZI sites.

The pL6 derived plasmids and pUF-Cas9 (19) were co-transfected (25 μg each) in the NF54 A11 clone (replacement of *ap2-g* 3' end) or the AP2G-GFP E1 clone (insertions of boundary elements). For that, ring stage parasites were transfected following the protocol described elsewhere (20) and maintained under WR99210 and DSM1 drug selection pressure. Parasite clones were obtained by limiting dilution. Treatment with 1 μM 5-Fluocytosine was applied as negative selection for parasites carrying the pL6 derived plasmid.

Gametocyte conversion assay and live cell imaging

Gametocytogenesis was performed as previously described (21). For live cell imaging bottom of imaging dishes (Ibidi) were covered with concanavalin A (5mg/mL) and incubated at 37°C for 30 min. Then, concanavalin A was carefully removed and dishes gently washed with sterile PBS. 500 μL of parasite cultures (3-5% parasitemia, 3-4% hematocrit), were added to the dish and let settle down for 10 minutes at 37°C. Unattached cells were washed out with PBS and finally covered with culture medium prepared with phenol red free RPMI 1640. Samples were visualized using a Deltavision Elite imaging system (GE Healthcare) and processed with the Fiji package (http://fiji.sc).

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

RNA was harvested from synchronized parasite cultures at indicated timepoints after lysis with 0.075% saponin in PBS followed by one wash in PBS and resuspension in Qiazol. Total RNA was extracted using the miRNeasy mini kit and performing on-column DNase treatment (Qiagen). Reverse transcription 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 specific primers. Starting quantity mean was normalised with a housekeeping gene (ubiquitin-conjugating enzyme, PF3D7_0812600). Starting quantity means of three replicates were extrapolated from a standard curve of serial dilutions of genomic DNA.

Chromatin immunoprecipitation and data analysis

ChIP was performed as previously described (22) with parasites synchronizedat 36 hpi. Sonicated chromatin (500 ng DNA content) was immunoprecipitated overnight with 0.5 μg of

anti-HP1 (Genscript) polyclonal rabbit antibodies (22) and incubated after with 25 μL of Dynabeads Protein G (Invitrogen) for two hours. Subsequent washing, cross-link reversion and DNA extraction were carried out as described before (22). Sequencing libraries were produced with the immunoprecipitated DNA using the MicroPlex Library Preparation Kit v2 (Diagenode) with the KAPA HiFi 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. A minimum of two biological replicates were analysed for each clone.

Sequencing reads were mapped to the P. falciparum genome (23) (PlasmoDB, v29) using Burrows-Wheeler Alignment tool (BWA-MEM) with default settings (24). PCR duplicates were removed. ChIP-seq data were normalized over input and likelihood ratio calculation and peak calling was performed using the MACS2 (25) software with default parameters and a false discovery rate (FDR) cut-off of 0.05.

Protein nuclear extract, probes and electrophoretic motility shift assay (EMSA)

Nuclei were obtained from asynchronous NF54 A11 parasites treated with saponin 0.15% and incubated in rotation for 1 hour at 4° C with low salt lysis buffer 1x (20 mM Tris HCl pH 7.5, 10 mM KCl, 2 mM DTT, 1.5 mM $MgCl_2$, 1% Triton X-100 and 1x proteinase inhibitor cocktail -PIC-). After spinning at 17,000 g for 20 min at 4° C, nuclei were extracted with high salt 1x lysis buffer (20 mM Tris HCl pH 7.5, 600 mM KCl, 2 mM DTT, 1.5 mM $MgCl_2$, 1% Triton X-100 and 1x PIC) for 30 min in rotation at 4° C and brief sonication. Samples were spun again and supernatant was taken and diluted four times in low salt lysis buffer without detergent. Protein concentration was measured by Bradford assay, performing a reference curve by increasing concentrations of BSA.

Probes for EMSA were obtained by direct PCR amplification of desired fragments with biotin labeled oligos (5' biotin modification in forward primer), from genomic DNA of NF54 A11 parasites. PCR products were separated in a 2% agarose gel, sliced and purified with Nucleospin PCR and Gel Clean-up Kit (Macherey-Nagel). Concentration was measured in a NanoDrop device.

To perform the gel shift assay, protein nuclear extracts (1 μg) were incubated with 5-20

fmols of biotin labeled probe in 1x EMSA buffer (20 mM Tris HCl pH 7.5, 60 mM KCl, 1 mM EDTA, 1 mM DTT, 2 mM $MgCl_2$, 25 μM ZnCl2, 0.1% Triton X-100, 5% glycerol, 200 μg/mL BSA) with 1 μg Poly(dI-dC) and 200 fmols of ssDNA oligo 5B1motF (26) as unspecific competitors. Binding reaction was performed at room temperature for 20 min. If specific competition was assayed, incubation with an unlabeled probe was conducted for 20 min at room temperature prior to addition of the labeled probe. After binding, samples were run in a 4-5% acrylamide gel in 0.5x TBE buffer, blotted to positively charged nylon membrane and developed using the Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher) as recommended by the manufacturer.

REFERENCES

1. WHO (2018) World Malaria Report
2. Poran,A., Nötzel,C., Aly,O., Mencia-Trinchant,N., Harris,C.T., Guzman,M.L., Hassane,D.C., Elemento,O. and Kafsack,B.F.C. (2017) Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. Nature 551, 95-99.
3. Painter,H.J., Chung,N.C., Sebastian,A., Albert,I., Storey,J.D. and Llinás,M. (2018) Genome-wide real-time in vivo transcriptional dynamics during Plasmodium falciparum blood-stage development. Nat Commun 9, 2656.
4. Oakley,M.S., Verma,N., Myers,T.G., Zheng,H., Locke,E., Morin,M.J., Tripathi,A.K., Mlambo, G. and Kumar,S. (2018) Transcriptome analysis based detection of Plasmodium falciparum development in Anopheles stephensi mosquitoes. Sci Rep 8, 11568.
5. Balaji,S., Babu,M.M., Iyer,L.M. and Aravind,L. (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Res 33, 3994-4006.
6. Duffy,M.F., Selvarajah,S.A., Josling,G.A. and Petter,M. (2014) Epigenetic regulation of the Plasmodium falciparum genome. Brief Funct Genomics 13, 203-216.
7. Cortés,A. and Deitsch,K.W. (2017) Malaria Epigenetics. Cold Spring Harb Perspect Med.
8. Abel,S. and Le Roch,K.G. (2019) The role of epigenetics and chromatin structure in transcriptional regulation in malaria parasites. Brief Funct Genomics
9. 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 associated with transcription levels in the human malaria parasite Plasmodium falciparum. BMC Genomics 15, 347.

10. 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, 2110-2124.
11. 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. Mol Microbiol 87, 1061-1073.
12. Lopez-Rubio,J.J., Mancio-Silva,L. and Scherf,A. (2009) Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5, 179-190.
13. Karmodiya,K., Pradhan,S.J., Joshi,B., Jangid,R., Reddy,P.C. and Galande,S. (2015) A comprehensive epigenome map of Plasmodium falciparum reveals unique mechanisms of transcriptional regulation and identifies H3K36me2 as a global mark of gene suppression. Epigenetics Chromatin 8, 32. 14. Scherf,A., Lopez-Rubio,J.J. and Riviere,L. (2008) Antigenic variation in Plasmodium falciparum. Annu Rev Microbiol 62, 445-470.
15. 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, 2596-2606.
16. Flueck,C., Bartfai,R., Volz,J., Niederwieser,I., Salcedo-Amaya,A.M., Alako,B.T., Ehlgen,F., Ralph,S.A., Cowman,A.F., Bozdech,Z., Stunnenberg,H.G. and Voss,T.S. (2009) Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog 5, e1000569.
17. Kafsack,B.F., Rovira-Graells,N., Clark,T.G., Bancells,C., Crowley,V.M., Campino, S.G., Williams, A.E., Drought,L.G., Kwiatkowski,D.P., Baker,D.A., Cortés,A. and Llinás,M. (2014) A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature 507, 248-252.
18. Fraschka,S.A., Filarsky,M., Hoo,R., Niederwieser,I., Yam,X.Y., Brancucci,N.M.B., Mohring, F., Mushunje,A.T., Huang,X., Christensen,P.R., Nosten,F., Bozdech,Z., Russell,B., Moon,R.W., Marti,M., Preiser,P.R., Bártfai,R. and Voss,T.S. (2018) Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites. Cell Host Microbe 23, 407-420.e8.
19. Ghorbal,M., Gorman,M., Macpherson,C.R., Martins,R.M., Scherf,A. and Lopez-Rubio,J.J.

(2014) Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. Nat Biotechnol 32, 819-821.
20. Hasenkamp,S., Merrick,C.J. and Horrocks,P. (2013) A quantitative analysis of Plasmodium falciparum transfection using DNA-loaded erythrocytes. Mol Biochem Parasitol 187, 117-120.
21. Fivelman,Q.L., McRobert,L., Sharp,S., Taylor,C.J., Saeed,M., Swales,C.A., Sutherland,C.J. and Baker,D.A. (2007) Improved synchronous production of Plasmodium falciparum gametocytes in vitro. Mol Biochem Parasitol 154, 119-123.
22. 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, 2951-2963.
23. 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, 498-511. 24. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.
25. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. and Liu,X.S. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137.
26. Voss,T.S., Mini,T., Jenoe,P. and Beck,H.P. (2002) Plasmodium falciparum possesses a cell cycle-regulated short type replication protein A large subunit encoded by an unusual transcript. J Biol Chem 277, 17493-17501.
27. Alano,P., Roca,L., Smith,D., Read,D., Carter,R. and Day,K. (1995) Plasmodium falciparum: parasites defective in early stages of gametocytogenesis. Exp Parasitol 81, 227-235.
28. Lin,N., Li,X., Cui,K., Chepelev,I., Tie,F., Liu,B., Li,G., Harte,P., Zhao,K., Huang,S. and Zhou,L. (2011) A barrier-only boundary element delimits the formation of facultative heterochromatin in Drosophila melanogaster and vertebrates. Mol Cell Biol 31, 2729-2741.

29. Wang,J., Lawry,S.T., Cohen,A.L. and Jia,S. (2014) Chromosome boundary elements and regulation of heterochromatin spreading. Cell Mol Life Sci 71, 4841-4852.
30. Cohen,A.L. and Jia,S. (2014) Noncoding RNAs and the borders of heterochromatin. Wiley Interdiscip Rev RNA 5, 835-847.
31. 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, 9414-9431.
32. Fan,Q., An,L. and Cui,L. (2004) Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling. Eukaryot Cell 3, 264-276.
33. 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 falciparum. Eukaryot Cell 6, 1219-1227.
34. Kamata,K., Hatanaka,A., Goswami,G., Shinmyozu,K., Nakayama,J., Urano,T., Hatashita, M., Uchida,H. and Oki,M. (2013) C-terminus of the Sgf73 subunit of SAGA and SLIK is important for retention in the larger complex and for heterochromatin boundary function. Genes Cells 18, 823-837.
35. Cui,L., Fan,Q., Cui,L. and Miao,J. (2008) Histone lysine methyltransferases and demethylases in Plasmodium falciparum. Int J Parasitol 38, 1083-1097.
36. Cui,L. and Miao,J. (2010) Chromatin-mediated epigenetic regulation in the malaria parasite Plasmodium falciparum. Eukaryot Cell 9, 1138-1149.
37. Josling,G.A., Petter,M., Oehring,S.C., Gupta,A.P., Dietz,O., Wilson,D.W., Schubert,T., Längst,G., Gilson,P.R., Crabb,B.S., Moes,S., Jenoe,P., Lim,S.W., Brown,G.V., Bozdech,Z., Voss, T.S. and Duffy,M.F. (2015) A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression. Cell Host Microbe 17, 741-751.
38. Gdula,D.A., Gerasimova,T.I. and Corces,V.G. (1996) Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. Proc Natl Acad Sci U S A 93, 9378-9383.
39. 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, 859-872.
40. 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, 451-463.
41. 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, 248-251.

References

[Agarwal et al., 2011] Agarwal, S., Kern, S., Halbert, J., Przyborski, J. M., Baumeister, S., Dandekar, T., Doerig, C., and Pradel, G. (2011). Two nucleus-localized cdk-like kinases with crucial roles for malaria parasite erythrocytic replication are involved in phosphorylation of splicing factor. *J Cell Biochem*, 112(5):1295–1310. (Cited on page 27.)

[Akhtar and Gasser, 2007] Akhtar, A. and Gasser, S. M. (2007). The nuclear envelope and transcriptional control. *Nat Rev Genet*, 8(7):507–517. (Cited on page 37.)

[Amit-Avraham et al., 2015] Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E., and 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):E982–91. (Cited on pages 38 and 57.)

[Andrews et al., 2008] Andrews, K. T., Tran, T. N., Lucke, A. J., Kahnberg, P., Le, G. T., Boyle, G. M., Gardiner, D. L., Skinner-Adams, T. S., and Fairlie, D. P. (2008). Potent antimalarial activity of histone deacetylase inhibitor analogues. *Antimicrob Agents Chemother*, 52(4):1454–1461. (Cited on page 18.)

[Ariey and Ménard, 2019] Ariey, F. and Ménard, D. (2019). An update on artemisinin resistance. *Methods Mol Biol*, 2013:141–149. (Cited on page 6.)

[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. (Cited on page 5.)

[Armelin-Correa et al., 2014] Armelin-Correa, L. M., Gutiyama, L. M., Brandt, D. Y., and Malnic, B. (2014). Nuclear compartmentalization of odorant receptor genes. *Proc Natl Acad Sci U S A*, 111(7):2782–2787. (Cited on page 48.)

[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. (Cited on page 5.)

[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 erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res*, 24(6):974–988. (Cited on pages 22 and 39.)

[Azzalin et al., 2007] Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing rna and rna surveillance factors at mammalian chromosome ends. *Science*, 318(5851):798–801. (Cited on page 56.)

[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. (Cited on pages 24 and 25.)

[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. (Cited on page 56.)

[Bannister and Dluzewski, 1990] Bannister, L. H. and Dluzewski, A. R. (1990). The ultrastructure of red cell invasion in malaria infections: a review. *Blood Cells*, 16(2-3):257–92; discussion 293. (Cited on page 9.)

[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 conundrum. *PLoS Pathog*, 10(9):e1004273. (Cited on page 6.)

[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. (Cited on page 11.)

[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. (Cited on page 55.)

[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*. (Cited on page 28.)

[Bell and Felsenfeld, 2000] Bell, A. C. and Felsenfeld, G. (2000). Methylation of a ctcf-dependent boundary controls imprinted expression of the igf2 gene. *Nature*, 405(6785):482–485. (Cited on page 46.)

[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 regulates zeb2/sip1 gene expression during snail1-induced epithelial-mesenchymal transition. *Genes Dev*, 22(6):756–769. (Cited on page 54.)

[Bernabeu and Smith, 2017] Bernabeu, M. and Smith, J. D. (2017). Epcr and malaria severity: The center of a perfect storm. *Trends Parasitol*, 33(4):295–308. (Cited on page 31.)

[Berriman et al., 2005] Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E., and El-Sayed, N. M. (2005). The genome of the african trypanosome trypanosoma brucei. *Science*, 309(5733):416–422. (Cited on page 44.)

[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. (Cited on page 24.)

[Borst, 2003] Borst, P. (2003). Mechanisms of antigenic variation: an overview. *Antigenic Variation*, ed. A Craig, A Scherf. London: Academic Press:1–16. (Cited on page 29.)

[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., 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. (Cited on page 14.)

[Bozdech et al., 2003] Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of plasmodium falciparum. *PLoS Biol*, 1:E5. (Cited on pages 12 and 13.)

[Branco and Pombo, 2007] Branco, M. R. and Pombo, A. (2007). Chromosome organization: new facts, new models. *Trends Cell Biol*, 17(3):127–134. (Cited on page 12.)

[Brancucci et al., 2012] Brancucci, N. M., Witmer, K., Schmid, C. D., Flueck, C., and Voss, T. S. (2012). Identification of a cis-acting dna-protein interaction implicated in singular var gene choice in plasmodium falciparum. *Cell Microbiol*, 14(12):1836–1848. (Cited on page 39.)

[Briquet et al., 2006] Briquet, S., Boschet, C., Gissot, M., Tissandié, E., Sevilla, E., Franetich, J., Thiery, I., Hamid, Z., Bourgouin, C., and Vaquero, C. (2006). High-mobility-group box nuclear factors of plasmodium falciparum. *Eukaryot Cell*, 5(4):672–682. (Cited on page 25.)

[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. (Cited on pages 55, 56, and 57.)

[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. (Cited on pages 55, 56, and 58.)

[Brockdorff et al., 1992] Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (1992). The product of the mouse xist gene is a 15 kb inactive x-specific transcript containing no conserved orf and located in the nucleus. *Cell*, 71(3):515–526. (Cited on page 47.)

[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. (Cited on page 37.)

[Brown and Brown, 1965] Brown, K. N. and Brown, I. N. (1965). Immunity to malaria: antigenic variation in chronic infections of plasmodium knowlesi. *Nature*, 208(5017):1286–1288. (Cited on page 29.)

[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). (Cited on pages 38, 57, and 179.)

[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 cell surface are targets for naturally acquired immunity to malaria. *Nat Med*, 4(3):358–360. (Cited on page 31.)

[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. (Cited on page 51.)

[Bunnik et al., 2018] Bunnik, E. M., Cook, K. B., Varoquaux, N., Batugedara, G., Prudhomme, J., Cort, A., Shi, L., Andolina, C., Ross, L. S., Brady, D., Fidock, D. A., Nosten, F., Tewari, R., Sinnis, P., Ay, F., Vert, J. P., Noble, W. S., and Le Roch, K. G. (2018). Changes in genome organization of parasite-specific gene families during the plasmodium transmission stages. *Nat Commun*, 9(1):1910. (Cited on page 23.)

[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 associated with transcription levels in the human malaria parasite plasmodium falciparum. *BMC Genomics*, 15:347. (Cited on page 17.)

[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 Stunnenberg, H. G. (2010). H2a.z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by h3k9ac and h3k4me3. *PLoS Pathog*, 6(12):e1001223. (Cited on pages 20 and 37.)

[Cabral and Wunderlich, 2009] Cabral, F. J. and Wunderlich, G. (2009). Transcriptional memory and switching in the plasmodium falciparum rif gene family. *Mol Biochem Parasitol*, 168(2):186–190. (Cited on page 41.)

[Calderwood et al., 2003] Calderwood, M. S., Gannoun-Zaki, L., Wellems, T. E., and 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(36):34125–34132. (Cited on pages 24, 38, and 57.)

[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. (Cited on pages 4 and 5.)

[Cam et al., 2005] Cam, H. P., Sugiyama, T., Chen, E. S., Chen, X., FitzGerald, P. C., and Grewal, S. I. (2005). Comprehensive analysis of heterochromatin- and rnai-mediated epigenetic control of the fission yeast genome. *Nat Genet*, 37(8):809–819. (Cited on page 54.)

[Campbell et al., 2010] Campbell, T. L., De Silva, E. K., Olszewski, K. L., Elemento, O., and Llinás, M. (2010). Identification and genome-wide prediction of dna binding specificities for the apiap2 family of regulators from the malaria parasite. *PLoS Pathog*, 6(10):e1001165. (Cited on page 25.)

[Canella et al., 2010] Canella, D., Praz, V., Reina, J., Cousin, P., and Hernandez, N. (2010). Defining the rna polymerase iii transcriptome: Genome-wide localization of the rna polymerase iii transcription machinery in human cells. *Genome Res*, 20(6):710–721. (Cited on page 170.)

[Carmo-Fonseca and Rino, 2011] Carmo-Fonseca, M. and Rino, J. (2011). Rna seeds nuclear bodies. *Nat Cell Biol*, 13(2):110–112. (Cited on page 53.)

[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. (Cited on page 54.)

[Cech and Steitz, 2014] Cech, T. R. and Steitz, J. A. (2014). The noncoding rna revolution-trashing old rules to forge new ones. *Cell*, 157(1):77–94. (Cited on page 49.)

[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. (Cited on page 55.)

[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 function identified in malaria parasites by comparative genomics and rna analysis. *RNA*, 13(11):1923–1939. (Cited on pages 55, 58, 59, and 60.)

[Chen and Shyu, 2011] Chen, C. Y. and Shyu, A. B. (2011). Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA*, 2:167–183. (Cited on page 27.)

[Chen and Manley, 2009] Chen, M. and Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*, 10(11):741–754. (Cited on page 27.)

[Chen et al., 2016] Chen, P. B., Ding, S., Zanghì, G., Soulard, V., DiMaggio, P. A., Fuchter, M. J., Mecheri, S., Mazier, D., Scherf, A., and Malmquist, N. A. (2016). Plasmodium falciparum pfset7: enzymatic characterization and cellular localization of a novel protein methyltransferase in sporozoite, liver and erythrocytic stage parasites. *Sci Rep*, 6:21802. (Cited on page 18.)

[Chess, 2016] Chess, A. (2016). Monoallelic gene expression in mammals. *Annu Rev Genet*, 50:317–327. (Cited on pages 45 and 47.)

[Chess et al., 1994] Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell*, 78(5):823–834. (Cited on pages 47 and 48.)

[Choi et al., 2006] Choi, S. W., Keyes, M. K., and Horrocks, P. (2006). Lc/esi-ms demonstrates the absence of 5-methyl-2'-deoxycytosine in plasmodium falciparum genomic dna. *Mol Biochem Parasitol*, 150(2):350–352. (Cited on page 21.)

[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. (Cited on pages 18, 35, 36, and 169.)

[Chujo et al., 2016] Chujo, T., Yamazaki, T., and Hirose, T. (2016). Architectural rnas (arcrnas): A class of long noncoding rnas that function as the scaffold of nuclear bodies. *Biochim Biophys Acta*, 1859(1):139–146. (Cited on page 169.)

[Chêne et al., 2012] Chêne, A., Vembar, S. S., Rivière, L., Lopez-Rubio, J. J., Claes, A., Siegel, T. N., Sakamoto, H., Scheidig-Benatar, C., Hernandez-Rivas, R., and Scherf, A. (2012). Pfalbas constitute a new eukaryotic dna/rna-binding protein family in malaria parasites. *Nucleic Acids Res*, 40(7):3066–3077. (Cited on page 25.)

[Clemson et al., 1996] Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). Xist rna paints the inactive x chromosome at interphase: evidence for a novel rna involved in nuclear/chromosome structure. *J Cell Biol*, 132(3):259–275. (Cited on page 47.)

[Clowney et al., 2012] Clowney, E. J., LeGros, M. A., Mosley, C. P., Clowney, F. G., Markenskoff-Papadimitriou, E. C., Myllys, M., Barnea, G., Larabell, C. A., and Lomvardas, S. (2012). Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell*, 151(4):724–737. (Cited on page 48.)

[Cobb et al., 2006] Cobb, R. M., Oestreich, K. J., Osipovich, O. A., and Oltz, E. M. (2006). Accessibility control of v(d)j recombination. *Adv Immunol*, 91:45–109. (Cited on page 55.)

[Coetzee et al., 2017] Coetzee, N., Sidoli, S., van Biljon, R., Painter, H., Llinás, M., Garcia, B., and Birkholtz, L. M. (2017). Quantitative chromatin proteomics reveals a dynamic histone post-translational modification landscape that defines asexual and sexual plasmodium falciparum parasites. *Sci Rep*, 7(1):607. (Cited on page 18.)

[Cortés et al., 2007] Cortés, A., Carret, C., Kaneko, O., Yim Lim, B., Ivens, A., and Holder, A. (2007). Epigenetic silencing of plasmodium falciparum genes linked to erythrocyte invasion. *PLoS Pathog*, 3(8):e107. (Cited on page 44.)

[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. (Cited on page 4.)

[Cowman and Crabb, 2006] Cowman, A. F. and Crabb, B. S. (2006). Invasion of red blood cells by malaria parasites. *Cell*, 124(4):755–766. (Cited on page 9.)

[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 Microbe*, 22(2):232–245. (Cited on pages 7 and 9.)

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

[Cox, 1959] Cox, H. W. (1959). A study of relapse plasmodium berghei infections isolated from white mice. *J Immunol*, 82(3):209–214. (Cited on page 29.)

[Crepaldi et al., 2013] Crepaldi, L., Policarpi, C., Coatti, A., Sherlock, W. T., Jongbloets, B. C., Down, T. A., and Riccio, A. (2013). Binding of tfiiic to sine elements controls the relocation of activity-dependent neuronal genes to transcription factories. *PLoS Genet*, 9(8):e1003699. (Cited on page 170.)

[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. (Cited on page 18.)

[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. (Cited on page 18.)

[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 falciparum. *Eukaryot Cell*, 6(7):1219–1227. (Cited on page 18.)

[Cunningham et al., 2010] Cunningham, D., Lawton, J., Jarra, W., Preiser, P., and Langhorne, J. (2010). The pir multigene family of plasmodium: antigenic variation and beyond. *Mol Biochem Parasitol*, 170(2):65–73. (Cited on page 30.)

[Dalmasso et al., 2011] Dalmasso, M. C., Sullivan, W. J., and Angel, S. O. (2011). Canonical and variant histones of protozoan parasites. *Front Biosci (Landmark Ed)*, 16:2086–2105. (Cited on page 20.)

[Dalton et al., 2013] Dalton, R. P., Lyons, D. B., and Lomvardas, S. (2013). Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell*, 155(2):321–332. (Cited on page 48.)

[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 chromatin assembly pathway. *BMC Biol*, 10:5. (Cited on page 18.)

[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 antibody modulation of cytoadherence of infected erythrocytes. *Proc Natl Acad Sci U S A*, 80(16):5075–5079. (Cited on page 10.)

[Day and Tuite, 1998] Day, D. A. and Tuite, M. F. (1998). Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *J Endocrinol*, 157(3):361–371. (Cited on page 26.)

[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. (Cited on page 52.)

[De Silva et al., 2008] De Silva, E. K., Gehrke, A. R., Olszewski, K., León, I., Chahal, J. S., Bulyk, M. L., and Llinás, M. (2008). Specific dna-binding by apicomplexan ap2 transcription factors. *Proc Natl Acad Sci U S A*, 105(24):8393–8398. (Cited on page 25.)

[Deitsch et al., 2001] Deitsch, K. W., Calderwood, M. S., and Wellems, T. E. (2001). Malaria. cooperative silencing elements in var genes. *Nature*, 412(6850):875–876. (Cited on page 38.)

[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. (Cited on pages 29 and 30.)

[Dieci et al., 2007] Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., and Pagano, A. (2007). The expanding rna polymerase iii transcriptome. *Trends Genet*, 23(12):614–622. (Cited on page 171.)

[Dixit et al., 2010] Dixit, A., Singh, P. K., Sharma, G. P., Malhotra, P., and Sharma, P. (2010). Pfsrpk1, a novel splicing-related kinase from plasmodium falciparum. *J Biol Chem*, 285(49):38315–38323. (Cited on page 27.)

[Doerig et al., 2015] Doerig, C., Rayner, J. C., Scherf, A., and Tobin, A. B. (2015). Post-translational protein modifications in malaria parasites. *Nat Rev Microbiol*, 13(3):160–172. (Cited on page 19.)

[Dondorp et al., 2009] Dondorp, A. M., Nosten, F., Yi, P., Das, D., Phyo, A. P., Tarning, J., Lwin, K. M., Ariey, F., Hanpithakpong, W., Lee, S. J., Ringwald, P., Silamut, K., Imwong, M., Chotivanich, K., Lim, P., Herdman, T., An, S., Yeung, S., Singhasivanon, P., Day, N., Lindegardh, N., Socheat, D., and White, N. J. (2009). Artemisinin resistance in plasmodium falciparum malaria. *N Engl J Med*, 361(5):455–467. (Cited on page 5.)

[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. (Cited on page 6.)

[Donze and Kamakaka, 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. (Cited on page 54.)

[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. (Cited on page 27.)

[Duffy et al., 2006] Duffy, M. F., Maier, A. G., Byrne, T. J., Marty, A. J., Elliott, S. R., O'Neill, M. T., Payne, P. D., Rogerson, S. J., Cowman, A. F., Crabb, B. S., and Brown, G. V. (2006). Var2csa is the principal ligand for chondroitin sulfate a in two allogeneic isolates of plasmodium falciparum. *Mol Biochem Parasitol*, 148(2):117–124. (Cited on page 31.)

[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 virulence genes in plasmodium falciparum. *Cell*, 121(1):13–24. (Cited on pages 18, 36, and 37.)

[Dzikowski et al., 2006] Dzikowski, R., Frank, M., and Deitsch, K. W. (2006). Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathog*, 2(3):e22. (Cited on pages 38 and 39.)

[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. (Cited on page 54.)

[Enderes et al., 2011] Enderes, C., Kombila, D., Dal-Bianco, M., Dzikowski, R., Kremsner, P., and Frank, M. (2011). Var gene promoter activation in clonal plasmodium falciparum isolates follows a hierarchy and suggests a conserved switching program that is independent of genetic background. *J Infect Dis*, 204(10):1620–1631. (Cited on page 39.)

[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. (Cited on pages 38 and 57.)

[Eshar et al., 2012] Eshar, S., Allemand, E., Sebag, A., Glaser, F., Muchardt, C., Mandel-Gutfreund, Y., Karni, R., and Dzikowski, R. (2012). A novel plasmodium falciparum sr protein is an alternative splicing factor required for the parasites' proliferation in human erythrocytes. *Nucleic Acids Res*, 40(19):9903–9916. (Cited on page 27.)

[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). Expression 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. (Cited on pages 54 and 55.)

[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. (Cited on page 54.)

[Fan et al., 2004] Fan, Q., An, L., and Cui, L. (2004). Plasmodium falciparum histone acetyltransferase, a yeast gcn5 homologue involved in chromatin remodeling. *Eukaryot Cell*, 3(2):264–276. (Cited on page 18.)

[Faria et al., 2019] Faria, J., Glover, L., Hutchinson, S., Boehm, C., Field, M. C., and Horn, D. (2019). Monoallelic expression and epigenetic inheritance sustained by a trypanosoma brucei variant surface glycoprotein exclusion complex. *Nat Commun*, 10(1):3023. (Cited on page 45.)

[Ferguson-Smith, 2011] Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet*, 12(8):565–575. (Cited on page 46.)

[Fernandez et al., 1999] Fernandez, V., Hommel, M., Chen, Q., Hagblom, P., and Wahlgren, M. (1999). Small, clonally variant antigens expressed on the surface of the plasmodium falciparum-infected erythrocyte are encoded by the rif gene family and are the target of human immune responses. *J Exp Med*, 190(10):1393–1404. (Cited on page 41.)

[Ferrigno et al., 2001] Ferrigno, O., Virolle, T., Djabari, Z., Ortonne, J. P., White, R. J., and Aberdam, D. (2001). Transposable b2 sine elements can provide mobile rna polymerase ii promoters. *Nat Genet*, 28(1):77–81. (Cited on page 54.)

[Figueiredo and Cross, 2010] Figueiredo, L. M. and Cross, G. A. (2010). Nucleosomes are depleted at the vsg expression site transcribed by rna polymerase i in african trypanosomes. *Eukaryot Cell*, 9(1):148–154. (Cited on page 44.)

[Figueiredo et al., 2002] Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J. C., and Scherf, A. (2002). A central role for plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J*, 21(4):815–824. (Cited on page 56.)

[Figueiredo et al., 2008] Figueiredo, L. M., Janzen, C. J., and Cross, G. A. (2008). A histone methyltransferase modulates antigenic variation in african trypanosomes. *PLoS Biol*, 6(7):e161. (Cited on page 45.)

[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. (Cited on page 14.)

[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. (Cited on page 57.)

[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. (Cited on page 50.)

[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. (Cited on pages 25 and 38.)

[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. (Cited on pages 18 and 36.)

[Fox et al., 2002] Fox, A. H., Lam, Y. W., Leung, A. K., Lyon, C. E., Andersen, J., Mann, M., and Lamond, A. I. (2002). Paraspeckles: a novel nuclear domain. *Curr Biol*, 12(1):13–25. (Cited on page 53.)

[Francia and Striepen, 2014] Francia, M. E. and Striepen, B. (2014). Cell division in apicomplexan parasites. *Nat Rev Microbiol*, 12(2):125–136. (Cited on page 6.)

[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. (Cited on pages 35 and 167.)

[Frank et al., 2006] Frank, M., Dzikowski, R., Costantini, D., Amulic, B., Berdougo, E., and Deitsch, K. W. (2006). Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite plasmodium falciparum. *J Biol Chem*, 281(15):9942–9952. (Cited on page 38.)

[Fraschka et al., 2016] Fraschka, S., Henderson, R. W., and Bártfai, R. (2016). H3.3 demarcates gc-rich coding and subtelomeric regions and serves as potential memory mark for virulence gene expression in plasmodium falciparum. *Sci Rep*, 6:31965. (Cited on pages 20 and 37.)

[Freitas-Junior et al., 2000] Freitas-Junior, L. H., Bottius, E., Pirrit, L., Deitsch, K. W., Scheidig, 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. falciparum. *Nature*, 407(6807):1018–1022. (Cited on pages 22 and 33.)

[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 propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*, 121(1):25–36. (Cited on pages 18, 35, 36, 37, and 167.)

[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. (Cited on page 31.)

[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. (Cited on page 6.)

[Fujioka and Aikawa, 2002] Fujioka, H. and Aikawa, M. (2002). Structure and life cycle. *Chem Immunol*, 80:1–26. (Cited on page 7.)

[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. (Cited on page 35.)

[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. (Cited on pages 14, 22, 33, 40, 55, 58, and 166.)

[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. (Cited on page 14.)

[Gdula et al., 1996] Gdula, D. A., Gerasimova, T. I., and Corces, V. G. (1996). Genetic and molecular analysis of the gypsy chromatin insulator of drosophila. *Proc Natl Acad Sci U S A*, 93(18):9378–9383. (Cited on page 170.)

[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. (Cited on pages 51 and 52.)

[Gissot et al., 2005] Gissot, M., Briquet, S., Refour, P., Boschet, C., and Vaquero, C. (2005). Pfmyb1, a plasmodium falciparum transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *J Mol Biol*, 346(1):29–42. (Cited on page 25.)

[Gissot et al., 2008] Gissot, M., Choi, S. W., Thompson, R. F., Greally, J. M., and Kim, K. (2008). Toxoplasma gondii and cryptosporidium parvum lack detectable dna cytosine methylation. *Eukaryot Cell*, 7(3):537–540. (Cited on page 21.)

[Glover et al., 2016] Glover, L., Hutchinson, S., Alsford, S., and Horn, D. (2016). Vex1 controls the allelic exclusion required for antigenic variation in trypanosomes. *Proc Natl Acad Sci U S A*, 113(26):7225–7230. (Cited on page 45.)

[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. (Cited on pages 39, 41, and 42.)

[Gong and Maquat, 2011] Gong, C. and Maquat, L. E. (2011). lncrnas transactivate stau1-mediated mrna decay by duplexing with 3' utrs via alu elements. *Nature*, 470(7333):284–288. (Cited on page 54.)

[Govindaraju et al., 2017] Govindaraju, G., Jabeena, C. A., Sethumadhavan, D. V., Rajaram, N., and Rajavelu, A. (2017). Dna methyltransferase homologue trdmt1 in plasmodium falciparum specifically methylates endogenous aspartic acid trna. *Biochim Biophys Acta Gene Regul Mech*, 1860(10):1047–1057. (Cited on page 21.)

[Goyal et al., 2012] Goyal, M., Alam, A., Iqbal, M. S., Dey, S., Bindu, S., Pal, C., Banerjee, A., Chakrabarti, S., and Bandyopadhyay, U. (2012). Identification and molecular characterization of an alba-family protein from human malaria parasite plasmodium falciparum. *Nucleic Acids Res*, 40(3):1174–1190. (Cited on page 25.)

[Grant et al., 2012] Grant, J., Mahadevaiah, S. K., Khil, P., Sangrithi, M. N., Royo, H., Duckworth, J., McCarrey, J. R., VandeBerg, J. L., Renfree, M. B., Taylor, W., Elgar, G., Camerini-Otero, R. D., Gilchrist, M. J., and Turner, J. M. (2012). Rsx is a metatherian rna with xist-like properties in x-chromosome inactivation. *Nature*, 487(7406):254–258. (Cited on page 47.)

[Grewal and Jia, 2007] Grewal, S. I. and Jia, S. (2007). Heterochromatin revisited. *Nat Rev Genet*, 8(1):35–46. (Cited on page 20.)

[Guizetti et al., 2016] Guizetti, J., Barcons-Simon, A., and Scherf, A. (2016). Trans-acting gc-rich non-coding rna at var expression site modulates gene counting in malaria parasite. *Nucleic Acids Res*, 44(20):9710–9718. (Cited on page 167.)

[Guizetti et al., 2013] Guizetti, J., Martins, R. M., Guadagnini, S., Claes, A., and Scherf, A. (2013). Nuclear pores and perinuclear expression sites of var and ribosomal dna genes correspond to physically distinct regions in plasmodium falciparum. *Eukaryot Cell*, 12(5):697–702. (Cited on page 38.)

[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. (Cited on page 35.)

[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 motifs uncovered among gene expression clusters in plasmodium falciparum. *Mol Biochem Parasitol*, 153(1):19–30. (Cited on page 24.)

[Gunasekera et al., 2004] Gunasekera, A. M., Patankar, S., Schug, J., Eisen, G., Kissinger, J., Roos, D., and Wirth, D. F. (2004). Widespread distribution of antisense transcripts in the plasmodium falciparum genome. *Mol Biochem Parasitol*, 136(1):35–42. (Cited on page 55.)

[Gómez-Díaz et al., 2017] Gómez-Díaz, E., Yerbanga, R. S., Lefèvre, T., Cohuet, A., Rowley, M. J., Ouedraogo, J. B., and Corces, V. G. (2017). Epigenetic regulation of plasmodium falciparum clonally variant gene expression during development in anopheles gambiae. *Sci Rep*, 7:40655. (Cited on page 32.)

[Günzl et al., 2003] Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Chung, H. M., Lee, P. T., and Lee, M. G. (2003). Rna polymerase i transcribes procyclin genes and variant surface glycoprotein gene expression sites in trypanosoma brucei. *Eukaryot Cell*, 2(3):542–551. (Cited on page 44.)

[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., Burrows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Corton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J., Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K. D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P., Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Rutter, S., Rutherford, K. M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J. E., Craig, A., Newbold, C., and Barrell, B. G. (2002). Sequence of plasmodium falciparum chromosomes 1, 3-9 and 13. *Nature*, 419(6906):527–531. (Cited on page 58.)

[Handunnetti et al., 1987] Handunnetti, S. M., Mendis, K. N., and David, P. H. (1987). Antigenic variation of cloned plasmodium fragile in its natural host macaca sinica. sequential appearance of successive variant antigenic types. *J Exp Med*, 165(5):1269–1283. (Cited on page 29.)

[Hansen et al., 2013] Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., and Kjems, J. (2013). Natural rna circles function as efficient microrna sponges. *Nature*, 495(7441):384–388. (Cited on page 55.)

[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 correlates with antisense rna levels. *Nucleic Acids Res*, 25(21):4296–4300. (Cited on page 54.)

[Helmby et al., 1993] Helmby, H., Cavelier, L., Pettersson, U., and Wahlgren, M. (1993). Rosetting plasmodium falciparum-infected erythrocytes express unique strain-specific antigens on their surface. *Infect Immun*, 61(1):284–288. (Cited on page 41.)

[Henikoff and Smith, 2015] Henikoff, S. and Smith, M. M. (2015). Histone variants and epigenetics. *Cold Spring Harb Perspect Biol*, 7(1):a019364. (Cited on page 20.)

[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., and Hernández-Rivas, R. (2019). Clipped histone h3 is integrated into nucleosomes of dna replication genes in the human malaria parasite plasmodium falciparum. *EMBO Rep*, 20(4). (Cited on page 19.)

[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. (Cited on page 40.)

[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 noncoding rnas. *Nature*, 456(7218):130–134. (Cited on page 51.)

[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 Microbiol*, 14(9):1391–1401. (Cited on pages 14 and 20.)

[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. *Mol Microbiol*, 87(5):1061–1073. (Cited on page 20.)

[Holoch and Moazed, 2015] Holoch, D. and Moazed, D. (2015). Rna-mediated epigenetic regulation of gene expression. *Nat Rev Genet*, 16(2):71–84. (Cited on page 49.)

[Hommel et al., 1983] Hommel, M., David, P. H., and Oligino, L. D. (1983). Surface alterations of erythrocytes in plasmodium falciparum malaria. antigenic variation, antigenic diversity, and the role of the spleen. *J Exp Med*, 157(4):1137–1148. (Cited on page 29.)

[Hongay et al., 2006] Hongay, C. F., Grisafi, P. L., Galitski, T., and Fink, G. R. (2006). Antisense transcription controls cell fate in saccharomyces cerevisiae. *Cell*, 127(4):735–745. (Cited on page 51.)

[Horrocks et al., 1998] Horrocks, P., Dechering, K., and Lanzer, M. (1998). Control of gene expression in plasmodium falciparum. *Mol Biochem Parasitol*, 95(2):171–181. (Cited on page 24.)

[Horrocks et al., 2004] Horrocks, P., Pinches, R., Christodoulou, Z., Kyes, S. A., and Newbold, C. I. (2004). Variable var transition rates underlie antigenic variation in malaria. *Proc Natl Acad Sci U S A*, 101(30):11129–11134. (Cited on page 35.)

[Horrocks et al., 2009] Horrocks, P., Wong, E., Russell, K., and Emes, R. D. (2009). Control of gene expression in plasmodium falciparum - ten years on. *Mol Biochem Parasitol*, 164(1):9–25. (Cited on page 24.)

[Howitt et al., 2009] Howitt, C. A., Wilinski, D., Llinás, M., Templeton, T. J., Dzikowski, R., and Deitsch, K. W. (2009). Clonally variant gene families in plasmodium falciparum share a common activation factor. *Mol Microbiol*, 73(6):1171–1185. (Cited on pages 37, 41, and 167.)

[Hozumi and Tonegawa, 1976] Hozumi, N. and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci U S A*, 73(10):3628–3632. (Cited on page 47.)

[Huang et al., 2013] Huang, X., Yuan, T., Tschannen, M., Sun, Z., Jacob, H., Du, M., Liang, M., Dittmar, R. L., Liu, Y., Liang, M., Kohli, M., Thibodeau, S. N., Boardman, L., and Wang, L. (2013). Characterization of human plasma-derived exosomal rnas by deep sequencing. *BMC Genomics*, 14:319. (Cited on page 55.)

[Hughes et al., 2007] Hughes, K., Wand, M., Foulston, L., Young, R., Harley, K., Terry, S., Ersfeld, K., and Rudenko, G. (2007). A novel iswi is involved in vsg expression site down-regulation in african trypanosomes. *EMBO J*, 26(9):2400–2410. (Cited on page 45.)

[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. (Cited on page 58.)

[Iengar and Joshi, 2009] Iengar, P. and Joshi, N. V. (2009). Identification of putative regulatory motifs in the upstream regions of co-expressed functional groups of genes in plasmodium falciparum. *BMC Genomics*, 10:18. (Cited on page 22.)

[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 Microbiol*, 10:1999–2011. (Cited on page 18.)

[Jain and Vale, 2017] Jain, A. and Vale, R. D. (2017). Rna phase transitions in repeat expansion disorders. *Nature*, 546(7657):243–247. (Cited on page 53.)

[Jensen et al., 2004] Jensen, A. T., Magistrado, P., Sharp, S., Joergensen, L., Lavstsen, T., Chiucchiuini, A., Salanti, A., Vestergaard, L. S., Lusingu, J. P., Hermsen, R., Sauerwein, R., Christensen, J., Nielsen, M. A., Hviid, L., Sutherland, C., Staalsoe, T., and Theander, T. G. (2004). Plasmodium falciparum associated with severe childhood malaria preferentially expresses pfemp1 encoded by group a var genes. *J Exp Med*, 199(9):1179–1190. (Cited on page 31.)

[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. (Cited on pages 18 and 36.)

[Joannin et al., 2008] Joannin, N., Abhiman, S., Sonnhammer, E. L., and Wahlgren, M. (2008). Sub-grouping and sub-functionalization of the rifin multi-copy protein family. *BMC Genomics*, 9:19. (Cited on page 41.)

[Jones, 2012] Jones, P. A. (2012). Functions of dna methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, 13(7):484–492. (Cited on page 20.)

[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 Microbiol*, 13(9):573–587. (Cited on pages 7 and 8.)

[Kassem et al., 2014] Kassem, A., Pays, E., and Vanhamme, L. (2014). Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in trypanosoma brucei. *Proc Natl Acad Sci U S A*, 111(24):8943–8948. (Cited on page 45.)

[Kaviratne et al., 2002] Kaviratne, M., Khan, S. M., Jarra, W., and R, P. P. (2002). Small variant stevor antigen is uniquely located within maurer's clefts in plasmodium falciparum-infected red blood cells. *Eukaryotic Cell*, Vol. 1, No. 6. (Cited on pages 39 and 42.)

[Keeley and Soldati, 2004] Keeley, A. and Soldati, D. (2004). The glideosome: a molecular machine powering motility and host-cell invasion by apicomplexa. *Trends Cell Biol*, 14(10):528–532. (Cited on page 9.)

[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. (Cited on page 14.)

[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. (Cited on page 17.)

[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. (Cited on page 51.)

[Khan et al., 2011] Khan, M., Vaes, E., and Mombaerts, P. (2011). Regulation of the probability of mouse odorant receptor gene choice. *Cell*, 147(4):907–921. (Cited on page 48.)

[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. (Cited on page 42.)

[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. (Cited on page 42.)

[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. (Cited on page 52.)

[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. (Cited on page 52.)

[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. (Cited on page 169.)

[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. (Cited on page 26.)

[Komaki-Yasuda et al., 2013] Komaki-Yasuda, K., Okuwaki, M., Nagata, K., Kawazu, S., and Kano, S. (2013). Identification of a novel and unique transcription factor in the intraerythrocytic stage of plasmodium falciparum. *PLoS One*, 8(9):e74701. (Cited on page 25.)

[Kornberg, 1974] Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and dna. *Science*, 184(4139):868–871. (Cited on page 15.)

[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. (Cited on page 12.)

[Kouzarides, 2007] Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4):693–705. (Cited on pages 17 and 18.)

[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. (Cited on pages 33 and 34.)

[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 Microbiol*, 50(4):1215–1227. (Cited on page 35.)

[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. (Cited on page 25.)

[Kyes et al., 2007] Kyes, S., Christodoulou, Z., Pinches, R., Kriek, N., Horrocks, P., and Newbold, C. (2007). Plasmodium falciparum var gene expression is developmentally controlled at the level of rna polymerase ii-mediated transcription initiation. *Mol Microbiol*, 63(4):1237–1247. (Cited on pages 22, 35, and 36.)

[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. (Cited on pages 35 and 41.)

[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. (Cited on page 41.)

[LaMonte et al., 2012] LaMonte, G., Philip, N., Reardon, J., Lacsina, J. R., Majoros, W., Chapman, L., Thornburg, C. D., Telen, M. J., Ohler, U., Nicchitta, C. V., Haystead, T., and Chi, J. T. (2012). Translocation of sickle cell erythrocyte microrrnas into plasmodium falciparum inhibits parasite translation and contributes to malaria resistance. *Cell Host Microbe*, 12(2):187–199. (Cited on page 57.)

[Langdon et al., 2018] Langdon, E. M., Qiu, Y., Ghanbari Niaki, A., McLaughlin, G., Weidmann, C. A., Gerbich, T. M., Smith, J. A., Crutchley, J. M., Termini, C. M., Weeks, K. M., Myong, S., and Gladfelter, A. S. (2018). mrna structure determines specificity of a polyq-driven phase separation. *Science*, 360(6391):922–927. (Cited on page 53.)

[Lanzer et al., 1992a] Lanzer, M., de Bruin, D., and Ravetch, J. (1992a). Transcription mapping of a 100 kb locus of plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates. *EMBO J*, 11(5):1949–1955. (Cited on page 13.)

[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. (Cited on page 24.)

[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. (Cited on page 29.)

[Latos et al., 2012] Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., Aumayr, K., Pasierbek, P., and Barlow, D. P. (2012). Airn transcriptional overlap, but not its lncrna products, induces imprinted igf2r silencing. *Science*, 338(6113):1469–1472. (Cited on page 51.)

[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. (Cited on pages 40 and 43.)

[Lavazec et al., 2007] Lavazec, C., Sanyal, S., and Templeton, T. J. (2007). Expression switching in the stevor and pfmc-2tm superfamilies in plasmodium falciparum. *Mol Microbiol*, 64(6):1621–1634. (Cited on pages 42 and 43.)

[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. (Cited on page 33.)

[Le Roch et al., 2003] Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., and Winzeler, E. A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, 301(5639):1503–1508. (Cited on page 12.)

[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. (Cited on page 16.)

[Lee, 2009] Lee, J. T. (2009). Lessons from x-chromosome inactivation: long ncrna as guides and tethers to the epigenome. *Genes Dev*, 23(16):1831–1842. (Cited on page 47.)

[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. (Cited on page 10.)

[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. (Cited on page 29.)

[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. (Cited on pages 51 and 54.)

[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 Microbiol*, 90(3):519–537. (Cited on pages 22 and 39.)

[Li et al., 2008] Li, F., Sonbuchner, L., Kyes, S. A., Epp, C., and Deitsch, 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. (Cited on pages 14 and 56.)

[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. (Cited on page 53.)

[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. (Cited on page 46.)

[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. (Cited on page 48.)

[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 Microbiol*, 66(6):1296–1305. (Cited on pages 18, 35, 36, 37, and 169.)

[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. (Cited on pages 20, 22, 35, 36, 37, and 169.)

[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. (Cited on page 6.)

[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. (Cited on pages 54 and 170.)

[Lyon, 1986] Lyon, M. F. (1986). X chromosomes and dosage compensation. *Nature*, 320(6060):313. (Cited on page 47.)

[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. (Cited on page 48.)

[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. (Cited on pages 55 and 56.)

[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. (Cited on page 24.)

[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. (Cited on page 49.)

[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. (Cited on page 48.)

[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. (Cited on page 53.)

[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 Microbiol*, 7(5):341–354. (Cited on pages 10 and 11.)

[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. (Cited on page 22.)

[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. (Cited on page 20.)

[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. (Cited on page 57.)

[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. (Cited on page 53.)

[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. (Cited on page 51.)

[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. (Cited on page 48.)

[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. (Cited on page 40.)

[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. (Cited on page 51.)

[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 Microbiol*, 62(1):72–83. (Cited on page 37.)

[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. (Cited on page 12.)

[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. (Cited on page 29.)

[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. (Cited on page 55.)

[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. (Cited on page 22.)

[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. (Cited on page 18.)

[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. (Cited on pages 14, 16, and 18.)

[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 Microbiol*, 78(4):883–902. (Cited on pages 18 and 36.)

[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. (Cited on pages 7, 31, and 32.)

[Miller et al., 1994] Miller, L. H., Good, M. F., and Milon, G. (1994). Malaria pathogenesis. *Science*, 264(5167):1878–1883. (Cited on pages 29 and 30.)

[Miller and Su, 2011] Miller, L. H. and Su, X. (2011). Artemisinin: discovery from the chinese herbal garden. *Cell*, 146(6):855–858. (Cited on page 5.)

[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. (Cited on page 36.)

[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. (Cited on pages 55 and 58.)

[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. (Cited on page 26.)

[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. (Cited on page 10.)

[Munroe and Lazar, 1991] Munroe, S. H. and Lazar, M. A. (1991). Inhibition of c-erba mrna splicing by a naturally occurring antisense rna. *J Biol Chem*, 266(33):22083–22086. (Cited on page 54.)

[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. (Cited on page 45.)

[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 Microbiol*, 17(8):1205–1216. (Cited on page 179.)

[Navarro and Gull, 2001] Navarro, M. and Gull, K. (2001). A pol i transcriptional body associated with vsg mono-allelic expression in trypanosoma brucei. *Nature*, 414(6865):759–763. (Cited on page 44.)

[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. (Cited on page 56.)

[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. (Cited on page 27.)

[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. (Cited on page 39.)

[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. (Cited on pages 54 and 170.)

[Olins and Olins, 1974] Olins, A. L. and Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science*, 183(4122):330–332. (Cited on page 15.)

[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. (Cited on page 24.)

[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. (Cited on page 54.)

[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. (Cited on page 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. (Cited on pages 33, 40, 58, and 59.)

[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 Microbiol*, 3(6):687–697. (Cited on pages 29, 40, 59, 60, and 168.)

[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 Microbiol*, 76(1):12–24. (Cited on page 58.)

[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. (Cited on page 5.)

[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. (Cited on pages 46 and 51.)

[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. (Cited on pages 58 and 60.)

[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. (Cited on page 55.)

[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. (Cited on page 18.)

[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. (Cited on page 41.)

[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 family show distinct subcellular localization and developmental expression patterns. *Mol Biochem Parasitol*, 156(1):51–61. (Cited on page 42.)

[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. (Cited on page 37.)

[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. (Cited on page 37.)

[Poinar, 2005] Poinar, G. (2005). Plasmodium dominicana n. sp. (plasmodiidae: Haemosporida) from tertiary dominican amber. *Syst Parasitol*, 61(1):47–52. (Cited on page 3.)

[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. (Cited on page 171.)

[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. (Cited on page 21.)

[Pollack et al., 1991] Pollack, Y., Kogan, N., and Golenser, J. (1991). Plasmodium falciparum: evidence for a dna methylation pattern. *Exp Parasitol*, 72(4):339–344. (Cited on page 21.)

[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. (Cited on page 49.)

[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. (Cited on page 21.)

[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. (Cited on page 17.)

[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. (Cited on page 16.)

[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. (Cited on pages 18 and 36.)

[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. (Cited on pages 46, 49, 51, and 53.)

[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. (Cited on pages 54 and 170.)

[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. (Cited on page 55.)

[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. (Cited on page 57.)

[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. (Cited on pages 22, 36, 37, 167, and 169.)

[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. (Cited on page 20.)

[Raulet et al., 1985] Raulet, D. H., Garman, R. D., Saito, H., and Tonegawa, S. (1985). Developmental regulation of t-cell receptor gene expression. *Nature*, 314(6006):103–107. (Cited on page 47.)

[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. (Cited on page 39.)

[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. (Cited on page 57.)

[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 tran-

scription termination and dual strand transcription of sirna loci in trypanosoma brucei. *PLoS Genet*, 12(1):e1005758. (Cited on page 45.)

[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. (Cited on page 24.)

[Rinn and Chang, 2012] Rinn, J. L. and Chang, H. Y. (2012). Genome regulation by long noncoding rnas. *Annu Rev Biochem*, 81:145–166. (Cited on page 50.)

[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. (Cited on page 39.)

[Rowley and Corces, 2018] Rowley, M. J. and Corces, V. G. (2018). Organizational principles of 3d genome architecture. *Nat Rev Genet*, 19(12):789–800. (Cited on page 22.)

[Royo and Cavaillé, 2008] Royo, H. and Cavaillé, J. (2008). Non-coding rnas in imprinted gene clusters. *Biol Cell*, 100(3):149–166. (Cited on page 46.)

[RTSS Clinical Trials Partnership, 2015] RTSS Clinical Trials Partnership, R. (2015). Efficacy and safety of rts,s/as01 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. (Cited on page 6.)

[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. (Cited on pages 26 and 167.)

[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. (Cited on page 24.)

[Sachs and Malaney, 2002] Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680–685. (Cited on page 11.)

[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. (Cited on pages 39 and 41.)

[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. (Cited on page 19.)

[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. (Cited on page 55.)

[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. (Cited on page 43.)

[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. (Cited on page 19.)

[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. (Cited on page 18.)

[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. (Cited on page 53.)

[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. (Cited on pages 14 and 15.)

[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. (Cited on pages 11, 30, and 44.)

[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. (Cited on pages 31, 34, and 40.)

[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. (Cited on pages 23 and 25.)

[Schieck et al., 2007] Schieck, E., Pfahler, J. M., Sanchez, C. P., and Lanzer, M. (2007). Nuclear run-on analysis of var gene expression in plasmodium falciparum. *Mol Biochem Parasitol*, 153(2):207–212. (Cited on pages 35 and 36.)

[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. (Cited on page 36.)

[Schoeftner and Blasco, 2008] Schoeftner, S. and Blasco, M. A. (2008). Developmentally regulated transcription of mammalian telomeres by dna-dependent rna polymerase ii. *Nat Cell Biol*, 10(2):228–236. (Cited on page 56.)

[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. (Cited on page 45.)

[Schwartz, 2016] Schwartz, S. (2016). Cracking the epitranscriptome. *RNA*, 22(2):169–174. (Cited on page 27.)

[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. (Cited on page 54.)

[Selsing, 2006] Selsing, E. (2006). Ig class switching: targeting the recombinational mechanism. *Curr Opin Immunol*, 18(3):249–254. (Cited on page 55.)

[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. (Cited on page 48.)

[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. (Cited on page 52.)

[Sharma and Shukla, 2017] Sharma, L. and Shukla, G. (2017). Placental malaria: A new insight into the pathophysiology. *Front Med (Lausanne)*, 4:117. (Cited on page 11.)

[Shevtsov and Dundr, 2011] Shevtsov, S. P. and Dundr, M. (2011). Nucleation of nuclear bodies by rna. *Nat Cell Biol*, 13(2):167–173. (Cited on page 53.)

[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. (Cited on page 20.)

[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. (Cited on page 48.)

[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. (Cited on page 20.)

[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. (Cited on pages 55, 56, and 58.)

[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. (Cited on pages 55 and 56.)

[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. (Cited on page 26.)

[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. (Cited on page 55.)

[Sleutels et al., 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. (Cited on page 46.)

[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. (Cited on page 30.)

[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. (Cited on pages 26 and 27.)

[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. (Cited on page 51.)

[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. (Cited on page 44.)

[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 plasmodium falciparum clinical isolates. *Science*, 347(6220):428–431. (Cited on page 5.)

[Struhl and Segal, 2013] Struhl, K. and Segal, E. (2013). Determinants of nucleosome positioning. *Nat Struct Mol Biol*, 20(3):267–273. (Cited on page 16.)

[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. (Cited on page 34.)

[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. (Cited on page 16.)

[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. (Cited on page 54.)

[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. (Cited on page 53.)

[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. (Cited on page 38.)

[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. (Cited on page 26.)

[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. (Cited on page 36.)

[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. (Cited on page 18.)

[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. (Cited on page 18.)

[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. (Cited on page 51.)

[Tu, 2011] Tu, Y. (2011). The discovery of artemisinin (qinghaosu) and gifts from chinese medicine. *Nat Med*, 17(10):1217–1220. (Cited on page 5.)

[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. (Cited on page 5.)

[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. (Cited on page 55.)

[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. (Cited on page 6.)

[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. (Cited on page 26.)

[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. (Cited on page 24.)

[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. (Cited on page 27.)

[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. (Cited on page 27.)

[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. (Cited on pages 14 and 57.)

[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. (Cited on page 14.)

[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. (Cited on pages 18 and 36.)

[Voss et al., 2006] Voss, T. S., Healer, J., Marty, A. J., Duffy, M. F., Thompson, J. K., Beeson, 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. (Cited on pages 37, 38, and 167.)

[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. (Cited on page 24.)

[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. (Cited on page 31.)

[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. (Cited on pages 39 and 42.)

[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. (Cited on page 52.)

[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. (Cited on page 47.)

[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. (Cited on page 55.)

[Wei et al., 2014] Wei, C., Xiao, T., Zhang, P., Wang, Z., Chen, X., Zhang, L., Yao, M., Chen, R., and Wang, H. (2014). Deep profiling of the novel intermediate-size noncoding rnas in intraerythrocytic plasmodium falciparum. *PLoS One*, 9(4):e92946. (Cited on page 58.)

[Wei et al., 2015] Wei, G., Zhao, Y., Zhang, Q., and 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. (Cited on pages 59, 167, and 170.)

[Weiner et al., 2011] Weiner, A., Dahan-Pasternak, N., Shimoni, E., Shinder, V., von Huth, 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. (Cited on page 23.)

[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. (Cited on page 16.)

[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. (Cited on page 6.)

[WHO, 2018] WHO (2018). World malaria report 2018. (Cited on pages 3, 4, 5, and 11.)

[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. (Cited on page 48.)

[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. (Cited on page 46.)

[Wu et al., 2008] Wu, J., Sieglaff, D. H., Gervin, J., and Xie, X. S. (2008). Discovering regulatory motifs in the plasmodium genome using comparative genomics. *Bioinformatics*, 24(17):1843–1849. (Cited on page 24.)

[Xue et al., 2008] Xue, X., Zhang, Q., Huang, Y., Feng, L., and Pan, W. (2008). No mirna were found in plasmodium and the ones identified in erythrocytes could not be correlated with infection. *Malar J*, 7:47. (Cited on page 57.)

[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. (Cited on page 53.)

[Yang et al., 2011] Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., Dorrestein, 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. (Cited on page 53.)

[Yang and Kuroda, 2007] Yang, P. K. and Kuroda, M. I. (2007). Noncoding rnas and intranuclear positioning in monoallelic gene expression. *Cell*, 128(4):777–786. (Cited on page 169.)

[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. (Cited on page 51.)

[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). (Cited on page 26.)

[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. (Cited on page 12.)

[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. (Cited on page 24.)

[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. (Cited on page 32.)

[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. (Cited on pages 25, 36, and 38.)

[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. (Cited on pages 27, 38, 59, and 169.)

[Zhang and Tycko, 1992] Zhang, Y. and Tycko, B. (1992). Monoallelic expression of the human h19 gene. *Nat Genet*, 1:40–44. (Cited on page 46.)

[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. (Cited on page 5.)

[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. (Cited on page 9.)

[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. (Cited on page 52.)