Genome Engineering and Functional Gene Regulation Tools for the Study of Malaria Parasites

By

Diana Alejandra Falla Castillo

BSc. Microbiology
Universidad de los Andes (2006)

MSc. Microbiology
Universidad de los Andes (2008)

Submitted to the Department of Biology in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Microbiology at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Signature of Author _____ Signature redacted

Diana Alejandra Falla Castillo

Microbiology Graduate Program

Certified by _____ Signature redacted

Jacquin C.Niles

Associate Professor of Biological Engineering

Supervisor

Accepted by _____ Signature redacted

Kristala Jones Prather

Arthur D. Little Professor of Chemical Engineering

Co-Director, Microbiology Graduate Program

Genome Engineering and Functional Gene Regulation Tools for the Study of Malaria Parasites

By

Diana Alejandra Falla Castillo

Submitted in partial fulfillment of the
requirement for the degree of Doctor of
Philosophy in Microbiology

ABSTRACT

Plasmodium falciparum is the causative agent of the most severe form of human malaria, a mosquito-borne disease that remains a major global health problem. The efforts to create new antimalarial drugs and effective vaccines have been significantly hindered by the lack of robust tools for performing functional genetics in *P. falciparum*. The identification and characterization of essential functions for parasite survival are fundamental steps towards the creation of effective antimalarial therapies. In this work, we developed an integrated set of gene editing and functional gene regulation tools that enable the study of essential and non-essential genes in blood stage parasites. We first created a robust and versatile conditional expression system that uses a fusion of endogenous translational regulatory elements and synthetic RNA-protein modules to regulate gene expression in the parasite. Using this system, we achieved tight regulation of expression of reporter and essential antimalarial genes. Next, we created an integrated strategy that utilizes our conditional system together with a CRISPR-Cas9 gene editing system to identify and characterize the function of an essential RNA-Binding protein (RBP). We first determined the essentiality of our target protein using a two-step approach, in which a merodiploid line conditionally expresses an ectopic copy of the RBP and the native gene is disrupted using CRISPR technologies. This approach was next streamlined into a single-step methodology to genetically modify native loci to regulate expression from their promoters. We performed biochemical and biological characterization of this essential protein, and established the role of this RBP in cell cycle progression and parasite schizogony. Finally, to expand the repertoire of *P. falciparum* target loci, we implemented the editing activity of CRISPR-Cpf1, and showed high efficiency in the disruption of non-essential genes and genes located in AT-rich regions. We also integrated the Cpf1 editing activity with our conditional system to achieve conditional regulation of native loci. This work combines genome-engineering technologies and regulatory systems designed to provide a robust platform for the identification and characterization of essential functions in human malarial parasites.

Thesis Supervisor: Jacquin C. Niles
Title: Associate Professor of Biological Engineering

To my dad, my brother and the memory of my mom. To my husband and my daughter.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to all the people who helped and supported me through my PhD studies. First of all, I thank my PhD advisor Jacquin Niles, for his mentorship and guidance through all these years. I was very fortunate for learning everything directly from Jacquin. He made sure I was getting the right scientific training at every single step of my research from the experimental design to the results discussion and the future work. He encouraged me and kept me motivated even when the results were not promising. I am especially grateful to Jacquin for being supportive and understanding during truly difficult times of my personal life. It has been an honor to be his student. I would also like to thank the members of my thesis committee: Leona Samson and Feng Zhang for their advice and continued interest in my projects.

I would also like to thanks so many collaborators I had inside and outside of MIT along the way, especially collaborators at the Zhang Lab in the Broad Institute for their useful input and for kindly providing materials that helped to push our projects forwards. To collaborators at the Sharp Lab in the Koch Institute, especially to Tim Kelly for his advice and for all his help on helping me design new protocols to be implemented on a very challenging organism. To collaborators at the Boyden Lab for sharing protocols and training me on different high-resolution techniques in microscopy and to the people in the MIT-BioMicro center facility and Whitehead-Kech facility for their technical support and good advice with my projects. Outside of MIT I am especially thankful with Karine LeRoch and Maggie Lu at University of California – Riverside for their enormous help not ontly at the technical level but also for their contribution on the data analysis and for knowing how to guide the project to a final interesting and exciting story.

I have had the great fortune to work with excellent people. I thank all the past and current members of the Niles Lab for the help and advice needed to complete this work. I am especially grateful with Jeff Wagner who recruited me to join the lab, for trusting me to work on his project and for giving me the perfect training I needed to have to survive in grad school. I also thank my friends Jim Abshire and Sebastian Nasamu not only for scientifically contributing to my research but also for their friendship and their support during my PhD. I also thank my lab mom Denise Macphail, the fabulous admin in our lab, who has always standed by me even during the most difficult times, who has supported me through hard times and has given me so much love since the day I joined the lab. To the new generation of amazing and talented postdocs in the lab, who were very supportive and helpful during the last year of my PhD; especially to Charisse Pasaje and Kyle Mclean who taught me many of the fundamental principles about the parasite biology and for their contribution during data analysis and conclusions.

Big thanks to what I call my "Micro Family", my 2011-classmates from the Microbiology program. I wouldn't have survived the first two years of grad school without their support and daily encouragement. Many thanks to Rob, Mark, Kat, Leonor and Katie, they were not only my friends but also the best collegues anyone could have asked for. To my MIT friends outside of the lab; Remi, Sara, Ana, Lindsay, the entire

Lourido Lab, especially to Sebastian, Saima and Diego, members from the Bell Lab and I might be missing many names here but many thanks to all of the people who also contributed to my education at a different level, through their supportive words and their meaningful advice.

I want to thank all my friends in Boston who have made these past six and half years a fun journey. To my dearest colombian friends at MIT and at Harvard, with whom I did not only shared fun moments, but also had the opportunity to work with in non-science related project that helped me to develop new leadership skills that will be fruitful in the future. Finally and most importantly, I want to thank my parents and my little brother. Thanks for their neverending support. Everything I am, I owe to my family, and they are my greatest source of inspiration. I especially would like to dedicate my studies and every single achievement of my life to the memory of my mom. She contributed to every single aspect of my life, she was my role model, she taught me the good and the bad of life, she pushed me hard when I needed it, she supported me, she guided me, she was proud of me at every single moment and most importantly she gave me her unconditional love during my entire life. The last year and half since she passed, has been the saddest time of my life but she continues giving me and my family the strength we need to continue with our lives but always having her memory and her heart in each of our daily actions. I also want to thank the family I have always dreamed of having. To my husband Andres for all his love, help and support during all these years. He is not only the smartest person, but also the most considerate, loving, caring and understanding person I know. He is my best friend, the best husband and now I know he will the most amazing father to our daughther who is on her way. I am very fortunate to have Andres as the person who I chose to spend the rest of my life with and I hope he and I can be the role model parents that our parents were for us. I am truly blessed for having Andres as my husband.

TABLE OF CONTENTS

ABSTRACT 3

ACKNOWLEDGMENTS 6

TABLE OF CONTENTS 8

LIST OF FIGURES 11

LIST OF TABLES 13

CHAPTER 1: INTRODUCTION 14

Malaria burden and pathogenesis 15

Biology of the *P. falciparum* life cycle 17

Antimalarials and drug resistance 21

Advances in molecular tools to manipulate *P. falciparum* 25

CRISPR basics and Class II systems 33

Type II CRISPR applications 38

Rationale and Overview 40

REFERENCES 43

CHAPTER 2: SYNTHETIC RNA-PROTEIN MODULES INTEGRATED WITH NATIVE TRANSLATION MECHANISMS TO CONTROL GENE EXPRESSION IN MALARIA PARASITES 54

ABSTRACT 55

INTRODUCTION 56

RESULTS AND DISCUSSION 58

Molecular design principles 58

TetR fusion proteins conditionally regulate translation 59

Enhanced regulation of gene expression in *P. falciparum* 62

Enhanced regulation is independent of 5'UTR context 66

Leaky expression is reduced with the TetR-DOZI system 68

Another native *Plasmodium* protein implicated in translational regulation in diverse organisms also mediates enhanced conditional regulation of translation. 71

Establishing gene essentiality in *Plasmodium falciparum* 72

METHODS 77

REFERENCES 81

SUPPLEMENTARY INFORMATION 85

CHAPTER 3: PROBING THE FUNCTION OF AN ESSENTIAL RNA-BINDING PROTEIN IN *P. FALCIPARUM* 90

ABSTRACT 91

INTRODUCTION 92

RESULTS 98

PfRRM1 shows inducible regulation and essentiality in engineered merodiploid parasite line 98

CRISPR-Cas9 mediated disruption of the PfRRM1 locus reveals its essentiality in blood stage parasites 103

Efficient disruption and conditional regulation of *PfRRM1* at the native promoter reveals a consistent essential phenotype on *P. falciparum* blood stages.9 104

PfRRM1 is essential for schizogony development in *P. falciparum* blood stage parasites 109

DISCUSSION 117

METHODS 122

REFERENCES 132

SUPPLEMENTARY INFORMATION 138

CHAPTER 4: EXPANDING THE COVERAGE OF CRISPR-GENE TARGETS IN *P. FALCIPARUM* USING CPF1 ENDONUCLEASES 144

ABSTRACT 145

INTRODUCTION 146

RESULTS 150

In vivo editing of native *P. falciparum* loci using Cpf1 endonucleases 150

Cpf1 displays efficient editing activity in AT-rich regions 154

Achieving inducible regulation of a native locus using Cpf1 editing activity 157

Cpf1 sgRNAs binding sites show a larger genome-wide coverage in *P. falciparum* than Cas9 sgRNAs 160

DISCUSSION 169
METHODS 173
REFERENCES 180
SUPPLEMENTARY INFORMATION 185
CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS 191
CONCLUSIONS AND FUTURE DIRECTIONS 191
REFERENCES 200

LIST OF FIGURES

Figure 1-1: Life cycle of *P. falciparum* in the human host and the mosquito vector 19

Figure 1-2: Conditional expression systems used in *P. falciparum* 29

Figure 1-3: Mechanistic differences between Class II CRISPR-Cas systems and their main effector proteins, Cas9 and Cpf1 37

Figure 2-1. Achieving inducible gene expression by integrating a synthetic protein-RNA interaction module with native translation control mechanisms 59

Figure 2-2. Various TetR fusion proteins enhance TetR aptamer-mediated translational regulation in both S. cerevisiae and P. falciparum 62

Figure 2-3. TetR-DOZI–aptamer module provides robust control of gene expression across the *P. falciparum* intraerythrocytic developmental cycle 66

Figure 2-4. Multi-dimensional synthetic control of translation achieves significant reduction in leaky expression 70

Figure 2-5. Establishing essentiality of a native parasite gene using TetR-DOZI with a genetically-encoded 10x aptamer array in the 3'UTR 74

Figure 2-S1. Sequence comparison of several protein homologs from various model and non-model organisms that have been implicated in translational regulation 86

Figure 2-S2. Flow cytometry histograms showing parasite developmental stage- and aTc-dependent EYFP expression controlled by TetRDOZI 87

Figure 2-S3. TetR fused to the *P. falciparum* CITH homolog improves the regulatory dynamic over TetR alone 87

Figure 3-1: Sequence similarity and subdomain structure conservation of TgRRM1 and PfRRM1 95

Figure 3-2: Inducible regulation of PfRRM1 on an engineered merodiploid parasite line 100

Figure 3-3: Native PfRRM1 disruption using CRISPR/Cas9 on engineered merodiploid line background 102

Figure 3-4: Development of integrated technology to target and inducibly regulate PfRRM1 at its native locus reveals consistent essential phenotype in *P. falciparum* blood stages 106

Figure 3-5: PfRRM1 is involved in nuclear processes and is essential for schizogony development 113

Figure 3-S1: Site-specific integration of the PfCRT:PfRRM1 expression plasmid into the chromosomal cg6 locus 138

Figure 3-S2: Chromosomal integration of new linear vector with the TetR-Dozi-aptamer regulatory components shows inducible regulation 139

Figure 3-S3: PfRRM1 knockdown parasites show arrested phenotype 72 hours post-invasion after aTc-withdrawal at 5h, 14h and 30h post-invasion 140

Figure 3-S4: PfRRM1 knockdown parasites display an arrested morphology 38 hours post-invasion at late throphozoite stage 141

Figure 3-S5: PfRRM1 knockdown parasites do not show a splicing defect in trophozoite stage 142

Figure 4-1. Cpf1 editing activity on the *PfEBA-175* locus in *P. falciparum* blood stage parasites 151

Figure 4-2: Cpf1 shows editing activity in AT-rich regions and can mediate translational control on targeted genes 157

Figure 4-3. Global sgRNA binding analysis on the *P. falciparum* genome 161

Figure 4-4: Analysis of the Cpf1 and Cas9 sgRNA binding sites distribution within the genes and across all genes on the *P. falciparum* genome 166

Figure 4-S1: Efficiency of Cpf1-editing activity on the PfEBA-175 locus 185

Figure 4-S2: Comparable results between Cpf1 and Cas9-editing activities to target *PfEBA-175* during transient transfections 186

Figure 4-S3: Comparison of different off-target scoring algorithms on filter stringency 187

Figure 4-S4: Distribution of off-target scores in total sgRNA database on total sgRNA binding sites analysis on *P.falciparum* 188

LIST OF TABLES

Table 2-S1. List of plasmids used in this study 88

Table 2-S2. List of primers used in this study 89

Table 3-S1: List of Oligos for plasmid construction 143

Table 3-S2: List of oligos used to detect integration events 143

Table 4-S1: Reference of the percentiles of On-Target and Off-Target Scores to select Cpf1 and Cas9 sgRNAs from the general database. 189

Table 4-S2: List of oligos used for plasmid construction 189

Table 4-S3: List of oligos used to detect integration events 190

CHAPTER 1

INTRODUCTION

Malaria burden and pathogenesis

Malaria is a mosquito-borne disease that affects over 212 million persons around the world, especially pregnant women and children under five years old. Nearly half of the global population is at risk for acquiring malaria, most of them inhabitants of 91 countries where active transmission has been reported (WHO, 2016). The parasite *Plasmodium falciparum* is the causative agent of the most severe form of human malaria, which can be accompanied by complications including cerebral malaria, severe malarial anemia and placental malaria (Rasti et al., 2004). Cerebral malaria and severe malarial anemia can also lead to lactic acidosis (an excessive lactate production and decrease in the pH of the tissues), causing respiratory distress and reducing oxygen delivery to tissues. This complication is often fatal in children (English et al., 1997). Although *P. falciparum* is the most clinically relevant agent of malaria, other *Plasmodium* species including *P. vivax*, *P. ovale*, *P. malariae*, *P. knowlesi* cause disease in humans, with *P. vivax* being the most broadly distributed species.

The principal clinical manifestations of malaria are a consequence of the parasite's invasion of human red blood cells. *P. falciparum*-infected red blood cells (iRBCs) become cytoadherent, stiffer, adopt a rounded form, and experience changes in the permeability of their membrane (Dondorp et al., 2000; Tiburcio et al., 2012). Changes in iRBCs are recognized as an abnormality by the host and those cells can be cleared in the spleen. However, iRBCs can bind to uninfected red blood cells (uRBCs) facilitating formation of rosettes and can also bind to the host endothelium causing obstruction of the of the microvasculature, and reduced blood flow to different tissues (Rasti et al., 2004). Consequently, oxygen levels are reduced in tissues leading to respiratory distress, anemia, and coma.

P. falciparum is an ancient parasite that has adapted to its human host for millions of years (Joy et al., 2003). Among these adaptations is the parasite's ability to avoid

clearance by the host immune system. Parasite ligands, such as the *P. falciparum* erythrocyte membrane 1 (PfEMP1), the repetitive interspersed families of polypeptides (RIFINs) and subtelomeric variable open reading frame (STEVOR), mediate adherence to host vascular endothelium and interactions with uRBCs and have been implicated in severe malaria (Wahlgren et al., 2017). These parasite ligands have a role in virulence and immunogenicity as they encode for antigenic variants expressed at the surface of iRBCs. For example, PfEMP1 is a membrane protein family that is inserted into knob-like projections on the surface of iRBCs that mediate adhesion (Kilejian, 1979). This protein is encoded by a group of 60 *var* genes (Gardner et al., 2002), but only one *var* gene is expressed at a time through a switching expression mechanism that the parasite uses to promote antigenic variation (Chen et al., 1998). The parasite's antigenic variation inhibits an effective immune response, enabling parasites to establish infection, replicate and induce pathogenesis. Nonetheless, after several rounds of parasite antigen exposure, the host can develop non-sterilizing clinical immunity due to the generation of a subclonal population of antibodies that can cross-react against the parasite receptors and confer protection (Chattopadhyay et al., 2003). Despite this response, the host is often not able reach the desired level of immunity, particularly in the case of infected children, who consequently may not survive the parasite infection.

Efforts to eradicate malaria through different control programs have been focused on clinical treatment using front-line antimalarial therapies. Despite the continous efforts to reduce the global burden of malaria over the last decade, the continuous emergence of parasite drug resistance, a lack of effective vaccines, and increases in insecticide resistance are responsible factors for the remaining 438,000 deaths per year (Achieng et al., 2017). New strategies and the development of more effective drugs and vaccines are urgently needed for immediate implementation in malaria eradication and control programs.

Biology of the *P. falciparum* life cycle

P. falciparum is transmitted by mosquito females of the genus *Anopheles.* This parasite has a complex life cycle, which involves the invasion of different cell types that require specific parasite-stage morphologies and adaptations (Fig 1-1). The cycle begins when sporozoites are released into the human host. Tens to hundreds of motile sporozoites are injected via mosquito bite into the blood stream to reach the liver. At this stage, there is minimal acquired immune response against sporozoites, probably due to the small number of parasites that are injected (Bijker et al., 2015). Once sporozoites reach the liver, they invade hepatocytes to initiate liver stage infection. Proteins such as SPECT (sporozoite microneme protein essential for traversal) (Ishino et al., 2004), SPECT2 (perforlin-like protein 1, PLP1) (Risco-Castillo et al., 2015), CelTOS (Cell traversal protein for ookinetes and sporozoites) (Bhanot et al., 2005) and GEST (gamete egress and sporozoite traversal protein) play roles in cell traversal, a process that parasites use to glide through the blood vessels and reach the liver. Once parasites are in the liver tissue, the circumsporozoite protein (CSP) and type I thrombospondin repeat protein (TSR) mediate an active process of parasite hepatocyte invasion (Herrera et al., 2015). In hepatocytes, infection progreses over a period of days until vesicles filled with about 40,000 merozoites are released into the blood stream (Sturm et al., 2006).

Once free merozoites are released into bloodstream, parasites rapidly invade erythrocytes starting the erythrocytic cycle. This invasion process consists of three steps: pre-invasion, active invasion and echinocytosis, which, all together, take place in about two minutes (Weiss et al., 2015). During the pre-invasion phase, the Merozoite Surface Protein-1 (MSP1) mediates binding to the erythrocyte (Lin et al., 2016). Pre-invasion also involves host cell cytoskeleton deformation, which is activated by the parasite's actomyosin (Weiss et al., 2015). At this step, proteins that belong to a group of erythrocyte binding-like proteins (EBLs) and *P. falciparum* reticulocyte-binding protein homologs (PfRhs) are required to interact with host cell receptors such as glycophorins A, B, C and the complement receptor (CR1). Although the function of EBLs and PfRHs appear to be

redundant, some members of these families play additional essential roles in subsequent steps of invasion, such as the mediation of the release of organelles called rhoptries (EBA-175, an EBL-family protein) and the activation of Ca^{2+} signaling pathways (PfRHs). (Cowman et al., 2016; Gao et al., 2013). After host cell deformation and parasite attachment, PfRh5 protein interacts with the host receptor Basigin. This interaction leads to parasite reorientation to the erythrocyte membrane and Ca^{2+} influx into the host cell, processes which are are essential for parasite invasion. Following parasite attachment and reorientation, proteins such as AMA1 and RON play a role in the irreversible attachment between the parasite and the erythrocyte. The RON complex binds to AMA1 on the surface of the merozoite (Besteiro et al., 2011), while rhoptry proteins are released into the host cell, mediating parasitophorous vacuole formation, and the merozoite is pushed in using its own actomyosin motor (Riglar et al., 2011). After invasion is completed, the parasite enters the parasitophorous vacuole and the membranes seal behind it. During the final step of echinocytosis, the iRBC morphology changes during several minutes, probably due to Ca^{2+} influx, forming spicules on the cell membrane and finally the iRBC resumes to its normal shape (Gao et al., 2013).

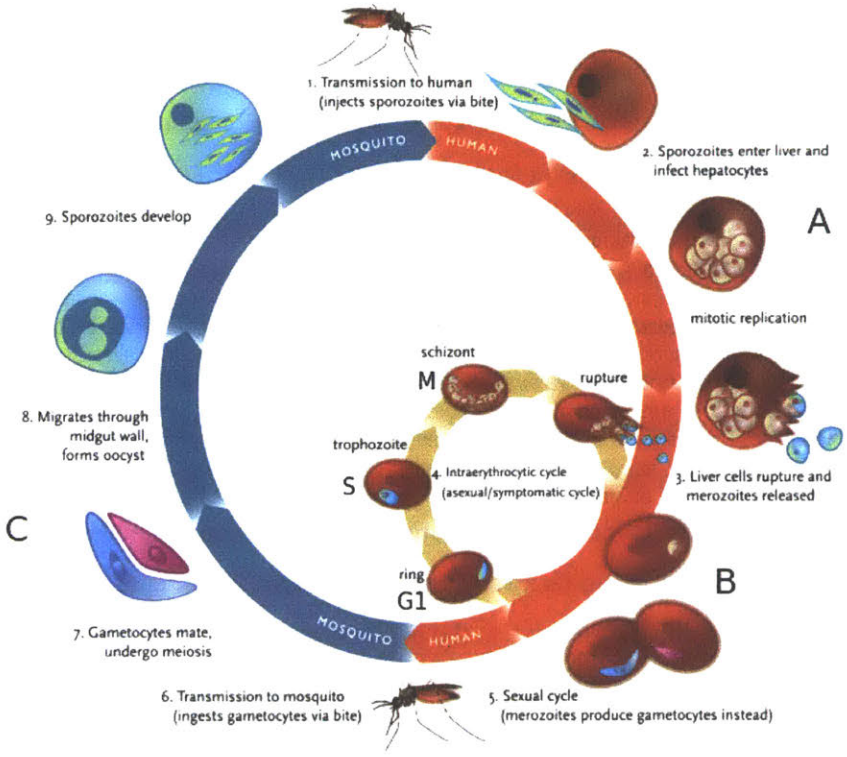


Figure 1-1: Life cycle of *P. falciparum* in the human host and the mosquito vector
After inoculation of sporozoites into the human blood stream, **A)** the parasites invade liver hepatocytes and replicate via mitosis, producing thousands of merozoites. Merozoites are released into the bloodstream to start the intraerythrocytic developmental cycle (IDC). **B)** The merozoites invade the erythrocytes forming the initial stage of the IDC, a ring during a "G1" phase. The parasite progresses through the cell cycle, goes through DNA synthesis at trophozoite stage (S phase) and then replicates through multiple rounds of mitotic divisions (M) during schizogony, causing a rupture of the iRBC. A portion of the merozoites start a new invasion cycle when colonizing another uRBC. Another portion of the merozoites released commit to sexual development producing gametes to start the sexual cycle. **C)** The mosquito ingests infectious gametocytes during a blood meal and the cycle in the mosquito vector begins. Gametocyte fertilization occurs within the mosquito and new sporozoites are produced and migrate to the salivary glands of the mosquito to start a new life cycle. Figure adapted from (Klein, 2013).

After erythrocytic invasion, the intraerythrocytic cycle begins. The most common clinical symptoms of malaria are associated with this stage of the life cycle, and include fever, anemia and neurological manifestations (Schofield, 2007). When a single merozoite invades an RBC and establishes an infection, it is called a ring. This ring contains a single copy of the genome (1N), and progresses through its cell cycle to become a trophozoite stage parasite by ~20 hours post-invasion. Onset of S phase occurs around 24 hours, and

is accompanied by DNA synthesis and an increase in the uptake of purine nucleobase precursors (Arnot et al., 2011). DNA synthesis and multiple rounds of nuclear division (within the same nuclear membrane) ensue, and lead to formation of multiple merozoites in subsequent phases. This process occurs over a ~6-10 hour window until the parasite progresses into schizogony. During schizogony, between 33-37 hours post-invasion, nuclei and organelles assemble during the M phase of the cell cycle. Late schizonts are produced around 38-42 hours after invasion. At this time, schizonts undergo nuclear fragmentation and the host cell membrane ruptures to release 16-22 merozoites into the blood stream to start a new invasion cycle (Bannister et al., 2000). The whole intraerythrocytic cycle takes 45-48 hours, although this time might vary depending on the parasite strain. The intraerythrocytic developmental cycle (IDC) constitutes one generation of the asexual stage of *P. falciparum*.

During the intraerythrocytic cycle, a portion of the parasites become committed to sexual differentiation and form female and male gametocytes. The molecular mechanisms that underlie this commitment are not well understood; however, it appears that during schizogony, merozoite daughter cells commit to either gametocyte development or remaining in the asexual form. Environmental stimuli, such as high parasitemia or exposure to drugs, seem to be sensed by the parasites, leading them to favor sexual commitment (Cowman et al., 2016). Epigenetic control and transcriptional regulation executed by the gametogenesis "master regulator" AP2-G, contribute to sexual differentiation (Kafsack et al., 2014). Gametocyte maturation takes about 11 days, during which parasites are sequestered in the bone marrow to avoid clearance by the immune system (Joice et al., 2014). When a mosquito bites an infected host with circulating mature gametocytes, the parasites return to the mosquito host, which allows them to begin their sexual cycle and the production of sporozoites to infect a new human host, thus starting the life cycle anew (Meibalan and Marti, 2017).

Antimalarials and drug resistance

Over the last century, chloroquine and Sulfadoxine-pyrimethamine were standard antimalarials included in malaria control programs. However, during the last few decades, they have proven ineffective due to widespread resistance (Trape et al., 1998). As a response to this, potent Artemisinin combination therapies (ACTs) were introduced as the WHO-recommended frontline treatment (WHO, 2016). Despite the effectiveness of ACT in reducing malaria burden, resistance in different countries of South Asia has been recently reported (Dondorp et al., 2010).

Existing antimalarials target metabolic pathways related to the asexual stages of the parasites. For example, quinolines such as chloroquine and amodiaquine interfere with hemozoin formation, a detoxification product made by crystallization of heme released during degradation of hemoglobin by the parasite to make available amino acids for protein synthesis. Chloroquine binds to heme to inhibit hemozoin to induce parasite toxicity, presumably via damage to cellular components (e.g. lipid cell membranes) caused by an increase in labile heme/chloroquine-heme complexes (Combrinck et al., 2013). Chloroquine-resistant parasites display mutations in the cloroquine resistance transporter (PfCRT) that localizes to the membrane of the digestive vacuole (Pascual et al., 2013). Other antimalarials that target metabolic processes at the asexual stage inhibit folate synthesis, the mitochondrial transport chain for pyrimidine biosynthesis, and protein synthesis in the apicoplast, a relic plastid organelle (Blasco et al., 2017).

Artemisinin (ART) is a sesquiterpene trioxane lactone derived from the leaves of the chinese wormwood plant *Artemisia annua* (Achieng et al., 2017). ACTs contain ART derivatives such as artemether, artesunate or dihydroartemisinin that act together by reducing parasite biomass by up to 10,000 fold every 48 hours. Dihydroartemisinin has a short half-life (1 hour) and multiple doses per day for few consecutive days might be required to eliminate the total number of parasites in an adult. Increasing the dose of ART as monotherapy could lead to the emergence of resistant parasites and thus the

combination with other drug components with better pharmacokinetics properties was adviced to stabilize the ART derives half-life to last longer until parasite clearance is achieved (Dondorp et al., 2009; Dondorp et al., 2010). Synthetic ART derivatives have shown activity on all erythrocytic parasite stages, including gametocytes (White et al., 2014). ART or ART-derivatives are presumably activated in the parasitophorous vacuole by the reduction of Fe^{3+} released from Heme from the human red blood cell hemoglobin, the major source of iron for parasites, to Fe^{2+}. ART activation generates carbon-centered radicals that can react with some of the parasite biomolecules, including heme that can trigger oxidative damage in the intracellular compartments, resulting in cell damage (Tilley et al., 2016). Studies of the mechanism of action of ART in asexual stages suggest that ART is highly active in the presence of Fe^{2+} heme species produced by hemoglobin degradation, especially at the trophozoite stage, during which hemoglobin degradation is highly active (Blasco et al., 2017). In contrast to other antimalarial drugs, ART is also active against ring forms, during which the hemoglobin degradation process is thought to begin (Xie et al., 2016). The activity of ART and ART-derivatives on trophozoite and ring stages indicates that this therapy could be used for both uncomplicated and severe malaria (White et al., 2014).

Emergence of ACT resistance was detected for the first time more than 10 years ago in Cambodia where resistant parasites showed a longer rate of clearance in patients treated with ACT (Noedl et al., 2008). Interestingly, clinical, epidemiological and *in vitro* studies showed resistance to ACT during ring stage, but not in trophozoites (Ashley et al., 2014; Miotto et al., 2015; Witkowski et al., 2013). Genome-wide studies showed a single point mutation in the Kelch-like protein K13 that is associated with resistant parasite isolates (Ariey et al., 2014). Moreover, gene-editing studies in cultured parasites demonstrated that introduction of mutations on PfK13 on wild type (and sensitive) parasites conferred a resistant phenotype to ART (Straimer et al., 2015). Compared to Asian isolates, the frequency of K13-mutations in Africa is very low and resistant parasites associated with the K13 mutation have not yet been established, perhaps due to acquired immunity and polyclonal infections that dominate over drug resistant parasites in the African population

(Menard et al., 2016). ART resistance associated with the PfK13 point mutation has spread on different genetic backgrounds. Furthermore, additional variants associated with different mutations have shown different degrees of ART resistance and different fitness costs (Takala-Harrison et al., 2015).

In Asia, ACT was replaced by artesunate-mefloquine therapy, however Asian parasites developed resistance to the artesunate-mefloquine therapy via expression of multiple copies of the *pfmdr1* gene, PfMDR1 is an ATP-binding cassette transporter located at the membrane of the parasite digestive vacuole that transports nutrients from the parasite's cytoplasm to the vacuole (Rohrbach et al., 2006). Overexpression of this gene and sequence variations have now been associated with mefloquine resistance (Uhlemann et al., 2007; Veiga et al., 2016). Consequently, a combination of dihydroartemisinin with piperaquine is currently the first line antimalarial combination used in Cambodia. Piperaquine is similar to chloroquine except for the presence of two chloroquine-like 4-aminoquinoline parts with a central linker (Blasco et al., 2017). Piperaquine is usually effective on chloroquine-resistant parasites.

Interestingly, a number of antimalarial drugs have targeted hemoglobin catabolism, which is both an essential pathway in the parasite for protein synthesis and at the same time a potentially toxic process for the parasite. Unfortunately, the consistent targeting of this key pathway appears to have led to the development of cross-resistance between different classes of drugs. The extensive use of drugs, such as chloroquine, has favored the selection of *pfcrt* mutant alleles that have spread globally and affected the efficiency of other antimalarial drugs. Mutations associated with PfCRT induce H^+-dependent efflux of the drug out of the digestive vacuole, which prevents chloroquine from binding to heme (Lehane et al., 2008). This effect can also inhibit the action of other drugs that target the same pathway, making parasites resistant to other drugs that target the same pathway (Ecker et al., 2012). Similarly, new mutations and variants observed in mefloquine-resistant parasites have been shown to inhibit the action of other antimalarials

and increase the level of resistance to chloroquine (Veiga et al., 2016). Recent genome-wide studies have produced evidence of novel mutations in the *pfcrt* and *pfmdr1* alleles of resistant parasites in Asia and Africa, suggesting that ACT is selecting for new multidrug-resistant parasites (Malaria, 2016), which might ultimately require replacement of ACT with new and more effective drugs.

Current efforts to identify new antimalarial drugs that target other essential metabolic pathways in the parasites are underway. High-throughput screening of chemical compounds from diverse libraries is a new approach that pharmaceutical companies and academic institutions have adopted to identify new compounds with antimalarial activity. Some promising compounds for which the molecular target has been identified include: Imidazopyrazines, DDD107498, Spiroindolones and Imidazolopiperazine. Imidazopyrazines targets a phosphatidylinositol 4-kinase type III beta (PI4KIIIβ), an enzyme that converts phosphatidylinositol (PI) to phosphatidylinositol 4-phosphate (PI4P). PI4P is essential for membrane integrity (McNamara et al., 2013). DDD107498 is a synthetic derivative of the 2,6-disubstituted quinoline-4-carboxamide and it targets the translation elongation factor 2 (PfeEF2), inhibiting protein synthesis (Baragana et al., 2015). Spiroindolones (e.g. NITD6090 target the P-type cation transporter ATPase4 (PfATP4). PfATP4 is a membrane transporter that regulates Na^+ homeostasis (Rottmann et al., 2010). GNF179 is an analog of imidazolopiperazine, a compound that showed activity against *P. falciparum* that potentially targets the cyclic amine resistance locus (PfCARL). PfCARL is a protein localized on the cis-Golgi apparatus involved in protein sorting and membrane trafficking. The mechanism of resistance to imidazolopiperazine was associated to this particular gene (Wu et al., 2011).

In spite of the progress on the identification of new compounds with antimalarial activity, the exact mechanism of action and potential for resistance to these compounds is not known. The lack of robust molecular tools to manipulate *P. falciparum* continues to be a challenge for forward genetic approaches. The validation of the discovered compounds

with antimalarial activity and the dissection of the biological aspects of their molecular targets are necessary to introduce a new effective therapeutic drug.

Advances in molecular tools to manipulate *P. falciparum*

The *P. falciparum* genome was sequenced in 2002 from a clone of the 3D7 strain (Gardner et al., 2002). The parasite genome is stored in 14 chromosomes and has a size of 22.8 Mb. The overall AT-composition is 80.6% and increases to 90% in intergenic regions. About 5,300 genes were predicted to encode for protein products, 54% of which were predicted to contain at least one intron. The mean genes length is about 2.3 kb, which is unexpectedly long compared to other organisms with a similar genome size, such as *S. cerevisiae* (Gardner et al., 2002). In addition, a mitochondrial genome of 6 kb was identified and a circular 35 kb plastid genome, encoding 30 proteins that correspond to the apicoplast. About 60% of the coding sequences do not show sequence homology to genes in other organisms, which has hampered the assignment of their putative functions. In addition, subtelomeric regions are observed in the chromosomes, which is a conserved feature in all *Plasmodium spp*. These regions encode for antigenic variants associated with the erythrocyte surface, such as the *var* genes (Rasti et al., 2004).

Recent advances in the study of the transcriptome and proteome of this parasite have helped to reveal information on the expression profile of specific genes at each stage of the asexual cycle. This information has provided a platform to interpret putative gene functions and has helped to push forward reverse genetics approaches in *P. falciparum* to investigate the molecular functions of gene products and their contribution to the parasite's infection biology. A more complete understanding of the parasite biology would facilitate identification of new targets for therapeutics and would increase our knowledge about drug resistance mechanisms. Approximately 50% of the genes in *P. falciparum* have no known function, and it is possible that a good fraction plays essential roles in the parasite. Hence, establishing more robust and versatile tools for inducible

gene regulation and gene-editing is necessary to decipher the function of the complete set of *P. falciparum* genes.

The following are the most relevant advances in the development of molecular tools in *P. falciparum* during the last few years, and their advantages and disadvantages.

DNA transfection of *in vitro* cultured asexual stages

P. falciparum can be transfected using plasmid and other DNA and electroporation. However, not much less progress has been in delivering other macromolecules, such as RNA and protein. Transfections are performed by direct electroporation of uninfected RBCs with plasmid DNA (Wu et al., 1995), and are relatively inefficient (Hasenkamp et al., 2012). Drug resistance genes (e.g. human dihydrofolate reductase (*DHFR*), Blastidicidin S deaminase (BSD), Yeast dihydroorotate dehydrogenase (yDHODH), neomycin phosphotransferase II from transposon Tn 5 (geneticin-G418)) can be used to select for stably transformed lines. The low transformation efficiency contributes to making this process relatively time consuming and subject to increased failures. In addition, the AT-richness of the parasite's genome poses a challenge in the construction of circular plasmid vectors because unwanted recombination events can happen between low complexity regions during cloning procedures. Upon transfection, DNA plasmids replicate as episomes, eventually forming large, episomally-maintained concatamers (O'Donnell et al., 2002). Plasmid segregation is uneven, which leads to heterogeneous expression of the plasmid content. There is also paucity of available drug markers and validated promoters for gene expression, and global transcription start and stop sites are largely unknown.

It is also possible to introduce, replace or mutate a specific gene. For this purpose, a homologous template must be introduced in the plasmid to drive chromosomal integration via recombination. Unfortunately, this process is inefficient and the parasites

have to be cycled on and off drug selection for several months to select for stable single-crossover events and eliminate episomes (Crabb et al., 2004). Fortunately, novel gene-editing techniques have accelerated this process significantly as we will discuss later on in this section. Spontaneous double-crossover recombination is an alternative to achieve gene modification. However, this occurs at low frequency, and requires the use negative selection markers, which do not work consistently in *P. falciparum* (de Koning-Ward et al., 2015). In *P. berguei*, linear DNA is used to increase the efficiency of single and double-crossover recombination events (Janse et al., 2006). Linearized plasmids may increase the recombination efficiency, although this has not been thoroughly addressed in *P. falciparum.*

In spite of the advances during the last 20 years to create and optimize genetic techniques, only about 500 out of the 5300 genes, have been successfully disrupted to study their function (de Koning-Ward et al., 2015). It is presumed that a large fraction of the genes that remain to be studied accomplish essential functions. Due to the haploid nature of the genome during blood stages, disruption and characterization of the function of essential genes requires the development of conditional expression systems that enable the controlled production of a particular gene product, as complete disruption of these genes result in inviable parasites.

Conditional expression systems

In some eukaryotic systems–including the parasitic trypanosomes responsible for causing African sleeping sickness–RNA interference has provided an effective and generalizable strategy for studying gene function globally (Shi et al., 2000). However, the absence of a native RNA interference machinery in *P. falciparum* (Baum et al., 2009) precludes using this approach, for characterizing gene function in the parasite. To overcome the limitations in targeting potentially essential genes, a few conditional gene expression systems have been developed (Fig 1-2). Despite their usefulness for initial gene function characterization efforts, these systems are limited in their ability to unrestrictedly regulate

any desired gene. Here, we summarize some of the most relevant conditional systems that were developed in *P. falciparum* (Fig 1-2).

The first transcriptional-level regulatory system described for *P. falciparum* was reported more than 10 years ago (Meissner et al., 2005). This strategy adopted a tetracycline-responsive transcriptional activator previously used in *Toxoplasma gondii* to regulate the transcriptional activity of the target gene (Fig 1-2A) (Meissner et al., 2005). The principal limitations of this system include high heterogeneity in the expression and an unpredictable loss of regulatory behavior (Meissner et al., 2005). Another conditional system that has been used to regulate expression at the transcriptional level in *P. falciparum* is the Tet-OFF system, which employs a transcriptional transactivator domain (TRAD). In this system, a TRAD-minimal promoter-TetO sequence module is installed at the 5'UTR of the gene of interest. TRAD, driven by expression from the targeted gene promoter binds to the TetO sequence in the absence of aTc and the targeted gene is transcribed. In the presence of aTc, TRAD cannot bind and transcription shuts off (Fig 1-2B) (Gilson et al., 2008; Meissner et al., 2005). This system was shown to be effective on episomes, but not for knockdown of native genes, probably due to inefficiency of the transactivators. Some efforts to improve this tool have been made, however the characterized *T. gondii* or mammalian transactivators do not work in *P. falciparum*. Thus, further studies might be required to identify endogenous transactivators.

Different approaches have been developed to abrogate expression of target protein by creating conditional deletions at chromosomal locations. In this approach a Cre recombinase or a site-specific FLP enzyme are expressed that have the ability to irreversibly excise genetic elements through the recombination of *loxP* (for Cre enzymes) or *frt* (for FLP enzymes) sequence sites flanking the target region (O'Neill et al., 2011). Both of these systems have shown spontaneous excision when the recombinases are constitutively expressed, so an inducible version in which the expression of Cre is regulated was developed using a diCRE system (a split version of CRE that can be dimerized using rapamycin) (Jullien et al., 2007) (Fig 1-2C). This system was applied to identify the function of the essential invasion protein AMA1 (Yap et al., 2014) and to

regulate the expression of SERA5, which is a protease secreted into the parasitophorous vacuole and possibly involved in merozoite egress (Yap et al., 2014). Both studies showed leaky expression or a low reduction in the amount of protein. The regulatory outcome, and thus the phenotype being investigated, of this system is likely dependent on where in the genome the recombination sites are installed and on the amount of recombinase expressed.

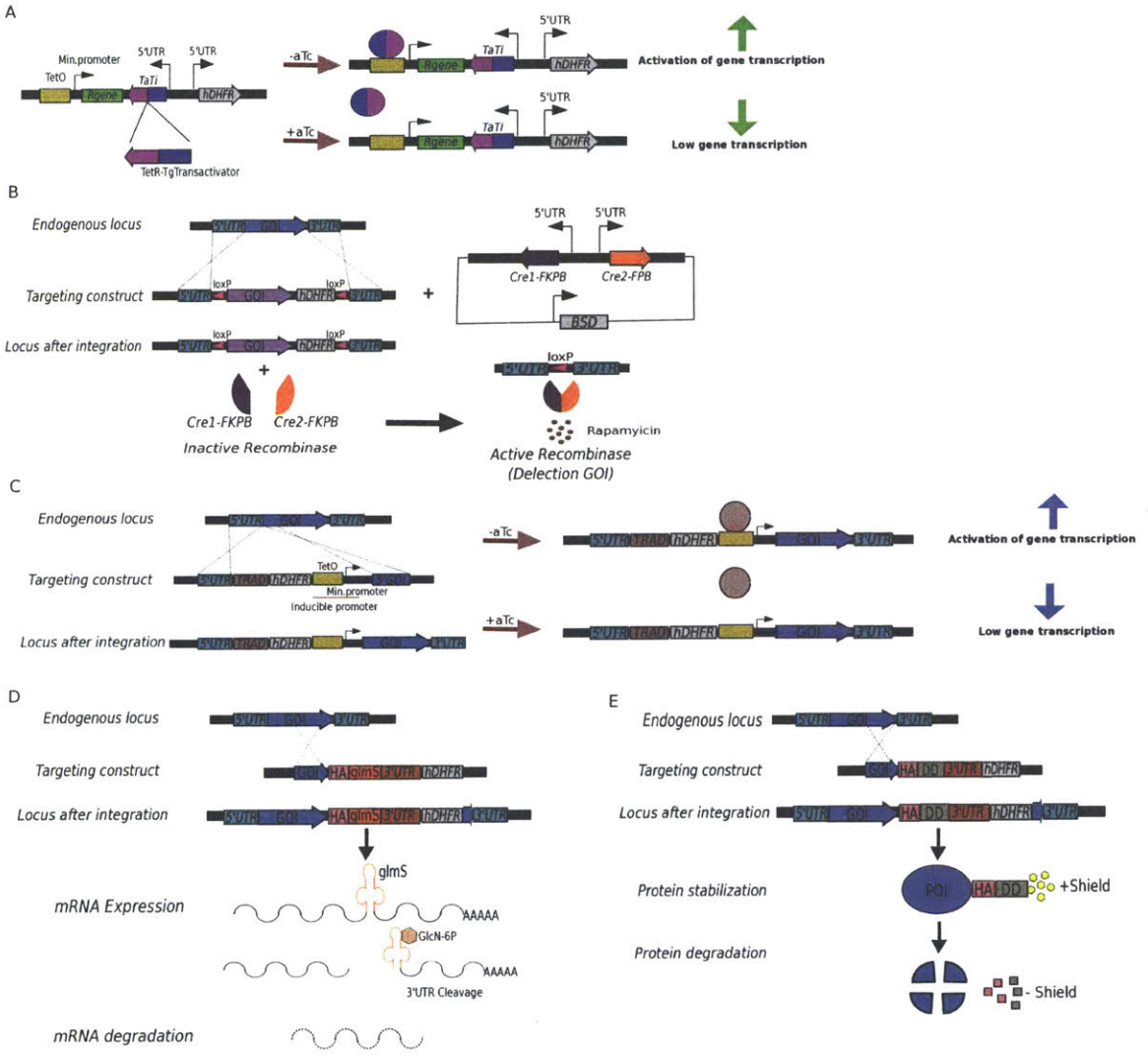


Figure 1-2: Conditional expression systems used in *P. falciparum*
A) Conditional regulation using a Tetracycline-analogue transgene expression system in *P. falciparum*. A *T. gondii* transactivating domain is N-terminal fused to a Tet Repressor (TaTi) to regulate the expression of a reporter gene (Rgene) from an aTc-regulatable PfCam minimal promoter. The construct includes an

hDHFR selective marker. Regulation is achieved on an episomal stable parasite population when in the absence of aTc, the transactivator TaTi interacts with the Tet Operator (TetO) and transcription of the reporter gene activated. In contrast, in the presence of aTc TaTi cannot interact with the operator sequence and the expression of the reporter gene is low. **B)** A conditional gene deletion system using an inducible Cre Recombinase (diCre)-lox recombination. A targeting construct includes a recoded version of the gene of interest (GOI-Purple) and a selective marker (hDHFR) flanked by loxP sites. The construct can be integrated into the genome using homologous regions to the UTRs of the gene of interest (GOI-Blue). A second plasmid is built with two separate cassettes driving the expression two separate inactive polypetides of Cre (Cre1 and Cre2). Each of them is fused to different rapamyicin proteins: FKPB (FK506-binding protein) and FPB (fibronectin binding protein). When rapamyicin is added the Cre polyptides form a functional and active Cre recombinase and this results in the excision of the GOI-hDHFR region between the loxP sites. **C)** Transcriptional knockdown using an aTc-inducible system and a transcriptional transactivation domain (TRAD). The targeting construct includes the TRAD gene and a human DHFR (hDHFR) selective marker. In addition, an inducible promoter that consists in a *P. falciparum* minimal promoter and a Tet Operator (TetO) sequence. The construct can be integrated via HR using homologous regions flanking TRAD-hDHFR and the inducible promoter. The left homologous region targets the 5'UTR of the gene of interest (GOI) and the right homologous region targets the 5' end of the GOI coding region. When the construct is integrated, the GOI native promoter regulates the expression of the TRAD-hDHFR cassette. In the other hand, the inducible promoter controls the GOI. In the absence of aTc, TRAD interacts with TetO and the GOI transcription increases. In contrast, in the absence of aTc cannot longer bind to TetO and the GOI transcription goes down. **D)** Post-transcriptional knockdown regulation using self-cleaving rybozymes. A targeting construct includes the glmS ribozyme downstream from a HA-tag to be installed after the stop codon of the gene of interest (GOI). The construct is integrated into the genome through single crossover with a homologous region that targets the 3' of the GOI so that the glmS is installed between the stop codon and the 3'UTR of the GOI. Addition of GlcN-6P activates the ribozyme and after self-cleavage the 3'UTR is detached from the mRNA, which leads to mRNA degradation. **E)** Post-transcriptional knockdown using the protein-RNA synthetic module TetR-DOZi-aptamer. The targeting construct includes the TetR-aptamers array (10X) downstream from a HA-tag to confirm protein regulation. In addition, it contains an expression cassette with the TetR/Dozi fusion protein, a renilla luciferase reporter gene and blasticidin (BSD) drug marker separated. The three genes are separated by a self-cleavable T2A peptide for polycistronic translation. The construct is integrated into the parasite's genome using a single homologous region that targets the 3' of the GOI, via single crossover. The aptamers and the regulatory machinery are installed downstream from the last amino acid of the protein. After transcription, the secondary structures of the TetR-aptamers interact tightly with the synthetic TetR-DOZI module and translation goes off. However, in the presence of aTc TetR looses the affinity for the aptamer and translation goes on. f) Post-translational knockdown using destabilization domains. A targeting construct includes the destabilization domain (DD) that can be fused to either the N or C terminus of the GOI. This DD is a mutant version of the rapamycin-FKB12 domain. The construct is integrated into the genome through single crossover with a homologous region that targets the 3' of the GOI. When the gene is translated, Shield interacts with the DD and stabilizes the protein of interest (POI). However, if Shield is removed degradation of the protein is induced.

A conditional system that uses self-cleaving ribozymes to post-transcriptionally regulate mRNA in *P. falciparum* has also been reported. This system requires installation of a *glmS* ribozyme downstream of the stop codon of the gene of interest, which undergoes self-cleaveage when the inducer glucosamine-6-phosphate (GlcN-6P) is added. This leads to 3'UTR removal and mRNA degradation (Fig 1-2D) (Prommana et al., 2013). The *glmS* system was observed to be efficient for knocking down reporter genes and for

targeting the translocon component PTEX150 (Elsworth et al., 2014). However, GlcN-6P can be toxic to parasites and like the previous inducible system described, the *glmS* inducible tool shows leaky expression and only 50-60% of knockdown is achieved.

Regulation at the post-translational level has also been attempted to inducibly regulate gene expression in *P. falciparum*. Conditional degradation of proteins can be achieved when the target protein is fused to a protein destabilization domain based on either FK506-binding protein (DD system) or the *Escherichia coli* DHFR destabilizing (DDD system). These domains are unstable and are readily degraded by the proteasome. This degradation can be reversed by the small molecule Shield 1 (for DD fusions) or folate analogue trimethoprim (for DDD fusions) (Armstrong and Goldberg, 2007; Muralidharan et al., 2011) (Fig 1-2E). These systems are reversible, and *P. falciparum* gene expression can be knocked down using these approaches (Azevedo et al., 2013; Beck et al., 2014; Dvorin et al., 2010; Farrell et al., 2012; Muralidharan et al., 2011; Muralidharan et al., 2012; Russo et al., 2009). Though promising, these tools have certain disadvantages. Leaky expression is observed, and application is mostly limited to cytosolic and nuclear proteins (non-exported or organellar) and/or proteins that tolerate N/C-terminal peptide fusions. In addition, some of these systems require use of specific parasite strain backgrounds. For instance, for the DDD system in the parental strain must be resistant (either naturally or through transgenic expression of the human DHFR gene) because trimethoprim is naturally toxic to the parasites.

Genome editing tools

The inefficiency of transfections in *P. falciparum* causes a delay in the selection of stable transfectant parasites. This time can be prolonged when selecting for epsiomal resistance makers, due to uneven segregation. To tackle this problem, advances to promote plasmid integration within the parasite genome have been implemented in *P. falciparum*. The *Bxb1* integrase system was developed for stable heterologous gene expression or complementation of gene knockouts because it involves modification of the same

dedicated locus every time. The *Bxb1* integrase catalyzes recombination between an *attB* site engineered into the parasite's genome and an *attP* site on a plasmid containing the transgene of interest (Nkrumah et al., 2006). This system requires the generation of a parental strain to introduce the *attB* site at a neutral locus (e.g. the *cg6* locus is commonly used for blood stage parasites).

Zinc Finger Nucleases (ZFN) have also been used in *P. falciparum*. These enzymes induce double strand-breaks in the genome that are preferentially repaired via homologous recombination due the absence of the canonical non-homologous end joining (NHEJ) pathway in the parasite's genome. The ZFN system was used to target the PfCRT locus, and resulted in rapid integration (2 weeks) of a homologous template (Straimer et al., 2012). Recently this gene-editing system was used to introduce mutations in the kelch protein K13 to study the mechanism of artemisinin resistance (Straimer et al., 2015).

A high-throughput genetic screening system based on a piggyBac transposable element has also been adapted to perform genetics in *P. falciparum*. This system requires the co-transfection of a plasmid that contains a selectable marker and inverted terminal repeats of the *piggyBac* element and a second plasmid that expresses the piggyBac transposase. The *piggyBac* element is excised and randomly inserted in the parasite's genome at TTAA target sites, which occur at high frequency in the *P. falciparum* genome (Balu et al., 2005). The outcome of this transfection is a parasite library that should contain at least one insertion per genome per clone. The disrupted locus can be identified via PCR using the inverted repeats as landing region. The efficiency of this screening is extremely low, with 1-10 clones recovered, targeting only non-essential genes.

Finally, CRISPR/Cas9 technologies have been a truly revolutionary approach to perform functional genetics in *P. falciparum*. The heterologous system expresses a *Streptococcus pyogenes* Cas9 endonuclease and a targeting single guide RNA (sgRNA). Cas9 induces a

single strand break that is repaired via homologous recombination when a template is provided. Two different systems have been proposed to transcribe the sgRNA in the parasite; one of them utilizes the U6 promoter (Ghorbal et al., 2014), which has previously not been used as an orthogonal system in *P. falciparum*, and the T7 promoter, which has been proven to robustly deliver non-coding RNA in the parasite (Wagner et al., 2014). Both systems have been used to target or introduce point mutations in native genes in blood stages of the parasite, showing high efficiencies (Bansal et al., 2016; Bryant et al., 2017; Cobb et al., 2017; Crawford et al., 2017; Ghorbal et al., 2014; Ng et al., 2016; Sidik et al., 2016; Sonoiki et al., 2017; Vanaerschot et al., 2017; Wagner et al., 2014). Compared to ZFN, the CRISPR/Cas9 technology represents a simpler approach, is more cost-effective, and more efficient. This system will likely replace other gene-editing techniques in *P. falciparum* in the near future, and we will expand upon the topic in the following section.

CRISPR basics and Class II systems

The diversity of bacteriophages, bacteria, and archaeal organisms and their constant interaction has led to the development of defense mechanisms within prokaryotes that include restriction-modification system and the recently characterized CRISPR system (Barrangou and Horvath, 2017). Unlike other defense mechanisms, CRISPR (clustered regularly interspaced short palindromic repeat) is an adaptive form of immunity. In bacteria, the system encompasses 3 stages: acquisition of foreign DNA, CRISPR (crRNA) biogenesis, and target interference (Sorek et al., 2013). During the acquisition stage, the foreign DNA is recognized, captured and incorporated as a new "spacer" next to the repeat sequences in the CRISPR locus. Small nucleotide sequences adjacent to the spacers (also known as protospacers) called PAMs (protospacer adjacent motif play an important role in the DNA acquisition stage. CRISPR loci belong to a highly diverse family, in which each CRISPR array contains short (usually conserved) repeat sequences and hypervariable spacers that defend the host from different phages and plasmids. The repeat sequences are usually 20-50 bp long, but the sequences and the length vary in

different CRISPR loci (Kunin et al., 2007). In addition to the repeat sequence and the spacer, there is also a diversity of *cas* genes adjacent to the CRISPR arrays. The different CRISPR/Cas systems can be placed into two distinct classes based on the effector protein, *cas1* and *cas2*, that are universally conserved (Haft et al., 2005). At the second stage the RNA Polymerase transcribes the CRISPR locus producing a pre-crRNA that further is processed into active crRNAs by the endonucleases activity. Lastly, during the third stage the crRNA forms a complex with the effector protein that can then recognize regions of incoming foreign DNA with high specificity by complementarity. This complex cleaves the foreign DNA acquiring and maintaining immunity against that particular phage or plasmid (Brouns et al., 2008). In contrast, if there is not complete pairing between the sequences or the PAM is absent, the phage can infect the host leading to lysis (Bhaya et al., 2011).

Selective pressure has driven the evolution of CRISPR-Cas systems and has diversified the structure and function of their *cas* genes into two main classes, Class I and Class II. The main difference between the two classes is the presence of a single (Class II) versus multiple subunit complexes (Class I). These two classes are divided into six types, depending on the Cas protein features and further subdivided into 32 sub-types (Koonin et al., 2017). The first characterized Class II CRISPR system includes the Cas9 effector protein that binds and cleaves double-strand DNA when its guide crRNA identifies a region of complementarity (Gasiunas et al., 2012; Jinek et al., 2012). Since the discovery and characterization of Cas9, it has been utilized extensively as a very powerful genome-editing technology in a wide diversity of organisms (Hsu et al., 2014). The successful outcome of a variety of Cas9 applications prompted exploration of other Class II systems. In the last two years, emergent Class II types and subtypes of CRISPR systems have been identified and characterized. Some of those systems have even shown gene-editing activity in eukaryote systems. These systems include: Type V with the Cas12 (also known as Cpf1 and C2c1) effectors, which have the ability to target DNA (Zetsche et al., 2015); and Type IV with Cas13 (also known as C2c2) effectors, which have the ability to target RNA (Shmakov et al., 2015).

In addition to differences in the organization of subunit complexes, there are also mechanistic differences between Type I and Type II CRISPR systems. The Cas9 ortholog from *Streptococcous pyogenes* (SpCas9) is the most studied effector protein in Type II systems. Cas9 is a multidomain protein that cleaves both strands of the DNA using the activity of the HNH and RuvC nuclease domains (Gasiunas et al., 2012). Cas9 is guided to target DNA when it finds a complementary strand to its crRNA. Cas9 also requires a *trans-acting* crRNA (tracrRNA) for its editing activity (Deltcheva et al., 2011). In bacteria, transcription of the CRISPR locus array generates a long precursor RNA (pre-crRNA). When the repetitive regions of this transcript interact by base pairing with the tracrRNA, they form an RNA duplex that is recognized by a host RNaseIII to process them into mature RNAs. The mature RNAs correspond to the sgRNAs or spacers that did not base pair with the tracrRNA; both tracRNA and sgRNA are needed to activate Cas9 (Fig 1-3) (Deltcheva et al., 2011; Murugan et al., 2017). When Cas9 is activated, it finds its target DNA by first recognizing the PAM site downstream from the non-complementary strand (Sternberg et al., 2014). After PAM recognition, the double stranded DNA target (dsDNA) unwinds to base pair with the incoming crRNA. If there is complementarity between the target DNA and the crRNA, a R-loop is formed upstream from the PAM site, followed by DNA cleavage. Cas9 remains bound to its target DNA until complementarity is confirmed by complete base pairing of the first 8-10 nucleotides at the 5' of the sgRNA and base pairing with the PAM site (Dagdas et al., 2017). Cas9 changes its catalytic conformation with its mobile HNH domain (Anders et al., 2014) and the RuvC domain regulates cleavage (Sternberg et al., 2015).

Type V systems have been classified into five sub-types according to their effector protein (Type V-A-E) (Koonin et al., 2017). These proteins are all variants of the Cas12 effectors, although only Cas12a (Cpf1) and Cas12b (C2c1) have been characterized (Yang et al., 2016; Zetsche et al., 2015). Type V effectors contain a RuvC domain; however, unlike the Type II systems, they lack an HNH domain. It is possible that an uncharacterized nuclease domain (Nuc) substitutes for the HNH domain in these proteins. (Yamano et al., 2016). Aside from the absence of the HNH domain, the structure of the Cas9 and Cas12 proteins are very similar. Both form a negatively

charged crRNA: DNA target heteroduplex and they both have domains for PAM recognition, although their structure and sequence differ.

Although two of the characterized Type V CRISPR systems subtypes (Cas12b and Cas12e), like Cas9, require a tracrRNA for activation, Cas12a (Cpf1) has no such requirement. Instead, Cpf1 needs only a single ~44 nt crRNA for activation (Zetsche et al., 2015). Furthermore, it has been demonstrated that Cpf1 processes its own pre-crRNA (Fonfara et al., 2016). Similar to Cas9, Type V effectors locate and locally unwind their DNA target upon recognition of a PAM site. Type VA and VB recognize a T-rich PAM located upstream of the non-complementary strand of the target (Fig 1-3). An R-loop is also formed when the Cpf1 crRNA interacts with the non-target strand of DNA to prevent reannealing. However, unlike Cas9, it has been observed that Cpf1-R-loop formation is less energetically stable than the Cas9, probably due to instability of the interaction between the effector protein and the non-target strand, in spite of the stable interaction that occurs at the distal region of the PAM site (Swarts et al., 2017; Yang et al., 2016). This might explain why some genome-editing studies have suggested more specific activity and reduced off-target effects during editing with Cpf1 as compared to Cas9 (Kim et al., 2016). Another difference between Cas9 and Cas12 endonucleases is the type of cleavage of the target DNA. Cas9 leaves a blunt end after cleavage proximal to the PAM site. In contrast, both Cas12a (Cpf1) and Cas12b generate a staggered cut at the end far (18 nt) from the PAM site (Yang et al., 2016; Zetsche et al., 2015). However, it remains unclear how Cpf1 cleaves both DNA strands using a single active site.

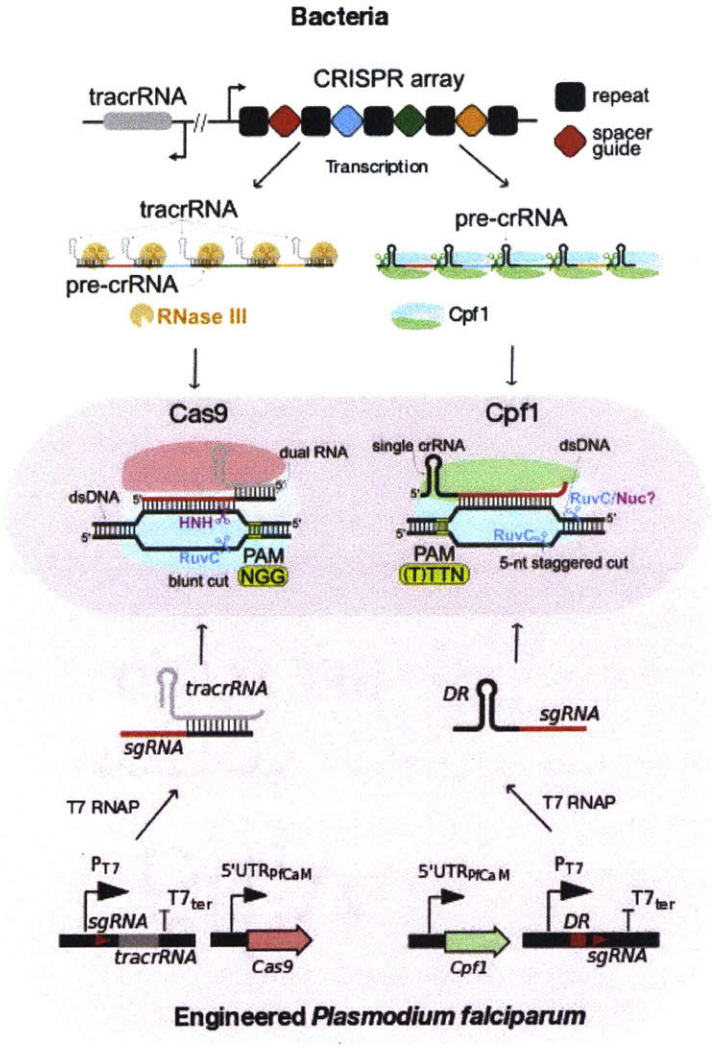


Figure 1-3: Mechanistic differences between Class II CRISPR-Cas systems and their main effector proteins, Cas9 and Cpf1

Top panel: After the pre-crRNA is transcribed from a CRISPR array is processed into mature crRNA by a host RNaseIII when it base pairs with the *trans-acting* crRNA (tracrRNA). The dual RNA complex interacts with Cas9 and guides it to its target DNA when it finds a "NGG" PAM binding site. Cas9 remains bound to the DNA target until both PAM site and spacer gRNA (sgRNA) complementary is confirmed and cleavage is induced. In contrast to Cas9, the pre-crRNA is self-processed by Cpf1. Cpf1 does not need a tracrRNA and the mature and single crRNA interacts with Cpf1. Cpf1 is guided to the target DNA when it recognizes a "TTTN" PAM binding site. Recognition of sgRNA complementarity leads to double strand break. *Bottom panel*: Implementation of Class II CRISPR-Cas9 systems to edit the *P. falciparum* genome. Cas9 and Cpf1 genes are expressed from a plasmid-based system using parasite 5'UTRs. Both the crRNA for Cas9 and Cpf1 are expressed from an orthogonal and validated T7 promoter in the parasite. Figure adapted from (Murugan et al., 2017).

Type II CRISPR applications

Differences in the mechanisms of action of the Class II systems, Cas9 and Cpf1, make them complementary tools for genome-editing purposes, especially in eukaryotes. The CRISPR-Cas9 system has been extensively used to edit mammalian cells, using both *Streptococcus thermophilus* and *Streptococcus pyogenes* Cas9 (Cong et al., 2013). It is possible to synthetically introduce all the different CRISPR-Cas9 components for DNA targeting. Starting from finding optimal strategies to express and deliver Cas9 in the cells to gateway strategies for the delivery and synthesis of one of multiple tracrRNA and sgRNAs. Endogenous PolIII promoters such as the U6 promoter and the H1 promoter have been used to express sgRNAs (Ranganathan et al., 2014). In mammalian cells, double-strand breaks are preferentially repaired via NHEJ (except during S-phase) rather than via homologous recombination (HR). However, recent studies have shown attempts to integrate foreign DNA using both pathways, showing extremely low rates of insertion when using the NHEJ pathway (Bachu et al., 2015). To increase the HR frequency, NHEJ major factors such as KU70 and DNA ligase IV are inhibited (Chu et al., 2015).

Advances in the establishment of CRISPR-Cas as a tool have expanded the repertoire of applications in biological systems. Important progress in this technology has been reported, especially in different mammalian cell lines in the last few years. These efforts have been oriented towards the development of a genome-engineering platform to target loci involved in human diseases for therapeutic purposes. Differences in the mechanisms of action between Cas9 and Cpf1 CRISPR systems make these two technologies complementary to each other. Cpf1 has shorter sgRNAs compared to Cas9, which makes Cpf1-sgRNA synthesis simpler. In addition, the Cpf1 PAM recognition site has a different sequence composition, which increases the range of genomic loci that can be targeted. Differences in the cleavage site and in how the cleavage is performed create another layer of additional applications in Cpf1. In contrast to Cas9, Cpf1 produces a staggered cut similar to the "sticky ends" of traditional restriction enzymes. This feature may allow an increase in the HR events compared to NHEJ and the disruption at the cut

site might contain fewer indels for further modifications using the same cut site (Zetsche et al., 2015). Another feature of Cpf1 is its ability to self-process its crRNA, which facilitates multiplex gene editing applications. This application has been demonstrated in mammalian cells (Zetsche et al., 2017) plants (Wang et al., 2017) and yeast (Verwaal et al., 2017)

The CRISPR-Cas9 system has become a very powerful technology to disrupt genes in the apicomplexa *P. falciparum.* As mentioned before, the genetic manipulation of this parasite is challenging and the availability of robust genetic tools is relatively lacking. As a consequence, the function of the majority of genes encoded by this parasite remains unknown. Efforts to implement and optimize the CRISPR-Cas9 technology have shown high efficiency gene disruption and have contributed to unveiling the function of different genes (Bansal et al., 2016; Bryant et al., 2017; Cobb et al., 2017; Crawford et al., 2017; Ghorbal et al., 2014; Ng et al., 2016; Sidik et al., 2016; Sonoiki et al., 2017; Vanaerschot et al., 2017; Wagner et al., 2014). On the other hand, genome editing using Cpf1 has not yet been reported in this parasite. Intriguingly, the *P falciparum* genome is highly AT-rich, a feature that could favor application of this technology in this organism.

In mammalian cells a recombinant system to express Cas9 has made this alternative the preferred method to efficiently edit loci, as plasmid-based expression systems have not shown the same efficiency. In *P. falciparum*, on the other hand, Cas9 is usually expressed from an expression plasmid since direct transfection of recombinant proteins or mRNA has not been efficient in this parasite. Using this method, high disruption efficiencies have been observed, even within the first intraerythrocytic cycle after transfection. The delivery of the other CRISPR components, such as the tracrRNA and sgRNA are also plasmid mediated. Similar to mammalian systems, the U6 promoter has been implemented (Ghorbal et al., 2014), but the orthogonal and broadly used T7 promoter (Wagner et al., 2014) has also shown strong activity in *P. falciparum* (Fig 1-3). In addition, it is possible to insert foreign DNA at the repair site, as HR is the dominant

mechanism for DNA repair in the absence of NHEJ-pathway proteins. A similar approach could be used to express Cpf1 and demonstrate its editing activity in the parasite (Fig 1-3)

Rationale and Overview

The overarching goal of this thesis is to develop an integrated set of molecular tools that enable the study of gene function in *P. falciparum.* Our work seeks to expand the genetic toolkit that is available for *P. falciparum* by using previous and newly developed technologies of the Niles lab. We focus our efforts on the creation of gene editing and conditional regulation systems that are designed to provide a robust platform for the identification and characterization of essential functions in the parasite.

In Chapter 2, we address the challenge of creating a robust and versatile conditional expression system in *P. falciparum.* We describe the generation of a system that regulates the expression of a native locus by fusing endogenous translational regulatory elements from the parasite to a synthetic RNA-protein module. This system is first validated using reporter genes in the parasite, which show that tight regulation of gene expression throughout the intraerythrocytic cycle is possible. Next, we demonstrate that the system is suitable to achieve stringent regulation of an important essential gene product, believed to be the target inhibited by the recently described spiroindolone antimalarial drug class. Furthermore, we demonstrate that, using this system, we are able achieve regulation of target genes independent of their subcellular localization, biochemical properties, and transcript levels.

Next, in Chapter 3, we seek to create an integrated strategy that uses the conditional system developed in Chapter 2 together with a CRISPR-Cas9 gene editing system to study the function of a putative RNA-binding protein (RBP). We first investigate the

essentiality of this RBP using a two-step approach that involves creation of a merodiploid line. In this approach, the native copy of the RBP is disrupted after a synthetic copy of the gene is conditionally expressed from an ectopic locus using a conditionally-regulated, synthetic promoter. Having proved the efficacy of the two-step approach, we go on to streamline this process by creating a single-step methodology to genetically modify the native locus of the protein and regulate its expression from its native promoter. The successful implementation of these technologies enables us to perform biological characterization of the function of this RBP, which include demonstrating its RNA-binding activity, subcellular localization, and phenotype when the protein is depleted. These experiments allow us to confirm the essentiality of this gene for blood stage parasites and further characterize its role in parasite schizogony and cell cycle progression.

The robustness and efficiency of CRISPR-Cas technologies motivated us to explore other genome-editing technologies that could expand the ability to manipulate the highly AT-rich *P. falciparum*. In particular, we became interested in the novel Cpf1 CRISPR endonuclease described by the Zhang lab, as this enzyme uses an AT-rich binding matrix for identification of its DNA target. We hypothesized that this feature might be useful for targeting *P. falciparum* genes located in extremely AT-rich loci. Therefore, in Chapter 4, we explore the the opportunities for expanding target loci manipulation in *P. falciparum* by validating the activity of various orthologs of the Cpf1 endonuclease. We describe efficient, Cpf1-mediated genome editing in *P. falciparum,* and show that this allows efficient disruption of non-essential genes and genes located in AT-rich regions. In addition, we integrate the Cpf1 editing activity with our conditional system from Chapter 2 to target and regulate native loci in the parasite using a similar approach described for Cas9 in Chapter 3. Finally, we perform bioinformatic analyses to demonstrate that the AT-rich binding matrix of Cpf1 expands the coverage of potential targets compared to Cas9. We conclude that Cpf1 and Cas9 are complementary CRISPR-Cas class II technologies that operate using different target site selection rules and that these differences can broaden genome engineering applications in *P. falciparum*.

Overall, this work establishes several new molecular tools to genetically manipulate blood stage *P. falciparum*. This work describes a genetic toolkit that can be used to assign function to many of the putative genes annotated in the *P. falciparum* genome, including essential and non-essential genes. The different technologies generated in this work are meant to aid in our understanding of basic parasite biology as well as help validate molecular targets with potential relevance for antimalarial drug and vaccine development.

REFERENCES

Achieng, A.O., Rawat, M., Ogutu, B., Guyah, B., Michael Ong'echa, J., D, J.P., and Kempaiah, P. (2017). Antimalarials: Molecular Drug Targets and Mechanism of Action. Curr Top Med Chem *17*, 2114-2128.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., *et al.* (2014). A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature *505*, 50-55.

Armstrong, C.M., and Goldberg, D.E. (2007). An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. Nat Methods *4*, 1007-1009.

Arnot, D.E., Ronander, E., and Bengtsson, D.C. (2011). The progression of the intra-erythrocytic cell cycle of Plasmodium falciparum and the role of the centriolar plaques in asynchronous mitotic division during schizogony. Int J Parasitol *41*, 71-80.

Ashley, E.A., Dhorda, M., Fairhurst, R.M., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J.M., Mao, S., Sam, B., *et al.* (2014). Spread of artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med *371*, 411-423.

Azevedo, M.F., Sanders, P.R., Krejany, E., Nie, C.Q., Fu, P., Bach, L.A., Wunderlich, G., Crabb, B.S., and Gilson, P.R. (2013). Inhibition of Plasmodium falciparum CDPK1 by conditional expression of its J-domain demonstrates a key role in schizont development. Biochem J *452*, 433-441.

Bachu, R., Bergareche, I., and Chasin, L.A. (2015). CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. Biotechnol Bioeng *112*, 2154-2162.

Balu, B., Shoue, D.A., Fraser, M.J., Jr., and Adams, J.H. (2005). High-efficiency transformation of Plasmodium falciparum by the lepidopteran transposable element piggyBac. Proc Natl Acad Sci U S A *102*, 16391-16396.

Bannister, L.H., Hopkins, J.M., Fowler, R.E., Krishna, S., and Mitchell, G.H. (2000). A brief illustrated guide to the ultrastructure of Plasmodium falciparum asexual blood stages. Parasitol Today *16*, 427-433.

Bansal, A., Ojo, K.K., Mu, J., Maly, D.J., Van Voorhis, W.C., and Miller, L.H. (2016). Reduced Activity of Mutant Calcium-Dependent Protein Kinase 1 Is Compensated in Plasmodium falciparum through the Action of Protein Kinase G. MBio *7*.

Baragana, B., Hallyburton, I., Lee, M.C., Norcross, N.R., Grimaldi, R., Otto, T.D., Proto, W.R., Blagborough, A.M., Meister, S., Wirjanata, G., *et al.* (2015). A novel multiple-stage antimalarial agent that inhibits protein synthesis. Nature *522*, 315-320.

Barrangou, R., and Horvath, P. (2017). A decade of discovery: CRISPR functions and applications. Nat Microbiol *2*, 17092.

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*, 3788-3798.

Beck, J.R., Muralidharan, V., Oksman, A., and Goldberg, D.E. (2014). PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. Nature *511*, 592-595.

Besteiro, S., Dubremetz, J.F., and Lebrun, M. (2011). The moving junction of apicomplexan parasites: a key structure for invasion. Cell Microbiol *13*, 797-805.

Bhanot, P., Schauer, K., Coppens, I., and Nussenzweig, V. (2005). A surface phospholipase is involved in the migration of plasmodium sporozoites through cells. J Biol Chem *280*, 6752-6760.

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet *45*, 273-297.

Bijker, E.M., Borrmann, S., Kappe, S.H., Mordmuller, B., Sack, B.K., and Khan, S.M. (2015). Novel approaches to whole sporozoite vaccination against malaria. Vaccine *33*, 7462-7468.

Blasco, B., Leroy, D., and Fidock, D.A. (2017). Antimalarial drug resistance: linking Plasmodium falciparum parasite biology to the clinic. Nat Med *23*, 917-928.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

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*.

Chattopadhyay, R., Sharma, A., Srivastava, V.K., Pati, S.S., Sharma, S.K., Das, B.S., and Chitnis, C.E. (2003). Plasmodium falciparum infection elicits both variant-specific and cross-reactive antibodies against variant surface antigens. Infect Immun *71*, 597-604.

Chen, Q., Fernandez, V., Sundstrom, A., Schlichtherle, M., Datta, S., Hagblom, P., and Wahlgren, M. (1998). Developmental selection of var gene expression in Plasmodium falciparum. Nature *394*, 392-395.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548.

Cobb, D.W., Florentin, A., Fierro, M.A., Krakowiak, M., Moore, J.M., and Muralidharan, V. (2017). The Exported Chaperone PfHsp70x Is Dispensable for the Plasmodium falciparum Intraerythrocytic Life Cycle. mSphere *2*.

Combrinck, J.M., Mabotha, T.E., Ncokazi, K.K., Ambele, M.A., Taylor, D., Smith, P.J., Hoppe, H.C., and Egan, T.J. (2013). Insights into the role of heme in the mechanism of action of antimalarials. ACS Chem Biol *8*, 133-137.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Cowman, A.F., Healer, J., Marapana, D., and Marsh, K. (2016). Malaria: Biology and Disease. Cell *167*, 610-624.

Crabb, B.S., Rug, M., Gilberger, T.W., Thompson, J.K., Triglia, T., Maier, A.G., and Cowman, A.F. (2004). Transfection of the human malaria parasite Plasmodium falciparum. Methods Mol Biol *270*, 263-276.

Crawford, E.D., Quan, J., Horst, J.A., Ebert, D., Wu, W., and DeRisi, J.L. (2017). Plasmid-free CRISPR/Cas9 genome editing in Plasmodium falciparum confirms mutations conferring resistance to the dihydroisoquinolone clinical candidate SJ733. PLoS One *12*, e0178163.

Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A., and Yildiz, A. (2017). A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci Adv *3*, eaao0027.

de Koning-Ward, T.F., Gilson, P.R., and Crabb, B.S. (2015). Advances in molecular genetic systems in malaria. Nat Rev Microbiol *13*, 373-387.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Dondorp, A.M., Kager, P.A., Vreeken, J., and White, N.J. (2000). Abnormal blood flow and red blood cell deformability in severe malaria. Parasitol Today *16*, 228-232.

Dondorp, A.M., Nosten, F., Yi, P., Das, D., Phyo, A.P., Tarning, J., Lwin, K.M., Ariey, F., Hanpithakpong, W., Lee, S.J., *et al.* (2009). Artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med *361*, 455-467.

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*, 272-280.

Dvorin, J.D., Martyn, D.C., Patel, S.D., Grimley, J.S., Collins, C.R., Hopp, C.S., Bright, A.T., Westenberger, S., Winzeler, E., Blackman, M.J., *et al.* (2010). A plant-like kinase in Plasmodium falciparum regulates parasite egress from erythrocytes. Science *328*, 910-912.

Ecker, A., Lehane, A.M., Clain, J., and Fidock, D.A. (2012). PfCRT and its role in antimalarial drug resistance. Trends Parasitol *28*, 504-514.

Elsworth, B., Matthews, K., Nie, C.Q., Kalanon, M., Charnaud, S.C., Sanders, P.R., Chisholm, S.A., Counihan, N.A., Shaw, P.J., Pino, P., *et al.* (2014). PTEX is an essential nexus for protein export in malaria parasites. Nature *511*, 587-591.

English, M., Muambi, B., Mithwani, S., and Marsh, K. (1997). Lactic acidosis and oxygen debt in African children with severe anaemia. QJM *90*, 563-569.

Farrell, A., Thirugnanam, S., Lorestani, A., Dvorin, J.D., Eidell, K.P., Ferguson, D.J., Anderson-White, B.R., Duraisingh, M.T., Marth, G.T., and Gubbels, M.J. (2012). A DOC2 protein identified by mutational profiling is essential for apicomplexan parasite exocytosis. Science *335*, 218-221.

Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature *532*, 517-521.

Gao, X., Gunalan, K., Yap, S.S., and Preiser, P.R. (2013). Triggers of key calcium signals during erythrocyte invasion by Plasmodium falciparum. Nat Commun *4*, 2862.

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.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586.

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.

Gilson, P.R., O'Donnell, R.A., Nebl, T., Sanders, P.R., Wickham, M.E., McElwain, T.F., de Koning-Ward, T.F., and Crabb, B.S. (2008). MSP1(19) miniproteins can serve as targets for invasion inhibitory antibodies in Plasmodium falciparum provided they contain the correct domains for cell surface trafficking. Mol Microbiol *68*, 124-138.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol *1*, e60.

Hasenkamp, S., Russell, K.T., and Horrocks, P. (2012). Comparison of the absolute and relative efficiencies of electroporation-based transfection protocols for Plasmodium falciparum. Malar J *11*, 210.

Herrera, R., Anderson, C., Kumar, K., Molina-Cruz, A., Nguyen, V., Burkhardt, M., Reiter, K., Shimp, R., Jr., Howard, R.F., Srinivasan, P., *et al.* (2015). Reversible Conformational Change in the Plasmodium falciparum Circumsporozoite Protein Masks Its Adhesion Domains. Infect Immun *83*, 3771-3780.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Ishino, T., Yano, K., Chinzei, Y., and Yuda, M. (2004). Cell-passage activity is required for the malarial parasite to cross the liver sinusoidal cell layer. PLoS Biol *2*, E4.

Janse, C.J., Ramesar, J., and Waters, A.P. (2006). High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite Plasmodium berghei. Nat Protoc *1*, 346-356.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Joice, R., Nilsson, S.K., Montgomery, J., Dankwa, S., Egan, E., Morahan, B., Seydel, K.B., Bertuccini, L., Alano, P., Williamson, K.C., *et al.* (2014). Plasmodium falciparum transmission stages accumulate in the human bone marrow. Sci Transl Med *6*, 244re245.

Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., *et al.* (2003). Early origin and recent expansion of Plasmodium falciparum. Science *300*, 318-321.

Jullien, N., Goddard, I., Selmi-Ruby, S., Fina, J.L., Cremer, H., and Herman, J.P. (2007). Conditional transgenesis using Dimerizable Cre (DiCre). PLoS One *2*, e1355.

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., *et al.* (2014). A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature *507*, 248-252.

Kilejian, A. (1979). Characterization of a protein correlated with the production of knob-like protrusions on membranes of erythrocytes infected with Plasmodium falciparum. Proc Natl Acad Sci U S A *76*, 4650-4653.

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H., and Kim, J.S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol *34*, 863-868.

Klein, E.Y. (2013). Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread. Int J Antimicrob Agents *41*, 311-317.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol *37*, 67-78.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol *8*, R61.

Lehane, A.M., Hayward, R., Saliba, K.J., and Kirk, K. (2008). A verapamil-sensitive chloroquine-associated H+ leak from the digestive vacuole in chloroquine-resistant malaria parasites. J Cell Sci *121*, 1624-1632.

Lin, C.S., Uboldi, A.D., Epp, C., Bujard, H., Tsuboi, T., Czabotar, P.E., and Cowman, A.F. (2016). Multiple Plasmodium falciparum Merozoite Surface Protein 1 Complexes Mediate Merozoite Binding to Human Erythrocytes. J Biol Chem *291*, 7703-7715.

Malaria, G.E.N.P.f.C.P. (2016). Genomic epidemiology of artemisinin resistant malaria. Elife *5*.

McNamara, C.W., Lee, M.C., Lim, C.S., Lim, S.H., Roland, J., Simon, O., Yeung, B.K., Chatterjee, A.K., McCormack, S.L., Manary, M.J., *et al.* (2013). Targeting Plasmodium PI(4)K to eliminate malaria. Nature *504*, 248-253.

Meibalan, E., and Marti, M. (2017). Biology of Malaria Transmission. Cold Spring Harb Perspect Med *7*.

Meissner, M., Krejany, E., Gilson, P.R., de Koning-Ward, T.F., Soldati, D., and Crabb, B.S. (2005). Tetracycline analogue-regulated transgene expression in Plasmodium falciparum blood stages using Toxoplasma gondii transactivators. Proc Natl Acad Sci U S A *102*, 2980-2985.

Menard, D., Khim, N., Beghain, J., Adegnika, A.A., Shafiul-Alam, M., Amodu, O., Rahim-Awab, G., Barnadas, C., Berry, A., Boum, Y., *et al.* (2016). A Worldwide Map of Plasmodium falciparum K13-Propeller Polymorphisms. N Engl J Med *374*, 2453-2464.

Miotto, O., Amato, R., Ashley, E.A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S.O., Dhorda, M., *et al.* (2015). Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet *47*, 226-234.

Muralidharan, V., Oksman, A., Iwamoto, M., Wandless, T.J., and Goldberg, D.E. (2011). Asparagine repeat function in a Plasmodium falciparum protein assessed via a regulatable fluorescent affinity tag. Proc Natl Acad Sci U S A *108*, 4411-4416.

Muralidharan, V., Oksman, A., Pal, P., Lindquist, S., and Goldberg, D.E. (2012). Plasmodium falciparum heat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers. Nat Commun *3*, 1310.

Murugan, K., Babu, K., Sundaresan, R., Rajan, R., and Sashital, D.G. (2017). The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. Mol Cell *68*, 15-25.

Ng, C.L., Siciliano, G., Lee, M.C., de Almeida, M.J., Corey, V.C., Bopp, S.E., Bertuccini, L., Wittlin, S., Kasdin, R.G., Le Bihan, A., *et al.* (2016). CRISPR-Cas9-modified pfmdr1 protects Plasmodium falciparum asexual blood stages and gametocytes against a class of piperazine-containing compounds but potentiates artemisinin-based combination therapy partner drugs. Mol Microbiol *101*, 381-393.

Nkrumah, L.J., Muhle, R.A., Moura, P.A., Ghosh, P., Hatfull, G.F., Jacobs, W.R., Jr., and Fidock, D.A. (2006). Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase. Nat Methods *3*, 615-621.

Noedl, H., Se, Y., Schaecher, K., Smith, B.L., Socheat, D., Fukuda, M.M., and Artemisinin Resistance in Cambodia 1 Study, C. (2008). Evidence of artemisinin-resistant malaria in western Cambodia. N Engl J Med *359*, 2619-2620.

O'Donnell, R.A., Freitas-Junior, L.H., Preiser, P.R., Williamson, D.H., Duraisingh, M., McElwain, T.F., Scherf, A., Cowman, A.F., and Crabb, B.S. (2002). A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of Plasmodium falciparum chromosomes. EMBO J *21*, 1231-1239.

O'Neill, M.T., Phuong, T., Healer, J., Richard, D., and Cowman, A.F. (2011). Gene deletion from Plasmodium falciparum using FLP and Cre recombinases: implications for applied site-specific recombination. Int J Parasitol *41*, 117-123.

Pascual, A., Madamet, M., Bertaux, L., Amalvict, R., Benoit, N., Travers, D., Cren, J., Taudon, N., Rogier, C., Parzy, D., *et al.* (2013). In vitro piperaquine susceptibility is not associated with the Plasmodium falciparum chloroquine resistance transporter gene. Malar J *12*, 431.

Prommana, P., Uthaipibull, C., Wongsombat, C., Kamchonwongpaisan, S., Yuthavong, Y., Knuepfer, E., Holder, A.A., and Shaw, P.J. (2013). Inducible knockdown of Plasmodium gene expression using the glmS ribozyme. PLoS One *8*, e73783.

Ranganathan, V., Wahlin, K., Maruotti, J., and Zack, D.J. (2014). Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. Nat Commun *5*, 4516.

Rasti, N., Wahlgren, M., and Chen, Q. (2004). Molecular aspects of malaria pathogenesis. FEMS Immunol Med Microbiol *41*, 9-26.

Riglar, D.T., Richard, D., Wilson, D.W., Boyle, M.J., Dekiwadia, C., Turnbull, L., Angrisano, F., Marapana, D.S., Rogers, K.L., Whitchurch, C.B., *et al.* (2011). Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. Cell Host Microbe *9*, 9-20.

Risco-Castillo, V., Topcu, S., Marinach, C., Manzoni, G., Bigorgne, A.E., Briquet, S., Baudin, X., Lebrun, M., Dubremetz, J.F., and Silvie, O. (2015). Malaria Sporozoites Traverse Host Cells within Transient Vacuoles. Cell Host Microbe *18*, 593-603.

Rohrbach, P., Sanchez, C.P., Hayton, K., Friedrich, O., Patel, J., Sidhu, A.B., Ferdig, M.T., Fidock, D.A., and Lanzer, M. (2006). Genetic linkage of pfmdr1 with food vacuolar solute import in Plasmodium falciparum. EMBO J *25*, 3000-3011.

Rottmann, M., McNamara, C., Yeung, B.K., Lee, M.C., Zou, B., Russell, B., Seitz, P., Plouffe, D.M., Dharia, N.V., Tan, J., *et al.* (2010). Spiroindolones, a potent compound class for the treatment of malaria. Science *329*, 1175-1180.

Russo, I., Oksman, A., Vaupel, B., and Goldberg, D.E. (2009). A calpain unique to alveolates is essential in Plasmodium falciparum and its knockdown reveals an involvement in pre-S-phase development. Proc Natl Acad Sci U S A *106*, 1554-1559.

Schofield, L. (2007). Intravascular infiltrates and organ-specific inflammation in malaria pathogenesis. Immunol Cell Biol *85*, 130-137.

Shi, H., Djikeng, A., Mark, T., Wirtz, E., Tschudi, C., and Ullu, E. (2000). Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA. RNA *6*, 1069-1076.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., *et al.* (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell *60*, 385-397.

Sidik, S.M., Huet, D., Ganesan, S.M., Huynh, M.H., Wang, T., Nasamu, A.S., Thiru, P., Saeij, J.P., Carruthers, V.B., Niles, J.C., *et al.* (2016). A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. Cell *166*, 1423-1435 e1412.

Sonoiki, E., Ng, C.L., Lee, M.C., Guo, D., Zhang, Y.K., Zhou, Y., Alley, M.R., Ahyong, V., Sanz, L.M., Lafuente-Monasterio, M.J., *et al.* (2017). A potent antimalarial benzoxaborole targets a Plasmodium falciparum cleavage and polyadenylation specificity factor homologue. Nat Commun *8*, 14574.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem *82*, 237-266.

Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. Nature *527*, 110-113.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.

Straimer, J., Gnadig, N.F., Witkowski, B., Amaratunga, C., Duru, V., Ramadani, A.P., Dacheux, M., Khim, N., Zhang, L., Lam, S., *et al.* (2015). Drug resistance. K13-propeller mutations confer artemisinin resistance in Plasmodium falciparum clinical isolates. Science *347*, 428-431.

Straimer, J., Lee, M.C., Lee, A.H., Zeitler, B., Williams, A.E., Pearl, J.R., Zhang, L., Rebar, E.J., Gregory, P.D., Llinas, M., *et al.* (2012). Site-specific genome editing in Plasmodium falciparum using engineered zinc-finger nucleases. Nat Methods *9*, 993-998.

Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., Krueger, A., Pollok, J.M., Menard, R., and Heussler, V.T. (2006). Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. Science *313*, 1287-1290.

Swarts, D.C., van der Oost, J., and Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. Mol Cell *66*, 221-233 e224.

Takala-Harrison, S., Jacob, C.G., Arze, C., Cummings, M.P., Silva, J.C., Dondorp, A.M., Fukuda, M.M., Hien, T.T., Mayxay, M., Noedl, H., *et al.* (2015). Independent emergence of artemisinin resistance mutations among Plasmodium falciparum in Southeast Asia. J Infect Dis *211*, 670-679.

Tiburcio, M., Niang, M., Deplaine, G., Perrot, S., Bischoff, E., Ndour, P.A., Silvestrini, F., Khattab, A., Milon, G., David, P.H., *et al.* (2012). A switch in infected erythrocyte deformability at the maturation and blood circulation of Plasmodium falciparum transmission stages. Blood *119*, e172-180.

Tilley, L., Straimer, J., Gnadig, N.F., Ralph, S.A., and Fidock, D.A. (2016). Artemisinin Action and Resistance in Plasmodium falciparum. Trends Parasitol *32*, 682-696.

Trape, J.F., Pison, G., Preziosi, M.P., Enel, C., Desgrees du Lou, A., Delaunay, V., Samb, B., Lagarde, E., Molez, J.F., and Simondon, F. (1998). Impact of chloroquine resistance on malaria mortality. C R Acad Sci III *321*, 689-697.

Uhlemann, A.C., McGready, R., Ashley, E.A., Brockman, A., Singhasivanon, P., Krishna, S., White, N.J., Nosten, F., and Price, R.N. (2007). Intrahost selection of Plasmodium falciparum pfmdr1 alleles after antimalarial treatment on the northwestern border of Thailand. J Infect Dis *195*, 134-141.

Vanaerschot, M., Lucantoni, L., Li, T., Combrinck, J.M., Ruecker, A., Kumar, T.R.S., Rubiano, K., Ferreira, P.E., Siciliano, G., Gulati, S., *et al.* (2017). Hexahydroquinolines

are antimalarial candidates with potent blood-stage and transmission-blocking activity. Nat Microbiol.

Veiga, M.I., Dhingra, S.K., Henrich, P.P., Straimer, J., Gnadig, N., Uhlemann, A.C., Martin, R.E., Lehane, A.M., and Fidock, D.A. (2016). Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. Nat Commun *7*, 11553.

Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S., and Roubos, J.A. (2017). CRISPR/Cpf1 enables fast and simple genome editing of Saccharomyces cerevisiae. Yeast.

Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014). Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. Nat Methods *11*, 915-918.

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*, 479-491.

Wang, M., Mao, Y., Lu, Y., Tao, X., and Zhu, J.K. (2017). Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System. Mol Plant *10*, 1011-1013.

Weiss, G.E., Gilson, P.R., Taechalertpaisarn, T., Tham, W.H., de Jong, N.W., Harvey, K.L., Fowkes, F.J., Barlow, P.N., Rayner, J.C., Wright, G.J., *et al.* (2015). Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during Plasmodium falciparum invasion of erythrocytes. PLoS Pathog *11*, e1004670.

White, N.J., Pukrittayakamee, S., Hien, T.T., Faiz, M.A., Mokuolu, O.A., and Dondorp, A.M. (2014). Malaria. Lancet *383*, 723-735.

WHO (2016). World Health Organization. WHO Global Malaria Programme. World Malaria Report 2016.

Witkowski, B., Amaratunga, C., Khim, N., Sreng, S., Chim, P., Kim, S., Lim, P., Mao, S., Sopha, C., Sam, B., *et al.* (2013). Novel phenotypic assays for the detection of artemisinin-resistant Plasmodium falciparum malaria in Cambodia: in-vitro and ex-vivo drug-response studies. Lancet Infect Dis *13*, 1043-1049.

Wu, T., Nagle, A., Kuhen, K., Gagaring, K., Borboa, R., Francek, C., Chen, Z., Plouffe, D., Goh, A., Lakshminarayana, S.B., *et al.* (2011). Imidazolopiperazines: hit to lead optimization of new antimalarial agents. J Med Chem *54*, 5116-5130.

Wu, Y., Sifri, C.D., Lei, H.H., Su, X.Z., and Wellems, T.E. (1995). Transfection of Plasmodium falciparum within human red blood cells. Proc Natl Acad Sci U S A *92*, 973-977.

Xie, S.C., Dogovski, C., Hanssen, E., Chiu, F., Yang, T., Crespo, M.P., Stafford, C., Batinovic, S., Teguh, S., Charman, S., *et al.* (2016). Haemoglobin degradation underpins the sensitivity of early ring stage Plasmodium falciparum to artemisinins. J Cell Sci *129*, 406-416.

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., *et al.* (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell *165*, 949-962.

Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. Cell *167*, 1814-1828 e1812.

Yap, A., Azevedo, M.F., Gilson, P.R., Weiss, G.E., O'Neill, M.T., Wilson, D.W., Crabb, B.S., and Cowman, A.F. (2014). Conditional expression of apical membrane antigen 1 in Plasmodium falciparum shows it is required for erythrocyte invasion by merozoites. Cell Microbiol *16*, 642-656.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., *et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., *et al.* (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotechnol *35*, 31-34.

CHAPTER 2

Synthetic RNA-protein modules integrated with native translation mechanisms to control gene expression in malaria parasites

Suresh M. Ganesan[1]

Alejandra Falla[1]

Stephen J. Goldfless[1]

Armiyaw S. Nasamu[1]

Jacquin C. Niles[1]

[1]Department of Biological Engineering, Massachusetts Institute of Technology,

Cambridge, MA 02139 USA

Published in Nature Communications 7, 10727 (2016)

ABSTRACT

Synthetic post-transcriptional regulation of gene expression is important for understanding fundamental biology and programming new cellular processes in synthetic biology. Previous strategies for regulating translation in eukaryotes have focused on disrupting individual steps in translation, including initiation and mRNA cleavage. In emphasizing modularity and cross-organism functionality, these systems are designed to operate orthogonally to native control mechanisms. Here, we introduce a broadly applicable strategy for robustly controlling protein translation by integrating synthetic translational control via a small molecule-regulated RNA-protein module with native mechanisms that simultaneously regulate multiple facets of cellular RNA fate. We demonstrate that this strategy reduces 'leakiness' to improve overall expression dynamic range, and can be implemented without sacrificing modularity and cross-organism functionality. We illustrate this in *S. cerevisae* and the non-model human malarial parasite, *Plasmodium falciparum*. Given the limited functional genetics toolkit available for *P. falciparum*, we establish the utility of this strategy for defining essential genes.

INTRODUCTION

The ability to precisely, reversibly and temporally regulate gene expression is essential for gaining insights into how biological systems function as well as designing biological circuits that direct novel cellular behaviors in synthetic biology applications. In model eukaryotic systems, this can be flexibly achieved at the transcriptional, translational and post-translational levels. When available, transcription-based regulatory systems are favored, as they provide access to large dynamic ranges that allow gene expression to be titrated to a level appropriate for the specific application (Forster et al., 1999; No et al., 1996; Rönicke et al., 1997). Increasingly, translational and post-translational control systems are being used to diversify the gene regulation toolkit and fine-tune expression. However, compared with transcription-based systems, the dynamic ranges attainable with these systems are relatively modest, and higher basal expression levels are observed in the repressed state (Ausländer et al., 2014; Goldfless et al., 2012; Suess et al., 2003; Win and Smolke, 2007; Wittmann and Suess, 2011). This limits their utility as robust and general solutions for controlling gene expression. Thus, straightforward strategies that extend the dynamic range of post-transcriptional control systems would significantly increase the flexibility with which these systems can be broadly used as effective tools.

Toward this goal, we have focused on enhancing the regulatory dynamic range attainable during synthetic control of translation. Current designs emphasize achieving function independent of host cell regulatory mechanisms. Thus, small molecule-regulated aptamers and riboswitches that directly block translation initiation, and natural or engineered ribozymes that induce *cis*-cleavage of target transcripts have dominated system design (Ausländer et al., 2014; Suess et al., 2003; Win and Smolke, 2007; Wittmann and Suess, 2011). These approaches enable modular systems that function independently of host cell mechanisms, and in principle, are broadly applicable in different organisms and cell types. While conceptually simple, this framework is restrictive, as it limits opportunities to achieve improved functionality through direct integration of synthetic parts with the diverse mechanisms cells have evolved to robustly regulate translation.

Natively, translation is controlled at multiple levels. Initiation via cap-dependent recruitment of factors critical for ribosome assembly, mRNA subcellular sequestration and turnover are key and highly regulated steps. Many proteins that play key roles in regulating these processes have been previously characterized (Besse and Ephrussi, 2008; Rocak and Linder, 2004). For example, eukaryotic translation requires recognition of the mRNA cap structure by eIF4E, which recruits the eIF4F complex and leads ultimately to ribosome assembly and translation. However, the efficiency of this process is modulated by eIF4E-binding proteins (4E-BPs), which block eIF4F recruitment and inhibit translation. Similarly, de-capping and de-adenylation enzymes, such as Dcp1p and Pop2p, respectively, in *S. cerevisiae*, control mRNA turnover have been characterized (Decker and Parker, 2012). In many instances, homologs of these characterized factors are found widely across eukaryotes, suggesting they play highly conserved roles. Thus, synthetic gene regulation schemes can potentially be significantly enhanced if effective strategies for integration with native translation control components can be devised. Furthermore, conservation of the core translation machinery and regulatory components across diverse organisms suggest that this should be attainable without sacrificing design modularity and portability between different cell types and organisms.

Herein, we describe a straightforward strategy for directly integrating synthetic and native translation regulatory components to achieve improved inducible control of gene expression. By fusing the Tet repressor protein (TetR) to various translation regulatory effectors, we achieve robust anhydrotetracycline (aTc)-dependent regulation of target transcripts via genetically-encoded TetR aptamers in either their 5'- or 3'- UTRs. The regulatory dynamic ranges achieved approach that of some transcription based approaches when these aptamer elements flank the target ORF. We also illustrate the generality of our approach across two unrelated organisms, namely the model yeast *S. cerevisiae* and the human malarial parasite, *P. falciparum*. In the latter organism, where few tools are available for studying gene function, we show that this technology can be used to stringently classify genes essential for parasite survival.

RESULTS AND DISCUSSION

Molecular design principles

Existing strategies for achieving synthetic control of eukaryotic translation rely mostly on aptamer-small molecule interactions (Suess et al., 2003), inducible ribozymes (Win and Smolke, 2007; Wittmann and Suess, 2011) and RNA-protein interactions (Ausländer et al., 2014; Goldfless et al., 2012; Saito et al., 2010). These control systems are 'uni-dimensional' in design in that the required RNA/aptamer element is strategically positioned within a target transcript to predominantly direct a single outcome, such as disruption of translation initiation or induced RNA degradation (Fig. 2-1A). Such designs emphasize modularity and portability, but likely at the expense of achieving maximal robustness, by limiting intimate coupling between the synthetic module and native regulatory mechanisms.

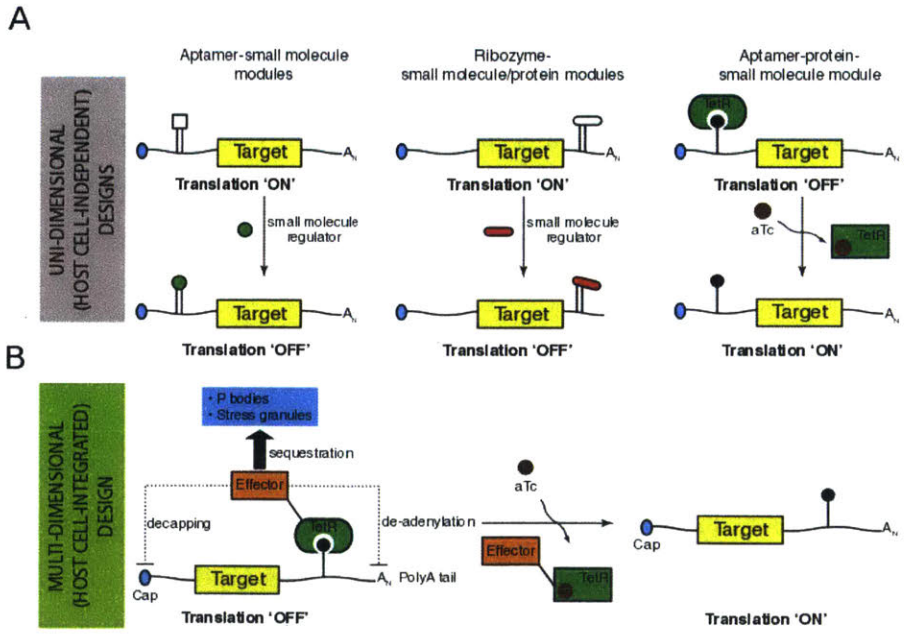


Figure 2-1. Achieving inducible gene expression by integrating a synthetic protein-RNA interaction module with native translation control mechanisms
(A) Several previously implemented approaches for post-transcriptional regulation of gene expression in eukaryotes are summarized. These are classified as 'uni-dimensional' since primarily one activity determines the availability or accessibility of a target transcript for translation. **(B)** Schematic of the proposed strategy for integrating a synthetic RNA-protein module with host cell translation control mechanisms to achieve 'multi-dimensional', post-transcriptional regulation of gene expression. Fusing the regulatory protein in the synthetic module to host cell factors involved in controlling RNA turnover, degradation and sequestration is proposed as an approach for extending the dynamic range and reducing leaky expression of the synthetic system. Direct control of gene expression is achieved by toggling the synthetic RNA-protein interaction via a small molecule inducer.

Examination of natural translation control mechanisms, however, reveals that a diverse set of activities simultaneously converge on a regulated transcript to achieve stringent control over its translation (Besse and Ephrussi, 2008; Rocak and Linder, 2004). We hypothesized that synthetic translation control systems that more closely recapitulate nature's 'multi-dimensional' strategy might inherently be more robust, and achieve significantly improved regulatory outcomes than existing uni-dimensional designs. To establish proof-of-concept, we reasoned that a broadly useful framework should consist of a foundational RNA-protein interaction module in which: (1) a small molecule directly controls the interaction, and allows for transcription-independent translational control; and (2) the protein component is readily engineered to create fusions that bridge the foundational RNA-protein module with native cellular translation control components (Fig. 2-1B). Successfully validating this approach would provide immediate access to diverse mechanisms used natively to stringently control the cellular fate of RNA, and establish a new and easily adaptable paradigm for designing multi-dimensional translation control systems with improved functional characteristics.

TetR fusion proteins conditionally regulate translation

We selected a previously defined aptamer-TetR protein module that is regulated by tetracycline analogs (Belmont and Niles, 2010) as the foundational framework for this study. This module has been used in a uni-dimensional context (Fig. 2-1A, *right panel*) to control translation in the model organism, *S. cerevisiae* (Goldfless et al., 2012) and the unrelated human malarial pathogen, *Plasmodium falciparum* (Goldfless et al., 2014). As

such, direct quantitative comparisons between uni- and multi-dimensional implementation can be made to rigorously quantify improvements in functionality achieved via the latter design. Critical to addressing generalizability of this concept, this selection also facilitates assessment of this multi-dimensional regulatory framework to deliver enhanced translational control in an unrelated organism.

In selecting TetR fusion partners, we emphasized two criteria that the selected proteins must satisfy. Namely, these must have: (1) been previously implicated in regulating translation initiation, mRNA stability and/or turnover or mRNA sequestration; and (2) clearly identifiable homologs in other eukaryotes, and in particular, *P. falciparum*. The latter criterion provides a stringent test of our central principle in a non-model organism where very little basic mechanistic evidence supporting specific post-transcriptional and translational regulatory mechanisms has been directly elucidated.

The success of this approach is predicated on identifying fusion proteins that preserve both the inducible interaction between TetR and its aptamer, and the partner's native role in regulating translation. We performed a focused screen in *S. cerevisiae* to identify potential candidates by creating fusions of TetR to Dcp1p, Dhh1p, Hrp1p, Pop2p, Upf3p, Caf20p, Eap1p and Cdc33p, which have all previously been implicated in some aspect of native translational regulation in yeast (Cosentino et al., 2000; de la Cruz et al., 1997; Decker and Parker, 2012; He and Jacobson, 2001; Kessler et al., 1997). We made a reporter construct in which either five or ten tandem TetR aptamers (referred to hereafter as 5x and 10x aptamer arrays) were positioned within the 3'UTR of a Venus YFP (vYFP) coding sequence (Belmont and Niles, 2012) (Fig. 2-2A) to identify TetR fusions that mediated doxycycline (Dox)-inducible changes in vYFP expression relative to TetR alone. Notably, TetR by itself displays no inherent inducible behavior when its cognate aptamers are installed in a 3'UTR context (Fig. 2-2A). This is consistent with studies of native translational control mediated by protein binding within the 3'UTR, where factors that modulate cap-dependent translation initiation, decapping,

polyadenylation/deadenylation or mRNA sequestration, as examples, must be recruited to achieve a regulatory outcome (Besse and Ephrussi, 2008; Kuersten and Goodwin, 2003). We did not examine the effects of placing tandem aptamers within the 5'UTR, since our previous data indicated that increasingly stable secondary structure within this region significantly decreases maximal expression without substantially improving the regulatory dynamic range (Goldfless et al., 2012).

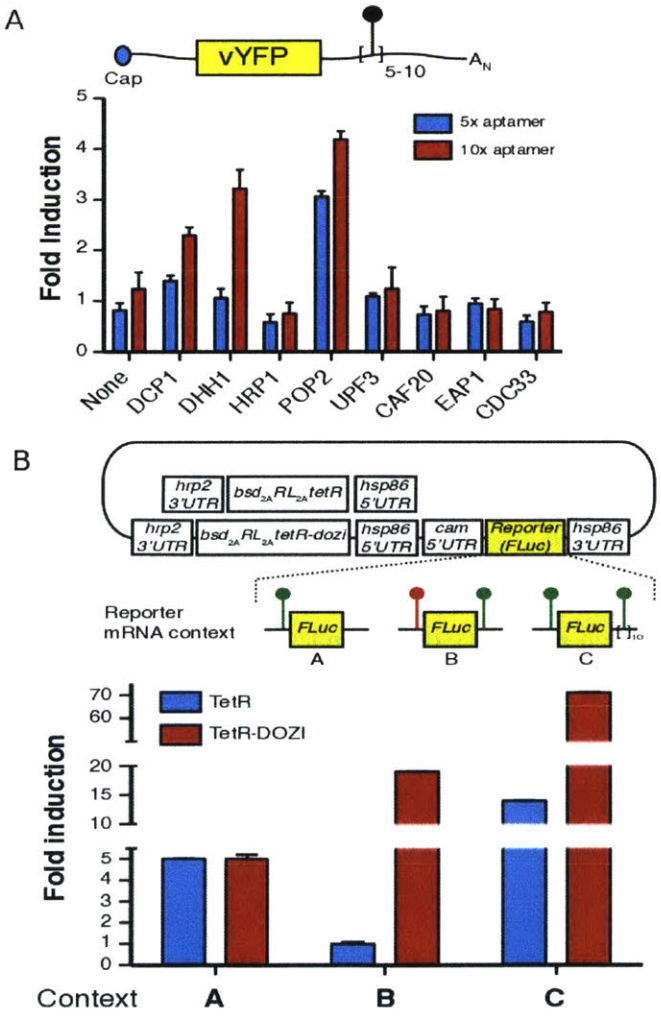


Figure 2-2. Various TetR fusion proteins enhance TetR aptamer-mediated translational regulation in both S. cerevisiae and P. falciparum

(**A**) In *S. cerevisiae*, various proteins fused to TetR were evaluated for their ability to enhance doxycycline-dependent regulation of a vYFP reporter via either five or ten tandem TetR aptamers positioned within the 3'-UTR. (**B**) In *P. falciparum*, regulated expression of a FLuc reporter by TetR and the TetR-DOZI fusion was tested. Three reporter contexts were examined, namely: (i) single TetR aptamer in the 5'-UTR only; (ii) single mutated TetR aptamer in the 5'-UTR and 10x TetR aptamers in the 3'-UTR; and (iii) single TetR aptamer in the 5'-UTR and 10x TetR aptamers in the 3'-UTR. Functional and mutated (no binding to TetR) aptamers are illustrated as green and red lollipops, respectively. In both panels, fold-induction is calculated as the ratio of reporter expression in the induced state (+ aTc) relative to that in the repressed state (- aTc). Data shown is the mean ± s.d. from biological triplicates, and is representative of 2 (*S. cerevisae*) and 3 (*P. falciparum*) independent experiments. * = $P < 0.05$; and *** = $P < 0.0001$ by t-Test.

In the 5x aptamer array reporter context, only TetR-Pop2p resulted in significant Dox-dependent regulation (~3-fold relative to TetR). In the 10x aptamer array reporter context, however, TetR-Dcp1p, TetR-Dhh1p and TetR-Pop2p all exhibited Dox-dependent regulation of vYFP expression (~2-4 fold relative to TetR). All fusions did not yield improved regulation, indicating that simply fusing any protein to TetR is insufficient to achieve an improvement in translational regulation. Overall, these screening data indicate that fusing specific proteins natively involved in regulating translation to TetR can significantly improve the regulatory dynamic range beyond that attainable with TetR alone, while preserving tetracycline analog-dependent control. These findings provided initial validation that integrating our synthetic RNA-protein module with host regulatory mechanisms could yield improved regulatory outcomes beyond that of the RNA-protein module alone when tested in the identical reporter context.

Enhanced regulation of gene expression in *P. falciparum*

We focused next on systematically understanding how a selected TetR fusion protein that improved translational regulation outcomes in a model system would function in the non-model eukaryotic pathogen, *P. falciparum*. Our rationale for this approach was twofold. First, native translational regulation in this organism is poorly understood. We reasoned that this choice would provide important insights into the broader applicability of our concept without a need for detailed prior knowledge of fundamental translation control mechanisms in the host organism. Secondly, the toolkit for robust conditional gene

expression in *P. falciparum* is small, with limitations in both dynamic range and basal leakiness (Armstrong et al., 2007; Goldfless et al., 2014; Meissner et al., 2005; Prommana et al., 2013) We reasoned, therefore, that a host-integrated synthetic control module with substantially enhanced regulatory dynamic range would also have immediate practical value as a functional genetics tool in this organism.

Of the three functional fusions identified during our yeast screens, we selected the TetR-Dhh1p context for detailed evaluation in *P. falciparum*. Several factors influenced this decision. First, we wished to use the *P. falciparum* homolog rather than assuming the yeast protein functions equivalently in *P. falciparum*. BLAST sequence analysis comparing the yeast and *P. falciparum* homologs of Dcp1p, Dhh1p and Pop2p revealed that Dhh1p and PF3D7_0320800 (DOZI) had very high sequence identity and similarity (68 % and 80%, respectively), and were similar in size (506 versus 433 amino acids, respectively) (Fig. S-1A). In contrast, Dcp1p and Pop2p and their *P. falciparum* homologs are much less conserved (Fig. S1B-C). The *P. falciparum* homologs are also large proteins with poorly conserved regions that potentially encode functionally important regulatory information. Therefore, creating functional fusions based on these proteins might entail initial screens to identify suitable fragments that preserve all critical biochemical information for efficient integration with the native machinery and TetR function upon fusion. We reasoned that a TetR-DOZI fusion closely reflected the context validated in yeast, and, together with its compact size, would be an ideal primary candidate to test in *P. falciparum*.

Second, the Dhh1p homolog from the related rodent malarial parasite, *P. berghei*, has been implicated in translational regulation. Deleting this protein (referred to as development of zygote inhibited or DOZI based on the observed phenotype) dysregulates the stability of many transcripts that are normally repressed during asexual stage development in blood and derepressed during sexual stage development in the mosquito vector (Mair et al., 2006). A detailed mechanism of how DOZI functions in this process

is lacking, though *in vitro* translation studies using the *P. falciparum* DOZI suggests a role in regulating eIF4E-dependent translation (Tarique et al., 2013). This is likely only a partial explanation of this protein's function *in situ*, however, since it seems to operate within a larger regulatory protein complex that is still to be functionally defined (Mair et al., 2010). Furthermore, the DDX6 helicase protein family, to which Dhh1p and DOZI belong, is implicated in regulating diverse aspects of RNA biology including decapping, degradation and sequestration (Decker and Parker, 2012; Ernoult-Lange et al., 2012; Rouya et al., 2014). Last, Dhh1p homologs in several model organisms including *C. elegans*, *D. melanogaster*, *X. laevis* and humans are all very highly conserved (Fig. S-1A). Therefore, proof-of-concept with this TetR fusion would support broad applicability beyond our two primary test organisms.

We first assessed whether TetR-DOZI functions to regulate translation in *P. falciparum*, and created several dual expression cassette reporter plasmids (Fig. 2-2B). One cassette uses a calmodulin promoter (*Pf*CAM) to transcribe a mRNA encoding firefly luciferase (FLuc) with regulatory TetR aptamers in the 5'UTR only, 3'UTR only or both 5'- and 3'-UTRs to test TetR-dependent translational regulation. The second cassette drives expression of TetR or TetR-DOZI, *Renilla* luciferase (RLuc), and the Blasticidin S deaminase (*bsd*) selection marker from a multi-cistronic transcript using the viral T2A 'skip' sequence (Wagner et al., 2013). The *RLuc* is used as an internal reference signal in our quantitative luciferase assays. We site-specifically integrated (Nkrumah et al., 2006) these reporter plasmids at the *cg6* chromosomal locus in *P. falciparum* NF54^{attB} (Adjalley et al., 2011), and obtained clonal parasites by limiting dilution. We grew parasites either in the absence or presence of anhydrotetracycline (aTc), and measured inducible FLuc expression in the various regulatory contexts described.

These data showed that both TetR and TetR-DOZI regulated FLuc expression by ~5 fold in an aTc-dependent manner when a single aptamer is present within the 5'UTR (Fig. 2-2B, Context A). Thus, TetR-DOZI functions indistinguishably from TetR in this context.

Next, we examined the degree of regulation achieved with TetR or TetR-DOZI when a 10x aptamer array is placed within the 3'UTR, and a non-functional mutated aptamer is placed in the 5'UTR of the FLuc reporter (Fig. 2-2B, Context B). With TetR, we observed no detectable aTc-dependent regulation of FLuc expression (fold regulation ~1), which is consistent with our observation in the yeast experiments (Fig. 2-2A). With TetR-DOZI, however, aTc regulated FLuc expression by ~19-fold (Fig. 2-2B). Thus, TetR-Dhh1p in yeast and TetR-DOZI in *P. falciparum* both substantially enhance aTc-inducible regulation of gene expression beyond what is attainable with TetR by itself.

Next, we tested whether combining regulation by a single aptamer in the 5'UTR and a 10x aptamer array in the 3'UTR (Fig. 2-2B, Context C) would synergize to increase the overall dynamic regulatory range attainable. With TetR alone, we observed ~8-14 fold aTc-dependent regulation of FLuc expression. This represented a ~2-3-fold increase in dynamic range over that observed during TetR regulation mediated via a single aptamer in the 5'UTR. Thus, while TetR did not directly mediate regulation via aptamers in the 3'UTR alone (Fig. 2-2B), an increase in fold-regulation occurred when 5'- and 3'-aptamers flank the reporter gene. Bridging interactions are believed to bring the 5'- and 3'- UTRs of eukaryotic transcripts into close proximity with each other during translation (Wells et al., 1998). This finding could reflect increased TetR recruitment by the 10x aptamer array to create locally high TetR concentrations in proximity to the 5'UTR, such that the fractional occupancy of the regulating 5'aptamer by TetR is higher than when no 3'UTR array is present and TetR can only be recruited from a relatively dilute cytoplasmic pool. Intriguingly, with TetR-DOZI, aTc-dependent regulation increased substantially to ~45-70 fold in the dual aptamer configuration (Fig. 2-2B). The improved regulatory outcome seen with TetR-DOZI relative to TetR is consistent with the gain of translation regulation via the 3'UTR aptamer array enabled by the DOZI component of the fusion protein.

Enhanced regulation is independent of 5'UTR context

Many genes exhibit highly regulated and presumably functionally important temporal expression profiles during the parasite's 48 hour intra-erythrocytic developmental cycle (IDC) (Bozdech et al., 2003; Le Roch et al., 2003). Therefore, we examined whether enhanced regulation is achieved at different time points within the IDC while using promoters having distinct temporal expression profiles. We selected the *P. falciparum* chloroquine resistance transporter (*Pf*CRT; PF3D7_0709000) and *Pf*CAM (PF3D7_1434200) promoters as representative candidates. These exhibit peak transcriptional activity in early (ring)- to mid- (trophozoite), and late (schizont) stages, respectively (Fig. 2-3A).

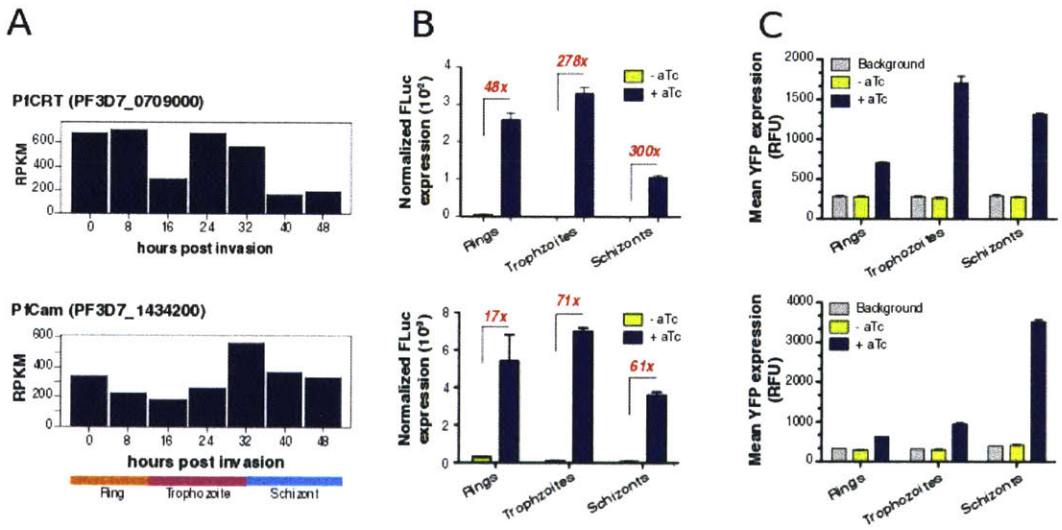


Figure 2-3. TetR-DOZI–aptamer module provides robust control of gene expression across the *P. falciparum* intraerythrocytic developmental cycle

(**A**) RNA-Seq data from PlasmoDB showing the stage-dependent transcription profile for two native parasite promoters, *Pf*CRT (PF3D7_0709000) and *Pf*CAM (PF3D7_1434200). (**B**) Stage-dependent FLuc expression normalized to RLuc expression (non-regulated internal reference) in the absence and presence of aTc for both *Pf*CRT (top) and *Pf*CAM (bottom) promoter contexts. The regulatory dynamic range (ratio of aTc-induced to basal expression) is shown in red for ring, trophozoite and schizont stage parasites. (**C**) Stage-dependent mean YFP expression in the absence and presence of aTc for both *Pf*CRT (top) and *Pf*CAM (bottom) promoter contexts. Both the FLuc and YFP reporters are controlled by a single TetR aptamer in the 5'UTR and ten tandem TetR aptamers in the 3'UTR. Fold-induction values were not calculated from the fluorescence reporter expression data as the signal-to-noise is relatively low, and in the repressed state the fluorescence signal is effectively at the autofluorescence background level. Data shown is the mean ± s.d. from biological triplicates, and is representative of 2-4 independent experiments. * = $P < 0.001$ by t-Test.

With both promoters, substantially enhanced regulation was observed when TetR aptamers in the 5′- and 3′- UTRs are used together. With the *Pf*CAM promoter, fold-expression upon aTc induction ranged from 20-70 fold across the IDC. Intriguingly, in the *Pf*CRT promoter context, the observed fold-expression was much higher and ranged from ~50-300 fold across the IDC (Fig. 2-3B). In general, the fold regulation was lowest in early stage parasites, but regulation by the TetR-DOZI fusion exceeded that by TetR alone (Goldfless et al., 2014). Interestingly, regulation increased more substantially in later stage parasites. This pattern of regulation is likely not simply due to delayed expression of the regulatory TetR-DOZI fusions, as the *Pf*HSP86 promoter driving expression of this construct is constitutively high across the IDC. We hypothesize that this phenomenon might instead be linked to stage-dependent activation of the native parasite translational regulatory pathway(s) in which the DOZI component of our fusion operates (Bozdech et al., 2003; Le Roch et al., 2003). Further studies will be required to precisely delineate the DOZI-dependent regulatory pathways in *P. falciparum.* However, it is worth emphasizing that even without a detailed fundamental understanding of these native mechanisms, we consistently achieve substantially enhanced regulation by integrating our synthetic system with this native host translational regulatory mechanism.

We also replaced FLuc with an EYFP reporter gene in both the *Pf*CRT and *Pf*CAM promoter contexts to confirm that other genes could stringently be regulated in this framework. In the repressed state (– aTc), no EYFP signal above background could be detected by flow cytometry for both *Pf*CRT and *Pf*CAM promoter contexts at any point during the IDC. However, aTc induction produced a substantial increase in EYFP signal at all IDC stages (Fig. 2-3C and Fig. 2-S2). In contrast, a single 5'-aptamer within the *Pf*CAM promoter context regulated by TetR alone results in noticeable fluorescence signal above background in the repressed state (Goldfless et al., 2014). Thus, the dual aptamer configuration with a host-integrated TetR-DOZI regulatory module stringently represses an orthogonal reporter beyond what is attainable with the standalone TetR module. These data indicate that: (1) DOZI-enhanced translational regulation is not restricted to a specific IDC stage; (2) enhanced regulation is achieved independent of the

promoter context and expression profile; and (3) adequate concentrations of TetR-DOZI are produced throughout the IDC to achieve robust regulation.

Leaky expression is reduced with the TetR-DOZI system

A frequent limitation of post-transcriptional regulatory schemes is the challenge of reducing leaky expression, as this can adversely impact the practical utility of these tools for both functional genetics and synthetic biology applications. Given the substantial increase in regulatory dynamic range associated especially with the dual aptamer configuration, we closely analyzed the basis for this outcome.

We observed that inclusion of a 10x aptamer array within the 3'UTR of the target transcript reduced maximal reporter expression by ~2-3 fold (Fig. 2-4A). This was independent of whether TetR or TetR-DOZI functioned as the regulator, and likely reflects the change introduced into the target transcript. We then compared fold reduction in FLuc expression in the repressed state (- aTc) using TetR with a single 5'UTR aptamer as a reference (Fig. 2-4B), and found that substituting TetR for TetR-DOZI in this context had no impact on leaky expression. In the 5'-mutant aptamer and 10x aptamer in the 3'UTR context, no reduction in leaky expression is observed with TetR as the regulator. In contrast, leaky expression is reduced 6-12 fold with TetR-DOZI. In the dual aptamer context, we observed a 7-10 fold reduction in leaky expression with TetR. In contrast, when TetR-DOZI is used as the regulator in this context, we observed a 17-30 fold reduction in leaky expression. Taken together, these data show that integrating the TetR-aptamer module with host translational mechanisms mediated through DOZI results in substantially improved regulation of gene expression, and this is predominantly achieved through a reduction in leaky expression. This feature is especially valuable in functional genetics applications where residual protein expression can mask phenotypes mediated by truly essential genes, or in synthetic biology applications where levels of critical components must be stringently regulated to achieve desired circuit performance. In these applications, the reduction in maximal protein expression observed by installing aptamer

arrays within the 3'-UTR can be further contextualized. In synthetic designs, the desired protein expression levels can be achieved by choosing appropriate promoters to pair with the aptamer-regulated 3'-UTR. In functional genetics applications, it is conceivable that installing a 3'-aptamer array in the context of a fixed, native promoter may prevent adequate maximal expression of essential genes and preclude recovery of viable organisms for further study. The extent to which this scenario is encountered will only be evident upon broader application of our technology. If encountered, this can be addressed by engineering 5'- and 3'-UTR aptamer regulated contexts wherein quantitatively appropriate promoter strength is selected to simultaneously afford functionally adequate maximal gene expression and stringently tunable regulation (Figs. 2-2 and 2-3). However, our PfATP4 data indicate that this system will be relevant for targeting and studying essential parasite genes, via engineering aptamer arrays into the 3'-UTR alone.

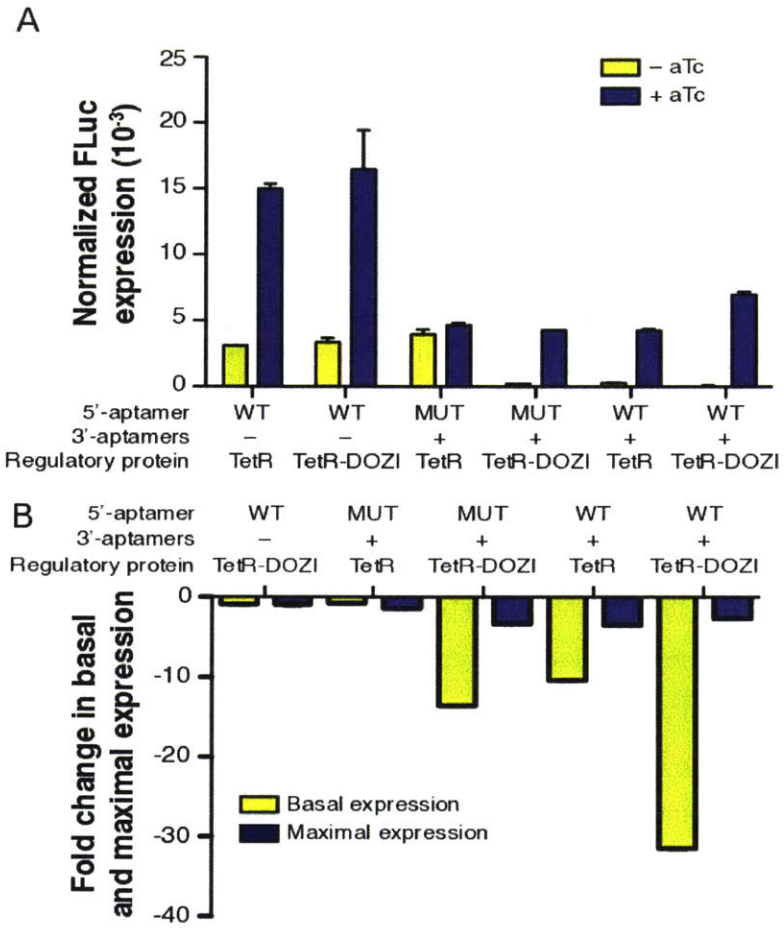


Figure 2-4. Multi-dimensional synthetic control of translation achieves significant reduction in leaky expression

(**A**) FLuc expression levels in the absence and presence of aTc normalized to RLuc expression (non-regulated internal reference) when TetR– and TetR-DOZI– are used to regulate FLuc expression via aptamers positioned within the 5'UTR, 3'UTR or both as shown. (**B**) Fold change in maximal FLuc expression upon induction (plus 0.5 μM aTc) and in basal FLuc expression (minus aTc) is shown relative to the context in which TetR regulates FLuc expression via a single aptamer in the 5'UTR. Data shown is the mean ± s.d. from biological triplicates, and is representative of 2-4 independent experiments. * = $P < 0.001$ by t-Test.

Our studies, along with previous data, also suggest a possible model for how regulation might be achieved in this system. We previously showed that a functional aptamer capable of interacting with TetR is absolutely required for aTc-dependent regulation of

gene expression, as when a mutated aptamer that does not interact with TetR is used, no aTc-dependent regulation is observed (Goldfless et al., 2012). Alternatively, if expressing the TetR fusion protein alone were sufficient to repress reporter gene expression, this would not be aTc-dependent. Furthermore, for the same TetR aptamer configuration, absolute FLuc expression levels in the presence of aTc were indistinguishable in the presence of TetR or TetR-DOZI (Fig. 2-4A). This indicates that any repression by direct interaction of the DOZI fusion component with the reporter is either minimal or does not have a significant effect on expression. This observation is analogous to our previous study in which TetR fusions to She2p and She3p in yeast were used to conditionally localize target transcripts to daughter cells in yeast (Belmont and Niles, 2012). Taken together, these data support a model in which TetR aptamers in a target transcript recruit the TetR-DOZI (or other fusion partner), and this facilitates downstream regulatory outcomes mediated through the fusion partner to occur in a transcript-specific manner. Future studies will be needed to quantitatively dissect which combination(s) of the various translational regulatory roles mediated by DOZI/DDX6 family proteins–such as degradation, deadenylation, and sequestration—contribute most to the TetR-DOZI-dependent regulation of targeted transcripts. This information could be useful for designing even more efficient and broadly applicable regulators, in addition to expanding pour currently poor fundamental understanding of translational regulation in asexual stage parasites specifically.

Another native *Plasmodium* protein implicated in translational regulation in diverse organisms also mediates enhanced conditional regulation of translation.

We next examined whether the extended regulatory dynamic range observed with the TetR-DOZI fusion in *P. falciparum* was limited to DOZI. We selected the *P. falciparum* CAR-1/Trailer Hitch Homolog (CITH) as another fusion partner to test for three key reasons. First, CAR-1 and Trailer Hitch in *C. elegans* and *D. melanogaster*, respectively, have been implicated as key translational regulators, and this protein family is widely conserved across eukaryotes (Decker and Parker, 2006). Second, like DOZI, CITH has been implicated in regulating translation in *Plasmodium berghei*. Both proteins seem to

function within the same translation regulatory complex, though additional orthogonal pathways cannot be excluded (Mair et al., 2010). Third, from a practical standpoint, CITH homologs are relatively small proteins, which we reasoned should produce fairly compact and biochemically well-behaved fusions to TetR.

We generated a reporter plasmid in which we replaced TetR-DOZI with a TetR-CITH fusion, and FLuc expression was regulated by aptamers in both 5'-and 3'-UTRs. We then compared aTc-dependent regulation achieved by TetR/5'-aptamer only (benchmark), and TetR, TetR-DOZI and TetR-CITH in the dual aptamer context. Our data show that TetR-DOZI and TetR-CITH both produced substantially improved translational regulation over TetR alone in either the single or dual aptamer context (Fig. 2-S3). These data show that translational regulators from distinct protein families can be successfully incorporated into our conceptual framework for successfully integrating synthetic and native translational regulation to achieve substantially improved conditional control over target gene expression.

Establishing gene essentiality in *Plasmodium falciparum*

Having established quantitative proof-of-concept for our proposed strategy, we sought to establish that this approach could have useful applications. To demonstrate this, we used our framework to directly establish for the first time the essential function of a difficult-to-study *P. falciparum* membrane protein, PfATP4 (PF3D7_1211900), which is of increasing interest as a new antimalarial drug target/resistance mechanism. This P-type ATPase protein has been proposed as the putative target of a new, potent and clinically promising antimalarial drug class, the spiroindolones (Rottmann et al., 2010; Spillman et al., 2013). However, the available evidence does not definitively demonstrate that loss of PfATP4 function should, in fact, impair parasite survival. Thus, a critical piece of information required for rigorously assigning mechanism of action for a new drug target could not be easily satisfied due to technological limitations. The availability of this information assumes even greater significance given recent data indicating that many

diverse chemical scaffolds converge on PfATP4 as a common direct or indirect mechanism of parasite sensitivity and/or resistance to these compounds (Flannery et al., 2014; Lehane et al., 2014).

To begin addressing this knowledge gap, we implemented our TetR-DOZI system to achieve conditional regulation of PfATP4 expression. Given the large size (~3.8 kb) of PfATP4, we genetically encoded a 10x TetR aptamer array in a 3'UTR context at the native locus as illustrated (Fig. 2-5A). Clonal parasites were genotyped by PCR (Fig. 2-5B) and sequencing of the diagnostic products, and Southern analysis (Fig. 2-5C). We also confirmed stringent aTc-dependent regulation of PfATP4 expression by Western blot analysis (Fig. 2-5D). Altogether, these data confirmed that we had successfully modified the PfATP4 genomic locus and could stringently control the expression of this integral membrane protein target (Rottmann et al., 2010).

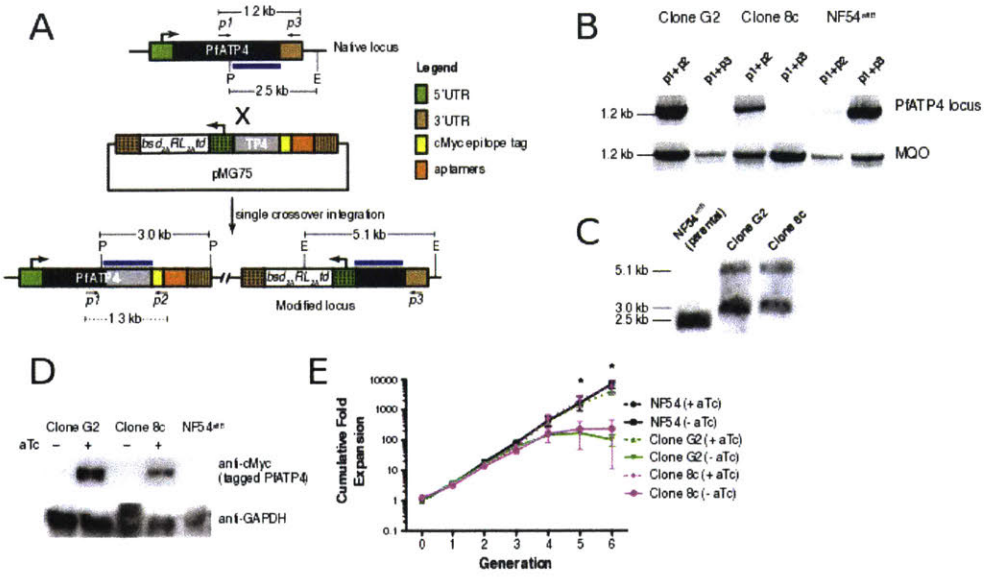


Figure 2-5. Establishing essentiality of a native parasite gene using TetR-DOZI with a genetically-encoded 10x aptamer array in the 3'UTR

(**A**) Schematic of the single crossover event to modify the native PfATP4 chromosomal locus to install a C-terminal epitope tag, a 10x aptamer array and the regulatory TetR-DOZI expression cassette in a single step. Diagnostic PCR primers to assess the status of the locus are designated *p1-p3*, and the expected sizes of PCR products derived from the native and modified locus are indicated. The blue bar corresponds to the probe used in Southern blot experiments in which genomic DNA was digested simultaneously with EcoR1 (E) and PstI (P). The expected diagnostic band sizes for the native and modified loci are indicated. (**B**) Diagnostic PCR analysis of two isolated clones (G2 and 8c) with modified PfATP4 loci. The malate:quinone oxidoreductase (MQO) gene is used as a positive control for the presence of genomic DNA in all samples tested. (**C**) Southern blot analysis for the parental NF54^{attB} strain and clones G2 and 8c. (**D**) Western blot using anti-cMyc antibody to detect aTc-dependent expression of PfATP4 in clones G2 and 8c. The parental NF54^{attB} strain (no cMyc-tagged PfATP4) is included as a negative control. (**E**) ATc-dependent growth of the parental NF54^{attB} strain and Clones G2 and 8c over multiple generations. Data shown is the mean ± s.d from biological triplicates, and is representative of three independent experiments. * = $P < 0.001$ by t-Test.

We next determined whether down-regulating PfATP4 expression adversely impacted parasite growth. We selected two isogenic clones that had been grown continuously in the presence of aTc to ensure maximal PfATP4 expression, and split these into minus and plus aTc conditions to monitor parasite growth over multiple IDCs. Over the first two IDCs, we observed no significant aTc-dependent difference in growth between the G2 and 8c clones. However, beginning in the third IDC and continuing thereafter, there was a significant decrease in growth of the G2 and 8c clones in the no aTc condition

compared to when aTc was present (Fig. 2-5E). As the parental NF54 line grew similarly both in the absence and presence of aTc, this eliminates an inherent proliferative effect of aTc on parasites as the explanation for the growth phenotype. Altogether, these findings establish, for the first time, that loss of PfATP4 function is indeed deleterious to blood stage parasite growth.

Here we demonstrate that synthetic RNA-protein modules can be effectively integrated with native cellular mechanisms for controlling translation, and in so doing, more robust regulation of target gene expression can be achieved. While preservation of modularity and cross-organism compatibility criteria have typically motivated designs intended to minimize interactions with native control mechanisms, we show that these properties can be preserved even while intentionally pursuing host cell integrated designs. A key outcome of this conceptual framework is a substantial reduction in leaky/basal expression, which has been a major challenge in leveraging robust translation control tools for functional genetics and synthetic biology applications in eukaryotes.

We envision this strategy will be broadly useful as the principles for implementation are straightforward and apply in unrelated organisms, as illustrated here. Many post-transcriptional regulatory proteins operating either through the same or distinct mechanisms have been described (Decker and Parker, 2012). These can serve as potential library components that can be systematically evaluated to establish a core set that can serve as 'gateways' for efficiently accessing specific modes of native translational regulation. By focusing on components with high protein sequence conservation and presumably regulatory mechanism(s) across different genera, a high rate of successful portability amongst organisms should be attainable. Novel and orthogonally regulated synthetic RNA-protein interaction pairs can be efficiently discovered using existing methods (Ausländer et al., 2014; Belmont and Niles, 2010). We envision that implementing such orthogonal units using the framework validated here will facilitate multiplexed, stringent and directly tunable control of translation to motivate increasingly

sophisticated and robustly controlled functional genetics and synthetic biology applications.

METHODS

Molecular cloning

Plasmids used in this study were cloned either by restriction/ligation, yeast homologous recombination, or Gibson assembly methods (Wagner et al., 2013). All restriction enzymes were purchased from New England Biolabs and reagents were from Sigma-Aldrich, Research Products International, GoldBio Technology, or Alfa Aesar, unless otherwise specifically mentioned. PCR amplification of parasite DNA was performed on either crude lysates from the parasite fractions or purified DNA templates using Hemo KlenTaq mixed 15:1 (v:v) with PfuTurbo (Agilent). Tables 2-S1 and 2-S2, respectively, summarize the plasmids and primers used in this work. Plasmid DNAs for transfections were prepared using the Xtra Midi Kit (Clontech). Plasmids were propagated in DH5α bacteria, and grown on solid or in liquid media at 30 ºC to minimize the potential for spontaneous truncations of the 3'-UTR TetR aptamer arrays.

Plasmid DNA constructs sequences and sources

GenBank files for the constructs used in this study are provided as supplementary files. *FLuc* is wild-type firefly luciferase DNA carrying the K549E mutation for cytosolic targeting in yeast.

Yeast inducible expression assays

The various TetR-containing regulatory proteins were cloned as C-terminal fusions to TetR in the YCpSUP yeast expression vector (Goldfless et al., 2012). We created a vYFP reporter plasmid with either 5 and 10 tandem repeats of the TetR aptamer 5-1.2 in the 3'-*UTR* by replacing the non-fluorescent vYFPΔ in vYFPΔ–5–1.2(5x) and vYFPΔ–5–1.2(10x)(Belmont and Niles, 2012), respectively, with vYFP. *S. cerevisiae* W303-1B cells harboring both a repressor and an appropriate reporter plasmid were grown to saturation at 30 ºC in Synthetic Defined Media #1 (SD1) [6.7 g/l yeast nitrogen base without amino acids (RPI), 20 mg/l adenine, 30 mg/l lysine, 100 mg/l leucine, 20 mg/l histidine] + 20g/l glucose. Cells were diluted 1:80 into SD1 + 20 g/l raffinose and grown for 4 h. Glucose (to repress TetR or TetR fusion protein expression) or galactose (to

induce TetR or TetR fusion expression) was added to 20 g/l, and cells were grown 16 h at 30 °C with shaking before measurement. To measure vYFP expression by flow cytometry, cells were grown in triplicate either in the presence of glucose (± 22 μM doxycycline) or galactose (± 22 μM doxycycline), and analyzed on a C6 Flow Cytometer (Accuri). For all samples, ~50,000 events were captured and vYFP fluorescence measured in the FL1 channel.

P.falciparum culture and transfection

The *P. falciparum* $NF54^{attB}$ parasites were a gift from David Fidock (Columbia University). These were grown in human erythrocytes at 5% hematocrit under 5% O_2 and 5% CO_2 in RPMI 1640 media supplemented with 5 g/l Albumax II (Life Technologies), 2 g/l $NaHCO_3$, 25 mM HEPES-K pH 7.4, 1 mM hypoxanthine and 50 mg/l gentamicin. For transfections, we used 50-60 μg of each plasmid per 200 μl packed red blood cells, adjusted to 50% hematocrit. We used a Bio-Rad Gene Pulser II to either directly electroporate DNA into ring stage parasites (Wu et al., 1995) or preload uninfected red blood cells (Deitsch et al., 2001). For direct electroporation of ring stage parasites, 0.2 cm electroporation cuvettes were loaded with 0.3 ml of parasitized RBCs at 50% hematocrit and 50 μg of plasmid DNA in incomplete cytomix solution which contains 120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4/KH_2PO_4, 25 mM HEPES, 2 mM EGTA, 5 mM $MgCl_2$. Electroporation conditions were 0.31 kV and 960 μF. For spontaneous DNA uptake, 8 square-wave pulses of 365 V for 1 ms, separated by 100 ms was used. Transfected parasites were selected with 2.5 mg/l Blasticidin-S beginning 4 days after transfection. In the case of *Bxb1*-mediated integration, no selection for the integrase-expressing plasmid was applied. In allelic replacement experiments, drug on/off cycles (28 days each) were completed until the desired integration event was detectable by PCR. Parasites were cloned by limiting dilution thereafter and genotyped by PCR, sequencing of the diagnostic product and Southern analysis.

Induction experiments and luciferase assays

Ring-stage parasites were tightly synchronized using 0.3 M alanine for three consecutive intra-erythrocytic developmental cycles (IDCs) before induction. Synchronized parasites were induced using 0.5 μM aTc at ring stage and allowed to go through a complete IDC prior to collecting ring, trophozoite and schizont stage samples for luciferase assays. Firefly and *Renilla* luciferase measurements were made with the Dual Luciferase Assay kit (Promega) on a GloMax 20/20 luminometer (Turner Biosystems).

Western blot

Approximately 10^6 late trophozoite stage parasites were lysed by saponin treatment (0.5 g/L saponin in phosphate-buffered saline). These lysates were prepared for western blot by heating in Laemmli sample buffer at 95 °C for 10 min. After separation by SDS-PAGE, proteins were transferred to a PVDF membrane and probed with a rabbit monoclonal antibody (71D10) against the c-Myc tag (Cell Signaling Technology, Catalog # 2278S; 1:2,000 dilution), or rabbit polyclonal antibody against GAPDH (Abcam, ab9485; 1:1,000 dilution). Blots were imaged using a horseradish peroxidase-coupled secondary antibody and SuperSignal West Femto substrate (Thermo Scientific, Catalog # 34095).

Flow cytometry

Ring-stage parasites were synchronized and analyzed by flow cytometry at the trophozoite stage. Cells were stained for nucleic acid content with 1 μM SYTO 61 (Life Technologies) and analyzed using an Accuri C6 flow cytometer (BD Biosciences). Enhanced YFP (EYFP) and SYTO 61 signals were monitored in the FL1 and FL4 channels, respectively. Parasites were selectively analyzed by gating on events with high FL4 signal intensity. Adding 0.5 μM aTc to either uninfected red blood cells or parasites not expressing EYFP had no effect on FL1-H signal intensity.

Southern blot

Southern blots were carried out using gDNA isolated from schizont stage parasites at ~2-5% parasitemia in a 30 ml, at 2% hematocrit culture using the QIAamp DNA blood mini kit (Qiagen). Red blood cells were lysed using a 0.1% saponin solution. Between 2-3 μg of gDNA was restriction enzyme–digested overnight with *PstI* and *EcoRI* (New England Biolabs) and probed with a PfATP4 PCR product obtained using primers pMG291 and pMG292. The probe was labeled with biotin-11-dUTPs using the Pierce Biotin Random Prime kit (Thermo Scientific). Blots were processed using the TurboBlotter kit (Whatman) for transfer and the North2South kit (Thermo Scientific) for development.

Growth assays

Parasites were synchronized to rings using 0.3 M alanine in 10 mM HEPES (pH 7.4) at Day 0, adjusted to 1% parasitemia, and seeded in 96-well microtiter plates in quadruplicate at 2% hematocrit in 200 μl of RPMI Complete media in the presence or the absence of aTc (0.5 μM). The parental NF54^{attB} strain and engineered PfATP4 lines were all included in the assay. Expansion was measured over six IDCs, and samples were analyzed every 48 hours to determine parasitemias at the end of each IDC. After each measurement, all cultures were split by the same dilution factor required to keep the pre-invasion parasitemia of the control lines below 2% to avoid culture over-expansion. Parasitemias were measured by incubating the cells with 1:1000 dilution of SYBR Green I for 15 minutes at 37 °C, prior to flow cytometry on an Accuri C6 instrument (BD Biosciences) and analysis of data collected in the FL1 channel.

REFERENCES

Adjalley, S.H., Johnston, G.L., Li, T., Eastman, R.T., Ekland, E.H., Eappen, A.G., Richman, A., Sim, B.K.L., Lee, M.C.S., Hoffman, S.L., *et al.* (2011). Quantitative assessment of Plasmodium falciparum sexual development reveals potent transmission-blocking activity by methylene blue. Proc Natl Acad Sci USA *108*, E1214-1223.

Armstrong, C.M., Armstrong, C.M., Goldberg, D.E., and Goldberg, D.E. (2007). An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. Nat Methods *4*, 1007-1009.

Ausländer, S., Stücheli, P., Rehm, C., Ausländer, D., Hartig, J.S., and Fussenegger, M. (2014). A general design strategy for protein-responsive riboswitches in mammalian cells. Nat Methods *11*, 1154-1160.

Belmont, B.J., and Niles, J.C. (2010). Engineering a direct and inducible protein-RNA interaction to regulate RNA biology. ACS Chem Biol *5*, 851-861.

Belmont, B.J., and Niles, J.C. (2012). Inducible control of subcellular RNA localization using a synthetic protein-RNA aptamer interaction. PLoS ONE *7*, e46868.

Besse, F., and Ephrussi, A. (2008). Translational control of localized mRNAs: restricting protein synthesis in space and time. Nat Rev Mol Cell Biol *9*, 971-980.

Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J.C., and DeRisi, J.L. (2003). The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol *1*, 85-100.

Cosentino, G.P., Schmelzle, T., Haghighat, A., Helliwell, S.B., Hall, M.N., and Sonenberg, N. (2000). Eap1p, a novel eukaryotic translation initiation factor 4E-associated protein in Saccharomyces cerevisiae. Mol Cell Biol *20*, 4604-4613.

de la Cruz, J., Iost, I., Kressler, D., and Linder, P. (1997). The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in Saccharomyces cerevisiae. Proc Natl Acad Sci USA *94*, 5201-5206.

Decker, C.J., and Parker, R. (2006). CAR-1 and trailer hitch: driving mRNP granule function at the ER? J Cell Biol *173*, 159-163.

Decker, C.J., and Parker, R. (2012). P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harb Perspect Biol *4*, a012286.

Deitsch, K., Driskill, C., and Wellems, T. (2001). Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. Nucleic Acids Res *29*, 850-853.

Ernoult-Lange, M., Baconnais, S., Harper, M., Minshall, N., Souquere, S., Boudier, T., Bénard, M., Andrey, P., Pierron, G., Kress, M., *et al.* (2012). Multiple binding of repressed mRNAs by the P-body protein Rck/p54. RNA *18*, 1702-1715.

Flannery, E.L., McNamara, C.W., Kim, S.W., Kato, T.S., Li, F., Teng, C.H., Gagaring, K., Manary, M.J., Barboa, R., Meister, S., *et al.* (2014). Mutations in the P-Type Cation-Transporter ATPase 4, PfATP4, Mediate Resistance to Both Aminopyrazole and Spiroindolone Antimalarials. ACS Chem Biol.

Forster, K., Helbl, V., Lederer, T., Urlinger, S., Wittenburg, N., and Hillen, W. (1999). Tetracycline-inducible expression systems with reduced basal activity in mammalian cells. Nucleic Acids Res *27*, 708-710.

Goldfless, S.J., Belmont, B.J., de Paz, A.M., Liu, J.F., and Niles, J.C. (2012). Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction. Nucleic Acids Res *40*, e64.

Goldfless, S.J., Wagner, J.C., and Niles, J.C. (2014). Versatile control of Plasmodium falciparum gene expression with an inducible protein-RNA interaction. Nat Commun *5*, 5329.

He, F., and Jacobson, A. (2001). Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. Mol Cell Biol *21*, 1515-1530.

Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A., and Moore, C.L. (1997). Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. Genes Dev *11*, 2545-2556.

Kuersten, S., and Goodwin, E.B. (2003). The power of the 3' UTR: translational control and development. Nat Rev Genet *4*, 626-637.

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., *et al.* (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Science *301*, 1503-1508.

Lehane, A.M., Ridgway, M.C., Baker, E., and Kirk, K. (2014). Diverse chemotypes disrupt ion homeostasis in the malaria parasite. Mol Microbiol *94*, 327-339.

Mair, G.R., Braks, J.A.M., Garver, L.S., Wiegant, J.C.A.G., Hall, N., Dirks, R.W., Khan, S.M., Dimopoulos, G., Janse, C.J., and Waters, A.P. (2006). Regulation of sexual development of Plasmodium by translational repression. Science *313*, 667-669.

Mair, G.R., Lasonder, E., Garver, L.S., Franke-Fayard, B.M.D., Carret, C.K., Wiegant, J.C.A.G., Dirks, R.W., Dimopoulos, G., Janse, C.J., and Waters, A.P. (2010). Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development. PLoS Pathog *6*, e1000767.

Meissner, M., Krejany, E., Gilson, P.R., de Koning-Ward, T.F., Soldati, D., and Crabb, B.S. (2005). Tetracycline analogue-regulated transgene expression in Plasmodium falciparum blood stages using Toxoplasma gondii transactivators. Proc Natl Acad Sci USA *102*, 2980-2985.

Nkrumah, L.J., Muhle, R.A., Moura, P.A., Ghosh, P., Hatfull, G.F., Jacobs, W.R., and Fidock, D.A. (2006). Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase. Nat Methods *3*, 615-621.

No, D., Yao, T.P., and Evans, R.M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. Proc Natl Acad Sci USA *93*, 3346-3351.

Prommana, P., Uthaipibull, C., Wongsombat, C., Kamchonwongpaisan, S., Yuthavong, Y., Knuepfer, E., Holder, A.A., and Shaw, P.J. (2013). Inducible knockdown of Plasmodium gene expression using the glmS ribozyme. PLoS ONE *8*, e73783.

Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. Nat Rev Mol Cell Biol *5*, 232-241.

Rönicke, V., Graulich, W., Mumberg, D., Müller, R., and Funk, M. (1997). Use of conditional promoters for expression of heterologous proteins in Saccharomyces cerevisiae. Meth Enzymol *283*, 313-322.

Rottmann, M., McNamara, C., Yeung, B.K.S., Lee, M.C.S., Zou, B., Russell, B., Seitz, P., Plouffe, D.M., Dharia, N.V., Tan, J., *et al.* (2010). Spiroindolones, a potent compound class for the treatment of malaria. Science *329*, 1175-1180.

Rouya, C., Siddiqui, N., Morita, M., Duchaine, T.F., Fabian, M.R., and Sonenberg, N. (2014). Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. RNA *20*, 1398-1409.

Saito, H., Kobayashi, T., Hara, T., Fujita, Y., Hayashi, K., Furushima, R., and Inoue, T. (2010). Synthetic translational regulation by an L7Ae-kink-turn RNP switch. Nat Chem Biol *6*, 71-78.

Spillman, N.J., Allen, R.J.W., McNamara, C.W., Yeung, B.K.S., Winzeler, E.A., Diagana, T.T., and Kirk, K. (2013). Na(+) regulation in the malaria parasite Plasmodium falciparum involves the cation ATPase PfATP4 and is a target of the spiroindolone antimalarials. Cell Host Microbe *13*, 227-237.

Suess, B., Hanson, S., Berens, C., Fink, B., Schroeder, R., and Hillen, W. (2003). Conditional gene expression by controlling translation with tetracycline-binding aptamers. Nucleic Acids Res *31*, 1853-1858.

Tarique, M., Ahmad, M., Ansari, A., and Tuteja, R. (2013). Plasmodium falciparum DOZI, an RNA helicase interacts with eIF4E. Gene.

Wagner, J.C., Goldfless, S.J., Ganesan, S.M., Lee, M.C., Fidock, D.A., and Niles, J.C. (2013). An integrated strategy for efficient vector construction and multi-gene expression in Plasmodium falciparum. Malar J *12*, 373.

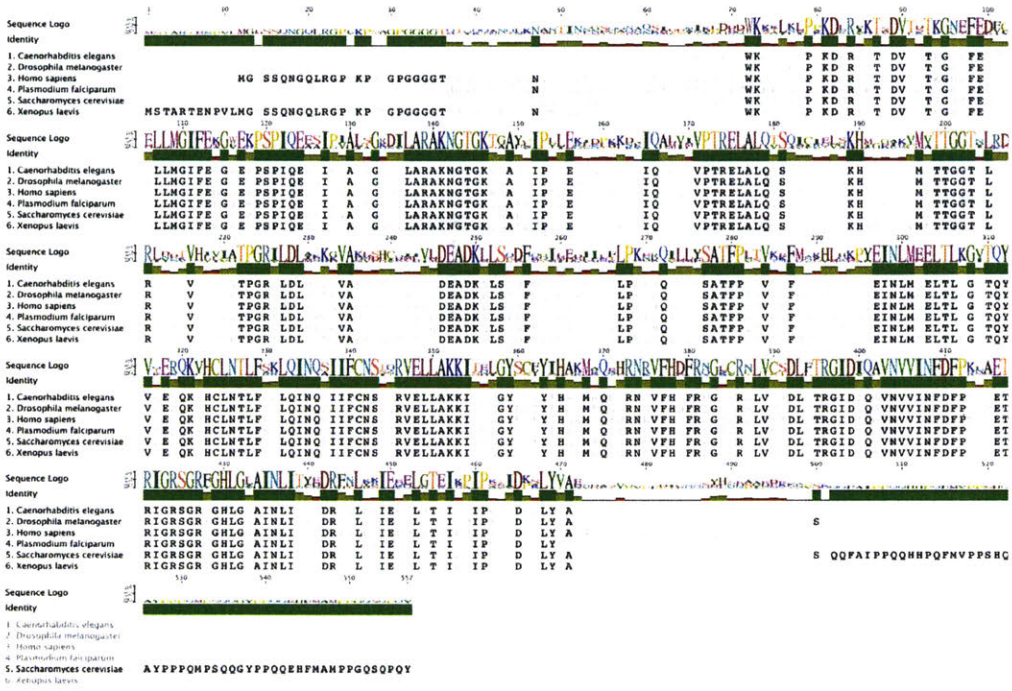
Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell *2*, 135-140.

Win, M.N., and Smolke, C.D. (2007). A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci USA *104*, 14283-14288.

Wittmann, A., and Suess, B. (2011). Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. Mol Biosyst *7*, 2419-2427.

Wu, Y., Sifri, C.D., Lei, H.H., Su, X.Z., and Wellems, T.E. (1995). Transfection of Plasmodium falciparum within human red blood cells. Proc Natl Acad Sci USA *92*, 973-977.

SUPPLEMENTARY INFORMATION



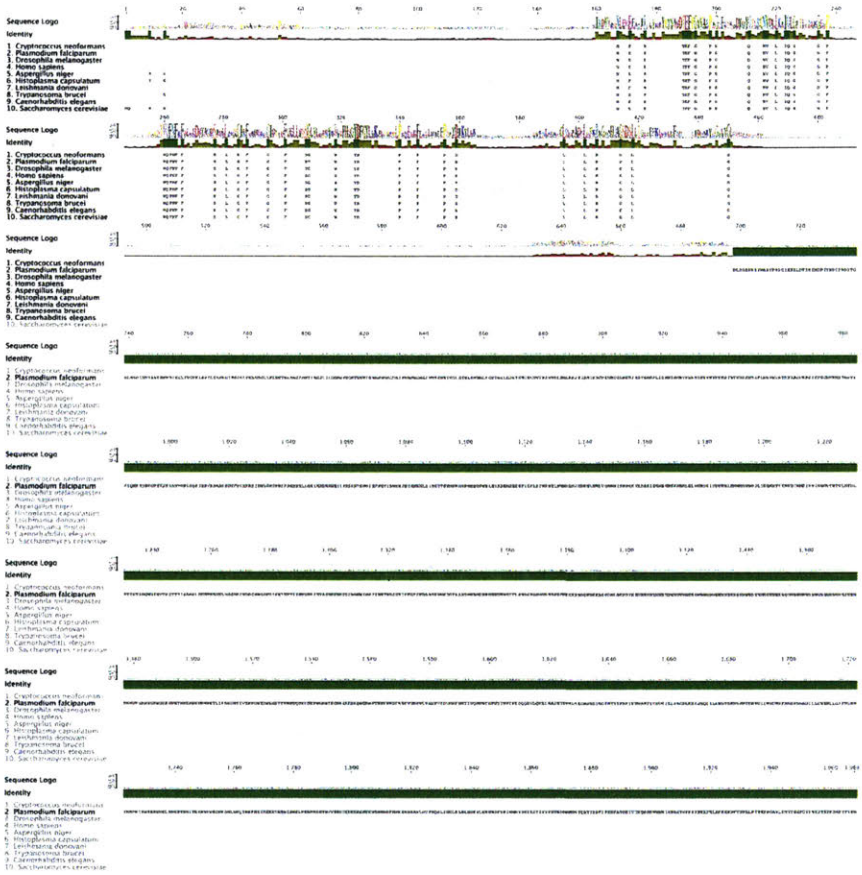


Figure 2-S1. Sequence comparison of several protein homologs from various model and non-model organisms that have been implicated in translational regulation
MUSCLE alignments of the (a) Dhh1p, (b) Dcp1p and (c) Pop2p homologs from *S. cerevisiae* with those from *C. elegans*, *D. melanogaster*, *H. sapiens*, *P.falciparum* and *X. laevis*.

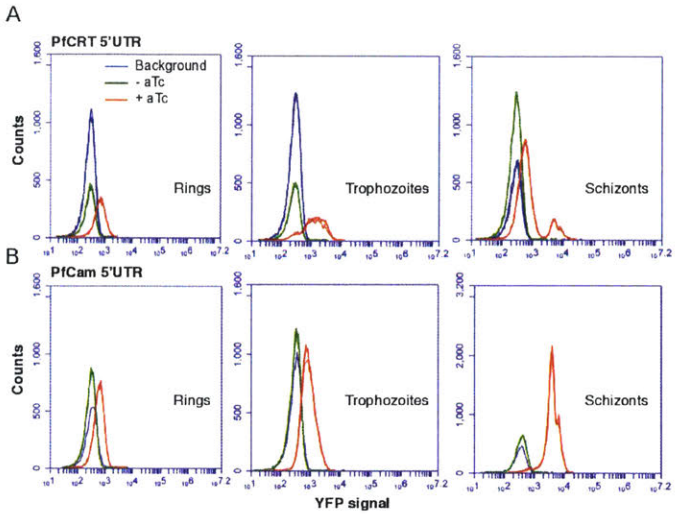


Figure 2-S2. Flow cytometry histograms showing parasite developmental stage- and aTc- dependent EYFP expression controlled by TetRDOZI

(A) PfCRT and **(B)** PfCAM promoters were used to drive expression of an EYFP reporter gene flanked by a single aptamer within the 5'UTR and ten tandem aptamers in the 3'-UTR.

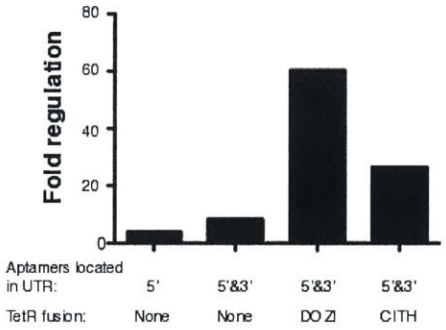


Figure 2-S3. TetR fused to the *P. falciparum* CITH homolog improves the regulatory dynamic over TetR alone

TetR-CITH regulation expression is compared with that of TetR and TetR-DOZI using a FLuc reporter flanked by a single aptamer within the 5'-UTR and ten tandem aptamers in the 3'-UTR. Regulation of FLuc by TetR via a single aptamer in the 5'UTR is included for reference.

Plasmid	Description	GenBank Accession
pMG32	PfCAM promoter driving a FLuc reporter with a single aptamer in the 5'UTR; TetR-DOZI$_{2A}$RLuc$_{2A}$BSD	
pMG53	PfCAM promoter driving a FLuc reporter with a single aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR$_{2A}$RLuc$_{2A}$BSD	
pMG54	PfCAM promoter driving a FLuc reporter with a single mutant aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR$_{2A}$RLuc$_{2A}$BSD	
pMG56	PfCAM promoter driving a FLuc reporter with a single aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR-DOZI$_{2A}$RLuc$_{2A}$BSD	
pMG57	PfCAM promoter driving a FLuc reporter with a single mutated aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR-DOZI$_{2A}$RLuc$_{2A}$BSD	
pMG62	PfCRT promoter driving a FLuc reporter with a single aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR-DOZI$_{2A}$RLuc$_{2A}$BSD	
pMG83	PfCAM promoter driving a FLuc reporter with a single aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR-CITH$_{2A}$RLuc$_{2A}$BSD	
pMG84	PfCRT promoter driving an EYFP reporter with a single aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR-DOZI$_{2A}$RLuc$_{2A}$BSD	
pMG92	PfCAM promoter driving an EYFP reporter with a single aptamer in the 5'UTR and 10 tandem aptamers in the 3'UTR; TetR-DOZI$_{2A}$RLuc$_{2A}$BSD	
pMG75	Allelic replacement plasmid to install a C-terminal epitope tag and 10 tandem aptamers in a 3'UTR context at the PfATP4 chromosomal locus to achieve regulated PfATP4 expression by TetR-DOZI.	
pSG372	PfCAM promoter driving a FLuc reporter with a single aptamer in the 5'UTR; TetR$_{2A}$RLuc$_{2A}$BSD	

Table 2-S1. List of plasmids used in this study

Primer	Sequence	Use
SMG305	tgtggtgatgttgaagaaaatccaggtccatggCCAGTGAGC AAGGGCGAGGAGCTGTTC	Forward primer used to clone EYFP into pMG84
SMG293	gatcttcttcgctaatcagtttctgttcggtaaccTTACTTGTAC AGCTCGTCCATGCCGAGAG	Reverse primer used to clone EYFP into pMG84
SMG307	GTATTAATGGAACAGAAGTAGCTAAAGGA GCATC	Forward primer for detecting integration at the PfATP4 locus
SMG291	ctaaatatatatccaatggccccttccgggcgcgccttaagCTCA AATATTCAAAAATTTGT ATG TTT CTT GCT TG	Forward primer for cloning the PfATP4 homologous region
SMG292	cacagatcttcttcgctaatcagtttctgttcggtaaccgcATTCTT AATAGTCATATATTTTCTTCTATATATAACC	Reverse primer for cloning the PfATP4 homologous region
SMG286	GTCAATTTTTAGAAGAAAGATATCCTGTAT TTAGA	Forward primer for amplifying the housekeeping MQO gene
SMG358	ATGAATAAGTGGAACATATGGAATAGATA AG	Reverse primer for amplifying the housekeeping MQO gene
SMG357	AAAAACAAGCAAAATTTTTACCACATGTA C	Reverse primer for detecting native PfATP4 locus
SMG342	actaaatatatatccaatggccccttccgggcgcgccCATTTTG TAAAAAAAATTAAAATATATTTATATAATA TTAT	Forward primer for amplifying PfCAM promoter/5'UTR
SMG343	ccgtccgtatcgaccttctctgcctggattgcttaagTGATATAT TTCTATTAGGTATTTATTATTATAAAATATA AATC	Reverse primer for amplifying PfCAM promoter/5'UTR
SMG255	actaaatatatatccaatggccccttccgggcgcgccTAGTAGT TGAGTGATTCTATATACATATAC	Forward primer for amplifying the PfCRT promoter
SMG256	ccgtccgtatcgaccttctctgcctggattgcttaagTGTTATAT GTAAGAAATTAAAAAAAACAAAATAATAA ATGAATG	Reverse primer for amplifying the PfCRT promoter
SMG173	ctggtggtggtggtggtggtggtggtactagtAGTTATAA AACCAATTGTACGAACT	Forward primer used to amplify

		dozi gene
SMG174	gttaataatgatcctctaccttcaccactaccgctagcGGTATAT AAGGATGGGTCAATTTCG	Reverse primer used to amplify *dozi* gene
SMG164	gtgggtctggtggtggtggtggtggtggtggtactagtTCATC TGTGTCAACCTTACCTTA	Forward primer used to amplify *cith* gene
SMG165	catgttaataatgatcctctaccttcaccactaccgctagc ATATGGTGGATATCTATTAAAAGGATTTTG ATTC	Reverse primer used to amplify *cith* gene

Table 2-S2. List of primers used in this study

Nucleotides shown in lowercase correspond to the homologous region needed for Gibson Assembly Cloning. Nucleotides in uppercase are complementary with the target gene/region.

CHAPTER 3

Probing the function of an essential RNA-binding protein in *P. falciparum*

ABSTRACT

RNA Binding Proteins (RBPs) play critical roles in regulating various post-transcriptional processes required for cell survival. Whole-genome studies have revealed that RBPs are highly abundant in *P. falciparum,* highlighting the importance of RNA metabolism and post-transcriptional regulation of gene expression in the biology of this parasite. Since RBPs participate in processes essential for survival in the majority of the cells, they may represent a novel class of therapeutic targets. However, the roles and essentiality of most of the annotated RBPs in the parasite genome are unknown. To address this, we used a highly efficient CRISPR/Cas9 genome editing method together with an anhydrotetracycline (aTc)-inducible TetR-aptamer translation control system to achieve conditional regulation of a putative RBP candidate gene as an initial proof-of-concept. The expression of this protein showed inducible regulation in the presence or in the absence of aTc when expressed from a synthetic promoter or from its native promoter. Knockdown of our RBP results in a severe impaired parasite growth in aTc-withdrawal growth assay, indicating this protein is essential for intraerythrocytic development. This phenotype is consistent with that observed in *Toxoplasma gondii* for the orthologous protein. However, while a global defect in splicing is observed upon disrupting this protein's function in *T. gondii*, no such defect is observed in *P. falciparum*. Knocking down expression of this protein in asexual blood stages of *P. falciparum* results in cell cycle arrest at late trophozoite stage, as determined by morphology, quantification of DNA content and analysis of nuclear markers. This protein has a nucleocytoplasmic distribution in parasites, and preliminary evidence suggests it associates with RNA in parasites, though the specific RNAs remain to be identified. Taken together, these data suggest that this is an essential RBP for blood stage parasites and it is crucial for parasite schizogony. More generally, we believe that integrative use of genome editing and conditional perturbation studies can be a useful approach for classifying RBP essentiality and opening avenues for detailed functional characterization of this important class of regulatory proteins.

INTRODUCTION

P. falciparum exhibits a complex life cycle that involves several phenotypic changes in the mosquito vector and the human host. Those changes are mediated by a tight regulation of cell-cycle dependent gene expression patterns (Le Roch et al., 2003). The enormous complexity of this parasite's life cycle is achieved with a small set of genes compared to other simpler organisms that harbor greater number of genes (Limenitakis and Soldati-Favre, 2011). To successfully achieve these multitasking functions, the parasite employs elaborate regulatory molecular mechanisms that facilitate the precise coordination of gene expression. Regulation of gene expression can operate at different levels, including transcriptional, post-transcriptional and translational regulation. The core transcriptional machinery is highly conserved with other eukaryote systems, but there is a relatively low number of transcription factors relative to the parasite's genome (Coulson et al., 2004; Hughes et al., 2010). Twenty-seven transcriptional factors that belong to the Apicomplexan-specific family (ApiAP2) have been described in *P. falciparum*, although it has been demonstrated that they do not significantly contribute to the global gene expression control (Coulson et al., 2004; De Silva et al., 2008; Painter et al., 2011).

Comparative studies between the parasite's transcriptional and proteomic profiles reveal disparities in RNA to protein concordance, suggesting that post-transcriptional mechanisms might play an important role in gene expression control (Bunnik et al., 2013; Foth et al., 2008). RNA-Binding Proteins (RBPs) are central to the different post-transcriptional processes of RNA metabolism from pre-mRNA splicing and polyadenylation to RNA modification, transport, translation and turnover. About 50% of the *P. falciparum* genes contain introns and only a small fraction of them (~4-5%) exhibit alternative splicing during the intraerythrocytic developmental cycle (Otto et al., 2010; Sorber et al., 2011). This again suggests that, post-transcriptional mechanisms must play an important role in expansion of the proteomic repertoire in the parasite. Recently, a total of 988 RBP candidates have been identified in the parasite's genome, however the

function of the majority of them is not known (Bunnik et al., 2016). In spite of their potential role in cell survival and that they are found in great abundance in the *P. falciparum* genome, this group of proteins remains understudied. In addition, whether this group of proteins has a functional redundancy in the parasite is not very well understood.

The majority of studies on RBPs in *P. falciparum* have assigned some of these proteins into the translational regulators category, with potential involvement in parasite development, sexual differentiation, gametogenesis and zygote formation. The Puf (Pumilio and fem-3 binding factor homolog), DOZI-CITH and Bruno/CELF proteins have been proposed to act as putative translational repressors during sexual stages based on their homology to translational repressors in *P. berguei* and low experimental evidence (Mair et al., 2006; Mair et al., 2010; Miao et al., 2010; Reddy et al., 2015; Wongsombat et al., 2014). However, the function of these proteins in the asexual stage is not understood yet. Recently, our group showed the contribution of PfDOZI to enhancing translational regulation when fused to a synthetic module using TetR-aptamer arrays in an aTc-dependent manner. The PfDOZI-TetR fusion showed an improvement of gene control regulation and less leaky expression in the system compared to TetR alone (Goldfless et al., 2014) in *P. falciparum* asexual stages. This might suggest addition of a translational partner to TetR could have facilitated the recruitment of TetR-transcripts complexes that showed a downstream improvement in the regulatory outcome (Ganesan et al., 2016). However, the mechanism of how specifically DOZI contributed to this regulation remains unknown. Other *P. falciparum* RBPs have been proposed as potential splicing and alternative splicing factors and their potential function associated to parasite development during asexual and sexual stages (Blythe et al., 2004; Eshar et al., 2012; Muhia et al., 2003; Singh et al., 2004). Despite these efforts, the molecular mechanisms of how these proteins can directly affect parasite development is not known.

To ensure progression through the different stages of their life cycle and final transmission, apicomplexans such as *P. falciparum* and the related *Toxoplasma gondii*,

utilize a very coordinated cell cycle division program that consists in iterative cycles of the G1, S and internal mitotic phases (Behnke et al., 2010). While *T. gondii* replicates intracellularly during the asexual stage following standard and synchronous G1/S/M phases that results in two daughter parasites after internal division (Gubbels et al., 2008), in *P. falciparum* the cycle phases are not clearly demarcated into distinct G1/S/M. Nuclear division in *P. falciparum* is asynchronous and consequently there is not binary division after each round of replication. It is known the parasites go through a G1 phase, followed by a DNA synthesis phase and later undergo multiple rounds of mitosis, however the exact timing of those phases is not well defined (Arnot et al., 2011). The parasite nucleus multiples several times to produce ~20-32 nuclei within a shared single cytoplasm to give rise to a multinucleated cell. This process is called schizogony. During late schizogony there is a rupture of the nuclear envelope, which releases new parasite daughters to start a new invasion cycle. The division rate of each nucleus within the nuclear envelope is not the same, and thus the outcome in the number of parasites produced in each division cycle varies. Recent studies have shown that *Plasmodium* cyclin kinases (PfCKs) are key players in regulating the cell cycle in asexual stages of *P. falciparum* (Solyakov et al., 2011; Tewari et al., 2010). Indirect or direct inhibition of certain PfCKs is associated with defects in schizogony and parasite development in the mosquito vector (Ganter et al., 2017). As of this date, no RBPs have been implicated in the regulation of schizogony in asexual stages of *P. falciparum*.

Recently, *Tg*RRM1 was identified as a novel and essential RBP involved in cell cycle regulation and splicing in asexual stages of *T. gondii* (Suvorova et al., 2013). Temperature sensitive mutants of TgRRM1 showed an arrested phenotype in G1 phase, and this was associated with a global splicing defect (Suvorova et al., 2013). As TgRRM1 interacts with the U4/U6.U5 tri-snRNP (small nuclear ribonucleo proteins) complex, disruption of its function within this complex may underlie the observed accumulation of unspliced mRNA (Suvorova et al., 2013). TgRRM1 contains a single and a highly conserved RRM (RNA-Recognition Motif domain) with other eukaryotes, including an RRM1-ortholog protein in *P. falciparum* (Fig 3-1A,B). Besides the N- and

C- terminal tails of the protein, the RRM domain has 93% of similarity with a putative RBP in *P. falciparum* (Suvorova et al., 2013), that we will name PfRRM1 (*PF3D7_1360100*). This protein is annotated in the *P. falciparum* genome as a "putative RNA-Binding protein" with unknown function (Aurrecoechea et al., 2009). No previous reports in the literature have described the function of this protein in blood stage *P. falciparum*.

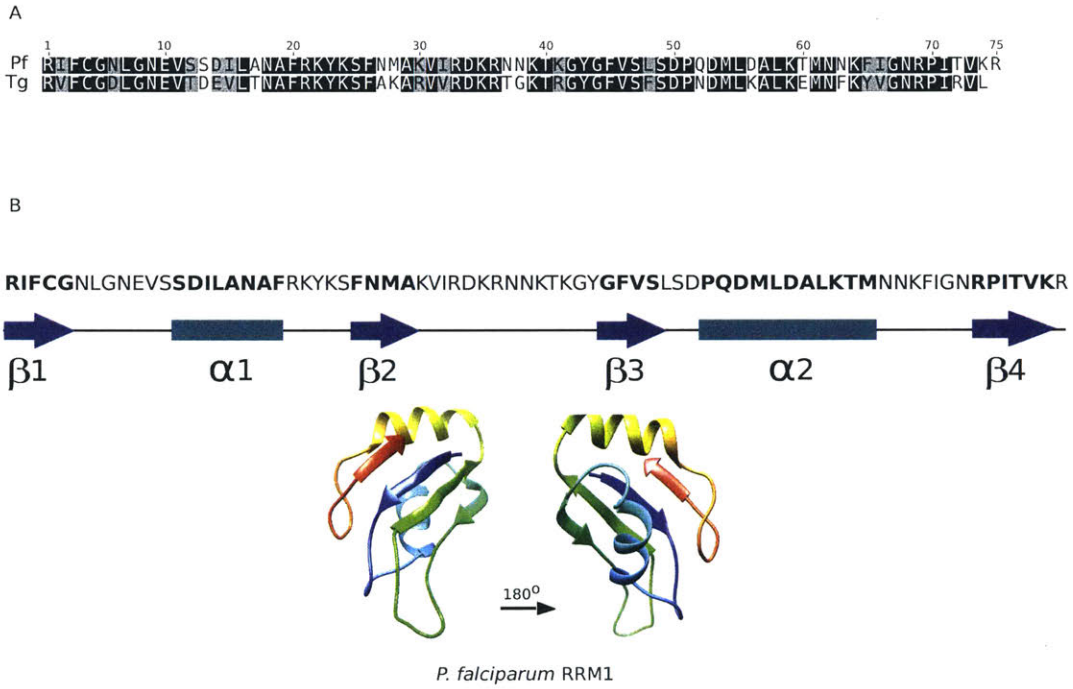


Figure 3-1: Sequence similarity and subdomain structure conservation of TgRRM1 and PfRRM1

A) Primary sequence similarity observed between the RRM1 domains of *P. falciparum* (Pf) and *T. gondii* (Tg). Black boxes indicate identities and gray boxes similarities between residues of the two orthologs. **B)** Predicted tertiary structure of the conserved RRM1 domain of the PfRRM1 protein. Conserved binding sequences are indicated in the blue and green boxes. The model was predicted using the primary sequence of the conserved RRM1 domain located at the 51-226 residues of the PfRRM1 protein using the best matching template 2mpu.1.A (37% similarity) in (http://swissmodel.expasy.org/) and edited on Chimera v1.11.1. This template was modeled using the HvGR-RBP1 protein, a glycine-rich RNA binding protein implicated in the regulation of barley leaf senescence and environmental adaptation in plants (Tripet et al., 2014). The structure shows the conserved four beta-sheets and two alpha-helixes observed in the TgRRM1 of *T. gondii* (Suvorova et al., 2013).

The level of conservation between the *T. gondii* and *P. falciparum* orthologs suggests that both proteins might have a similar function in these two closely related parasites. To demonstrate the functional similarity, Suvorova et al. performed complementation assays using wild type PfRRM1 and the human ortholog (RBM42) in mutant *Tg*RRM1 lines. Interestingly, the human ortholog, but not PfRRM1, restored the expected wild type phenotype to levels similar to the control experiment using native *T. gondii* copy. However, by creating a chimeric *Pf*RRM1 that included the N- and C- terminal extensions of *Tg*RRM1, rescue of the wild type phenotype was possible (Suvorova et al., 2013). These results suggest that although the principal RRM1 domain is highly conserved, other structural features of these orthologs are required to accomplish their function in their native hosts. More direct evidence of a possible essential role of PfRRM1 in *P. falciparum* comes from examination of a *piggybac* mutant containing an insertion into the 5'UTR of PfRRM. These parasites exhibit an attenuated asexual blood stage growth phenotype (Balu et al., 2010), but this mutant was not characterized further.

Defining the essentiality of a gene in *P. falciparum* is difficult due to the limited availability of generalizable tools for studying genes that are required to parasite survival in blood stages. The development of new technologies to control the expression of essential genes is crucial for a greater understanding of essential cellular processes of the parasite that ultimately could lead to the identification of new therapeutic targets.

Here we demonstrate the function of PfRRM1 as an essential RBP involved in cell cycle progression during the intraerythrocytic parasite development. We integrate use of CRISPR/Cas9 gene-editing and TetR aptamer conditional expression technologies to regulate PfRRM1 expression for further functional studies. We show that we can regulate the expression of PfRRM1 from either a synthetic promoter or its native promoter, and validate that PfRRM1 is essential for parasite survival. In addition, we determined that PfRRM1 regulates schizogony. We show that depletion of PfRRM1 does not cause a global splicing defect, which contrasts with the observation accompanying with loss of

TgRRM1 function in *T. gondii*. From a technical standpoint, we also showed how we streamlined our technology from a two-step approach to a single-step methodology to efficiently perform functional screens in *P. falciparum* that target essential and non-essential genes. Our integrated technology has the potential to serve as a platform for determining gene essentiality and performing downstream functional studies in *P. falciparum*.

RESULTS

PfRRM1 shows inducible regulation and essentiality in engineered merodiploid parasite line

Previous reports indicate that PfRRM1 might be an essential gene for *P. falciparum* blood stages (Balu et al., 2010; Suvorova et al., 2013). In addition, unsuccessful attempts to disrupt this gene in our group using conventional genetic tools suggest that this protein might be crucial for parasite survival, thus we hypothesized that PfRRM1 is essential. The genetic manipulation of this parasite is challenging due the lack of robust and efficient technologies to target essential genes. In addition the haploid genome of blood stage parasites constrains a straightforward methodology to target essential genes. Hence, before attempting disruption of the native PfRRM1, we engineered a merodiploid parasite line on a parental line that contains an *attB* site at the neutral *cg6* locus that allows chromosomal site-specific integration mediated by a Bxb1 integrase (Nkrumah et al., 2006). The merodiploid line expresses a TetR-DOZI–regulated copy of PfRRM1 expressed from the PfCRT promoter/5'UTR integrated at the *cg6* locus. Previously, we characterized the inducible activity of the *P. falciparum* chloroquine resistance transporter promoter (*PfCRT*; PF3D7_0709000). Using fluorescent and luminescence reporter genes, it was shown that the *PfCRT* promoter exhibited peak expression activity during trophozoite stages (24-32 hours post-invasion) (Ganesan et al., 2016). This pattern of expression is consistent with that predicted for PfRRM1 (24 hours post-invasion based on transcriptome profiling data available via PlasmoDB (Aurrecoechea et al., 2009). For this reason, we decided to use the *PfCRT* promoter to drive the transcription of the merodiploid copy of PfRRM1.

To do this, we built the pMG79:PfCRT-PfRRM1 construct to overexpress a recoded version of PfRRM1 under the activity of the *PfCRT* promoter. This PfRRM1 synthetic copy is flanked at the 5' and at the 3' end by a single TetR-aptamer and a 10x array TetR-aptamer respectively. In addition, a HA-epitope was included to tag the protein at the N-

terminus region. Previously, we showed the enhanced regulatory outcome, when using the TetR-DOZI module and the double aptamer regulation context (Ganesan et al., 2016). On the same construct a TetR/DOZI/Renilla Luciferase/Blasticidin cassette is expressed from the 5'UTR Hsp86 promoter (Fig 3-2A). The Renilla Luciferase reporter gene is used as an internal reference reporter to monitor transfections growth and for further quantitative parasite growth assays.

Next, we integrated the PfCRT: PfRRM1 construct into the *NF54attB* parental strain using a second Bxb1-expressing vector (Fig 3-2A) and obtained parasite clones via limiting dilution to assess conditional expression of PfCRT:PfRRM1. We confirmed integration of the plasmid at the *cg6* locus by PCR with sequencing (Fig 3-S1) and grew parasites either in the absence or presence of anhydrotetracycline (aTc) to perform western blot analysis. We confirmed stringent, aTc-dependent regulation of PfCRT_PfRRM1 expression, similar to what it was observed on the quantitative reporter gene assays (Fig 3-2B) (Ganesan et al., 2016). To rule out any possible fitness cost associated with overexpression of PfRRM1, we performed growth assays in the presence or absence of aTc. Using flow cytometry to quantify parasitemia over 5 cell divisions, we did not observe any growth defect when over-expressing PfRRM1 (Fig 3-2C).

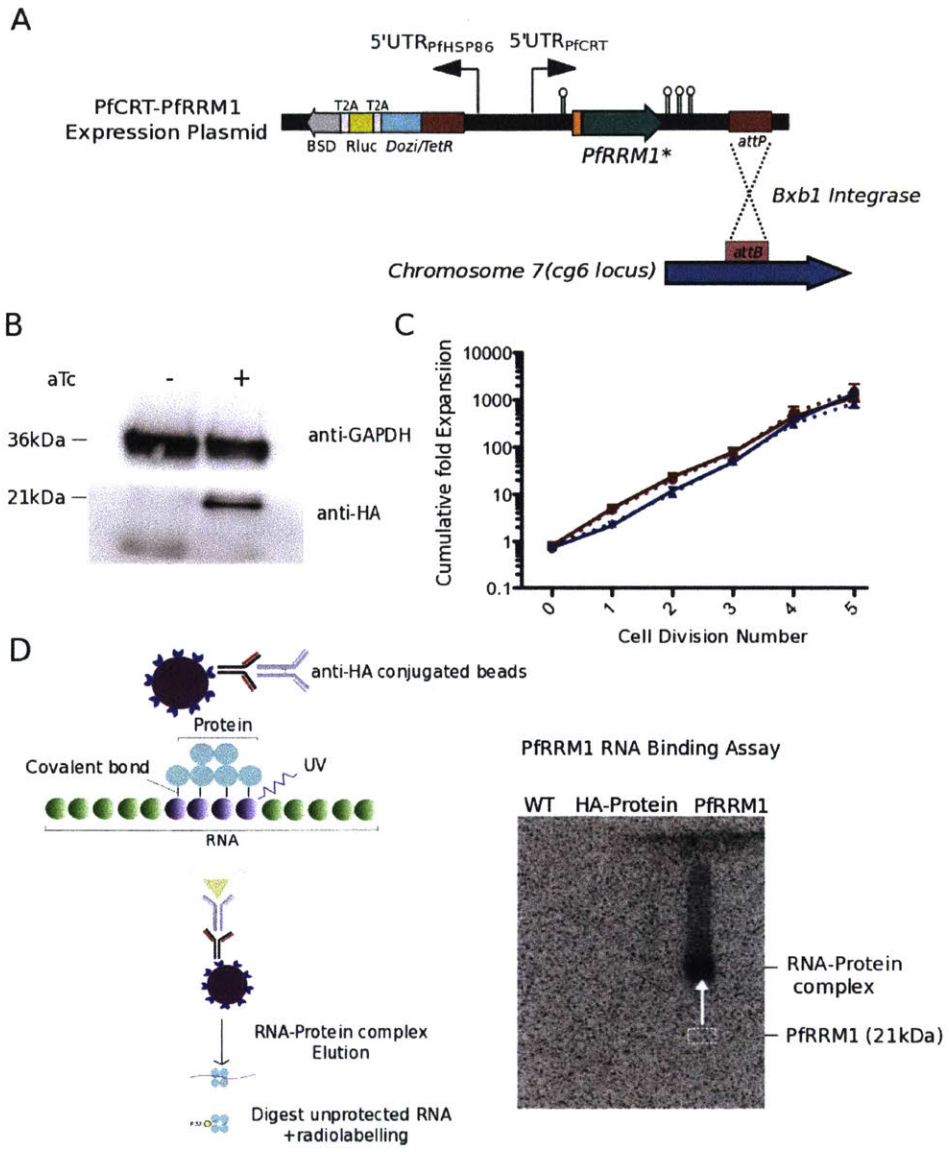


Figure 3-2: Inducible regulation of PfRRM1 on an engineered merodiploid parasite line

A) A genetically encoded system was designed to express a regulated recoded copy of PfRRM1 transcribed from the synthetic *PfCRT* promoter. A synthetic fragment encoding PfRRM1 was cloned between a 1X and a 10X TetR-aptamer array at the 5' and at the 3' respectively. The TetR/DOZI-Renilla Luciferase-Blasticidin expression cassette was included and it is driven by the promoter PfHSP86 on the sample plasmid. A 1X-HA tag was included at the N-terminus of the protein to confirm expression profile. For site-specific chromosomal integration an *attP* site was incorporated on the plasmid to recombine at the engineered and pre-existent *attB* site at the neutral cg6 locus (Nkrumah et al., 2006). **B)** Western blot protein profiling shows a tight inducible regulation of PfCRT:PfRRM1 in an aTc dependent-manner. An anti-HA antibody was used to detect tagged PfRRM1 (21kDa) and an anti-GAPDH was used to detect endogenous GAPDH (36kDa) as a loading control. **C)** Overexpression of a putative essential RBP did not confer a fitness cost to the parasites when both PfCRT:PfRRM1 (in the presence of aTc) and native PfRRM1 were expressed. Synchronized ring parasites were exposed to – and +aTc conditions and growth

was monitored for 5 generations. Wild type parasites were used as a control. **D)** PfCRT:PfRRM1 shows RNA-Binding activity. UV cross-linked parasites were subject to pull down assays using conjugated anti-HA agarose beads. Recovered RNA-protein complexes were P^{32} labeled and transferred to a nitrocellulose membrane to detect RNA binding activity. An HA-tagged non-RNA Binding protein and wild type parasite (with no HA tags) lysates were included as negative controls. The RNA-protein complex is about 41kDa in size, 20kDa bigger than PfRRM1 alone (21kDa).

PfRRM1 has been annotated as a putative RNA-binding protein, although the RNA-binding activity of this protein has not yet been established. To determine whether PfRRM1 interacts with RNA in parasites, we performed anti-HA pull-down assays on the engineered *PfRRM1* to determine if PfRRM1 displays RNA-binding activity. In addition to the PfCRT:PfRRM1 parasites, we included two negative controls in the experiment; a HA-tagged membrane protein previously characterized in our lab that should not have RNA-binding activity and a wild type negative control with no HA-tag. Parasite cultures at late stage were irradiated with UV light to preserve nucleic acid-protein interactions and a pull down was performed using HA-conjugated agarose beads to enrich for the *PfRRM1*-RNA complex. The pull-down complex from the three samples was radiolabelled and transferred to a nitrocellular membrane to detect RNA binding activity. RNA-binding activity was detected in the *PfRRM1* lane, not in the negative controls lines (Fig 3-2D). These results confirmed the PfRRM1 RNA-binding activity in the parasites, however the identity of the RNA-interacting species with this protein is yet to be determined.

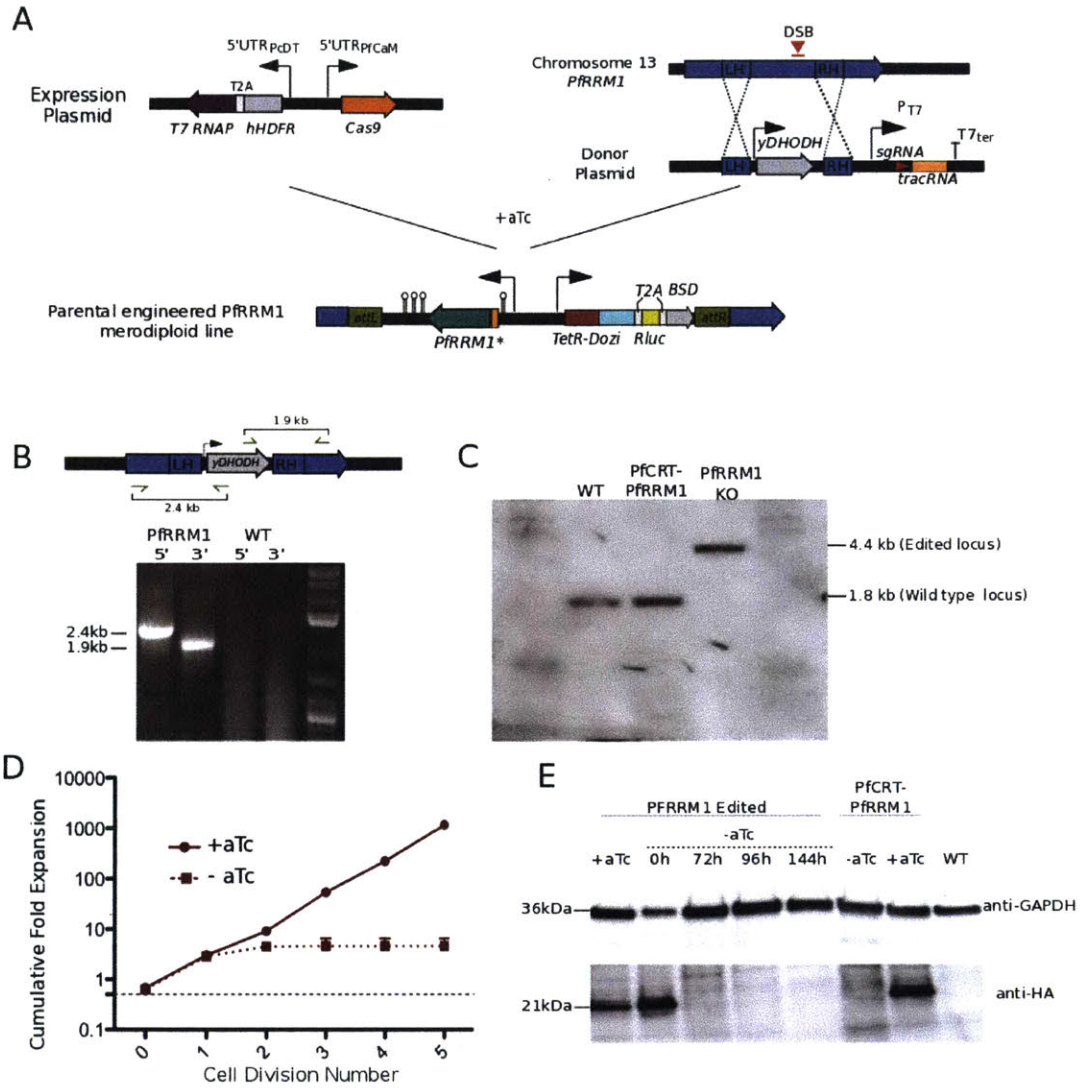


Figure 3-3: Native PfRRM1 disruption using CRISPR/Cas9 on engineered merodiploid line background

A) A genetically encoded system was designed to constitutively express Cas9 and T7RNAP from a single expression plasmid. Using a pfYc base plasmid, T7RNAP and hDHFR coding regions were cloned under the control of the PcDT promoter and Cas9 was expressed from PfCAM with a strong promoter activity. A pUF-based donor plasmid was used to clone a homologous template to repair Cas9 double-strand break using left and right regions that target the endogenous PfRRM1. A crRNA expression cassette was included on the same vector to deliver the PfRRM1-sgRNA using the T7 promoter. Cas9-expression and donor plasmids are co-transfected on late stage PfCRT:PfRRM1 clonal population parasites in the presence of aTc. **B)** Endogenous PfRRM1 gene is edited using Cas9. Integration event on stable parasite line was selected using DSM1 to obtain yDHODh resistant parasites. Specific primers hybridize on the yDHODh coding region and outside of the designed PfRRM1 homologous regions. Integration events at the 5' and 3' ends confirmed the disrupted locus of PfRRM1 in edited parasites but not in wild type parasites (negative control). **C)** Cas9 editing efficiency to target PfRRM1 was confirmed in 100% of the population using southern blot as a proxy for the frequency loci in the population. Edited PfRRM1, non-edited PfCRT:PfRRM1 merodiploid and wild type parasite gDNA samples were enzyme-digested and subject to

biotin-labeled probe hybridization. Editing was confirmed only on PfRRM1 Cas-edited parasites by showing a shift on the band size (4.4 kb edited PfRRM1, 1.8kb wild type) that represents the integration event of the donor plasmid after DNA repair. **D)** PfRRM1 is essential for parasite proliferation in asexual stages. Parasite growth was monitored via flow cytometry after aTc was withdrawn from PfRRM1 ring stage-edited parasites. A growth defect was observed during the second cell division and with no evidence of parasite expansion after 5 cell divisions. Dotted line indicates the detection limit for flow cytometry. **E)** PfRRM1 protein knockdown is observed after 72 hours of aTc withdrawal, which is consistent with the timing of the parasite's growth defect starting at the second cell division. A time course experiment using – aTc conditions showed PfRRM1 protein knockdown at different time points after 72 hours of aTc withdrawal on edited parasites. Protein profile was compared to PfCRT-PfRRM1 non-edited (and parental strain) in – and + aTc conditions and wild type parasites (negative control). An anti-HA antibody was used to detect tagged PfRRM1 (21kDa) and an anti-GAPDH was used to detect endogenous GAPDH (36kDa) as a loading control.

CRISPR-Cas9 mediated disruption of the PfRRM1 locus reveals its essentiality in blood stage parasites.

We used a CRISPR/Cas9 approach (Wagner et al., 2014) to delete the native *PfRRM1* locus in our merodiploid line (Fig. 3-3A). The recoded second copy of *PfRRM1* is immune to Cas9 cleavage by the sgRNA used to target the native locus. Edited parasites were recovered from the transfection and disruption of the *PfRRM1* native locus was confirmed by the amplification of the donor-vector integration event at the 5' and 3'-ends of the recombination sites (Fig 3-3B). To determine Cas9-efficiency on the *PfRRM1* edited population, we performed a southern blot analysis on the parental $NF54^{attB}$, *PfCRT_PfRRM1* merodiploid and *PfRRM1* knockout parasite gDNA lines, and confirmed that ~100% of the population was edited at the native locus (Fig 3-3C). These results are consistent with the previously documented Cas9-efficiency when targeting non-essential genes in *P. falciparum* (Ghorbal et al., 2014; Wagner et al., 2014)

To determine the essentiality of PfRRM1, we performed growth assays aimed at detecting a possible fitness cost/loss of viability associated to with PfRRM1 knockdown in clonal parasites. We synchronized parasites to ring stage, split them into –aTc (*PfCRT_PfRRM1* is no longer expressed) and +aTc conditions, and monitored parasite

growth over multiple cell cycles using flow cytometry. We observed a slight growth defect after one cell cycle, but a more severe defect after two cell cycles for parasites grown without aTc compared to with aTc (Fig 3-3D). The timing of when the phenotype of the edited parasites is observed on the growth assays is positively correlated with the knockdown of the protein ~72 hours after aTc withdrawal, when the protein is no longer detected (Fig 3-3E). These results confirmed that PfRRM1 is essential for growth of blood stage *P. falciparum*.

Here we showed how we have combined a post-transcriptional regulation system with a CRISPR-Cas9 genome-editing technology to identify and study an essential RNA-binding protein in *P. falciparum*. Other RBPs have been shown to be essential role for asexual proliferation of blood stage parasites (Eshar et al., 2012; Vembar et al., 2015), however to our knowledge PfRRM1 is the first essential RBP that has been edited using CRISPR-Cas9 technologies in asexual stages of *P. falciparum*. This new strategy can be implemented to study other essential genes on this parasite, overcoming the limitation of genome-editing tools to target only non-essential genes on the haploid genome of the parasite. This approach also serves as a versatile and convenient method to describe physicochemical and biochemical features of essential and non-essential proteins in the parasite.

Efficient disruption and conditional regulation of *PfRRM1* at the native promoter reveals a consistent essential phenotype on *P. falciparum* blood stages.9

The approach described above is a two-step strategy that requires the incorporation of a recoded copy of the putative essential gene driven by a synthetic promoter in an ectopic location. Although this strategy is highly valuable and offers a valid method to identify essential genes, the fact that the copy is in a genomic context different than the native locus affects the interpretation of the mechanisms that underlie its essentiality. To overcome this limitation, we developed an integrated technology that targets and regulates the expression of a gene at its native locus in a single editing step. For this

purpose, we created a *P. falciparum* strain derivative of NF54attB that constitutively expresses the Cas9 endonuclease and the *hDHFR*-T7RNAP gene fusion from the cg6 locus, which eliminates the need of co-transfecting an expression plasmid when performing gene disruption, To achieve a one-step gene editing that involves the introduction of the inducible regulation elements in the native locus is necessary to incorporate in a single construct all the elements previously present in the donor plasmid in addition to the TetR-DOZI/Blasticidin (BSD)/Renilla (Rluc) luciferase expression cassette and the 10X aptamer array (Fig. 3-4A). Due to the AT-richness of the *P. falciparum* genome, the use of large conventional circular plasmids often results in unwanted recombination and/or deletions events. To circumvent this problem, we combined all the CRISPR-Cas9 elements and the TetR/DOZI components in a single and stable pJazz linear vector that allows efficient cloning of low complexity regions, fragments with expanded repeats and stably maintains up to 30-40kb inserts (Godiska et al., 2010). The pJazz system consists of a linear plasmid, originally isolated from a lysogenic phage, that stably replicates in *E. coli* and prevents supercoiling and concatamer formation. The pJazz system has been used in eukaryote systems, such as mammalian cells, showing robust expression of several transcription units in a single linear construct (Guye et al., 2013).

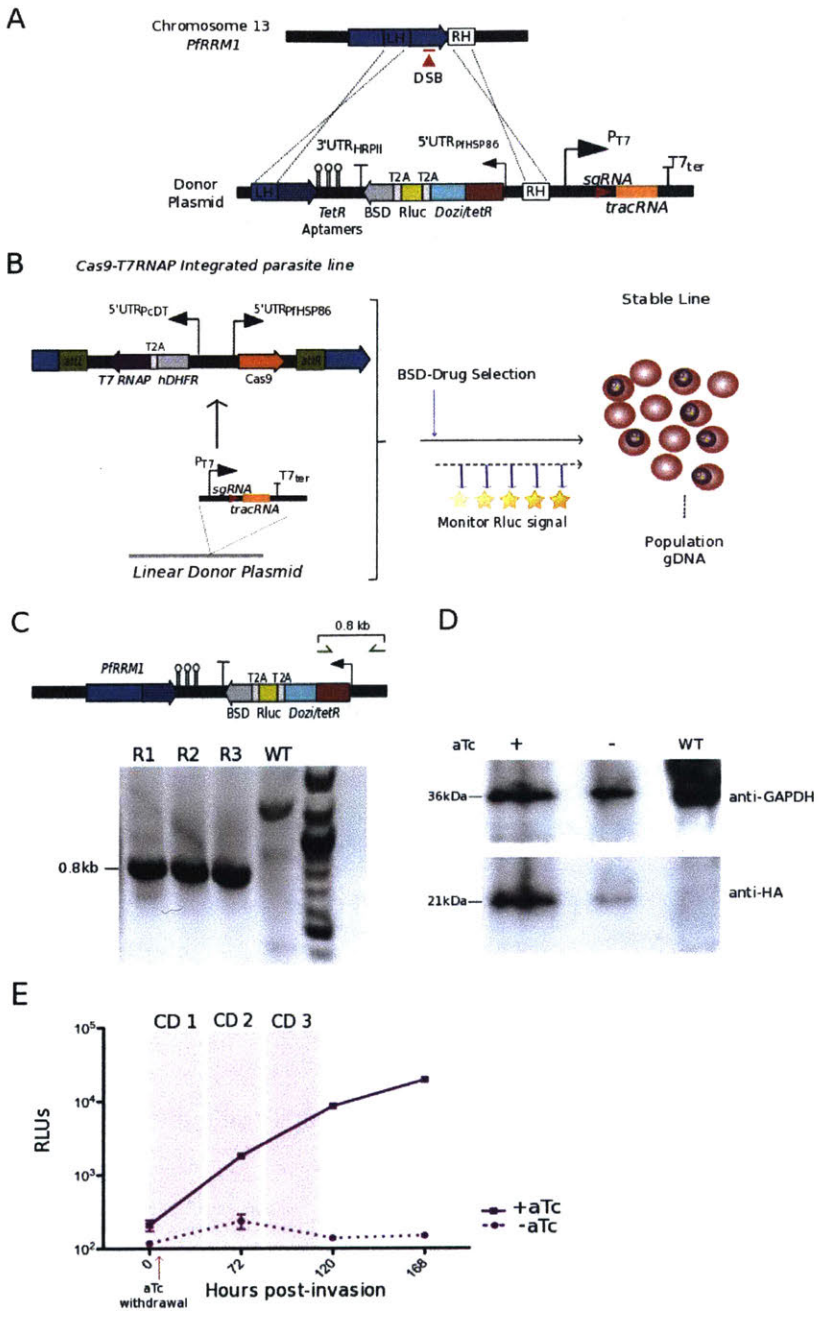


Figure 3-4: Development of integrated technology to target and inducibly regulate PfRRM1 at its native locus reveals consistent essential phenotype in *P. falciparum* blood stages

A) Development of a new strategy that integrates the TetR-DOZI-aptamer system and CRISPR/Cas9 elements into a single unit. The TetR-DOZI/Renilla Luciferase/Blastidicidin expression cassette and the 10X TetR-aptamer array were transferred in a single step from our previous conventional circular vector into the pJazz linear vector. A base plasmid was built (pSN053) with modular features to easily adapt this new donor plasmid to target any *P. falciparum* gene.. Left homologous region was ligated to a synthetic recoded DNA fragment homologous to approximately ~200 bp of the 3' end of the gene to prevent Cas9-

recutting. A 3X-HA tag was included upstream of the TetR-aptamer array to tag the protein at the C-terminus. Right homologous region is homologous to the 3'UTR of the native gene. The Cas9 T7 cRNA cassette was built into the donor vector first outside of the homologous region, to clone any sgRNA into a unique restriction site using a pair of oligos. **B)** Integrating Cas9 genetic components and inducible system in blood stage parasites to edit the endogenous copy of PfRRM1. The pSN053-PfRRM1 donor plasmid was transfected into a previously developed Cas9/T7RNAP-integrated parasite line. Renilla luciferase signal was monitored regularly along the time of the transfection. **C)** Cas9 shows editing activity on the PfRRM1 endogenous locus. *Top panel:* edited PfRRM1 locus with integration event of the donor plasmid. The coordinates of primer hybridization sites and the expected PCR integration products at the 3' end of the targeted region are indicated. *Bottom panel*: Amplification products that correspond to the editing event on the PfRRM1 locus are shown at the expected size. The experiment was performed in 3 (Replicates 1-3) independent transfections that showed the same results. The integration event was detected for the first time 2 weeks post-transfection. **D)** Western blot protein profiling demonstrates a tight, aTc-dependent regulation of PfRRM1 on edited parasites. An anti-HA antibody was used to detect tagged PfRRM1 (21kDa) and an anti-GAPDH was used to detect endogenous GAPDH (36kDa) as a loading control. **E)** Edited PfRRM1 parasites showed a growth defect phenotype 72 hours after aTc withdrawal, consistent with the previous phenotype described using PfCRT: PfRRM1 edited parasites. Synchronized ring parasites were exposed to – and + aTc conditions and renilla luciferase signal was monitored for 3 cell divisions (CD), starting 72 hours after the aTc withdrawal.

Our group has shown stable site-specific chromosomal integration of linear vectors and a robust translational regulation in an aTc-dependent manner along the parasite life cycle using reporter genes (Fig 3-S2), similar to the regulation observed using conventional plasmids (Ganesan et al., 2016). We set out to target the native PfRRM1 locus using the CRISPR-Cas9 system and a pJazz-based vector as a donor plasmid, to integrate the TetR/DOZI components during DNA repair at the Cas9 cut site, and regulate the expression of PfRRM1 when expressed from its native promoter. For this, we created a pSN053 linear vector derivative that includes homologous regions that target the native PfRRM1. The homologous regions flank the TetR-DOZI/aptamer regulatory machine and a Cas9 PfRRM1-sgRNA was cloned into a pre-existent T7 promoter expression cassette outside of the homologous regions (Fig 3-4A). In addition, a 3X-HA tag was included upstream of the aptamer array to tag the protein at the C-terminus and facilitate further identification. The resulting plasmid was 21kb in size and displayed stable replication in *E. coli* with no evidence of deletions or genomic rearrangements.

To test the one-step editing system we transfected the pSN053-PfRRM1 donor plasmid into late blood stage parasites in the presence of aTc (Fig 3-4B). Integration events were observed in stable BSD and Rluc-positive parasite lines from different experimental

replicates as early as two weeks post-transfection. PCR analysis and sequencing of the 5' and the 3' ends of the disrupted *PfRRM1* locus confirmed the effectiveness of the system in introducing regulatory elements at the native locus in a single editing step (Fig 3-4C) Protein profiling of the edited parasite lines revealed a tight regulation of PfRRM1 expression from its native promoter, similar to the regulation observed in the merodiploid line using the synthetic promoter (Fig 3-4D). In spite of the tight regulation, a faint band of the PfRRM1 protein is observed in the –aTc conditions. This leaky expression is expected due to the presence of only one TetR-aptamer array installation at the 3' end of the targeted gene. Higher levels of repression might be achieved by installing TetR-aptamer arrays at both the 3' UTR and 5' UTR regions of the gene (Ganesan et al., 2016).

We first demonstrated the essentiality of the PfRRM1 protein for asexual proliferation by using a PfRRM1 knockdown merodiploid line. Those parasites showed a growth defect at the second IDC when aTc was withdrawn at ring stage. We hypothesized that our new parasite line targeting the native PfRRM1 with regulation at its native promoter should show the same essential phenotype. To confirm this finding, we performed an aTc withdrawal growth assay using synchronized ring stage parasites where we monitored parasite growth during 3 cell divisions using the renilla luciferase signal. Consistent with our previous observations, PfRRM1 knockdown parasites showed a growth defect 72 hours after aTc withdrawal, confirming the essentiality of PfRRM1 for intraerythrocytic developmental growth in *P. falciparum* (Fig 3-4E). These results demonstrate that Cas9-based gene editing technologies can be used to probe the essentiality of genes in *P. falciparum* by the incorporation of translational regulatory elements into the native locus of the gene. In this case, the construction of a stable line of parasites where inducible regulation of the PfRRM1 gene is possible enables us to perform experiments that generate data from a more relevant biological context, and consequently allows a more meaningful interpretation of the mechanism of action of this protein in asexual stages in *P. falciparum*.

PfRRM1 is essential for schizogony development in *P. falciparum* blood stage parasites

The *T. gondii* ortholog (TgRRM1) of PfRRM1 showed an essential phenotype in asexual tachyzoite forms of the parasite. Further characterization of this phenotype using temperature-sensitive mutants of TgRRM1 demonstrated that this RBP is an essential cell cycle regulator for *T. gondii* (Suvorova et al., 2013). TgRRM1 mutants showed an arrested phenotype during G1 phase, which is characterized by the appearance of high proportion of haploid parasites (1N) with no evidence of DNA duplication(Suvorova et al., 2013). Although there is a high conservation in the secondary structure of the RRM domains of the *T.gondii* and the *P. falciparum* RRM1 proteins (Fig 3-1B), complementation assays in *T. gondii* with PfRRM1 failed to restore the wild type phenotype, which suggests that even thought these two proteins are homologous and are essential for survival, they might accomplish regulatory functions of cell cycle in different ways (Suvorova et al., 2013). We already have shown the essentiality of this protein for blood stage parasites and to move forward into establishing the function of this protein, we hypothesized that PfRRM1 plays a role in cell cycle regulation in asexual stages of *P. falciparum*.

The mechanisms of cell cycle division and regulation of cell cycle progression in *P. falciparum* are largely unknown, especially the timing of transition between the different phases that occur throughout the intraerythrocytic cycle. Some nuclear markers such as α-tubulin and telomerase-based markers have been described to show nuclear division and mitotic divisions. The use of these nuclease markers have helped to find an association between the dynamic parasite nuclear morphology, cytokinesis and the different cell cycle phases that *P. falciparum* undergoes during the asexual stage (Arnot et al., 2011; Ganter et al., 2017). Some Plasmodium cycline or cycline-dependent kinases have shown to have a role in cell cycle regulation during the asexual proliferation (Solyakov et al., 2011; Tewari et al., 2010). In particular, some of those cycline-dependent kinases (PfCKs) have shown to be essential for schizogony and development in the mosquito

vector (Ganter et al., 2017; Kern et al., 2014). Others have shown inhibition of the isoprenoid biosynthesis also causes intracellular parasites to arrest during schizogony (Howe et al., 2013). Even though there is not much evidence about direct or indirect inhibition of cell cycle regulators in *P. falciparum* that affect cell cycle progression, it seems that presumably a multifactorial network of elements involved in different pathways might be contribute to the different checkpoint pathways that govern cell cycle regulation in *P. falciparum*.

To get insights into the role of PfRRM1 in cell cycle regulation, we first determined the cellular localization of this protein using immunofluorescence microscopy. PfRRM1-3XHA localizes at the nuclear periphery of the parasite, showing a partial co-localization with DAPI - a nuclear stain- at rings and trophozoite stages. As the parasites progress into the schizont stage, the protein showed a more co-localized diffuse signal that overlaps with the nucleus (Fig 3-5A). These findings are consistent with the localization of TgRRM1 and the PfCKs involved in schizogony development. This localization is also consistent with the localization of the human ortholog RBM42 (Fukuda et al., 2009). Other RBPs in *P. falciparum* have also been localized to the nuclear compartment or have been described as shuttle proteins that move between the cytoplasm and the nucleus (Eshar et al., 2012; Vembar et al., 2015). This dual-compartment pattern is similar to what has been reported for some RBPs in mammalian cells (Caceres et al., 1997). The dynamics of localization of the PfRRM1 protein during cell cycle transitions suggest an involvement of this protein in nuclear processes.

Growth assays on PfRRM1 knockdown parasites showed a growth defect during the second cell division when aTc is withdrawn at ring stage after parasites invade (Fig. 3-4E). Immunofluorescence experiments for cellular localization of PfRRM1 revealed high abundance of the protein during ring stage, which indicated a relatively high abundance of the protein during the stage of aTc-removal in our standard growth assays. This residual protein expression might allow the parasites to progress and survive during the

first cycle and only when they start a new invasion cycle, the parasites show a lethal phenotype. This hypothesis was confirmed after we performed aTc-removal growth assays at different time points after parasite invasion during the first IDC, which resulted in a growth defect at the same 72 hours in the second cycle showing a relative standard stability of the protein along the cell cycle (Fig 3-S3).

To determine the role PfRRM1 at a specific phase of the cell cycle, we modified our growth assays by removing aTc *before* parasite invasion so no PfRRM1 protein would have been synthesized by moment of infection. In this way we were able to monitor growth during the next two IDCs and identify the exact point at which the arrest happens. To do this, we tightly synchronized and perfomed percoll purification of late stage parasites at 39 hours post-invasion, right before they start a new invasion cycle. Also at this point aTc was removed to prevent further expression of PfRRM1. Parasite growth was monitored at different time points that corresponded to the main asexual stages along two cell divisions (6, 18, 25, 38, 56 and 82 hours after aTc-removal) using giemsa smears and quantifying DNA content by flow cytometry as previously reported (Ganter et al., 2017; Theron et al., 2010). PfRRM1 knockdown displayed an arrested phenotype 38 hours after invasion; this time the growth defect was observed during the first cell division possibly due to the absence of residual PfRRM1 from the previous cycle (Fig 3-5B). This arrested phenotype was confirmed by observing morphological changes (Fig 3-S4) and differences in DNA content (frequency of 1N parasites vs >1N parasites) between – and +aTc conditions (Fig 3-5B). Nuclear division (>1N) in *P. falciparum* should occurs at S phase, around ~30 hours post-invasion. This division should continue for about 10 hours, until the multinucleated cell enters into schizogony and fragments the nuclear membrane into individual 1N nuclei at around 42 hours post-invasion (Arnot et al., 2011). PfRRM1 knockdown parasites showed an increase of >1N parasites 38 hours post-invasion, however they get arrested at that time point without progressing into schizogony. Hence, no individual nuclei and subsequent mitotic cell daughters are released to start a new invasion cycle, contrary to what was observed in parasites in the presence of aTc.

An arrested phenotype during the first cell division contrasts with our previous observations of a lethal phenotype displayed during the second cell divison on PfRRM1 knockdown parasites when aTc is removed at ring stage after invasion. To confirm our cell cycle experiment results and the PfRRM1 knockdown phenotype during the first cell division, we repeated the cell cycle experiment and monitored growth at a higher resolution by taking samples at more frequent time points after aTc withdrawal only during the first cell division. Similar to our previous experiment, we monitored parasite morphology and quantified DNA content along the first cell division but this time at 6, 20, 25, 20 and 42 hours post-invasion. PfRRM1 knockdown parasites showed an arrested phenotype at the trophozoite-to-schizont transition, with parasites failing to enter into the schizogony phase, 42 hours post invasion (Fig 3-5C), confirming our results from the previous cell cycle experiment. PfRRM1 parasites at –aTc conditions failed to perform nuclear separation and to progress into further mitotic division of schizogony, even though the parasites seemed to go through nuclear division during 30-38 hours post-invasion. These results suggest that PfRRM1 might play an important role during schizogony development in asexual stages of *P. falciparum*

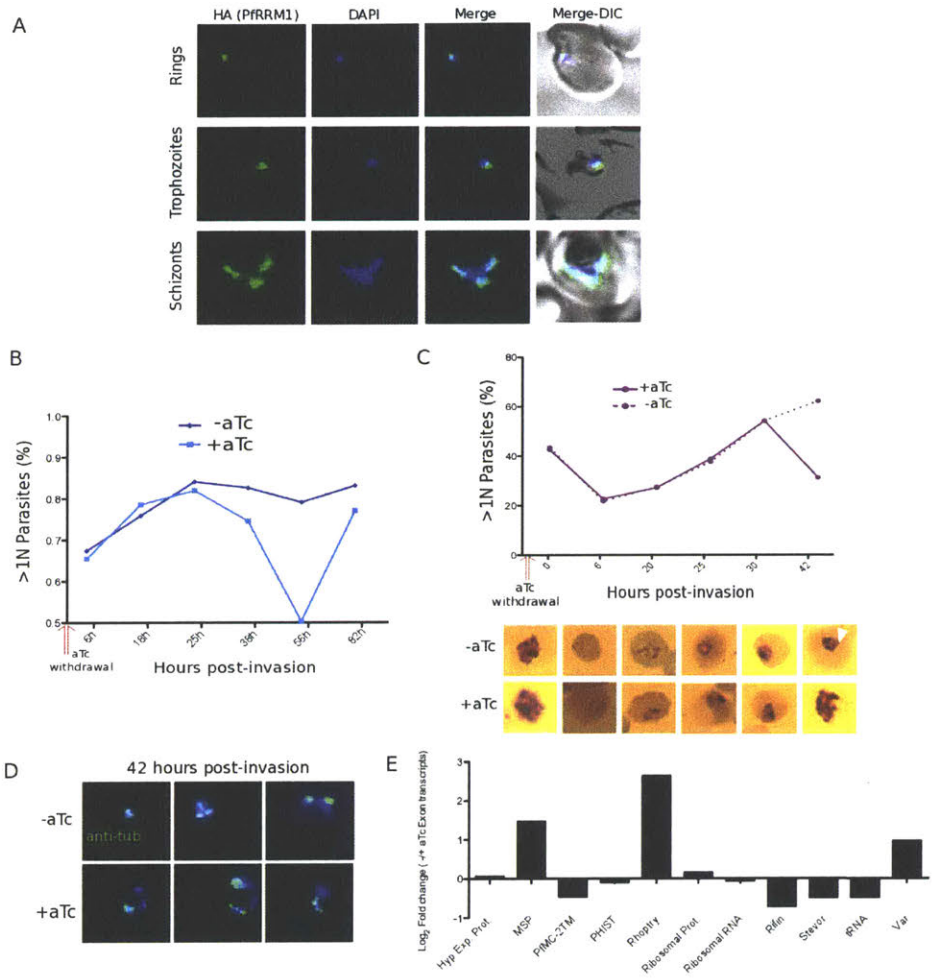


Figure 3-5: PfRRM1 is involved in nuclear processes and is essential for schizogony development

A) Immunofluourescence detection of PfRRM1 in asexual stages of *P. falciparum*. The protein localizes to the periphery of the nucleus in ring and trophozoite stages, however the signal overlaps with the entire nucleus during schizont stage. DIC=differential interference contrast. **B)** PfRRM1 edited parasites show an arrested phenotype before entering into schizogony at 38 hours post-invasion. Higher resolution cell cycle experiments to quantify the DNA content at each IDC during 2 generations were performed. Synchronous late stage parasites (~39 hours post-invasion) were percoll-purified and aTc was withdrawn before invasion. DNA content was monitored in subsequent time points using flow cytometry for the – and +aTc conditions during 2 IDCs. DNA synthesis and nuclear division was confirmed 25 hours post-invasion in PfRRM1 knockdown parasites, however they failed to enter into the schizogony phase (38 hours) and to produce individual daughter cells (1N) to start a new invasion cycle, compared to parasites in +aTc conditions. **C)** PfRRM1 parasites showed a stalled phenotype within the first IDC after aTc is withdrawn before invasion in –aTc conditions. *Top panel*: Knockdown parasites enter into S phase and go through at least one round of nuclear division without progressing into schizont stage at 38-42 hours post invasion. Parasite samples at – and + aTc conditions were monitored during several time points and split each time for flow cytometry and giemsa smears. *Bottom panel*: Parasite morphology on giemsa smears revealed arrested phonotype at

late trophozoite stage with no sign of nuclear fragmentation in schizont stage. **D)** Immunofluorescence detection of spindle structures in PfRRM1 knockdown parasites. Nuclear distortion and nuclear division (2N) is observed in stalled trophozoites of knockdown parasites 42 hours post-invasion compared to schizonts in +aTc conditions. Schizonts in +aTc conditions showed the expected punctuated pattern on the anti-tub antibody that confirms nuclear fragmentation. **E)** Exon transcripts analysis in trophozoites subject to – and + aTc conditions showed a difference in expression of merozoite-surface proteins (MSP) and rophtries, a group of proteins that serve as a dynamic marker of cell cycle progression. These proteins are required for invasion. These results are consistent with an arrested phenotype at trophozoite stage.

An arrested phenotype at the S-M transition phase of the intracellular cell cycle in PfRRM1 knockdown parasites differs from the G1 phase arrested-phenotype observed in the ortholog TgRRM1 mutant parasites in *T.gondii*. These two ortholog proteins are highly conserved, they both displayed essential phenotypes and similar localization characteristics in the cell. In spite of these similarities, the timing of when these cell cycle regulators are required during cell cycle progression is different. The discrepancy might be due to the differences in the transcriptional and post-transcriptional regulation mechanisms that exist in both parasites, reflected on the phenotypic responses that that these parasites display when these factors are depleted. In fact, TgRRM1 was characterized as a splicing factor, showing a severe splicing defect in about ~70% of the transcripts during the G1-S transition phase when TgRRM1 is depleted in *T.gondii* parasites (Suvorova et al., 2013). To investigate a potential similar effect on splicing, we analyzed differences in the Intron/Exon ratio between transcripts of RNA isolated from late trophozoite stage parasites under – and + aTc conditions. Our preliminary results showed no significant perturbation on splicing in PfRRM1 knockdown parasites displayed by an equivalent ratio of number of intron to exon reads (Fig S5). These results suggest PfRRM1 might not have a direct effect on splicing functions of the parasite. The depletion of this factor does not result in a splicing defect presumably because *P. falciparum* might stop the transcriptional machinery when PfRRM1 is not present before transcripts are misspliced. This is a hypothesis that will be investigated in future work.

Previous reports have shown the essential role of other cell cycle regulators during the asexual stage of *P. falciparum* that mirror the phenotype observed in the PfRRM1 knockdown parasites. Inhibition of cyclin-dependent kinases, such as PfCRK4 or Cyclin-

dependent kinase-like kinases (PfCLKs) has shown defects on DNA replication during schizogony in asexual *P. falciparum* parasites (Ganter et al., 2017; Kern et al., 2014). Specifically, depletion of PfCRK4 showed abnormalities on nuclear markers staining analyzed when establishing schizogony arrestment on PfCRK4 depleted parasites. Thus, we examined the fluorescent nuclear α-tubulin marker to detect spindle structures and duplication of centriole plaques (CP) to determine nuclear division and nuclear integrity along the *P. falciparum* cell cycle, as it has been previously described (Arnot et al., 2011). PfRRM1 knockdown parasites showed a nuclear distortion and a lack or only a single CP division in stalled trophozoites compared to the multiple CP observed in the progressed schizonts at +aTc conditions, in parasites analyzed 42 hours post-invasion after aTc removal (Fig 3-5D). The first two top panels showed the nuclear distortion with absence of CP division in the –aTc parasites, the last top panel showed a single round of division in those parasites (2N) compared to the panel below with multiple CPs in the progressed schizont parasite. The single round of division probably corresponds to the >1N parasites detected on the DNA content experiment. The nuclear distortion and lack of CP division phenotype using the α-tubulin nuclear marker phenotype is similar to the one observed for *P. falciparum* PfCRK4 depleted parasites, which fail to enter into schizogony (Ganter et al., 2017) and is also similar to the TgRRM1 mutant parasites of *T.gondii* that display arrested division in the centrosome nuclear marker during the G1-S phase transition (Suvorova et al., 2013).

Expression of merozoite-surface proteins (MSP) and rophtries are a dynamic cellular marker that characterize schizogony in *P. falciparum*. Transcripts that encode for MSPs and rophtries are highly abundant during schizont stage, as they are important invasion organelles, and they are used to identify the parasite stage (Bozdech et al., 2003). We performed an RNA-seq experiment on late trophozoite stage parasites to analyze the expression of the MSPs and rhoptries genes and we observed a marked fold change in the transcripts of those genes in PfRRM parasite at – and +aTc conditions (Fig 3-5E). These results confirmed the arrested phenotype at the trophozoite-to-schizont we observed using nuclear and cellular markers. Interestingly, PfCRK4 depleted parasites also lacked MSPs

and rhoptries organelles at the ultrastructre of electron microscopy images (Ganter et al., 2017), confirming the absence of schizogony on the arrested parasites.

DISCUSSION

The development of genetic tool kits that integrate different technologies to study a broader set of genes in the parasite genome is urgently needed to identify new therapeutic targets. Here we identified a new essential RNA-Binding Protein for asexual proliferation and schizogony in *P. falciparum* using an approach that combines gene-editing technologies and a conditional regulation system. CRISPR/Cas9 technologies have been extensively used to efficiently edit genomes of different eukaryote systems. Our group implemented the Cas9 system to efficiently edit non-essential genes showing complete gene-disruption in 100% of the population (Wagner et al., 2014). Here, we showed the Cas9-editing activity on an essential gene at a similar efficiency, and in addition we incorporated regulatory elements of our previously developed inducible system (Ganesan et al., 2016). The TetR-DOZI system has shown consistent tight regulation of gene expression along the parasite's asexual life cycle using reporter and native genes. In this study we showed tight regulation of PfRRM1 using double TetR-aptamer and single-aptamer installation that shows consistent levels of regulation in each case. The level of translational control achieved by this system *in vivo* proves to be a high precision tool to study genes involved in essential processes for the malaria parasite.

Our genetic tool kit has evolved from the generation of an engineered merodiploid parasite line expressing a regulated, synthetic and functional copy of an essential gene that serves as a platform for forward genome engineering applications, into an integrated circuit with editing and regulatory capacities to target an essential gene (or any other gene) and regulate its expression at its native locus. Both approaches revealed the essentiality of PfRRM1 for asexual stages of *P. falciparum* consistently. This finding is also consistent with the phenotype associated to *T.gondii* mutant parasites for the essential RNA-Binding protein TgRRM1 (Suvorova et al., 2013), which shows high levels of similarity with PfRRM1 including its RNA-binding activity.

The implementation of linear vector systems that helped us to transition into our integrated approach allowed us to incorporate various expression cassettes and elements into a single construct. We did not observe genomic rearrengements or deletions during plasmid construction, even when the cloning included low complexity regions such as the Plasmodium UTRs or secondary structures such as the aptamer arrays or the T7 terminator from the crRNA expression cassette. The utilization of linear plasmid technologies to build large circuits overcomes some of the challenges associated to the genetic manipulation of *P. falciparum* and facilitates the construction of new integrated technologies that provide more reliable genetic platforms for genetic screenings. For instance, our integrated technology using CRISPR/Cas9 and TetR-DOZI will serve as a screening platform to characterize other RBPs in *P. falciparum* and have a better understanding of the post-transcriptional mechanisms in the parasite.

P. falciparum has a very complex life cycle that includes many stages in the mosquito vector and in the human host. The parasites relies post-transcriptional mechanisms to expand its proteomic repertoire and adapt to the different environment using a relatively small genome. RBPs are key role players during these processes, yet the function of the majority of these proteins is unknown, even though they comprise ~20% of the coding parasite genome. Certain RBPs have been found to be associated to cell cycle progression in sexual and asexual stages (Blythe et al., 2004; Eshar et al., 2012; Muhia et al., 2003; Singh et al., 2004), nonetheless some aspects of the mitotic cycle in the asexual stage are not yet well understood. In this work, we demonstrate that PfRRM1 is an essential RBP for asexual proliferation in *P. falciparum* and depletion of this protein causes impairment in schizogony development. To our knowledge, this is the first report of an RBP involved in schizogony in *P. falciparum*.

The mitotic cell cycle during asexual stages in *P. falciparum* differs from other systems, including *T. gondii*, in spite of the relatedness of these two apicomplexa. During *P. falciparum* mitosis the nuclear envelope is retained (closed mitosis) until all the daughter

genomes are formed, during the M phase, after multiple rounds of mitotic events (De Souza and Osmani, 2009). This structure will generate a single multinucleated cell that eventually will form individual daughter cells or merozoites with their own nuclear and cell membrane. However, in contrast to the closed mitosis in *T.gondii,* as the multinucleated cell develops the nuclei do not divide synchronously and the outcome on the number of cell daughters does not follow a simple geometric expansion. These characteristics on the *P. falciparum* cell cycle make the assignment of conventional mitotic phases on this parasite very challenging specially the timing at which each of these phases occur (Arnot et al., 2011; Bannister et al., 2000; Gerald et al., 2011). Transcriptome data along the blood stage cycle and more recently the identification of nuclear markers have helped to describe and identify the major *P. falciparum* cell cycle phases (Arnot et al., 2011; Bozdech et al., 2003). On the other hand, the asexual mitotic cell cycle that occurs in *T.gondii* is very simple as it follows a traditional G1, S and mitotic phase in which two parasite daughters are generated in each mitotic division (Gubbels et al., 2008). This synchronicity has helped to define major cell cycle checkpoints on this parasite (Gubbels et al., 2008). In this study, we applied the nuclear markers previously defined to identify the contribution of PfRRM1 to cell cycle progression and to define the specific timing when this RBP might have an important role.

Cell cycle checkpoints in *P. falciparum* have not been very well defined, yet schizogony progression seems to be an important transition during the cycle that might require the operation of many factors from different pathways to check for the fidelity of this process. To this date, evidence shows that PfCKs seem to be major contributors of schizogony regulation in *P. falciparum.* In contrast, the synchronicity of *T.gondii* cell cycle regulation has facilitated the definition of major cell cycle checkpoints on this parasite. TgRMM1 is a major regulator of the G1-S checkpoint in *T.gondii* (Suvorova et al., 2013). We speculate PfRRM1 could function under the pathway of a major schizogony checkpoint and could represent a major determinant factor in cell cycle progression in asexual stages in *P. falciparum.*

Differences in the regulation of post-transcriptional processes such as splicing, between the apicomplexa *T. gondii* and *P. falciparum* might have also contributed to the differences found between TgRRM1 and PfRRM1. *P. falciparum* has a larger repertoire of RBPs diversity than *T.gondii* (Bunnik et al., 2016), which might also contribute to differences in RNA processing in response to certain stimuli that maintain RNA homeostasis in the cell. Despite the conservation of most of the splicing machinery between both organisms, there are certain splicing-associated factors in *T. gondii* that are missing in *P. falciparum,* a feature that might account for the observed differences in addition to the low similarity between the splicing factor proteins of both organisms (Suvorova and White, 2014). Disruption of splicing at G1 phase in TgRRM1 mutant parasites was a surprising outcome, since this contrasts with the constitutive splicing that occurs in the majority of eukaryotic systems (Braunschweig et al., 2013; Suvorova and White, 2014). Furthermore, TgRRM1 directly interacts with the U4/U6.U5 complex and when TgRRM1 is depleted this complex could be destabilized and interrupt splicing, however this interruption does not explain the arrested phenotype observed in the parasites at G1 phase in *T.gondii* parasites (Suvorova et al., 2013; Suvorova and White, 2014). In *P. falciparum* the majority of the splicing events and the mRNA processing factors seem to be abundantly expressed during the S phase or after (Bunnik et al., 2016) including PfRRM1. The fact we did not observe a splicing defect when PfRRM1 is knockdown might be explained with the differences in splicing machinery and splicing regulation found in *P. falciparum* compared to *T.gondii*. We do not discard the possibility that PfRRM1 is a splicing factor, however it might not have a direct effect on the global splicing on *P. falciparum* when this factor is depleted and the response on splicing processes might be operating at a different level of regulation.

In *P. falciparum* cyclin kinases have been found to be important regulators or to be part of the main components of the mitotic spindle division in *P. falciparum*, a requirement for nuclear division and schizogony development, such as PfCKs, centrin proteins NIMA-related kinases and Aurora-A-related kinases (Doerig et al., 2002; Halbert et al., 2010; Mahajan et al., 2008; Reininger et al., 2009; Reininger et al., 2011). Most of these

proteins are potentially essential for blood stage parasites and the function of those genes has not been elucidated. In *T.gondii*, RBPs such as TgRRM1 might be directly influencing cyclin kinases pathways that regulate cell cycle on this parasite (Suvorova and White, 2014) and since TgRRM1 directly interacts with the spliceosome of this parasite, splicing was compromised when this factor was depleted. In *P. falciparum*, the PfRRM1 interacting molecules remain unknown but no effect on splicing is observed. Whether this protein operates upstream or downstream of PfCKs regulators or whether PfRRM1 can directly affect the mRNA homeostasis in the cell is currently unknown.

The identification of protein and RNA interacting molecules of PfRRM1 is essential to determine the molecular mechanism of action of this protein. The application of conditional systems to study the genes that interact with this factor is required for further biological characterization. In this study, we targeted and regulated the expression of an essential cell cycle RBP and the technology developed could be broadly used to target other essential cell cycle regulators. The biochemical and biological evidence we showed and the phenotype we established when PfRRM1 is knockeddown using our integrated platform, enabled us find certain similarities with the function of other cell cycle regulators such as with its ortholog TgRRM1 or with the PfCKs to extrapolate the function of PfRRM1 in asexual stages of *P. falciparum*. However, further characterization of PfRRM1 is required to define the molecular mechanism of action of this protein.

METHODS

PfRRM1 alignment and model prediction

PfRRM1 is a putative RNA-Binding *P.falciparum* with unknown function that has been annotated on the PlasmoDB database (*PF3D7_1360100*). The protein sequence RRM1_RNA1 recognition motif (positions 51-226) was downloaded from the database and the sequence was compared to the TgRRM1 domain obtained in the ToxoDBA database (*TGGT1_203080*), residues 123-206. The two sequences were aligned in Genious v8 using a global alignment with free end gaps and the Blossom62 matrix cost. The predicted tertiary structure of PfRRM1 was modeled in Swiss-Model (http://swissmodel.expasy.org/) using the best matching template,2mpu.1.A. The predicted structure was visualized and edited on Chimera v1.11.1.

Plasmid construction

- PfCRT:PfRRM1 merodiploid generation

The base vector pMG79 was used as a starting point to generate a PfCRT:PfRRM1 expression vector. The list of primers used for the PfCRT:PfRRM1 construction is listed on Table 3-S1. Briefly, the pMG79 base vector contained the TetR-DOZI/Rluc/BSD expression cassette under the control the PfHsp86 promoter. The pMG79 configuration was arranged in a head-to-head context and contained modular features for a second expression in which the PfCRT promoter was included upstream from pre-existent TetR 5' and 3'10X aptamer arrays. A synthetic DNA fragment (IDT gblock) was cloned into the unique NcoI and BstEII restriction sites via Gibson.

- Cas9-T7RNAP expression plasmid

The Cas9 expression plasmid was previously built in our lab. To generate the Cas9 expression plasmid, we used our pfYC3-base vector (Wagner et al., 2013) with an *hDHFR* (Human dihydrofolate reductase) to clone the Cas9 gene. The Cas9 gene was subcloned from a plasmid provided by Feng Zhang's group, into the *XhoI* and *FseI*

unique restriction sites of the base-plasmid, using Gibson assembly, under the 5'UTR PfCAM promoter (*P.falciparum calmodulin*). T7RNAP was cloned previously, downstream from hDHFR marker using a T2A self-cleavage peptide for multi-cistronic protein expression, under the control of the 5'UTR PcDT. The Cas9 and hDHFR-2A-T7RNAP expression cassettes were cloned using the previously described head-to-head arranged configuration on the plasmid to optimize transcriptional activity (Wagner et al., 2013). An *aTTP* site was included into the SalI unique restriction site for cg6 chromosomal integration. The Cas9 integrated line was generated in previous work in our lab (unpublished work).

-PfRRM1 donor plasmids and Cas9-crRNA synthesis

1. pUF-1-based donor plasmid construction: a donor plasmid targeting the PfRRM1 native locus was built using the pUF-1 base plasmid (Ganesan et al., 2011) as a starting point. Homologous regions used as a template to repair Cas9-induced cleavage were amplified from parasite gDNA. PCR amplification reactions were performed using a 15:1 (v:v) mixture of Hemo KlenTaq polymerase and Vent polymerase (New England BioLabs) in Hemo KlenTaq Buffer. Elongation was carried out at 60 °C for 2 min/kb of DNA to be amplified and PCR products were confirmed on a 1% agarose gel. Left homologous region for PfRRM1 was purified and cloned into the *SacII* and *SpeI* unique restriction sites of the plasmid using Gibson assembly, using oligos listed on Table 3-S1. The right homologous regions were amplified and cloned into the *EcoRI* and *NcoI* unique restriction sites. The Cas9-crRNA was cloned using a pair of complementary oligos in a Klenow Fragment (3'→5' exo-) NEB and cloned into a unique *NotI* restriction site of the pUF-1 base plasmid. Oligos used to synthesize and clone homologous regions and gene specific-sgRNA are listed on Table 3-S1.
2. pSN053-based donor plasmid construction: donor plasmid to target native PfRRM1 was built into the pSN053 base plasmid (unpublished work). This base plasmid contained an expression cassette driving the expression of a single transcript encoding for TetR/DOZI/Renilla Luciferase/Blastocidin selective

marker for multi-cistronic protein expression using the 5'UTR and 3'UTR *PfHSP86*. Our group has previously shown the functionality of this expression cassette to regulate the expression of reporter and native *P. falciparum* genes using this plasmid on *BSD-resistant* and Renilla luciferase positive parasites. The plasmid was designed with modular features to easily assemble and replace gene-specific pieces to target any parasite locus. Left homologous regions were first amplified and gel-purified, followed by a Gibson that included the purified PCR product and a synthetic DNA fragment (IDG gblock) containing a recoded portion of the 3' end of the gene that prevented Cpf1 from recutting at the target site. Gibson reaction with the assembled product was then re-amplified and the purified prouct was cloned into the *FseI* and *BsiWI* restriction sites of the base plasmid. Right homologous regions correspond to the 3' UTR of each locus and were amplified, purified and cloned into *I-CeuI* unique restriction site of the pSN053 base plasmid. Oligos used to amplify left and right homologous regions of each locus are listed on Table 3-S1. Cas9-sgRNA was cloned using complementary oligos into an *AflII* restriction site at the pre-existent T7 promoter expression cassette included in the pSN053-base plasmid.

Parasite culture and transfections

P. falciparum parental parasite strain NF54attB and Cas9-T7RNAP integrated line were grown in 5% O_2, 5%CO_2 and 37C conditions in RPMI-1640 medium supplemented with 2.5g/L AlbumaxII (Life Technologies), 2g/L sodium bicarbonate, 25mM HEPES, pH 7.4, 1mM hypoxanthine and 50mg/L gentamicin. The Cas9-T7RNAP was maintained on drug selection using WR99210 at 2.5uM. Single and double plasmid transfections were performed in late stage parasites using the spontaneous uptake method (Deitsch et al., 2001) with 50ug of maxi-prepped DNA of each plasmid and eight square-wave electroporation pulses of 365 V for 1 ms each, separated by 0.1 s. Single and double drug selection was applied on the fourth day after transfection using 2.ug/mL of BSD (Blasticidine S), 1.5 uM DSM1 and 2.5 uM of WR99210, depending on the plasmids used on each experiment. DSM1 is a *Plasmodium* DHODH-specific inhibitor (Phillips et

al., 2008), BSD has also been shown to be a *P. falciparum* selective marker when using the blasticidin S deaminase of *Aspergillus terreus* (Mamoun et al., 1999) and WR99210 is an antifolate and pyrimethamine analog that inhibits the dihydrofolate reductase enzyme (DHFR) . Co-transfections on the PfCRT-PfRRM1 merodiploid parasites and single transfections on the Cas9-T7RNAP parasites transfected with pSN053 donor plasmid, were maintained on 0.5 uM of aTc (anhydrotetracycline) at all times and Renilla Luciferase signal was monitored periodically during the transfection period until it reached ~10^6 Renilla Luciferase Units (RLUs), at which parasites were detectable by Giemsa smear. Renilla Luciferase signal was measured using the Promega Renilla Glo kit, following the manufacturer's instructions on a GloMax 20/20 luminometer (Turner Biosystems). Stable parasite lines were further analyzed for specific editing events.

Luciferase assays

Renilla luciferase signal was measured by lysing infected red blood cells with 1X of the passive lysis buffer and exposing them to the specific Renilla Luciferase substrate included in the Renilla luciferase Assay kit (Promega). Signal was measured on a GloMax 20/20 luminometer (Turner Biosystems) or in a Promega Glomax plate reader using 5 seconds of exposure time.

Analysis of PfRRM1 edited parasites

Parasite genomic DNA (gDNA) was isolated from stable parasite lines. The samples were purified using the QIAamp DNA blood mini kit (Qiagen). Identification of editing events in each DNA sample was performed using PCR and the Hemo KlenTaq polymerase (New England BioLabs) in Hemo KlenTaq Buffer Oligos. The expected integration events at the 5' and the 3' ends after recombination at the target site, were amplified using P1 (hybridizes upstream to the left homologous region of the targeted gene on the native locus) and P3 (hybridizes on the yDHODh or TetR/DOZI/Rluc/BSD expression cassettes) paired oligos, and P2 (hybridizes on the yDHODh or

TetR/DOZI/Rluc/BSD expression cassettes) and P4 paired oligos (hybridizes downstream to the right homologous region of the targeted gene on the native locus). Oligos to amplify 5' and 3' integration events at each targeted site are listed on Table 3-S2.

Southern Blot

Southern blot were carried out using genomic DNA isolated from a 50mL flask at 2% hematocrit and ~15-20% parasitemia, using the QIAamp DNA blood mini kit (Qiagen). RBCs were lysed using a 0.15% saponin solution. Between 2-3 ug of genomic DNA was enzyme digested overnight *NcoI* (New England Biolabs), and probed with a PfRRM1 PCR prouduct using primers AF138 and AF139. The probe was labeled with biotin-11-dUTPs using the Pierce Biotin Random Prime kit (Thermo Scientific). Blots were processed using the Turboblotter kit (Whatman) for transferring and the North2South kit (Thermo Scientific) for development.

Western Blot

To establish inducible regulation of PfCRT:PfRRM1 and PfRRM1 parasites, approximately 10^6 late stage parasites were lysed using 0.15% saponin in 1X PBS after being grown for at least one IDC in the presence or absence of aTc (except during the PfCRT:PfRRM1 time course experiment). NF54attB WT parasites were included as negative controls. Lysates were prepared for western blot by heating in Laemmli sample buffer at 95 °C for 10 min. After separation by SDS-PAGE, proteins were transferred to a PVDF membrane and probed with a mouse monoclonal antibody against HA-tag clone 6E2 (Cell Signaling Technology) at a 1:1000 dilution, or rabbit polyclonal antibody against GAPDH (Abcam, ab9485) at a 1:1000 dilution. Blots were imaged using a horseradish peroxidase-coupled secondary antibody (1:5000 dilution) and SuperSignal West Femto substrate (Thermo Scientific, Catalog # 34095).

Growth Assays

- Flow cytometry-based assays

Ring-stage parasites were tightly synchronized using 0.3 M alanine in 10mM HEPES (pH 7.4) at day 0. Parasites were adjusted to 1% parasitemia and seeded in 96-well microtiter plates in quadruplicate at 2% hematocrit in 200uL of RPMI complete media in the presence or in the absence of aTc (0.5uM). Expansion of parasitemia was measured for 5 IDCs and samples were analyzed every 48 hours. After each measurement, all cultures were split by the same dilution factor required to keep the pre-invasion parasitemia belo 2% to avoid culture over-expansion. Parasites were stained for nucleic acid content with SYBR green I using a dilution 1:5000 for 15 min at 37C. Parasites were selectively analyzed on an Accuri C6 Instrument (BD Biosciences) by gating on events with high FL1 signal intensity. A sample of uninfected red blood cells was included as control to get for the infected population.

-Renilla Luciferase-based assay

Ring-stage parasites were tightly synchronized using a sorbitol 5% solution. On day 0 parasites were adjusted to ~1000 RLUs and were washed at least 3 times with RPMI complete media and seeded in 96-well microtiter plates in triplicate at 0.5% hematocrit in 200uL of RPMI complete media in the presence or in the absence of aTc (0.5uM). Parasites from each time point were seeded on different plates and were incubated without disruption for the time assigned to each plate (IDC 1, 2 or 3). Parasitemia was analyzed lysing infected red blood cells with 1X of the passive lysis buffer and exposing them to the specific Renilla Luciferase substrate included in the Renilla luciferase Assay kit (Promega). Renilla luciferase signal was measured on the plate using the Promega Glomax plate reader using 5 seconds of exposure time.

Crosslinked Immunoprecipitation assays to detect RNA-Binding Activity

Parasites were expanded to 4x50mL large flasks and maintained at 2% hematocrit. Parasites were extracted using 0.15% saponin and washed with 1X PBS. Parasites were resuspended in ~2-3mLs of PBS and transferred to a petri dish. Total protein concentration (used as an input the crosslinking experiment) was calculated using the BCA method. Parasite lysates were UV irradiated on ice using 600,000 uJ/cm2. Parasite lysate was washed with cold 1X PBS and lysed using NET-2 lysis buffer (50mM Tris-HCl, 150mM NaCl, 1% IGEPAL, 2 mM DTT, 1X of 7X cocktail Protease inhibitor Roche, Ribolock, 0.05% SDS, 0.5% Triton) and samples were incubated on ice for 1 hour. After lysis, a pre-clearing of the parasite lysate was carried out using a secondary and non-specific IgG anti-rabbit antibody to avoid unspecific binding. The lysate was incubated with the antibody for 1 hour at 4C. Beads were washed and conjugated first with a rabbit anti-Rat for 2 hours, then were conjugated rat anti-HA for another two hours at 4C. Removal of non-conjugated was performed using a buffer exchange with the NET2 lysis buffer. Parasite lysate was incubated with the conjugated beads in the presence of RNase (to remove unprotected RNA) and DNase, as well as with protease inhibitors. Incubation of the parasite lysate with the conjugated beads was performed overnight at 4C. After protein-RNA pull down, several beads washings are performed along with a second DNase treatment. *RNA labeling:* Conjugated beads bound to protein-RNA complexes were treated with PNK, PNK buffer and P^{32} isotope for 20 mins at 37C. Next, beads were prepared with loading buffer to run into a SDS-Page gel. Samples were run and then transfer into a nitrocellulose membrane. Blot was exposed to film. RNA Binding activity was detected only on the lane with PfRRM1 parasite sample but not in the negative controls.

Immunofluorescence assays

Immunofluorescence assays were carried out as described previously (Tonkin et al., 2004). Briefly, blood stage parasites were fixed using 4% paraformaldehyde, 0.0075% glutaraldehyde in 1X PBS for 30 mins at room temperature. Fixed cells were washed once with PBS and then permeabilized with 0.1% Triton X-100/PBS for 10 min NS

washed once with PBS. Next, cells were blocked with 3% BSA/PBS for one hour. After blocking, cells were exposed to a mouse monoclonal antibody against HA-tag clone 6E2 (Cell Signaling Technology) at a 1:500 dilution or to a mouse anti-α-tubulin (1:2,000, Sigma-Aldrich, clone B-5-1-2) in blocking solution and were incubated at room temperature for 1 hour. Cells were washed at least 3 times with PBS and then incubated with an anti-mouse IgG (H+L), F(ab')2 Fragment (Alexa Fluor® 488 Conjugate) from Cell Signaling Technology at a 1:500 dilution in blocking solution for 45mins at room temperature in dark conditions. Cells were washed at least 3 times with PBS and were incubated with DAPI (4'6-diamidino-2-phenylindole) at a 1:1000 dilution in PBS for 10 mins. Finally, cells were spotted onto pre-coated slides with 0.1% of polyethyleneimine and incubated at room temperature for 30 mins. Once the cells were attached to the slide, they were mounted with Prolong media and covered with a #1.5 coverslip. Samples were let dry overnight before imaging on a GE HC DeltaVision Elite Widefield Deconvolution system with a X100 objective. For each field, a 30-image-Z-stack was deconvoluted with the Softworx software and maximum intensity of Z-projections, false coloring and image merging were performed with Fiji software.

DNA content analysis and cell cycle experiments

Tightly synchronized parasites using sorbitol 5% at each parasite cycle before the experiment were purified using percoll-gradient, previously described (Rivadeneira et al., 1983). Briefly, late stage parasites (39 hours post-invasion) from a 50mL flask at 2% hematocrit were pelleted and deposited on a solution of 65% Percoll in wash media. A gradient centrifugation was performed at 1500g for 10 mins with no break to achieve effective separation of the different parasite stages from the uninfected red blood cells in the percoll solution. Late stage parasites were removed from the top of the multilayer solution and transferred to a new tube for washes with RPMI to remove excess of Percoll (and remove aTc for growth assays). Resuspended and purified late stage parasites were confirmed on a giemsa smear. Next, parasites were added to a new fresh 10mL cultures in – and +aTc conditions and the parasites were allowed to invade for ~4-5 hours. After the invasion, parasites were treated with 5% sorbitol and remove residual late stage

parasites and tightly synchronize the culture to rings. The sorbitol synchronization time was taken as Time 0 in cell cycle experiments and a sample for giemsa staining and flow cytometry were taken from the culture. Typically a sample of ~0.5mL from the culture was taken at each time point for flow cytometry and cells were fixed in 4% paraformaldehyde/PBS for at least 24 hours at 4C. At the end of the experiment after all time points were collected, fixed samples for flow cytometry were washed with 0.1% Triton-X100/PBS for 1 hour at room temperature. Cells were washed after for at least 3-4 times with PBS and incubated for 20 mins at each wash to remove excess of hemoglobin. After washing is complete, cells were resuspended in ~200 uL of PBS and were seeded in a 96-well microtiter plate and stained with SYBR green I at a 1:5000 dilution. Cells were incubated for 15 mins at 37C and were selectively analyzed on an Accuri C6 Instrument (BD Biosciences) by gating on events with high FL1 signal intensity. Gated infected red blood cells could be separately gated for 1N vs >1N parasites based on the fluorescent signal of their DNA content.

RNA isolation for RNA-seq experiments

Parasites were tightly synchronized for several cycles using sorbitol 5% and expanded to ~4 x 50mL large flasks at ~15% parasitemia at 2% hematocrit. Late stage parasites were washed from aTc and cultures at – and + aTc were allowed to progress until trophozoite stage of the second IDC. Two large flasks per condition per strain were used for RNA isolation. Parasites were pelleted and resuspended in 5 volumes of pre-warmed TRIzol® LS (37°C, Life Technologies 10296-010). Next, samples were incubated at 37°C for 5 min to ensure the complete de-proteinization of nucleic acids. Samples were stored at -80C until further processing. For processing, samples were thawed on ice and for each 5 mL of TRIzol® LS that was used, we added 1 mL of chloroform. After a centrifugation for 30 mins at 12,000g at 4C, the upper aqueous layer was transferred to a clean tube and 0.8 volume of pre-chilled isopropanol were added to precipitate the RNA. Next, samples were spinned at 12,000 for 30 mins at 4C and the supernatant was removed. The pellet was washed with 75% ethanol and air-dried for 5 min. Finally the pellet was resuspended in 30-100 µL of RNase-free non-DEPC treated water (Ambion AM9930) and the samples

were heated at 60°C for 10 min. RNA concentration was confirmed in the nanodrop and integrity of the RNA was analyzed in a BioAnalyzed. DNase treatment and ribosomal RNA depletion were performed before cDNA library preparation. Library preparation, sequencing and data analysis were performed in collaboration with Maggie Lu and Karine LeRoch at the facilities of University of California-Riverside and methods used were described previously (Lu et al., 2017).

REFERENCES

Arnot, D.E., Ronander, E., and Bengtsson, D.C. (2011). The progression of the intra-erythrocytic cell cycle of Plasmodium falciparum and the role of the centriolar plaques in asynchronous mitotic division during schizogony. Int J Parasitol *41*, 71-80.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., *et al.* (2009). PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res *37*, D539-543.

Balu, B., Singh, N., Maher, S.P., and Adams, J.H. (2010). A genetic screen for attenuated growth identifies genes crucial for intraerythrocytic development of Plasmodium falciparum. PLoS One *5*, e13282.

Bannister, L.H., Hopkins, J.M., Fowler, R.E., Krishna, S., and Mitchell, G.H. (2000). A brief illustrated guide to the ultrastructure of Plasmodium falciparum asexual blood stages. Parasitol Today *16*, 427-433.

Behnke, M.S., Wootton, J.C., Lehmann, M.M., Radke, J.B., Lucas, O., Nawas, J., Sibley, L.D., and White, M.W. (2010). Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of Toxoplasma gondii. PLoS One *5*, e12354.

Blythe, J.E., Surentheran, T., and Preiser, P.R. (2004). STEVOR--a multifunctional protein? Mol Biochem Parasitol *134*, 11-15.

Bozdech, Z., Llinas, 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.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. Cell *152*, 1252-1269.

Bunnik, E.M., Batugedara, G., Saraf, A., Prudhomme, J., Florens, L., and Le Roch, K.G. (2016). The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. Genome Biol *17*, 147.

Bunnik, E.M., Chung, D.W., Hamilton, M., Ponts, N., Saraf, A., Prudhomme, J., Florens, L., and Le Roch, K.G. (2013). Polysome profiling reveals translational control of gene expression in the human malaria parasite Plasmodium falciparum. Genome Biol *14*, R128.

Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L., and Krainer, A.R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J Cell Biol *138*, 225-238.

Coulson, R.M., Hall, N., and Ouzounis, C.A. (2004). Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. Genome Res *14*, 1548-1554.

De Silva, E.K., Gehrke, A.R., Olszewski, K., Leon, I., Chahal, J.S., Bulyk, M.L., and Llinas, M. (2008). Specific DNA-binding by apicomplexan AP2 transcription factors. Proc Natl Acad Sci U S A *105*, 8393-8398.

De Souza, C.P., and Osmani, S.A. (2009). Double duty for nuclear proteins--the price of more open forms of mitosis. Trends Genet *25*, 545-554.

Deitsch, K., Driskill, C., and Wellems, T. (2001). Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. Nucleic Acids Res *29*, 850-853.

Doerig, C., Endicott, J., and Chakrabarti, D. (2002). Cyclin-dependent kinase homologues of Plasmodium falciparum. Int J Parasitol *32*, 1575-1585.

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*, 9903-9916.

Foth, B.J., Zhang, N., Mok, S., Preiser, P.R., and Bozdech, Z. (2008). Quantitative protein expression profiling reveals extensive post-transcriptional regulation and post-translational modifications in schizont-stage malaria parasites. Genome Biol *9*, R177.

Fukuda, T., Naiki, T., Saito, M., and Irie, K. (2009). hnRNP K interacts with RNA binding motif protein 42 and functions in the maintenance of cellular ATP level during stress conditions. Genes Cells *14*, 113-128.

Ganesan, S.M., Falla, A., Goldfless, S.J., Nasamu, A.S., and Niles, J.C. (2016). Synthetic RNA-protein modules integrated with native translation mechanisms to control gene expression in malaria parasites. Nat Commun *7*, 10727.

Ganesan, S.M., Morrisey, J.M., Ke, H., Painter, H.J., Laroiya, K., Phillips, M.A., Rathod, P.K., Mather, M.W., and Vaidya, A.B. (2011). Yeast dihydroorotate dehydrogenase as a new selectable marker for Plasmodium falciparum transfection. Mol Biochem Parasitol *177*, 29-34.

Ganter, M., Goldberg, J.M., Dvorin, J.D., Paulo, J.A., King, J.G., Tripathi, A.K., Paul, A.S., Yang, J., Coppens, I., Jiang, R.H., *et al.* (2017). Plasmodium falciparum CRK4 directs continuous rounds of DNA replication during schizogony. Nat Microbiol *2*, 17017.

Gerald, N., Mahajan, B., and Kumar, S. (2011). Mitosis in the human malaria parasite Plasmodium falciparum. Eukaryot Cell *10*, 474-482.

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.

Godiska, R., Mead, D., Dhodda, V., Wu, C., Hochstein, R., Karsi, A., Usdin, K., Entezam, A., and Ravin, N. (2010). Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. Nucleic Acids Res *38*, e88.

Goldfless, S.J., Wagner, J.C., and Niles, J.C. (2014). Versatile control of Plasmodium falciparum gene expression with an inducible protein-RNA interaction. Nat Commun *5*, 5329.

Gubbels, M.J., White, M., and Szatanek, T. (2008). The cell cycle and Toxoplasma gondii cell division: tightly knit or loosely stitched? Int J Parasitol *38*, 1343-1358.

Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R. (2013). Rapid, modular and reliable construction of complex mammalian gene circuits. Nucleic Acids Res *41*, e156.

Halbert, J., Ayong, L., Equinet, L., Le Roch, K., Hardy, M., Goldring, D., Reininger, L., Waters, N., Chakrabarti, D., and Doerig, C. (2010). A Plasmodium falciparum transcriptional cyclin-dependent kinase-related kinase with a crucial role in parasite proliferation associates with histone deacetylase activity. Eukaryot Cell *9*, 952-959.

Howe, R., Kelly, M., Jimah, J., Hodge, D., and Odom, A.R. (2013). Isoprenoid biosynthesis inhibition disrupts Rab5 localization and food vacuolar integrity in Plasmodium falciparum. Eukaryot Cell *12*, 215-223.

Hughes, K.R., Philip, N., Starnes, G.L., Taylor, S., and Waters, A.P. (2010). From cradle to grave: RNA biology in malaria parasites. Wiley Interdiscip Rev RNA *1*, 287-303.

Kern, S., Agarwal, S., Huber, K., Gehring, A.P., Strodke, B., Wirth, C.C., Brugl, T., Abodo, L.O., Dandekar, T., Doerig, C., *et al.* (2014). Inhibition of the SR protein-phosphorylating CLK kinases of Plasmodium falciparum impairs blood stage replication and malaria transmission. PLoS One *9*, e105732.

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., *et al.* (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Science *301*, 1503-1508.

Limenitakis, J., and Soldati-Favre, D. (2011). Functional genetics in Apicomplexa: potentials and limits. FEBS Lett *585*, 1579-1588.

Lu, X.M., Batugedara, G., Lee, M., Prudhomme, J., Bunnik, E.M., and Le Roch, K.G. (2017). Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite Plasmodium falciparum. Nucleic Acids Res *45*, 7825-7840.

Mahajan, B., Selvapandiyan, A., Gerald, N.J., Majam, V., Zheng, H., Wickramarachchi, T., Tiwari, J., Fujioka, H., Moch, J.K., Kumar, N., *et al.* (2008). Centrins, cell cycle regulation proteins in human malaria parasite Plasmodium falciparum. J Biol Chem *283*, 31871-31883.

Mair, G.R., Braks, J.A., Garver, L.S., Wiegant, J.C., Hall, N., Dirks, R.W., Khan, S.M., Dimopoulos, G., Janse, C.J., and Waters, A.P. (2006). Regulation of sexual development of Plasmodium by translational repression. Science *313*, 667-669.

Mair, G.R., Lasonder, E., Garver, L.S., Franke-Fayard, B.M., Carret, C.K., Wiegant, J.C., Dirks, R.W., Dimopoulos, G., Janse, C.J., and Waters, A.P. (2010). Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development. PLoS Pathog *6*, e1000767.

Mamoun, C.B., Gluzman, I.Y., Goyard, S., Beverley, S.M., and Goldberg, D.E. (1999). A set of independent selectable markers for transfection of the human malaria parasite Plasmodium falciparum. Proc Natl Acad Sci U S A *96*, 8716-8720.

Miao, J., Li, J., Fan, Q., Li, X., Li, X., and Cui, L. (2010). The Puf-family RNA-binding protein PfPuf2 regulates sexual development and sex differentiation in the malaria parasite Plasmodium falciparum. J Cell Sci *123*, 1039-1049.

, 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*, 22014-22022.

Nkrumah, L.J., Muhle, R.A., Moura, P.A., Ghosh, P., Hatfull, G.F., Jacobs, W.R., Jr., and Fidock, D.A. (2006). Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase. Nat Methods *3*, 615-621.

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.

Painter, H.J., Campbell, T.L., and Llinas, M. (2011). The Apicomplexan AP2 family: integral factors regulating Plasmodium development. Mol Biochem Parasitol *176*, 1-7.

Phillips, M.A., Gujjar, R., Malmquist, N.A., White, J., El Mazouni, F., Baldwin, J., and Rathod, P.K. (2008). Triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors with potent and selective activity against the malaria parasite Plasmodium falciparum. J Med Chem *51*, 3649-3653.

Reddy, B.P., Shrestha, S., Hart, K.J., Liang, X., Kemirembe, K., Cui, L., and Lindner, S.E. (2015). A bioinformatic survey of RNA-binding proteins in Plasmodium. BMC Genomics *16*, 890.

Reininger, L., Tewari, R., Fennell, C., Holland, Z., Goldring, D., Ranford-Cartwright, L., Billker, O., and Doerig, C. (2009). An essential role for the Plasmodium Nek-2 Nima-related protein kinase in the sexual development of malaria parasites. J Biol Chem *284*, 20858-20868.

Reininger, L., Wilkes, J.M., Bourgade, H., Miranda-Saavedra, D., and Doerig, C. (2011). An essential Aurora-related kinase transiently associates with spindle pole bodies during Plasmodium falciparum erythrocytic schizogony. Mol Microbiol *79*, 205-221.

Rivadeneira, E.M., Wasserman, M., and Espinal, C.T. (1983). Separation and concentration of schizonts of Plasmodium falciparum by Percoll gradients. J Protozool *30*, 367-370.

Singh, N., Preiser, P., Renia, 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*, 589-599.

Solyakov, L., Halbert, J., Alam, M.M., Semblat, J.P., Dorin-Semblat, D., Reininger, L., Bottrill, A.R., Mistry, S., Abdi, A., Fennell, C., *et al.* (2011). Global kinomic and phospho-proteomic analyses of the human malaria parasite Plasmodium falciparum. Nat Commun *2*, 565.

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*, 3820-3835.

Suvorova, E.S., Croken, M., Kratzer, S., Ting, L.M., Conde de Felipe, M., Balu, B., Markillie, M.L., Weiss, L.M., Kim, K., and White, M.W. (2013). Discovery of a splicing regulator required for cell cycle progression. PLoS Genet *9*, e1003305.

Suvorova, E.S., and White, M.W. (2014). Transcript maturation in apicomplexan parasites. Curr Opin Microbiol *20*, 82-87.

Tewari, R., Straschil, U., Bateman, A., Bohme, U., Cherevach, I., Gong, P., Pain, A., and Billker, O. (2010). The systematic functional analysis of Plasmodium protein kinases identifies essential regulators of mosquito transmission. Cell Host Microbe *8*, 377-387.

Theron, M., Hesketh, R.L., Subramanian, S., and Rayner, J.C. (2010). An adaptable two-color flow cytometric assay to quantitate the invasion of erythrocytes by Plasmodium falciparum parasites. Cytometry A *77*, 1067-1074.

Tonkin, C.J., van Dooren, G.G., Spurck, T.P., Struck, N.S., Good, R.T., Handman, E., Cowman, A.F., and McFadden, G.I. (2004). Localization of organellar proteins in Plasmodium falciparum using a novel set of transfection vectors and a new immunofluorescence fixation method. Mol Biochem Parasitol *137*, 13-21.

Tripet, B.P., Mason, K.E., Eilers, B.J., Burns, J., Powell, P., Fischer, A.M., and Copie, V. (2014). Structural and biochemical analysis of the Hordeum vulgare L. HvGR-RBP1 protein, a glycine-rich RNA-binding protein involved in the regulation of barley plant development and stress response. Biochemistry *53*, 7945-7960.

Vembar, S.S., Macpherson, C.R., Sismeiro, O., Coppee, 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*, 212.

Wagner, J.C., Goldfless, S.J., Ganesan, S.M., Lee, M.C., Fidock, D.A., and Niles, J.C. (2013). An integrated strategy for efficient vector construction and multi-gene expression in Plasmodium falciparum. Malar J *12*, 373.

Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014). Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. Nat Methods *11*, 915-918.

Wongsombat, C., Aroonsri, A., Kamchonwongpaisan, S., Morgan, H.P., Walkinshaw, M.D., Yuthavong, Y., and Shaw, P.J. (2014). Molecular characterization of Plasmodium falciparum Bruno/CELF RNA binding proteins. Mol Biochem Parasitol *198*, 1-10.

SUPPLEMENTARY INFORMATION

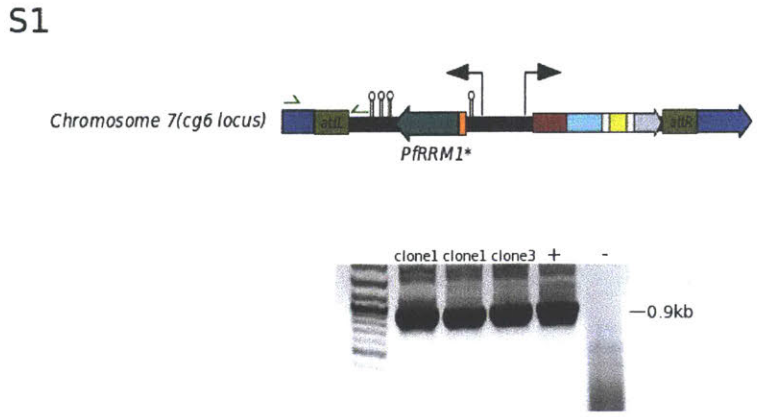


Figure 3-S1: Site-specific integration of the PfCRT:PfRRM1 expression plasmid into the chromosomal cg6 locus

Top panel: cg6-modified locus upon PfCRT:PfRRM1 plasmid integration via Bxb1 integrase. Specific primers were designed to hybridize on either the plasmid backbone or the native locus. *Bottom panel*: Parasites were co-transfected with the PfCRT:PfRRM1 and a marker-free Bxb1 integrase expression plasmids. Integration event was confirmed on the parasite stable population and subsequently parasites were cloned via limiting dilution. All clonal populations show the expected integration event at the cg6 locus when compared to a positive control (PfCRT:YFP integrated plasmid (Ganesan et al., 2016) and a negative wild type gDNA.

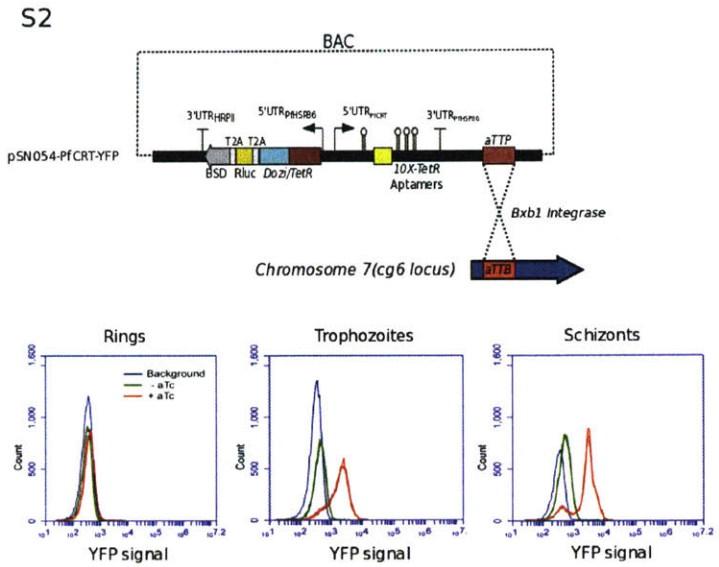


Figure 3-S2: Chromosomal integration of new linear vector with the TetR-Dozi-aptamer regulatory components shows inducible regulation

Previous work in our lab has showed that the transfer of the inducible regulatory components into the new pJazz system shows consistent and stable expression of the different elements. A YFP-fluorescent reporter was cloned into the pSN054 base plasmid and it was flanked by the dual-aptamer configuration. The YFP gene is expressed from a PfCRT promoter with major activity during trophozoite stages. The pSN054-PfCRT-YFP plasmid was rescued into a circular BAC system for site-specific chromosomal integration into the engineered aTTB site using the Bxb1 integrase (Nkrumah et al., 2006). An integrated parasite population was rescued after selection and cloned via limiting dilution. The clonal population was confirmed for positive YFP expression, showing tight inducible regulation in – and + aTc conditions consistently along the parasite cycle.

S3

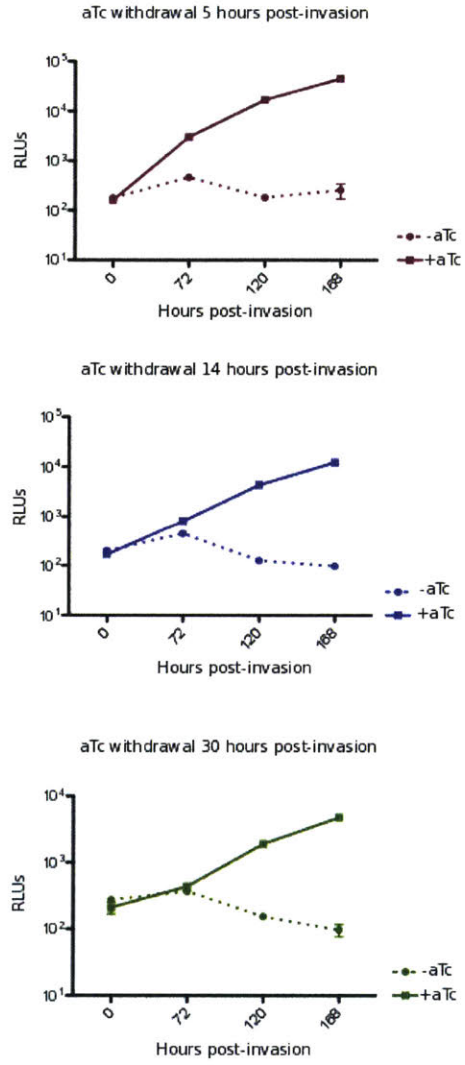


Figure 3-S3: PfRRM1 knockdown parasites show arrested phenotype 72 hours post-invasion after aTc-withdrawal at 5h, 14h and 30h post-invasion

Renilla luciferase-based growth assays were performed on PfRRM1 edited parasites in – and + aTc conditions with aTc withdrawn at different time points post-invasion. Growth was monitored for more than 3 cell divisions and an arrested phenotype was confirmed at 72 hours post-invasion in all cases.

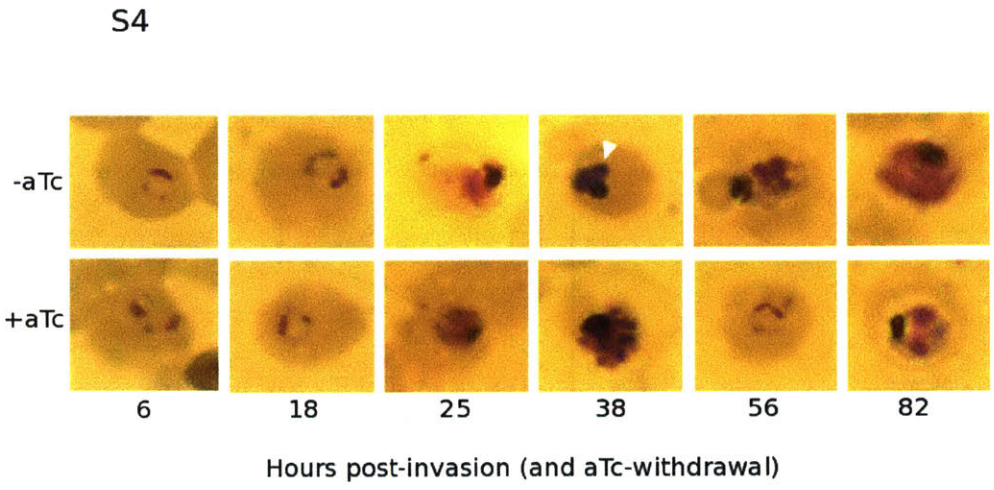


Figure 3-S4: PfRRM1 knockdown parasites display an arrested morphology 38 hours post-invasion at late throphozoite stage

Parasite morphology on giemsa smears was monitored at different time points after aTc-withdrawal during 2 cell divisions (in a parallel experiment to the cell cycle and DNA quantification experiment) and an arrested phenotype was observed 38 hours post-invasion. Parasites were stalled at late throphozoite stage without progressing to schizont stage. No signs of nuclear fragmentation or newly invaded red blood cells on the next cycle were observed in PfRRM1 knockdown parasites.

S5

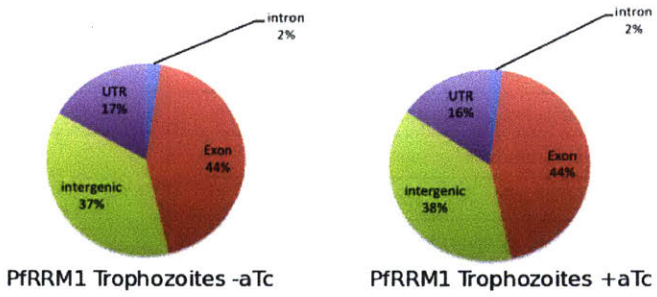


Figure 3-S5: PfRRM1 knockdown parasites do not show a splicing defect in trophozoite stage

Intron analysis was carried out at trophozoite stage parasites in – and + aTc conditions. No differences were observed in intron retention in the transcripts analyzed at both conditions. Transcripts from untranlated regions (UTRs), intergenic regions along with mRNA transcripts were included in the analysis.

Oligo Name	Sequence (5'-3')	Description
SMG145	catacccatacgatgttccagattacgctgttaacatgcgagagaacgataacactgaac	Fwd oligo to clone synthetic and recoded PfRRM1 into pMG79.
SMG146	ttcttcgctaatcagtttctgttcggtaaccttagttcagtgtatctttgttgatgagcc	Rv oligo to clone synthetic and recoded PfRRM1 into pMG79.
AFPfRRM1KO1-L	tccaatggccccttccgcggttttgtatctctaattatttttattaataaatataaatataaatataaattttttttttttttttttttttttaaatgac	Fwd Gibson oligo to amplify and clone PfRRM1 left homologous into the pUF-1 base plasmid
AFPfRRM1KO2-L	aaaatgcttaagacagatcttcggactagtatttgttgttctgtattatcattttccctc	Rv Gibson oligo to amplify and clone PfRRM1 left homologous into the pUF-1 base plasmid
AFPfRRM1KO1-R	ataagaacatatttattaaatctagaattcatgaaaaatctaataacaaaaaggtataaacaaataatttatacaatg	Fwd Gibson oligo to amplify and clone PfRRM1 right homologous into the pUF-1 base plasmid
AFPfRRM1KO2-R	atcaggcgccagcctaggagttccatggaggaaataaatttaaaaaaacatatacctaaatattcgacac	Rv Gibson oligo to amplify and clone PfRRM1 right homologous into the pUF-1 base plasmid
AFPfRRM1gRNA-F	taatacgactcactatagg**atcagatatgacaacttcaa**gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaact	Fwd Gibson oligo to synthesize and clone PfRRM1 gRNA into pUF-base plasmid
AFPfRRM1gRNA-R	caaaaaaccccctcaagacccgtttagaggccccaaggggttatgctagcaccgactcggtgccacttttcaagttgataacggactagcct	Rv Gibson oligo to synthesize and clone PfRRM1 gRNA
AF550	tgttgaagaaaatccaggtccagcgatcgctcaaaggatgaaaaatctaataacaaaaag	Fwd Gibson primer to amplify and clone left homologous region of PfRRM1 into pSN053 base plasmid
AF551	ttgagtgcatccagcatatcctgaggatcactaagactaacgaaaccatatcctttggtc	Rv Gibson primer to amplify and clone left homologous region of PfRRM1 into pSN053 base plasmid
AF554	tatacgaagttatcgtaactataacttatataacgatattgtatgataattttaataaaaatatatcatc	Fwd Gibson primer to amplify and clone right homologous region of PfRRM1 into pSN053 base plasmid
AF555	tcgtattattcgctaccttaggaccacacctctaaaaaggttacgtg	Rv Gibson primer to amplify and clone right homologous region of PfRRM1 into pSN053 base plasmid
AF556	gtggtcctaaggtagcgaataatacgactcactata**taaccgtaaaaagaagtaga**	Fwd Gibson oligo to synthesize and clone PfRRM1 gRNA into pSN053-base plasmid
AF557	gactagccttatttttaacttgctatttctagctctaaaactctacttctttttacggtta	Rv Gibson oligo to synthesize and clone PfRRM1 gRNA into pSN053-base plasmid

Table 3-S1: List of Oligos for plasmid construction

*Bold font corresponds to the PfRRM1 Cas9-gRNA.

Oligo Name	Sequence (5'-3')	Description
AF148	attattatttttttttgttaatattatacaatatacctaggatgtacccatacgatgttccagattacgctgctagcagggaaaatgataatacagaacaac	Fwd oligo to check for PfRRM1 disruption on the PfCRT:PfRRM1 line at the 5' of the PfRRM1 locus
AF323	ttaaatgctgttcaacttcccac	Rv oligo specific to the yDHODh selective marker to detect integration at the 5' end using when using the pUF:PfRRM1 donor plasmid.
AF192	cattgtgagtacataaatatattatataactcgaccttaagttagttcaaggtatctttatttattaatctttc	Rv oligo to check for PfRRM1 disruption on the PfCRT:PfRRM1 line at the 3' of the PfRRM1 locus
AF324	acagccagtttaactaccaagttcttg	Fwd oligo specific to the yDHODh selective marker to detect integration at the 3' end using when using the pUF:PfRRM1 donor vector
AF651	aaaaaatataataataataataataataaagtttcagcatac	Fwd oligo to check for PfRRM1 disruption at the 5' end of locus
AF412	ccctttatcatcatcatctttataatcg	Rv oligo specific to the TetR/DOZI backbone to detect integration at the 5' end when using pSN053 based donor plasmids
AF652	gtaaggcagctatgatttatttgtattcttg	Rv oligo to check for PfRRM1 disruption at the 3' end of locus
AF266	ggaatactaaatatatatccaatggcccc	Fwd oligo specific to theTetR/DOZI backbone to detect integration at the 3' end when using pSN053 based donor plasmids
AF08	catttgaattattgctcaacgct	Fwd P2 primer to check for integration at the cg6 locus (Nkrumah et al., 2006)
AF23	gctcgaattccgggtttgtacc	Rv P4 primer to check for integration at the cg6 locus. Hybridizes on the plasmid backbone.
AF138	cttaatggctactcataatcc	Fwd primer to synthesize southern blot probe from gDNA.
AF139	ataaattatacatatattgttatatacagc	Rv primer to synthesize southern blot probe from gDNA.

Table 3-S2: List of oligos used to detect integration events

CHAPTER 4

Expanding the coverage of CRISPR-gene targets in *P. falciparum* using Cpf1 endonucleases

ABSTRACT

Robust genetic tools are needed to perform functional genetics at a large scale on malaria parasites and identify potential antimalarial targets. CRISPR/Cas class II systems are powerful, single-effector protein editing tools that that can efficiently target DNA. Cas9-utilizing approaches have been widely characterized as high efficiency-editing systems in a wide variety of organisms, including *Plasmodium falciparum*. Cas9 requires an NGG motif for inducing targeted DNA strand breaks. However, the *P. falciparum* genome is ~80-90% AT-rich, and this can place constraints on finding suitable Cas9 cleavage sites compatible with the desired genomic manipulation. Recently, the Cpf1 endonuclease from novel class II CRISPR systems was characterized as an effector protein having a TTTN- binding motif. This opens the possibility for leveraging this system to increase the flexibility with which a wider range of target sites in the AT-rich P. *falciparum* genome can be accessed/manipulated. Here we report using Cpf1 to edit different *P. falciparum* loci with high efficiency, and independently of target sequence composition. Through bioinformatics analyses, we show on a genome-wide scale there is significantly more high specificity Cpf1- compared to Cas9- binding sites in the *P. falciparum* genome. Many sites are located proximal to the 5'- and 3'-termini of coding sequences, which should enable functional manipulations, such as epitope tagging and installation of genetically-encoded elements to regulate protein expression. We illustrate this by using Cpf1-editing technology to simultaneously install TetR-aptamer and epitope tags at the 3'-end of selected gene loci. Altogether, Cpf1 and Cas9 are highly complementary in accessing diverse loci for genome editing applications given the highly skewed AT-rich *P. falciparum* genome composition.

INTRODUCTION

The ability to genetically manipulate *P. falciparum* is key for the identification and characterization of essential genes for parasite survival and for understanding the molecular basis of drug resistance. Available genetic tools for the targeted manipulation of genes in this parasite include conventional single and double crossover recombination methods, site-specific genomic integration in the *cg6* locus via Bxb1 integrase 55 (Nkrumah et al., 2006), Zinc-finger nucleases (Straimer et al., 2012) and more recently the CRISPR-Cas9 system from *Streptococcus pyogenes* (Ghorbal et al., 2014; Wagner et al., 2014). The latter has shown to be cost-effective and more efficient compared to the other genetic approaches and, therefore, has been become an important tool for *P. falciparum* genetic studies during the last years (Bansal et al., 2016; Bryant et al., 2017; Cobb et al., 2017; Crawford et al., 2017; Ghorbal et al., 2014; Ng et al., 2016; Sonoiki et al., 2017; Vanaerschot et al., 2017; Wagner et al., 2014).

Genetic manipulation of blood stages in *P. falciparum* using conventional genetic tools has been challenging compared to other more tractable parasite model systems. Low efficiency in transfections, paucity of robust selection markers, uneven segregation of episomes, concatamer formation and inefficiency to maintain linear plasmid that favors chromosomal integration (de Koning-Ward et al., 2015; O'Donnell et al., 2001) are contributing factors that hamper the establishment of stable cell lines, making the process highly time consuming and laborious. The recent development of CRISPR-Cas9 technologies for this parasite has overcome some of those factors making gene disruption more efficient (Ghorbal et al., 2014; Wagner et al., 2014). This technology has helped to unveil the function of many genes since its implementation in asexual blood stage parasites (Bansal et al., 2016; Bryant et al., 2017; Ng et al., 2016; Sidik et al., 2016; Sonoiki et al., 2017). The effector protein Cas9, which belongs to the class II CRISPR bacterial system, needs to recognize an NGG motif on the target DNA that precedes the binding site of the 20-base CRISPR RNA (crRNA) to induce a double-strand break (Mali et al., 2013). In spite of the demonstrated utility of Cas9 for gene editing in this parasite,

the overall AT-content of its genome ranges from 80-90% (Gardner et al., 2002) and the identification of optimal NGG motifs within target genes of interest constrains the use of Cas9-based editing technologies in *P. falciparum*.

Class II CRISPR endonucleases are broadly distributed in diverse bacterial and archaeal lineages. However, only few of them have shown *in vivo* editing properties in eukaryotes (Koonin et al., 2017). The robustness and efficacy of protein effectors such as Cas9 to genetically manipulate a broad variety of biological systems, has led to the exploration of additional protein effectors within the Class II category. Recently, another endonuclease with gene-editing capabilities was recently described. Cpf1 is an enzyme that belongs to the subtype V of the Class II category (Zetsche et al., 2015). There are some mechanistic differences between Cas9 and Cpf1. Cpf1 does not require a trans-activating crRNA (tracrRNA) and it is activated using only a single guide RNA (sgRNA). In addition, Cpf1 recognizes a different binding site or PAM (Protospacer adjacent motif) in the genome. While Cas9 recognizes a "NGG" PAM motif Cpf1 recognizes a "TTTN" binding matrix, an important feature that could be leveraged to facilitate applications in engineering AT-rich genomes, such as in *P. falciparum.* Cpf1 cleaves DNA to produce a staggered cut with a 5 nucleotide (nt) 5' overhang, another major difference with Cas9 blunt cleavage. The Cpf1 cleavage site is more distal to the PAM (18 nt on the non-targeted strand from the PAM and 23 nt from the targeted strand) (Zetsche et al., 2015), than the Cas9 cut site, which is 3 nt upstream from the PAM) (Jinek et al., 2012). In contrast to Cas9, Cpf1 is able to self-process its pre-crRNAs into a mature and functional crRNAs (Zetsche et al., 2017), another major mechanistic difference with Cas9.

Sequence and structural requirements between Cpf1 and Cas9 also differ. Compared to the elaborated secondary RNA structure of sgRNA that interacts with Cas9, the Cpf1-interacting sgRNA is simpler. This sgRNA forms a single stem loop structure in the direct repeat region (DR) that Cpf1 requires for binding. Mutations in the DRs are tolerated as long as they do not disrupt the RNA duplex they form. These direct repeats

are structurally conserved among different bacterial species. However, each Cpf1-endonuclease has its cognate direct repeat sequence. Given the strong conservation, these direct repeats can be used interchangeably, as shown most extensively for Cpf1 isolated from *Francisella novicida* (Zetsche et al., 2015). This property together with the pre-crRNA self-processing mechanism that Cpf1 is able to perform, has facilitated applications on multiplex editing in mammalian cells (Zetsche et al., 2017), plants (Wang et al., 2017) and yeast (Verwaal et al., 2017).

The genetic manipulation of *P. falciparum* relies in the generation of double-strand breaks (DSB) that endonucleases such as Cas9 and Cpf1 make in the genome. In most of the eukaryotes, those DSBs can be repaired through either a homologous recombination (HR) pathway or an "error-prone" non-homologous end-joining (NHEJ) pathway. Each of those pathways requires a set of specialized proteins and substrates to complete a specific series of tasks during the DSBs repair. The HR pathway utilizes a homologous template as substrate, such as a sister chromatid or a donor plasmid (Lee et al., 2014). The *P. falciparum* genome contains almost all the conserved ortholog proteins involved in a canonical HR pathway when compared to humans or yeast. For instance, the complex Rad51-Mre11 that initially senses DSBs has homologous proteins is present in the parasite's genome. On the other hand, the NHEJ pathway does not require a homologous template and it ligates broken ends together, increasing the chance of generation of indels. Bioinformatic analyses have confirmed the absence of NHEJ associated proteins in the *P.falciparum* genome (Gardner et al., 2002). Moreover, *in vivo* studies have supported those findings showing the absence of NHEJ products after parasites were treated with endonucleases (Straimer et al., 2012). Conserved and crucial proteins such as the heterodimer Ku70/80 that binds broken ends in other systems are absent in the parasite's genome. A recent study suggested that *P. falciparum* might repair DSBs using an alternative mechanism such as a microhomology-mediated end joining (MMEJ) pathway, which is NHEJ-independent (Heinberg et al., 2013). However, the frequency of those events is very low and larger deletions besides the locus analyzed, were not examined (Lee et al., 2014)

Here we describe the use of Cpf1 as an alternative genome-editing tool to manipulate native loci in *P. falciparum* asexual blood stages. We test the activity of Cpf1 orthologs from *Lachnospiraceae bacterium* (LbCpf1) and *Francisella novicida* (FnCpf1) and show that highly efficient editing occurs when a homologous template is provided, and is detectable within a single cellular division cycle. Importantly, we demonstrate that CRISPR-Cpf1 can successfully target highly AT-rich regions that could not previously be modified using CRISPR-Cas9. Through genome-wide bioinformatics analyses, we predict that Cpf1 has a larger number of specific target sites compared to Cas9 that can be used to achieve various genetic modifications. Altogether, we show that CRISPR-Cpf1 is a useful complement to the *P. falciparum* functional genetics toolkit.

RESULTS

In vivo editing of native *P. falciparum* loci using Cpf1 endonucleases

We used a two-plasmid system to facilitate: (1) expression of CRISPR-Cpf1 endonuclease components; and (2) simultaneously provide a DNA template for homology directed repair (Fig 1a). The first plasmid is a pfYC3-based vector (Wagner et al., 2013) that uses human dihydrofolate reductase (*hDHFR*) as a selection marker. This expression plasmid utilizes a *P. falciparum* calmodulin 5'UTR (CAM) to express the Cpf1 orthologs and a PcDT 5'UTR (Crabb et al., 1997) to polycistronically express the T7 RNA polymerase (T7RNAP) and human DHFR (*hDHFR*) selective marker (Fig. 4-1A). The second plasmid is based on the pUF-1 vector for *P. falciparum* (Ganesan et al., 2011) and serves as a donor vector to repair the Cpf1-gRNA-induced DNA double-strand break. For this, regions homologous to the 5'- (left homologous region) and the 3'- (right homologous region) ends of the target gene flank a *ydhodh* expression cassette were included. The selective marker allows for selection of editing events as well as for remaining non-integrated episomes. In addition, the donor plasmid contains a T7 promoter-dependent expression cassette for producing Cpf1-sgRNA (Fig 4-1A). Our group has previously established the orthogonal transcriptional activity of T7 RNAP in *P. falciparum* and that the stem-loop of the T7 terminator does not interfere Cas9 cleavage *in vitro* or *in vivo* (Wagner et al., 2014).

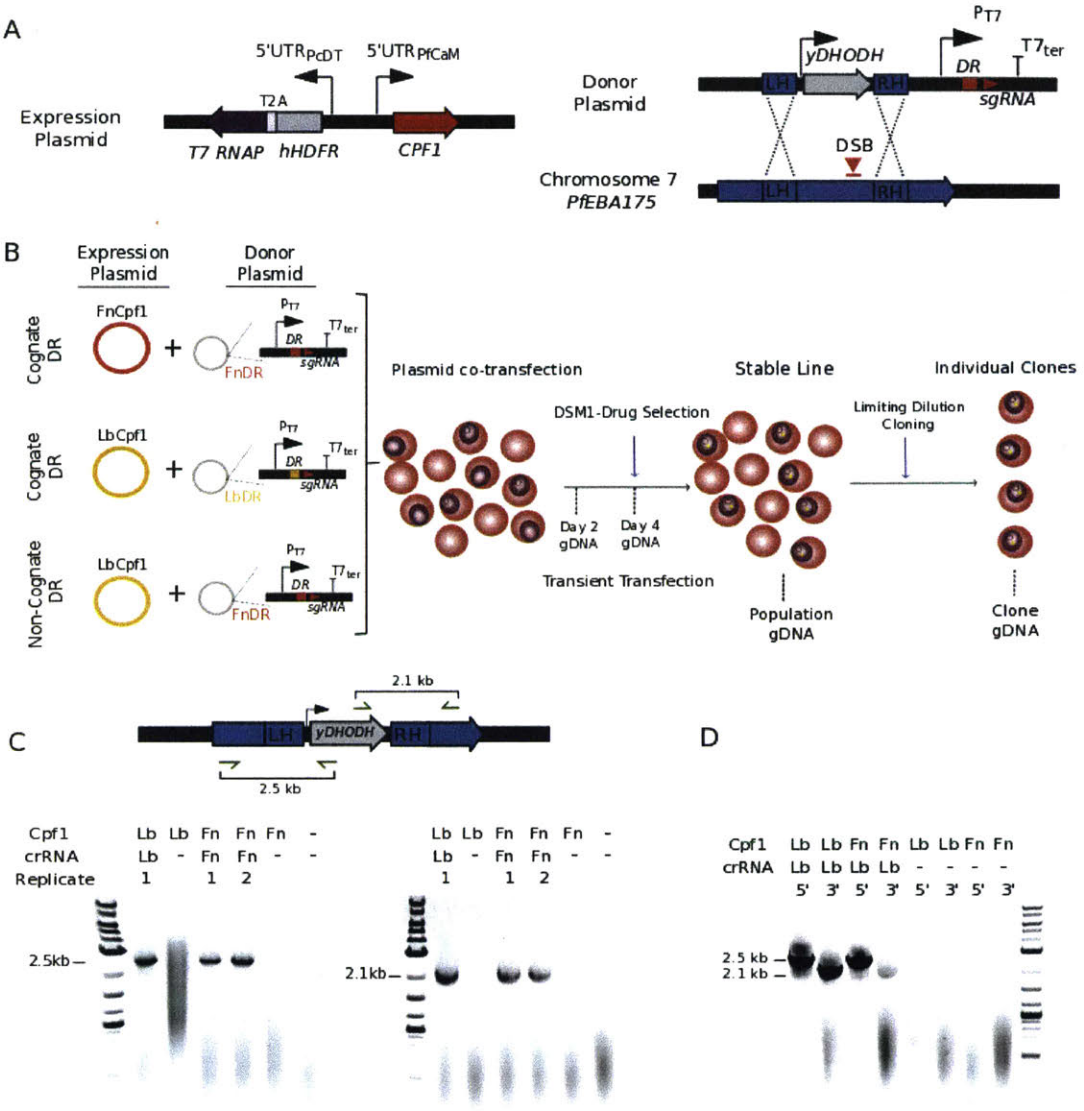


Figure 4-1. Cpf1 editing activity on the *PfEBA-175* locus in *P. falciparum* blood stage parasites

A) Genetically encoded system to edit non-essential genes composed of a Cpf1-T7RNAP expression plasmid and a crRNA donor plasmid. The Cpf1-T7RNAP plasmid includes the genes that encode for either *FnCpf1* or *LbCpf1* bacterial orthologs, and the T7RNAP gene to transcribe the Cpf1-sgRNA. The donor plasmid contains homologous regions flanking a *yDHODH* expression cassette to repair Cpf1-double strand break and a T7 promoter expression cassette driving the transcription of a Cpf1-specific Direct Repeat (DR), the sgRNA and the T7 terminator. This plasmid was used a base-plasmid with modular features to replace variable pieces and target different parasite loci. Coordinates indicating Cpf1-cut site on the *PfEBA175* locus and the designed homologous regions are indicated on the native locus diagram. **B)** Integrating Cpf1 genetic components in blood stage parasites to induce Cpf1-editing on the *PfEBA175* locus. Late stage NF54attB parasites are co-transfected with either the *FnCpf1* expression vector and the *Fn-EBA-175* or *Lb-EBA-175* donor plasmids, or with the *LbCpf1* expression vector and the *Lb-EBA-175* donor plasmid. Genomic DNA from transient and stable parasite lines of each line was isolated for further

analysis. **C)** Transformant parasites show editing on the *PfEBA175* locus with either FnCpf1 or LbCpf1 endonucleases using their cognate DR. *Top panel*: Edited *PfEBA-175* locus and coordinates of primer hybridization locations and expected PCR products to confirm homologous recombination at the 5' and at the 3' end of the targeted region are indicated on the diagram. *Bottom panel*: Amplification products that correspond to the editing events on the *PfEBA-175* locus at the 5' end (*left panel*) or 3' end of the targeted locus (*right panel*) using FnCpf1 or LbCpf1 and their cognate DR on the donor plasmids, but not on the no crRNA or wild type control lanes (no Cpf1, no crRNA). **D)** Cpf1 editing activity on the *PfEBA-175* locus using interchangeable DR. FnCpf1 and LbCpf1 expression vectors show editing activity at the 5' and at the 3' ends using LbDR on the donor plasmid to target the *PfEBA-175,* but not when there is no crRNA (negative control). The experiment was repeated at least three times in independent transfections, confirming the editing activity of LbCpf1 and FnCpf1 on the *EBA-175* locus.

To demonstrate the editing activity of Cpf1, we chose a previously Cas9-validated target as a proof-of-concept (Wagner et al., 2014). The erythrocyte-binding antigen (*PfEBA-175*; PF3D7_0731500) is a non-essential parasite invasion ligand that binds to glycophorin A on erythrocytes (Mayer et al., 2001). Donor plasmids targeting *EBA-175* were built using homologous regions of ~500bp, and a sgRNA targeting positions 723-742 bp from the start codon on the non-targeted strand selected. The left and right homologous regions were 225 bp upstream and 22 bp downstream from the cut site, respectively. Cpf1 endonuclease cleavage activity depends on the recognition of the DR stem loop that preserves the RNA duplex for DNA cleavage. Previous studies have described cognate DR sequences for each bacterial Cpf1 orthologs (Zetsche et al., 2015). To test the activity of the two Cpf1 orthologs, we built donor plasmids with Lb-DR (*Lb-EBA-175* plasmid) and Fn-DR (*Fn-EBA-175* plasmid) sequences and the same gRNA targeting the *EBA-175* locus.

We defined an editing event as integration of a selection marker cassette at the cut site of the target gene when DNA is repaired via HR. Late blood stage NF54*attB P. falciparum* parasites were co-transfected with each variant of Cpf1-T7RNAP expression plasmid and their cognate donor plasmid (Fig 4-1B). Transfections showed evidence of editing by days 2 and 4 post-transfection only when FnCpf1 or LbCpf1 and a sgRNA that recognizes the *EBA-175* locus was used, but not when parasites where co-transfected with a plasmid lacking sgRNA (Fig 4-1C). Editing at the *EBA-175* locus was verified by

amplifying and sequencing a region corresponding to the y*dhodh* expression cassette and flanking regions at the 5' and the 3' ends of the targeted locus (Fig 4-1C).

In vitro and *in vivo* studies showed the strong structural conservation of the Cpf1 DR sequences and demonstrated the nuclease activity of FnCpf1 and LbCpf1 when the DR sequence is interchanged with others Cpf1-ortholog DR (Verwaal et al., 2017; Wang et al., 2017; Zetsche et al., 2015; Zetsche et al., 2017). Hence, we tested the editing activity of LbCpf1 and FnCpf1 using only the *LbCpf1 DR* sequence in the crRNA (*Lb-EBA-175* plasmid) to target the *EBA-175* locus (Fig. 4-1B). Editing activity of the *EBA-175* locus was observed with either cognate or non-cognate DR sequences, consistent with previous studies showing interchangeable use of Cpf1 DR sequences (Fig.4-1D). To simplify our experimental design, we decided to use *Lb-DR* sequences for subsequent experiments.

To assess the efficiency of *EBA-175* locus disruption by LbCpf1 and FnCpf1, we obtained individual clones by limiting dilution and determined the frequency of edited parasite clones. Remarkably, all clones tested (21/21 from each endonuclease transfection) showed editing of the EBA-175 locus FnCpf1, confirming a high frequency editing at the population level (Sup Fig 4-S1). Taken together, these results we demonstrated that both LbCpf1 and FnCpf1 endonucleases are functional in *P. falciparum* blood stages, and can be used successfully to directly disrupt non-essential target loci. Importantly, we detected editing events during the first cell cycle post-transfection, and achieved high gene disruption frequencies, comparable to those observed previously for the Cas9 endonuclease in *P. falciparum* (Sup Fig 4-S2) (Ghorbal et al., 2014; Wagner et al., 2014).

Cpf1 displays efficient editing activity in AT-rich regions

The Cpf1 binding matrix is characterized by the presence of a TTTN PAM motif, which is highly frequent motif in the AT-rich *P. falciparum* genome. Given the frequency of these regions, we chose a highly AT-rich target for which previous attempts of gene editing using the CRISPR-Cas9 system had proven unsuccessful. A separate ongoing Cas9-screening project in our lab to target RNA-Binding Proteins in the parasite, showed a particular uneditable AT-rich gene (unpublished work). The recalcitrant locus encoded for the pentatricopeptide repeat domain-containing protein (PPR). This protein belongs to a large family of RNA-binding proteins involved in several aspects of RNA metabolism, such as splicing, editing, RNA processing, stability and translation (Manna, 2015). PPR proteins act as important post-transcriptional regulators and play key roles in gene expression control, especially in plants chloroplast (Small et al., 2013). However, while the *P. falciparum* genome encodes two PPRs, their function(s) are unknown. One of the PfPPR (*PF3D7_1406400*) proteins is annotated in the parasite genome as a putative RNA binding protein with no homology to other organisms and possible function in organellar compartments. PfPPR is a 608 aa protein encoded in a region that has 80% AT content and for which no successful disruption attempts have been reported in *P. falciparum* or other *Plasmodium* species. We decided to test the ability of our Cpf1 editing tool to target and edit the PfPPR gene as a proof of concept of the usefulness of this system for targeting AT-rich regions.

The homology-direct repair pathway we use in the parasite to repair Cpf1 DSBs, offers the possibility of installing additional DNA elements at a target DNA locus that can be useful for studying gene function. To achieve this, we changed the configuration of the donor vector to include the components of our previously reported anhydrotetracycline-inducible TetR-DOZI-aptamer translation control system (Ganesan et al., 2016), which enables regulation of gene expression (Fig 4-2A). This donor vector contains left and right homologous regions that were designed 327 bp and 115 bp away from the cut site and that will repair Cpf1-induced double strand breaks via homologous recombination, similar to our *EBA-175* donor plasmid. The left homologous region was designed to be a

chimeric DNA fragment made of parasite gDNA and a synthetic DNA sequence recoded to prevent Cpf1 from re-cutting the edited locus. The right homologous region corresponds to the native 3'UTR of the PPR gene. Both of these homologous regions flank the TetR-DOZI regulatory elements and the Blasticidin S resistance marker expression cassette (Fig 4-2A). In addition, a 10x TetR-aptamer array was included in the donor vector to be installed at the 3' end of the gene after the last amino acid of the targeted protein. The Cpf1-sgRNA was designed at the 3' end of the PPR gene, at a region close to where previous Cas9-sgRNAs were designed. The Cpf1-sgRNA was designed at position 1713 from the start codon to facilitate the aptamer installation at the end of the gene during DNA repair. The PPR-crRNA expression cassette was included outside of the homologous regions and the sgRNA was designed on position 1714 - 1724 from the start codon of the PPR gene (Fig 4-2A).

To demonstrate the ability of Cpf1 to target AT-rich regions and our capacity to install additional regulatory elements during recombination repair, NF54^{attB} parasites (Nkrumah et al., 2006) were co-transfected with the LbCpf1-T7RNAP expression plasmid and the Lb-PfPPR donor vector in the presence of anhydrotetracycline (aTc) (Fig 4-2B). PCR analysis of stable Blasticidin-resistant parasites confirmed the expected integration event of the donor plasmid at the 5' and 3' ends of the recombination events on the disrupted *PfPPR* native locus (Fig 4-2C). The experiment, repeated 3 times in independent transfections, yielded in every instance stable parasite lines that were edited at this locus, which demonstrates the reproducible use of Cpf1 to edit an AT-rich region that could not be modified using Cas9. Furthermore, this experiment demonstrates that, as with Cas9, CRISPR-Cpf1 facilitates installation of additional genetic information during DNA repair. Future experiments aimed at the identification of a phenotype associated to the PfPPR disruption will shed the light on the function of this protein in parasite development.

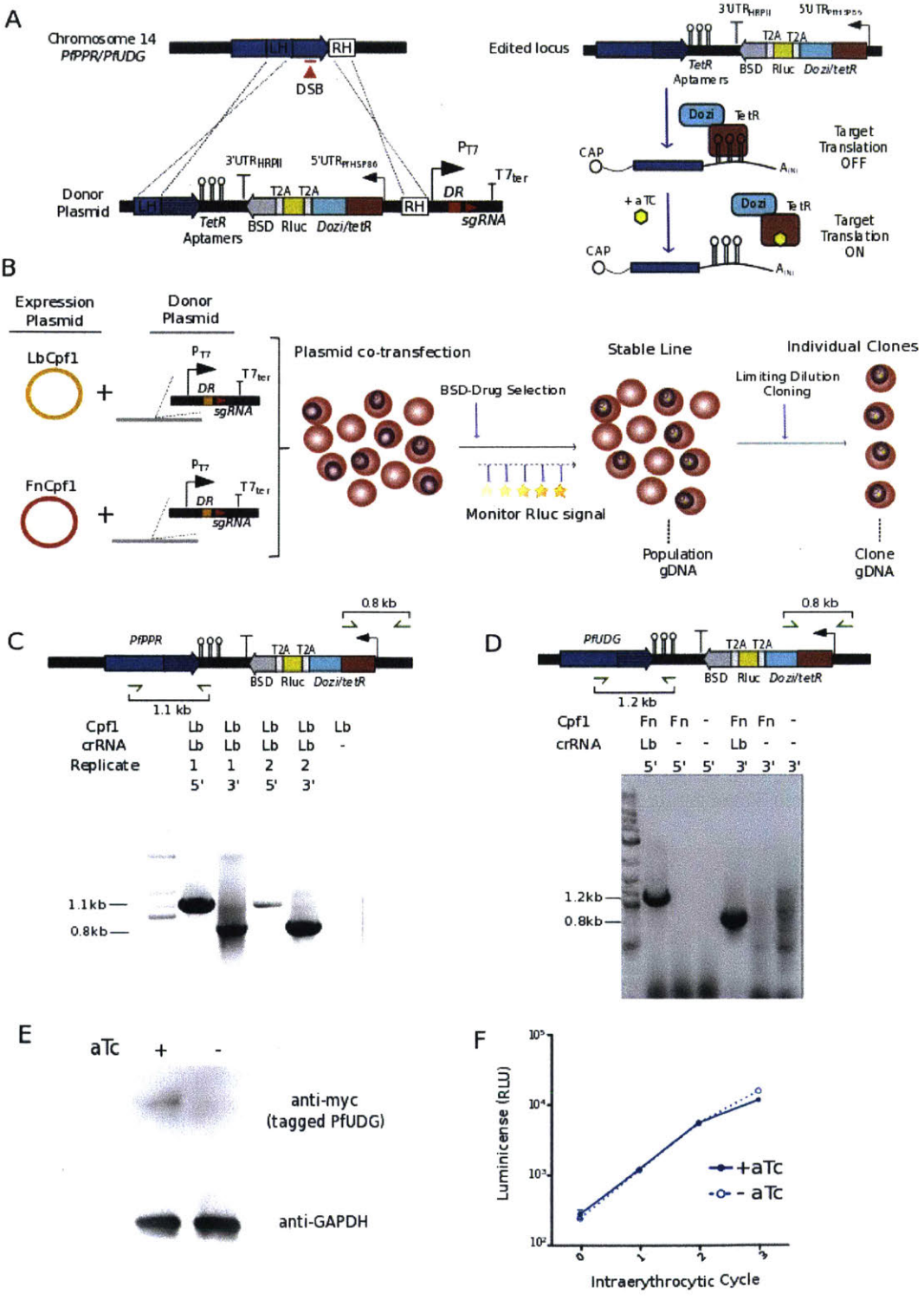

A
Chromosome 14
PfPPR/PfUDG
LH
RH
DSB
Donor Plasmid
3'UTR HRPII
5'UTR PfHSP86
T2A T2A
PT7
T7ter
DR
sgRNA
TetR Aptamers
BSD Rluc Dozi/tetR
Edited locus
Dozi
TetR
CAP
Target Translation OFF
+ aTc
Target Translation ON
B
Expression Plasmid
Donor Plasmid
LbCpf1
FnCpf1
Plasmid co-transfection
BSD-Drug Selection
Monitor Rluc signal
Stable Line
Individual Clones
Limiting Dilution Cloning
Population gDNA
Clone gDNA
C
PfPPR
0.8 kb
1.1 kb
Cpf1 Lb Lb Lb Lb Lb
crRNA Lb Lb Lb Lb -
Replicate 1 1 2 2
5' 3' 5' 3'
1.1kb
0.8kb
D
PfUDG
1.2 kb
Cpf1 Fn Fn - Fn Fn -
crRNA Lb - - Lb - -
5' 5' 5' 3' 3' 3'
1.2kb
0.8kb
E
aTc + -
anti-myc (tagged PfUDG)
anti-GAPDH
F
Luminicense (RLU)
+aTc
- aTc
Intraerythrocytic Cycle

Figure 4-2: Cpf1 shows editing activity in AT-rich regions and can mediate translational control on targeted genes

A) Genetically encoded system using an integrated platform of a new donor plasmid configuration to inducibly control gene expression on Cpf1-disrupted genes. A TetR/DOZI linear donor plasmid was designed with homologous regions to target either the *PfPPR* and *PfUDG* loci that are flanking the regulatory machinery for translational control. Late stage NF54attB parasites are co-transfected with an LbCpf1 expression plasmid and Lb-PPR donor plasmid to target the *PfPPR* locus and with an FnCpf1 expression plasmid and Lb-UDG donor plasmid to target the *PfUDG* locus. At the Cpf1-cut site, the 10X TetR-aptamer array and the TetR/DOZI/BSD/Rluc expression cassette, included in the donor plasmids are integrated during DNA repair via homologous recombination. To characterize the function of the targeted genes, the expression of the native promoter of each gene can be controlled in either a repressed state (-aTc) when the TetR/DOZI synthetic module interacts with the 10X TetR-aptamer array or the induced state (+aTc) when the TetR/DOZI module is removed. The Rluc gene is used to monitor parasite growth. The T7 promoter expression cassette driving the transcription of a Cpf1-specific Direct Repeat (DR), the sgRNA and the T7 terminator is outside of the homologous region. Coordinates indicating Cpf1-cut site and the designed homologous regions are indicated on the native locus diagram. **B)** Integrating Cpf1 genetic components and inducible system in blood stage parasites to edit low complexity regions and control the expression of a putative antimalarial target. *Top panel*: LbCpf1 expression plasmid is co-transfected with an Lb-PPR donor plasmid and *Bottom panel*: FnCpf1 expression plasmid is co-transfected with an Lb-UDG donor plasmid. Genomic DNA from BSD-resistant parasites was isolated for further analysis. **C)** LbCpf1 shows editing activity on the PfPPR low complexity region. *Top panel:* edited PfPPR locus with integration event. The coordinates of primer hybridization locations and expected PCR products to confirm homologous recombination at the 5' and at the 3' end of the targeted region are indicated. *Bottom panel*: Amplification products that correspond to the editing event on the *PfPPR* locus are shown at the expected size. **D)** FnCpf1-mediated cleavage integrates regulatory elements at the disrupted PfUDG gene. *Top panel:* edited PfUDG locus with integration event. The coordinates of primer hybridization locations and expected PCR products to confirm homologous recombination at the 5' and at the 3' end of the targeted region are indicated. *Bottom panel:* Amplification products that correspond to the editing event on the *PfUDG* locus are shown at the expected size. **E)** Western Blot showing inducible regulation of the PfUDG expression in –aTc and +aTc conditions using an anti-myc antibody, a tag included at the C-terminus of the protein. The GAPDH loading control was included in each sample. **F)** PfUDG is not essential for *P. falciparum* blood stages. Renilla luciferase-based growth assay on PfUDG edited parasites in –aTc and +aTc conditions. Parasites were monitored during 3 IDCs starting at synchronous ring stage parasites.

Achieving inducible regulation of a native locus using Cpf1 editing activity

Previous work from our group has shown how CRISPR-Cas9 editing technologies can be integrated with our TetR-DOZI-aptamer inducible system to regulate the expression of native *P. falciparum* loci throughout intra-erythrocytic parasite development (Ganesan et al., 2016). Our group has used this integrated system to target and regulate a variety of proteins in the parasite, and determine their essentiality for parasite survival. Features included in the developed TetR-DOZI-aptamer system, such as the constitutive expression of a renilla luciferase, enable monitoring of parasite expansion in aTc removal growth assays. In the presence of aTc, the native targeted locus is expressed and thus the renilla luciferase signal is high, indicating active protein translation and thus parasite

viability. In contrast, in the absence of aTc the TetR-DOZI module remains bound to the RNA aptamer blocking protein translation and if the protein is essential for parasite, the renilla luciferase signal should decrease, suggesting a growth defect associated to the protein knockdown. Therefore, we sought to establish that Cpf1 could also efficiently substitute into our framework for installing a 10x-TetR aptamers at the 3' end of native parasite genes to regulate their expression as a first step towards studying their more detailed biology.

As a proof-of-concept, we selected the *P. falciparum* uracil DNA glycosylase (*Pfung*). The PfUNG is a key enzyme for DNA repair involved in base excision repair (BER), however disruption of UNG enzymes has not shown essentiality in other organisms (Robertson et al., 2009). BER corrects DNA damage by recognizing DNA lesions such as mismatched bases (Krokan and Bjoras, 2013). The BER is presumably an essential pathway for parasite survival as it prevents the accumulation of mutations during DNA replication. In *P. falciparum*, only one *ung* has been identified, and the function of this protein has not yet been investigated in detail. Previous studies suggest that this protein could be a potential target for the design of new antimalarial drugs, as a bacteriophage Uracil glycosylate inhibitor-PBS1, a known UNG inhibitor in bacteria inhibited parasite growth (Suksangpleng et al., 2014). However, no detailed efforts were made to establish that inhibition of *Pf*ung function is the dominant cause of parasite death upon treatment with PBS1.

To achieve inducible regulation of *Pfung*, we designed a donor vector for homology-directed repair similar to the design used in the section above. The left homologous region amplified was a 528 bp DNA fragment and it was designed 20 bp upstream from the Cpf1-cut site. The right homologous region was a 496 bp fragment located 464 bp downstream from the cut site. To facilitate protein detection by Western blot and regulation of expression, a *myc* tag upstream of the 10X-aptamer array was included in the construct to be installed at the C-terminus of the *PfUNG* protein. The T7 UDG-

crRNA cassette was designed to target the +549–569 nt region from the start codon of the gene (Fig 4-2A). Late stage NF54^{attB} parasites were co-transfected with the FnCpf1-T7RNAP expression plasmid and the P*fung* donor plasmid in the presence of aTc. PCR analysis of stable Blasticidin S-resistant parasites and sequencing to confirm the junctions between the native locus and the repair construct demonstrated successful targeting of the *Pfung* locus (Fig 4-2B and 2D). We confirmed that aTc-dependent regulation of PfUNG protein expression is achieved as expected by the genetic modifications installed (Fig 4-2E).

To determine the essentiality of *Pfung* in blood stage parasites, we performed knockdown (aTc-withdrawal) growth assays over at least 3 cell divisions. Parasite survival in the presence or absence of aTc was monitored using constitutively expressed Renilla luciferase reporter signal introduced with the donor vector (Fig 4-2A) (Ganesan et al., 2016). Interestingly, we observed no defect in parasite growth under knockdown conditions, even though *Pf*UNG protein levels were substantially depleted (Fig. 4-2F). These results indicate that depletion of PfUDG does not lead to an obvious growth defect in short-term viability assays performed under normal cell culture conditions. Further experiments will be required to determine whether there are long-term fitness costs associated with PfUNG knockdown, or if loss of PfUNG expression results in parasites becoming hypersensitive to conditions favoring increased uracil incorporation into genomic DNA. The stable parasite lines constructed here will be useful for future experiments aimed at determining the role of this protein in DNA damage repair.

The validation of Cpf1-genome editing in *P. falciparum* genome holds promise to significantly expand the available toolkit for performing functional studies by increasing the flexibility with which different genomic regions can be specifically manipulated. This will facilitate better identification and characterization of essential genes, towards the goal of discovering novel antimalarial therapeutic targets.

Cpf1 sgRNAs binding sites show a larger genome-wide coverage in *P. falciparum* than Cas9 sgRNAs

We have shown that Cpf1 can be implemented in *P. falciparum* to edit native loci with efficiencies and kinetics similar to Cas9. Given that AT-rich genomic regions in the parasites could be selectively targeted for modification by Cpf1, we wished to gain a comparative understanding of the genome-wide availability of Cpf1 versus Cas9 target sites in the ~80% AT-rich *P. falciparum* genome. Previously, no systematic comparative analysis of Cas9 and Cpf1 has been reported for *P. falciparum*, thus we anticipate such an analysis will help facilitate selection of Cpf1-gRNA versus Cas9- sgRNA designs for achieving different genome manipulation outcomes while aiming to maximize specificity.

We performed a genome-wide computational analysis to characterize potential Cas9 and Cpf1 target sites across the *P. falciparum* genome. We started the analyses by defining specific criteria that would help to select good quality Cpf1 sgRNAs. A total of 5,712 genes from the *P. falciparum* genome database were analyzed and Cpf1- and Cas9-binding sites were identified at the 5' and at the 3' –ends of each gene. We delimited the regions at those ends by taking only the 500bp region upstream and downstream from the start and from the stop codon. We rationalized those were the regions of our interest to install and introduce DNA modifications at each target gene for regulatory purposes. A collection of 1,098,142 possible sgRNAs was identified based on each endonuclease canonical PAM recognition site and the information regarding to their sequence composition and coordinates of their location were stored in a new *P. falciparum* sgRNA database (Fig 4-3). We found that the number of Cpf1 sgRNAs outnumbered outnumbered the Cas9 sgRNAs (Fig 4-3A). The total of sgRNAs identified for Cas9 and Cpf1 were equally distributed around the 5' and 3' ends of each gene analyzed.

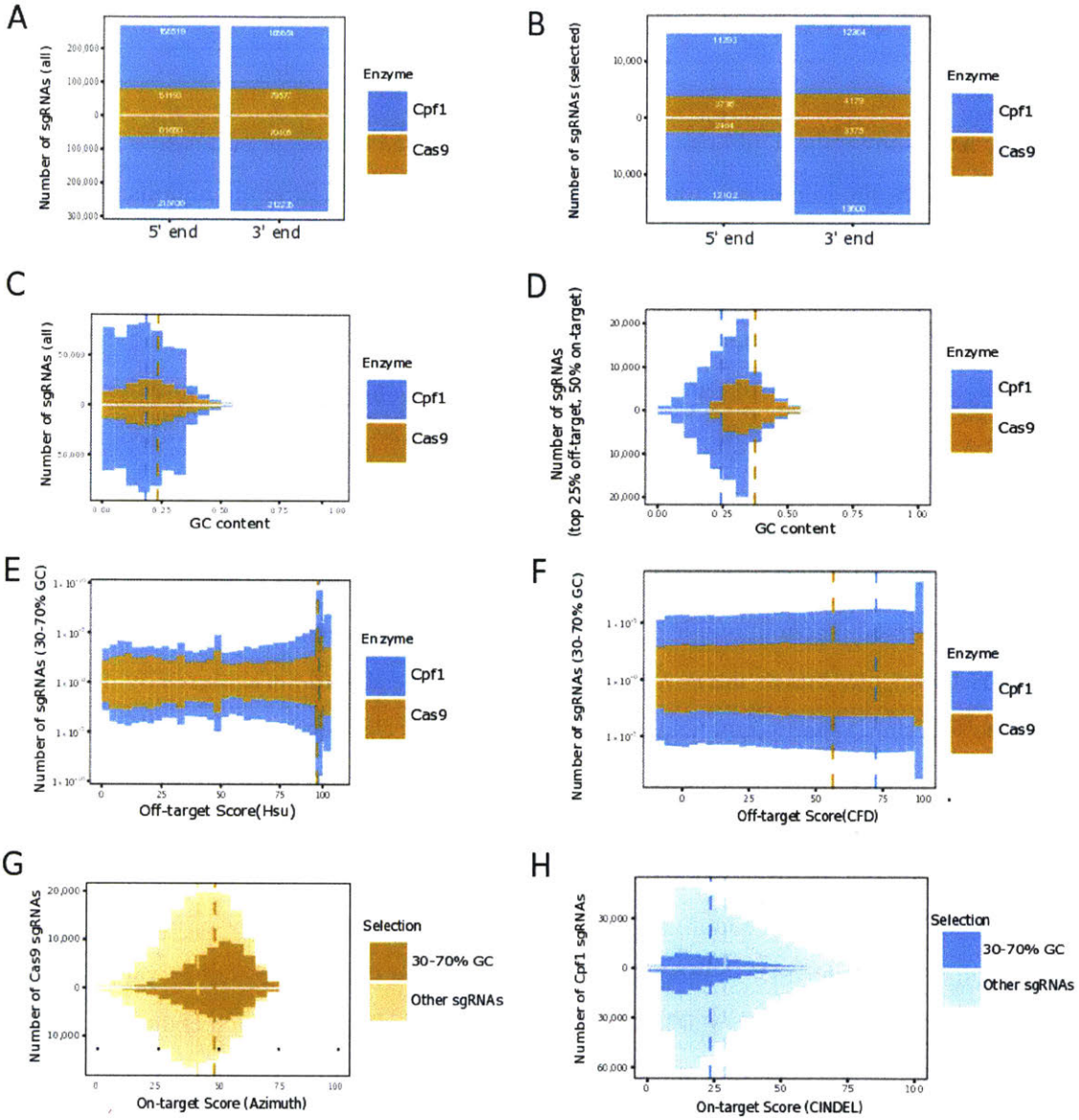


Figure 4-3. Global sgRNA binding analysis on the *P. falciparum* genome

Cpf1 shows a larger number of binding sites on the parasite's genome than Cas9. Bars above the horizontal axis indicate sgRNAs on the sense strand and bars below the horizontal indicate sgRNAs on the antisense strand. Dotted lines indicate averages of the given data series. **A)** A total of 1,098,142 sgRNAs with Cas9 or Cpf1 PAM sequences were found within the -/+ 500bp ranges at the 5' and the 3' ends of each target. **B)** Total of all selected Cas9 and Cpf1 sgRNAs within the 30-70% GC content, top 25% Hsu off-target score and top 50% on-target score. Percentage of GC content in all **C)** and selected gRNAs, on the top 25% off-target and 50% on target scores **D)**. Distribution of off-target scores in all **E)** and selected gRNAs after GC content filtering **F)**. Distribution of on-target scores in all **(E)** and selected gRNAs after GC content filtering

To predict the specificity of the identified sgRNAs for both Cas9 and Cpf1 endonucleases on the *P. falciparum* genome, we used criteria previously characterized for Cas9 and Cpf1 sgRNAs on mammalian cells since no on-target or off-target scoring systems have been characterized for *P. falciparum* sequences. We used the following criteria to score and select specific Cas9/Cpf1 sgRNAs: 1) Sequences with 30-70% GC content; 2) top with 25% Hsu (Hsu et al., 2013) or CFD (Doench et al., 2014) off-target score system; and 3) the 50% Azimuth (Doench et al., 2016) and CINDEL (Kim et al., 2017b) on-target score. Importantly the Azimuth on-target score system was developed for Cas9 sgRNAs and the CINDEL on-target score system was developed for Cpf1 sgRNAs. Both off-target scoring systems here (Hsu and CFD) were originally developed for Cas9. No reports to quantify Cpf1-off target effects are available yet.

The different scoring systems included in our algorithm were developed using different approaches to examine the Cas9 activity on mammalian cells. Each of those algorithms utilizes a different experimental framework to assign values and quantify either the on-target or the off-target effect of Cas9, based on predetermined criterion established during experimental set up. The Hsu system was developed based on the Cas9 cleavage activity that results from a sgRNA with single (or multiple) mismatches and the resulting number of indels on different on undesired targets, which in this case would be counted as the off-target effect (Hsu et al., 2013). In contrast, the CFD system was developed correlating the sequence and distribution preferences of the sgRNA to generate either an on-target or an off-target effect (Doench et al., 2014). The Azimuth on-target scoring system quantifies the Cas9 off-target activity from an sgRNA library generated after a positive and negative selection screens on human and mouse cells. The off-target activity is incorporated into an algorithm to optimize sgRNA libraries and maximize on-target activity (Doench et al., 2016). Lastly, the CINDEL system was recently developed to assess Cpf1-editing activity on human cells using AsCpf1 and LbCpf1 (orthologs commonly used on human cells). This system records indel frequencies from a library of >11,000 sgRNAs delivered into human cells (Kim et al., 2017b).

We developed a platform that includes those criteria to rank optimal Cas9 and Cpf1 sgRNAs as it compares the results of the selected sgRNAs from each scoring system. Our experimental evidence using Cas9 and Cpf1 sgRNAs designed on platforms that utilize algorithms implemented on human cells suggests that these scoring systems can be applied to the *P. falciparum* genome.

Using the 1,098,142 sgRNAs database as input, we selected 63,113 sgRNAs by applying the above criteria, resulting in 97% reduction (Figure 4-3B). Table 4-S1 shows quartile ranges for the on-target and off-target scores of the different sgRNA database subsets. Values of the on-target and off-score included in the quartiles were calculated after selecting sgRNAs with the 30-70% GC content. Selected sgRNAs were within top 25% off-target score and top 50% on-target quartiles. Interestingly, when using the CFD score or the single Hsu score off-target to select sgRNAs for either Cpf1 or Cas9, we found no significant differences in the selected sgRNAs (Sup Fig 4-S3). The number of sgRNAs around the 5' end resulted in a larger reduction possibly because of the AT-richness of the promoter sequences in that region and the selected sgRNAs were low-scored due to the low GC content. Low quality on Cas9 sgRNAs associated with low GC content has been previously established (Brick et al., 2008; Hsu et al., 2013; Ruvalcaba-Salazar et al., 2005).

To identify differences between Cas9 and Cpf1 binding sites at each DNA strand, we analyzed each DNA strand independently. Our analysis showed that Cas9 binding sites are more frequently found in the sense strand, while Cpf1 sgRNAs were more abundant in the antisense strand. The trend of Cas9 and Cpf1 binding sites can be appreciated in all panels of Fig 4-3. This behavior can be partially explained based on the *P. falciparum* codon usage within the coding regions, consulted in the HIVE-CUTs database (Athey et al., 2017). Given the abundance of TAAN PAM regions in the antisense strand, we observed that in the *P. falciparum* genome AA-containing codons are more frequent (34.697% of total codons) compared to TT-containing codons (15.919%) and the AAA

codons nearly triple the TTT codons (9.569% to 3.646%, respectively). On the other hand, the frequencies of GG- and CC-containing codons are almost equally distributed in the parasite's genome (3.188% and 3.779%, respectively). This could explain the difference in strand preference of Cas9 sgRNAs with respect to Cpf1 sgRNAs binding sites. In addition, previous reports have shown the notable bias in the *P. falciparum* genome for G nucleotides in the first (G: 22.19%, C: 9.58%) and third (G: 9.58%, C: 7.82%) bases of each codon (Saul and Battistutta, 1988). This composition makes the sense-strand NGG PAM sequences span in two codons at a higher probability.

Overall Cas9 sgRNAs showed a higher average GC content than the Cpf1 sgRNAs overall (Fig 4-3C,D), an expected result based on the binding sites that this endonuclease recognizes. After applying off-target and on-target filters, the GC content of both Cas9 and Cpf1 increased since the algorithms consider a 30-70% GC content as a parameter to select good quality sgRNAs. Hence, high GC content sgRNAs are less likely to have similar sequences on this particular genome reducing the predicted off-target effect.

Off-target scores using the Hsu and CFD algorithms had an approximate maximum of 100 (all the scores were normalized to 100, being 100 the best score) for both Cpf1 and Cas9 sgRNAs (Fig 4-3E,F). Both the Hsu and CFD scoring algorithms assign weights to the number and position of mismatches, while the CFD algorithm adds penalties for each mismatched nucleotide (Doench et al., 2016). Hence, when applying the CFD scores, sgRNAs showed a larger distribution in the lower percentiles than the Hsu score system. Although the CFD and the Hsu scoring systems were developed to assign values to Cas9-edited mammalian cells, in general the scores are heavily influenced by the number of similar sequences in each genome's off-target sgRNA database. As a consequence, the distribution of the sgRNAs with their off-target scores without filtering for GC content resulted in an additional peak of scores near the minimum score for both scoring systems, corresponding mostly to AT-rich repetitive sequences. For this reason, we included in our analysis an organism- and enzyme-unweighted metric based solely on the number of

mismatches using the Hsu algorithm with an equal-valued position mismatch matrix for use with Cpf1 guides (Fig 4-S4). Importantly, our group has empirically validated the Hsu scores in the design of Cas9 and Cpf1 sgRNAs in *P. falciparum* (scoring system implemented by default on Benchling (www.benchling.com) (unpublished work). For Cpf1 sgRNAs, the Hsu score was calculated with unweighted position matrices.

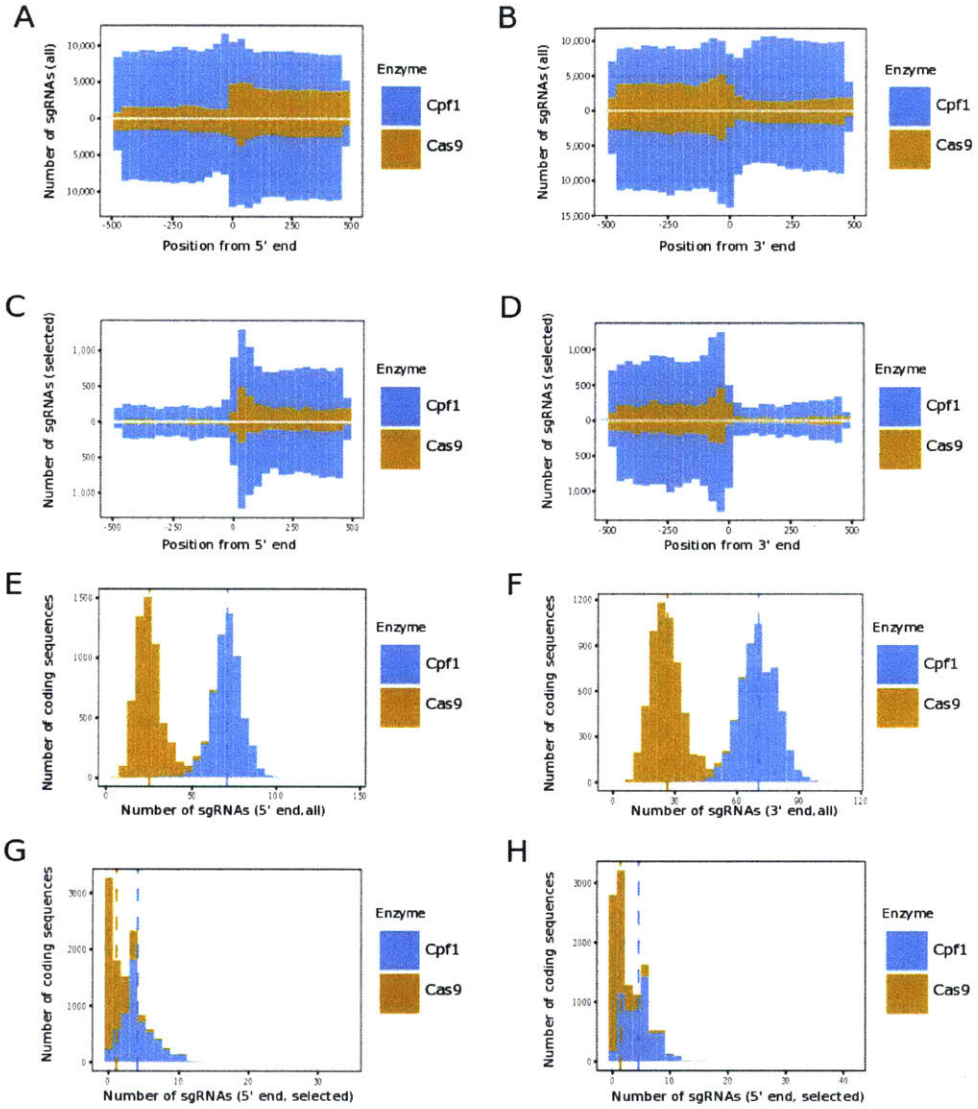


Figure 4-4: Analysis of the Cpf1 and Cas9 sgRNA binding sites distribution within the genes and across all genes on the *P. falciparum* genome

Cpf1 sgRNA binding sites showed a preferable gene-3' end distribution across the parasite's genome. Bars above and below the horizontal axis indicate all and selected sgRNA binding sites on the sense and anti-sense strand respectively. Dotted lines indicate averages of the given data series. Selected sgRNAs within the -/+ 500bp ranges at the 5' and the 3' ends of each gene were analyzed. Distribution of all sgRNAs at the 5' end **(A)** and selected gRNAs **(C)** within at the top 30–70% GC content, top 25% off-target scores, and top 50% on-target scores. Distribution of all sgRNAs at the 3' end **(B)** and selected gRNAs **(D)** within at the top 30–70% GC content, top 25% off-target scores, and top 50% on-target scores. Number of total sgRNAs per gene at the 5' **(E)** and at the 3' end **(F)** was calculated, as well as the number of selected gRNAs at the 5' **(G)** and at the 3' end **(H)**.

On-target scores were also normalized to 100. The Azimuth on-target scores increase for Cas9 sgRNAs when filtering for their GC content (Fig 4-3G), given that the GC content is associated with on-target activity (Doench et al., 2016; Doench et al., 2014). Conversely, the GC content filter decreased the CINDEL on-target scores of Cpf1 sgRNAs (Fig 4-3H). Importantly, for Cpf1 gRNAs, higher scores outside of the 30-70% GC content range is likely to be an artifact of the CINDER algorithm, as the authors note that the score calculation was developed solely on 30–70% GC sgRNAs, and caution against use of this metric with other sgRNA compositions (Kim et al., 2017b). Interestingly, the two on-target scores included in our analysis for both Cas9 and Cpf1 sgRNAs skewed in opposite directions on the x-axis, reflecting opposite score trends for each algorithm (Fig 4-3G,H). This might be due to the fact AT-rich Cpf1 sgRNAs are more abundant, but have lower quality, which gives low on-target CINDEL scores, while the opposite is true for Cas9 sgRNAs. However, the two on-target score algorithms are not comparable since they use different metric systems, thus the percentile value (instead of the score value) was used as a cutoff to select good on-target sgRNAs in our database.

Another aspect we analyzed for our selection of Cpf1 and Cas9 sgRNAs was the distribution of the binding sites within genes and across all genes in the parasite's genome (Fig 4-4). We first analyzed the distribution of the total number of sgRNAs for both Cpf1 and Cas9 with respect to the 5' (Fig 4-4A) and 3' (Fig 4-4B) ends of each target region. Next, we selected sgRNAs using our described criteria at the

5' (Fig 4-4C) and 3' ends (Fig 4-4D) of the target regions. Comparable to our previous analyses, we observed an overall reduction in the number of sgRNAs after selection. The distribution of Cpf1 sgRNAs was augmented around the 3' end of the coding region and outside of it, compared to the distribution of Cas9 sgRNAs. The distribution of the selected Cas9 sgRNAs was higher within the coding region, given its higher GC content.

Overall, we were able to confirm that the number of Cpf1 PAM motifs associated to sgRNA sequences exceeds the number of Cas9 sgRNAs in the *P. falciparum* genome,

even after applying quality filters. At a final count, we found 4,080 genes (71.43% of total genes) that contain at least one Cas9 sgRNA fulfilling the characteristics of 30–70% GC content, top 25% off-target scores, and top 50% on-target scores included in our analysis. In contrast, 5,673 genes (99.32% of total genes) were found to contain at least one Cpf1 sgRNA. Only 23 genes (0.3909 %) contained no satisfactory sgRNAs binding sites by either enzyme (either at the 5' or 3' end of the target region) (Fig 4-4 E,H). Those genes include 10 conserved proteins with unknown function, 6 apicoplast ribosomal proteins, 2 tRNAs sequences, 1 5S ribosomal RNA, 1 small nucleolar RNA, and 1 unspecified product. Graphs including the information for all 5712 genes can be viewed online as supplementary files, available at http://bit.ly/2eVme4C.

The criteria used in our analysis showed that the quality of Cpf1 gRNAs might be lower compared the Cas9 sgRNAs. However, our results suggest that the high abundance of Cpf1 binding sites across the *P. falciparum* genome offers more opportunities to edit a higher number of loci using this nuclease compared to Cas9. Our genome-wide *in silico* prediction analysis identifies Cpf1 sgRNAs that have a predicted greater probability of mediating a specific editing outcome. However, other sgRNAs rejected by the applied filters may also mediate specific genome-editing outcomes. Importantly, the experimental data used to inform our predictions are obtained from mammalian cells, which have distinct genome composition and DNA repair mechanisms than *P. falciparum*. A systematic analysis, combining large-scale experimental data from *P. falciparum* with computational analysis, would help refine our Cpf1 and Cas9 sgRNA prediction models to facilitate providing more accurate guidelines for sgRNA selection in *P. falciparum*.

DISCUSSION

CRISPR-Cas systems have become the method of choice for genetically modifying different types of cells and even multicellular organisms. CRISPR-Cas9 utilizes a single effector protein to target DNA. The Cas9 endonuclease has been also used to inhibit or activate genes, modify epigenetic marks and more recently, this protein has been engineered to inducibly control its editing function (Nunez et al., 2016). CRISPR-Cas9 is a powerful gene-editing tool that has accelerated gene function studies in large scale, especially in challenging organisms with limited genetic tools available, including *P. falciparum*. The CRISPR-Cpf1 system has been recently described as an alternative DNA editing tool that recognizes a different PAM, making it more suitable for organisms with an AT-rich genome. Nevertheless, in organisms such as marine bacteria, plants and human cells, Cpf1 has shown some advantageous features over Cas9, such as having less toxicity and reduced off-target effects (Jiang et al., 2017; Kim et al., 2017a; Kleinstiver et al., 2016; Ungerer and Pakrasi, 2016) while other reports have shown Cpf1 to be is as effective as Cas9 technologies (Yin et al., 2017).

We have established the CRISPR-Cpf1 system, as an alternative tool to edit native loci independent of their sequence complexity in blood stages of *P. falciparum*. We have demonstrated that Cpf1 DNA editing is as effective as Cas9 in this parasite when editing the parasite receptor *PfEBA-175*. Similar to Cas9, Cpf1 acts within the first cell division cycle of the parasite and yields virtually complete editing of a cell population based on surveying isolated clones. We observed editing activity using both LbCpf1 and FnCpf1. LbCpf1 has shown robust genome editing activity in mammalian cells (Agudelo et al., 2017; Kim et al., 2017b; Kim et al., 2016; Kleinstiver et al., 2016; Zetsche et al., 2015), whereas FnCpf1 has been more broadly used as an editing tool in plants and other organisms (Endo et al., 2016; Jiang et al., 2017; Ungerer and Pakrasi, 2016; Zaidi et al., 2017). Additionally, we showed that Cpf1-DRs are interchangeable between LbCpf1 and FnCpf1 and editing efficiency is not affected, consistent with previous studies (Zetsche et al., 2015). Using an approach for co-transfecting Cpf1 with homology-directed repair

donor vectors, we have shown the ability of this system to target native loci, at least one of which occurred in a low complexity region and was recalcitrant to editing using Cas9. Operationally, donor-vector assembly for Cpf1 is simpler than Cas9. Since the crRNA is shorter, we are able to generate T7 sgRNA expression cassettes using a single synthetic DNA fragment, which streamlines donor-vector construction.

Based on our *in silico* modeling we showed that the ability of Cpf1 to recognize AT-rich PAM motifs increases the number of potential binding sites across the genome at a similar specificity with respect to Cas9, expanding the coverage of genes that can be targeted. This newly implemented system will facilitate genome editing and pave the way to study gene function in a large-scale platform.

CRISPR-Cpf1 represents a powerful genome-editing tool to perform functional genetic screenings in *P. falciparum* and it offers an alternative platform in addition to existing Cas9 based-technologies to conduct reverse genetics on this parasite. Our group has recently shown how we can we efficiently integrate the editing activity of CRISPR-Cas9 systems using our TetR-aptamer translation control system to target and conditionally regulate the expression of any native gene (unpublished work). In this work, we showed how using Cpf1 editing activity, we were able to introduce DNA modifications at the cut site of the gene target via homologous recombination and inducibly regulate its expression. Such DNA modifications can be built during donor vector assembly and they were included within the homologous regions used to repair DNA cleavage. Elements such as peptide tags, TetR-aptamer arrays and TetR-DOZI cassette are necessary to profile protein expression and study the function of the selected gene. The implementation of this new Cpf1-integrated strategy to target native loci independent of their sequence complexity constitutes an important achievement to study a broader spectra of loci that will help to study essential aspects of the parasite biology, validate compounds with antimalarial activity and establish a new platform for the discovery of new targets for the development of new drugs and vaccines.

Genome-wide analysis showed how we could optimize Cpf1-sgRNAs design on the *P. falciparum* genome using criteria adopted from Cas9-based algorithms that might help to obtain a positive editing outcome with a higher on-target and lower off-target effects. Our Cpf1 library with selected sgRNAs showed a larger number of binding sites across the genome and that those sgRNAs were preferentially found at the anti-sense strand of the coding genes and skewed towards the 3' end of the targets. Interestingly, the majority of our successful editing events included in experimental data, were performed using Cpf1-sgRNAs designed on the anti-strand showing better on-target scores than other sgRNAs. However, other Cpf1-sgRNAs experimentally tested, were found on different regions of the target gene, out of the pattern showed by typical Cpf1-binding sites and some of them outside of the stringent parameters implemented, equally effective. This fact showed the feasibility of using this the Cpf1 editing system in genomes such as the one found in *P. falciparum*, where the frequency of available binding sites is so large that even after applying filters to select high quality sgRNAs the number of reminding cut sites is superior compared to Cas9-binding sites.

Our method offers a source to design improved sgRNAs that increase the probability of having a positive editing event. Our library of selected sgRNAs with good targeting scores could be highly valuable to perform large-scale screens gene editing experiments on this parasite at a higher success rate. Multiplex editing experiments could be one of the applications when using a *P.falciparum*-specific Cpf1 library of sgRNAs.

The demonstration of the Cpf1-efficacy enhances the utility of CRISPR-Cas class II systems in *P. falciparum* to perform functional genetics on blood stages on this parasite. The implementation of Cpf1-editing technologies complements our previous findings using Cas9 for genome engineering and presents an alternative system to target and regulate the expression of native loci that is independent of their sequence composition, protein-cellular localization and biochemical properties. We envision the potential of our complete genetic-tool kit as a highly robust and scalable platform to achieve high-

throughput genetic screenings associated with a specific phenotype in the parasite towards the identification of essential genes for parasite development.

METHODS

Genome-wide survey for Cas9 and Cpf1 sgRNA analysis

Genomic DNA sequences for the *Plasmodium falciparum* 3D7 genome (v.3.0, building on the work of Gardener et al., (Gardner et al., 2002) were obtained from the PlasmoDB project repository (Aurrecoechea et al., 2009). All gene sequences in the genome, including genomic sequences from mitochondrial and apicoplast chromosomes, were included in the analysis. For every gene sequence, regions between 500 bp upstream and 500 bp downstream of both the start and stop of the coding sequence —the target region for CRISPR-mediated editing— were searched for single guide RNAs (sgRNAs). Each of the noncoding regions, were taken target region independently, and the 5' and 3' ends of the sequence were taken as extremes. Searching was performed by identifying the frequency of Protospacer Adjacent Motif (PAM) sequences on the form of NGG for Cas9 sgRNAs (Doench et al., 2016; Doench et al., 2014) and TTTV for Cpf1 sgRNAs (Kim et al., 2017b; Zetsche et al., 2015). Each sgRNA was characterized using a variety of metrics, including position with respect to the start or stop of the target region, strand orientation with respect to coding strand of the gene, GC content and different on-target and off-target scores. On-target scoring for Cas9 sgRNAs was calculated using the Azimuth algorithm, based on experimental *Streptococcus pyogenes* Cas9 (SpCas9) activity data from the work of Doench et al. (Doench et al., 2016) in murine and human cells. The CINDEL algorithm, developed by Kim et al. (Kim et al., 2017b) through logistic regression of similar *Acidaminococcus* sp. BV3L6 Cpf1 activity data in human cells, was used for Cpf1 sgRNAs. To calculate off-target scores, a genome-wide sgRNA database for *P. falciparum* was generated, including sequences for alternative PAMs sequences shown to display measurable off-target activity such as NAG for Cas9 (Hsu et al., 2013) and CTTV for Cpf1 (Kim et al., 2017b). Two different algorithms were used to calculate Cas9 off-target scores of *P. falciparum* sgRNAs. The first, developed by Hsu et al. (Hsu et al., 2013) and it is based on human cell targeting specificity data using SpCas9. The second, the Cutting Frequency Determination (CFD) score developed by Doench et al. (Doench et al., 2016), and it is based on the same data as the Azimuth on-target metric.

To the best of our knowledge, no metric for sgRNA off-target activity has been developed based on Cpf1 experimental data. Hence, Cas9 sgRNA off-target algorithms were used to calculate Cpf1-specific metrics. Both the Hsu and CFD scoring algorithms assign weights to the number and position of mismatches, and the CFD algorithm includes mismatch identity into the calculation as well. Even though these weights are fit to data from mammalian cells edited with SpCas9, the number of similar sequences in each genome's off-target sgRNA database heavily influences the score calculation. Thus, an organism- and enzyme-agnostic scoring metric, based solely on the number of mismatches, was also included in the calculation using the Hsu algorithm with an equal-valued position mismatch matrix to calculate the off-target score on Cpf1 sgRNAs. Sequences, location, and metrics for all sgRNAs were stored to generate a *P. falciparum* sgRNA database. A second database of selected sgRNAs was created by choosing sgRNAs with 30–70% GC content, top 25% off-target Hsu scores and on-target scores in the top 50%. Given that off-target and on-target scoring metrics are largely derived from potentially useful sgRNAs with 30–70% GC content, the scoring percentile values were calculated from the results of the GC filter and not the total database. Hsu scores have been empirically validated in designing successful sgRNA guides in *P. falciparum* (unpublished work). For Cpf1 sgRNAs, the Hsu score was calculated using unweight position matrices. Additionally, sgRNAs were aggregated by gene, and the number of sgRNAs for each was calculated to generate a gene sgRNA database. All sequence manipulation and analyses were implemented in Python (v.2.7.8), while graphical and statistical analyses were done using R (v.3.4.1). The source code for the entire pipeline will be made available at https://github.com/pablocarderam/Pfal_CRISPR.

Plasmid construction

Cpf1 expression plasmids

To generate the different Cpf1 orthologs expression vectors, we used our pfYC3-base vector (Wagner et al., 2013) with an *hDHFR* (Human dihydrofolate reductase) to clone

the genes encoding the FnCpf1 and LbCpf1 proteins. Both genes were subcloned from plasmids provided by Feng Zhang's group, into the *XhoI* and *FseI* unique restriction sites of the base-plasmid, using Gibson assembly, under the 5'UTR PfCAM promoter (*P.falciparum calmodulin*). To clone the Cpf1 genes, we used synthetic adaptors with homologous regions to the 5' and 3' ends of each Cpf1 coding sequence and a unique *ApaI* restriction site to avoid PCR reactions on a long gene and to prevent possible introduction of mutations during the amplification. To build the synthetic adaptors, we used complementary oligos, through a Klenow Fragment (3'→5' exo-) NEB reaction reaction. Annealed oligos and DNA synthesis were confirmed on an agarose gel electrophoresis and purified using the Qiaex II gel extraction kit. The purified synthetic DNA fragment was cloned via Gibson into the *XhoI* and *FseI*, followed by a sequencing confirmation. Next, the pfYC3-plasmid was digested with *ApaI* included in the adaptor to receive purified Cpf1 gene fragments released from huFnCpf1 and huLbCpf1 plasmids provided by the Zhang group, and the fragments were assembled via Gibson (see list of oligos Table S2). T7RNAP was cloned previously, downstream from hDHFR marker using a T2A self-cleavage peptide for multi-cistronic protein expression, under the control of the 5'UTR PcDT. The Cpf1 and hDHFR-2A-T7RNAP expression cassettes were cloned using the previously described head-to-head arranged configuration on the plasmid to optimize transcriptional activity (Wagner et al., 2013).

Donor plasmids and crRNA synthesis

1. pUF-1-based donor plasmids construction: donor plasmids targeting the *PfEBA-175*, *PfFC* and *PfHRPII* native loci were built using the pUF-1 base plasmid (Ganesan et al., 2011) as a starting point. Homologous regions used as a template to repair Cpf1-induced cleavage were amplified from parasite gDNA. PCR amplification reactions were performed using a 15:1 (v:v) mixture of Hemo KlenTaq polymerase and Vent polymerase (New England BioLabs) in Hemo KlenTaq Buffer. Elongation was carried out at 60 °C for 2 min/kb of DNA to be amplified and PCR products were confirmed on a 1% agarose gel. Left homologous regions for all the targeted loci were purified and cloned into the

SacII and *SpeI* unique restriction sites of the plasmid using Gibson assembly, using oligos listed on Table S2. The right homologous regions were amplified and cloned into the *EcoRI* and *NcoI* unique restriction sites. The Cpf1-spacer (sgRNA) to target FC and HRPII loci were cloned in a single step using self-complementary oligos in a Klenow Fragment (3'→5' exo-) NEB and cloned into a unique *BsmBI* restriction site generated in a previously incorporated T7-Cpf1 expression cassette that included Cpf1-Direct Repeat sequences, T7 promoter and T7 terminator, as part of the pUF-1 base vector. To clone and synthesize EBA-175 sgRNA we used longer complementary oligos using the same Klenow synthesis reaction that included all the expression components including; T7 promoter, Cpf1-direct repeat sequence, EBA-175 sgRNA and T7 terminator. We cloned the synthetic fragment generated into the *NotI* unique restriction site of the pUF-1 vector. Oligos used to synthesize and clone gene specific-sgRNA are listed on Table S2.

2. pSN053-based donor plasmids construction: donor plasmid to target native *PfPPR* and *PfUNG* were built into the pSN053 base plasmid (unpublished work). This base plasmid contained an expression cassette driving the expression of a single transcript encoding for TetR/DOZI/Renilla Luciferase/Blastocidin selective marker for multi-cistronic protein expression using the 5'UTR and 3'UTR *PfHSP86*. Our group has previously shown the functionality of this expression cassette to regulate the expression of reporter and native *P. falciparum* genes using this plasmid on *BSD-resistant* and Renilla luciferase positive parasites. It was designed with modular features to easily assemble and replace gene-specific pieces to target any parasite gene. Left homologous regions were first amplified and gel-purified, followed by a Gibson that included the purified PCR product and a synthetic DNA fragment (IDG gblock) containing a recoded portion of the 3' end of the gene that prevented Cpf1 from recutting at the target site. Gibson reaction with the assembled product was then re-amplified and the purified prouct was cloned into the *FseI* and *BsiWI* restriction sites of the base plasmid. Right homologous regions correspond to the 3' UTR of each locus and were amplified,

purified and cloned into *I-CeuI* unique restriction site of the pSN053 base plasmid. Oligos used to amplify left and right homologous regions of each locus are listed on Table S2. T7-Cpf1 expression cassettes with all the components were synthesized in a single IDT gblock and cloned in a single step at the *I-CeuI* reconstituted from the right homologous regions and the *I-SceI* unique restriction sites of the plasmid, via Gibson. Designed sgRNAs sequences to target *PfPPR* and *PfUNG* loci included on oligos used to synthesize donor plasmids (Table S2).

Parasite culture and transfections

P. falciparum parental parasite strain NF54attB was grown in 5% O_2, 5%CO_2 and 37C conditions in RPMI-1640 medium supplemented with 2.5g/L AlbumaxII (Life Technologies), 2g/L sodium bicarbonate, 25mM HEPES, pH 7.4, 1mM hypoxanthine and 50mg/L gentamicin. Double plasmid co-transfections were performed in late stage parasites using the spontaneous uptake method (Deitsch et al., 2001) with 50ug (and 25ug for multiplex editing experiments) of maxi-prepped DNA of each plasmid and eight square-wave electroporation pulses of 365 V for 1 ms each, separated by 0.1 s. Single selection on the co-transfected parasites was applied on the fourth day after transfection using 1.5 uM DSM1 or 2.ug/mL of BSD (Blasticidine S), depending on the donor plasmid used in each experiment. DSM1 is a *Plasmodium* DHODH-specific inhibitor (Phillips et al., 2008) and BSD has also been shown to be a *P. falciparum* selective marker when using the blasticidin S deaminase of *Aspergillus terreus* (Mamoun et al., 1999). Parasites transfected with pSN053 donor plasmid, were maintained on 0.5 uM of aTc (anhydrotetracycline) at all times and Renilla Luciferase signal was monitored periodically during the transfection period until it reached ~10^6 Renilla Luciferase Units (RLUs), at which parasites were detectable by Giemsa smear. Renilla Luciferase signal was measured using the Promega Renilla Glo kit, following the manufacturer's instructions on a GloMax 20/20 luminometer (Turner Biosystems). Parasites that were transfected with pUF-1 plasmids, were monitored only by Giemsa smear within a typical

4- to 6-week transfection period, stable parasite lines were further analyzed for specific editing events.

Analysis of Cpf1-edited parasites

Parasite infected red blood cells from transient transfections (Day 2 and day 4 post-transfection) and from stable lines were lysed using 0.15% Saponin in 1X PBS. Parasite genomic DNA (gDNA) was isolated and purified using the QIAamp DNA blood mini kit (Qiagen). Identification of editing events in each DNA sample was performed using PCR and the Hemo KlenTaq polymerase (New England BioLabs) in Hemo KlenTaq Buffer Oligos. The expected integration events at the 5' and the 3' ends after recombination at the target site, were amplified using P1 (hybridizes upstream to the left homologous region of the targeted gene on the native locus) and P3 (hybridizes on the yDHODh or TetR/DOZI/Rluc/BSD expression cassettes) paired oligos, and P2 (hybridizes on the yDHODh or TetR/DOZI/Rluc/BSD expression cassettes) and P4 paired oligos (hybridizes downstream to the right homologous region of the targeted gene on the native locus). Oligos to amplify 5' and 3' integration events at each targeted site are listed on Table S3. For multiplex editing experiments, PCR reactions were performed on each DNA sample for each individual expected edited gene.

Samples with no-gRNA and NF54 negative controls were included in all 5' and 3'- end PCR reactions.

Renilla luciferase-based growth assays

To determine differences in growth rate between Cpf1-edited parasites at the *PfUNG* locus and establish gene essentiality, blood stage parasites were synchronized to rings using 5% Sorbitol. Parasites were washed off aTc at least 4-5 times using regular RPMI-supplemented media with no aTc. After washing, parasites were diluted and adjusted to ~10^3 RLUs and seeded in a 96-well microtiter plates in triplicate at 0.5% hematocrit in 200 µl of RPMI Complete media in the presence or the absence of aTc (0.5 µM). Four

different plates were set for the different time points of collection, including Time 0, IDC 1, 2 and 3 so the parasites would not get disturbed during the course of the experiment. All samples were processed simultaneously at the end of the experiment and the Renilla Luciferase signal was measured on a Promega Glomax plate reader.

Western Blot

To establish inducible regulation in the expression of the *PfUNG* edited parasites, approximately 10^6 late stage parasites were lysed using 0.15% saponin in 1X PBS after being grown for at least one IDC in the presence or absence of aTc. NF54attB WT parasites were included as negative controls. Lysates were prepared for western blot by heating in Laemmli sample buffer at 95 °C for 10 min. After separation by SDS-PAGE, proteins were transferred to a PVDF membrane and probed with a rabbit monoclonal antibody against Myc-tag (used to tag the UNG protein) Cell signal (71D10) at a 1:2000 dilution, or rabbit polyclonal antibody against GAPDH (Abcam, ab9485) at a 1:1000 dilution. Blots were imaged using a horseradish peroxidase-coupled secondary antibody (1:5000 dilution) and SuperSignal West Femto substrate (Thermo Scientific, Catalog # 34095).

REFERENCES

Agudelo, D., Duringer, A., Bozoyan, L., Huard, C.C., Carter, S., Loehr, J., Synodinou, D., Drouin, M., Salsman, J., Dellaire, G., *et al.* (2017). Marker-free coselection for CRISPR-driven genome editing in human cells. Nat Methods *14*, 615-620.

Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L.V., Katneni, U., Simonyan, V., and Kimchi-Sarfaty, C. (2017). A new and updated resource for codon usage tables. BMC Bioinformatics *18*, 391.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., *et al.* (2009). PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res *37*, D539-543.

Bansal, A., Ojo, K.K., Mu, J., Maly, D.J., Van Voorhis, W.C., and Miller, L.H. (2016). Reduced Activity of Mutant Calcium-Dependent Protein Kinase 1 Is Compensated in Plasmodium falciparum through the Action of Protein Kinase G. MBio *7*.

Brick, K., Watanabe, J., and Pizzi, E. (2008). Core promoters are predicted by their distinct physicochemical properties in the genome of Plasmodium falciparum. Genome Biol *9*, R178.

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*.

Cobb, D.W., Florentin, A., Fierro, M.A., Krakowiak, M., Moore, J.M., and Muralidharan, V. (2017). The Exported Chaperone PfHsp70x Is Dispensable for the Plasmodium falciparum Intraerythrocytic Life Cycle. mSphere *2*.

Crabb, B.S., Triglia, T., Waterkeyn, J.G., and Cowman, A.F. (1997). Stable transgene expression in Plasmodium falciparum. Mol Biochem Parasitol *90*, 131-144.

Crawford, E.D., Quan, J., Horst, J.A., Ebert, D., Wu, W., and DeRisi, J.L. (2017). Plasmid-free CRISPR/Cas9 genome editing in Plasmodium falciparum confirms mutations conferring resistance to the dihydroisoquinolone clinical candidate SJ733. PLoS One *12*, e0178163.

de Koning-Ward, T.F., Gilson, P.R., and Crabb, B.S. (2015). Advances in molecular genetic systems in malaria. Nat Rev Microbiol *13*, 373-387.

Deitsch, K., Driskill, C., and Wellems, T. (2001). Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. Nucleic Acids Res *29*, 850-853.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., *et al.* (2016). Optimized sgRNA design to

maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol *34*, 184-191.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol *32*, 1262-1267.

Endo, A., Masafumi, M., Kaya, H., and Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. Sci Rep *6*, 38169.

Ganesan, S.M., Falla, A., Goldfless, S.J., Nasamu, A.S., and Niles, J.C. (2016). Synthetic RNA-protein modules integrated with native translation mechanisms to control gene expression in malaria parasites. Nat Commun *7*, 10727.

Ganesan, S.M., Morrisey, J.M., Ke, H., Painter, H.J., Laroiya, K., Phillips, M.A., Rathod, P.K., Mather, M.W., and Vaidya, A.B. (2011). Yeast dihydroorotate dehydrogenase as a new selectable marker for Plasmodium falciparum transfection. Mol Biochem Parasitol *177*, 29-34.

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.

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.

Heinberg, A., Siu, E., Stern, C., Lawrence, E.A., Ferdig, M.T., Deitsch, K.W., and Kirkman, L.A. (2013). Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in Plasmodium falciparum. Mol Microbiol *88*, 702-712.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., *et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., Sun, B., Chen, B., Xu, X., Li, Y., *et al.* (2017). CRISPR-Cpf1 assisted genome editing of Corynebacterium glutamicum. Nat Commun *8*, 15179.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Kim, H., Kim, S.T., Ryu, J., Kang, B.C., Kim, J.S., and Kim, S.G. (2017a). CRISPR/Cpf1-mediated DNA-free plant genome editing. Nat Commun *8*, 14406.

Kim, H.K., Song, M., Lee, J., Menon, A.V., Jung, S., Kang, Y.M., Choi, J.W., Woo, E., Koh, H.C., Nam, J.W., *et al.* (2017b). In vivo high-throughput profiling of CRISPR-Cpf1 activity. Nat Methods *14*, 153-159.

Kim, Y., Cheong, S.A., Lee, J.G., Lee, S.W., Lee, M.S., Baek, I.J., and Sung, Y.H. (2016). Generation of knockout mice by Cpf1-mediated gene targeting. Nat Biotechnol *34*, 808-810.

Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J., and Joung, J.K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat Biotechnol *34*, 869-874.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol *37*, 67-78.

Krokan, H.E., and Bjoras, M. (2013). Base excision repair. Cold Spring Harb Perspect Biol *5*, a012583.

Lee, A.H., Symington, L.S., and Fidock, D.A. (2014). DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum. Microbiol Mol Biol Rev *78*, 469-486.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Mamoun, C.B., Gluzman, I.Y., Goyard, S., Beverley, S.M., and Goldberg, D.E. (1999). A set of independent selectable markers for transfection of the human malaria parasite Plasmodium falciparum. Proc Natl Acad Sci U S A *96*, 8716-8720.

Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. Biochimie *113*, 93-99.

Mayer, D.C., Kaneko, O., Hudson-Taylor, D.E., Reid, M.E., and Miller, L.H. (2001). Characterization of a Plasmodium falciparum erythrocyte-binding protein paralogous to EBA-175. Proc Natl Acad Sci U S A *98*, 5222-5227.

Ng, C.L., Siciliano, G., Lee, M.C., de Almeida, M.J., Corey, V.C., Bopp, S.E., Bertuccini, L., Wittlin, S., Kasdin, R.G., Le Bihan, A., *et al.* (2016). CRISPR-Cas9-modified pfmdr1 protects Plasmodium falciparum asexual blood stages and gametocytes against a class of piperazine-containing compounds but potentiates artemisinin-based combination therapy partner drugs. Mol Microbiol *101*, 381-393.

Nkrumah, L.J., Muhle, R.A., Moura, P.A., Ghosh, P., Hatfull, G.F., Jacobs, W.R., Jr., and Fidock, D.A. (2006). Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase. Nat Methods *3*, 615-621.

Nunez, J.K., Harrington, L.B., and Doudna, J.A. (2016). Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. ACS Chem Biol *11*, 681-688.

O'Donnell, R.A., Preiser, P.R., Williamson, D.H., Moore, P.W., Cowman, A.F., and Crabb, B.S. (2001). An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected Plasmodium falciparum parasites. Nucleic Acids Res *29*, 716-724.

Phillips, M.A., Gujjar, R., Malmquist, N.A., White, J., El Mazouni, F., Baldwin, J., and Rathod, P.K. (2008). Triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors with potent and selective activity against the malaria parasite Plasmodium falciparum. J Med Chem *51*, 3649-3653.

Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: Base excision repair: the long and short of it. Cell Mol Life Sci *66*, 981-993.

Ruvalcaba-Salazar, O.K., del Carmen Ramirez-Estudillo, M., Montiel-Condado, D., Recillas-Targa, F., Vargas, M., and Hernandez-Rivas, R. (2005). Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions. Mol Biochem Parasitol *140*, 183-196.

Saul, A., and Battistutta, D. (1988). Codon usage in Plasmodium falciparum. Mol Biochem Parasitol *27*, 35-42.

Sidik, S.M., Huet, D., Ganesan, S.M., Huynh, M.H., Wang, T., Nasamu, A.S., Thiru, P., Saeij, J.P., Carruthers, V.B., Niles, J.C., *et al.* (2016). A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. Cell *166*, 1423-1435 e1412.

Small, I.D., Rackham, O., and Filipovska, A. (2013). Organelle transcriptomes: products of a deconstructed genome. Curr Opin Microbiol *16*, 652-658.

Sonoiki, E., Ng, C.L., Lee, M.C., Guo, D., Zhang, Y.K., Zhou, Y., Alley, M.R., Ahyong, V., Sanz, L.M., Lafuente-Monasterio, M.J., *et al.* (2017). A potent antimalarial benzoxaborole targets a Plasmodium falciparum cleavage and polyadenylation specificity factor homologue. Nat Commun *8*, 14574.

Straimer, J., Lee, M.C., Lee, A.H., Zeitler, B., Williams, A.E., Pearl, J.R., Zhang, L., Rebar, E.J., Gregory, P.D., Llinas, M., *et al.* (2012). Site-specific genome editing in Plasmodium falciparum using engineered zinc-finger nucleases. Nat Methods *9*, 993-998.

Suksangpleng, T., Leartsakulpanich, U., Moonsom, S., Siribal, S., Boonyuen, U., Wright, G.E., and Chavalitshewinkoon-Petmitr, P. (2014). Molecular characterization of Plasmodium falciparum uracil-DNA glycosylase and its potential as a new anti-malarial drug target. Malar J *13*, 149.

Ungerer, J., and Pakrasi, H.B. (2016). Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. Sci Rep *6*, 39681.

Vanaerschot, M., Lucantoni, L., Li, T., Combrinck, J.M., Ruecker, A., Kumar, T.R.S., Rubiano, K., Ferreira, P.E., Siciliano, G., Gulati, S., *et al.* (2017). Hexahydroquinolines are antimalarial candidates with potent blood-stage and transmission-blocking activity. Nat Microbiol.

Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S., and Roubos, J.A. (2017). CRISPR/Cpf1 enables fast and simple genome editing of Saccharomyces cerevisiae. Yeast.

Wagner, J.C., Goldfless, S.J., Ganesan, S.M., Lee, M.C., Fidock, D.A., and Niles, J.C. (2013). An integrated strategy for efficient vector construction and multi-gene expression in Plasmodium falciparum. Malar J *12*, 373.

Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014). Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. Nat Methods *11*, 915-918.

Wang, M., Mao, Y., Lu, Y., Tao, X., and Zhu, J.K. (2017). Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System. Mol Plant *10*, 1011-1013.

Yin, X., Biswal, A.K., Dionora, J., Perdigon, K.M., Balahadia, C.P., Mazumdar, S., Chater, C., Lin, H.C., Coe, R.A., Kretzschmar, T., *et al.* (2017). CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene EPFL9 in rice. Plant Cell Rep *36*, 745-757.

Zaidi, S.S., Mahfouz, M.M., and Mansoor, S. (2017). CRISPR-Cpf1: A New Tool for Plant Genome Editing. Trends Plant Sci *22*, 550-553.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., *et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., *et al.* (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotechnol *35*, 31-34.

SUPPLEMENTARY INFORMATION

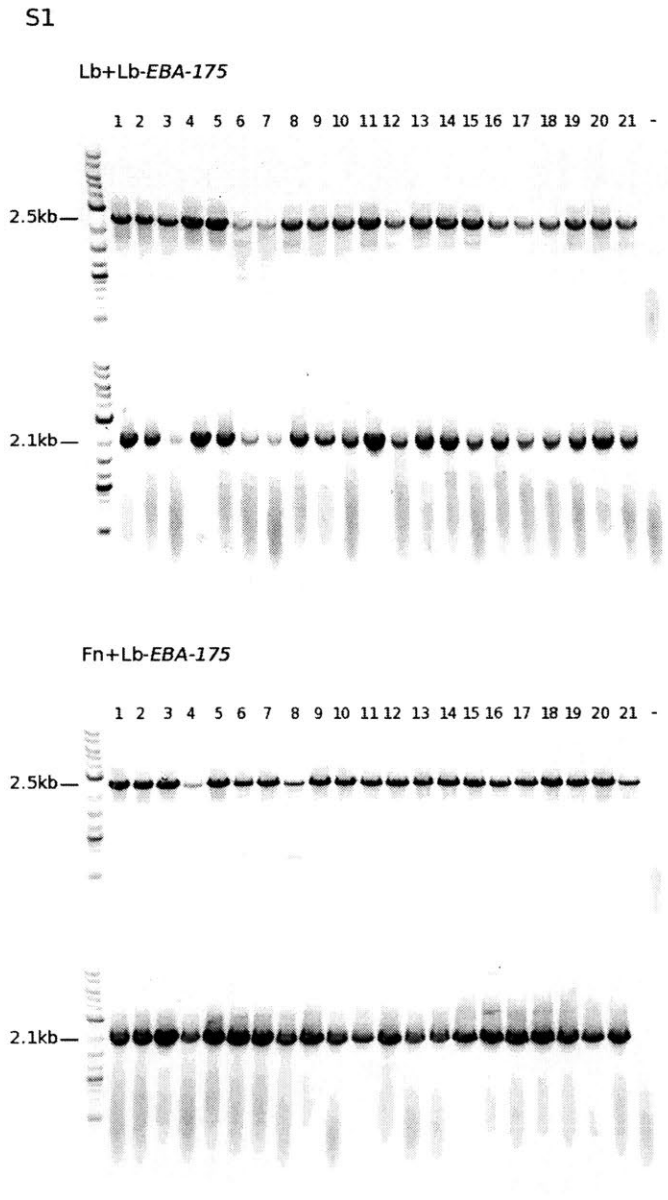


Figure 4-S1: Efficiency of Cpf1-editing activity on the PfEBA-175 locus
EBA-175 edited parasites using either FnCpf1 or LbCpf1 were cloned using limited dilution. All the clones (21/21) using both endonucleases show the expected integration event at the 5' (Expected amplification product of 2.5 kb) and 3' (Expected amplification product of 2.1 kb) of the disrupted locus, which confirms editing on 100% of the population.

S2

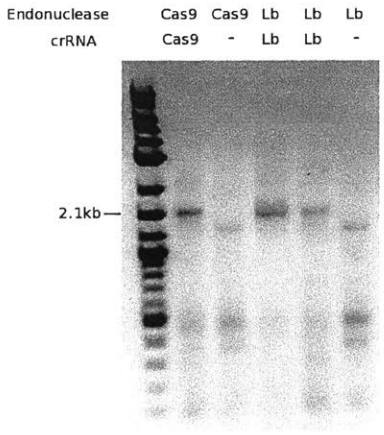


Figure 4-S2: Comparable results between Cpf1 and Cas9-editing activities to target *PfEBA-175* during transient transfections

Late stage NF54attB parental strain parasites were co-transfected with the LbCpf1 expression plasmid and Lb-EBA-175 donor plasmid or a no-crRNA plasmid (negative control). Similarly, a Cas9-EBA-175 donor plasmid was transfected on a Cas9-integrated parasite line, generated in our group or a no-gRNA plasmid (used as a control). Genomic DNA was isolated at Day 2 post-transfection and the agarose gel shows amplification products of the expected editing event at the 3' end of the disrupted *PfEBA-175* with the *yDHODH* expression cassette integrated.

S3

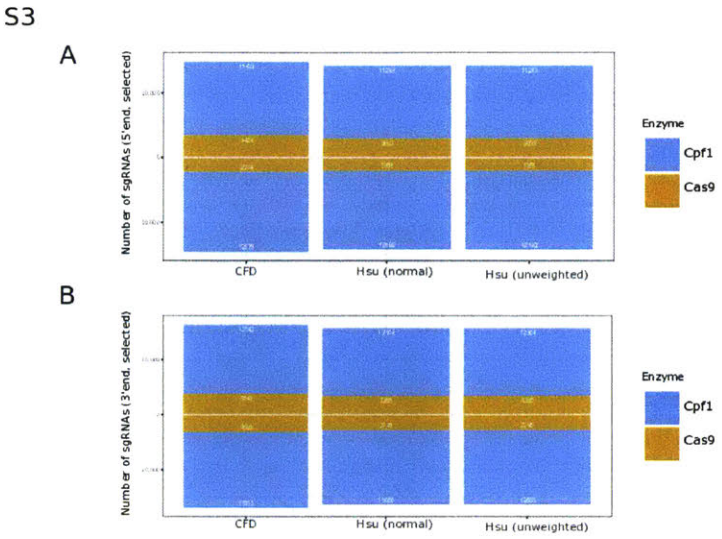


Figure 4-S3: Comparison of different off-target scoring algorithms on filter stringency

Bars above the horizontal axis indicate sgRNAs on the sense strand of the gene they are found on, while bars below the horizontal indicate sgRNAs on the antisense strand. Hsu scoring system resulted in a slightly more stringent filter than CFD scores for sgRNAs around both the 5' **(A)** and 3' **(B)** ends, particularly for Cas9 sgRNAs. However, Cpf1 sgRNAs outnumbered Cas9 sgRNAs by a similar amount, independently of the algorithm used. Normal (weighted) and unweighted position matrices did not affect filtering when using Hsu scores.

S4

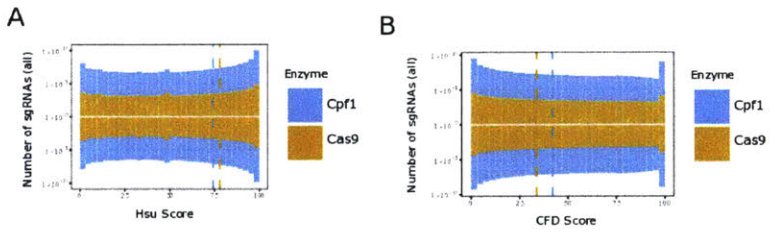


Figure 4-S4: Distribution of off-target scores in total sgRNA database on total sgRNA binding sites analysis on *P.falciparum*

Bars above the horizontal axis indicate sgRNAs on the sense strand of the gene they are found on, while bars below the horizontal indicate sgRNAs on the antisense strand. Dotted lines indicate averages of the given data series. Off-target sgRNA scores concentrate towards the two extremes of the distribution for both the Hsu **(A)** and CFD **(B)** algorithms.

Enzyme	GC Content	Score Algorithm	0%	25%	50%	75%
Cas9	All	Azimuth	-9.598	31.03	41.62	52.12
Cpf1	All	CINDEL	1.224	15.71	25.36	39.38
Cas9	All	Hsu	0.0152	66.58	93.76	98.95
Cpf1	All	Hsu agnostic	0.01548	59.16	96.15	99.8
Cas9	30> GC>70	Azimuth	-5.302	39.84	49.5	57.48
Cpf1	30> GC>70	CINDEL	1.224	12.38	20.44	31.73
Cas9	30> GC>70	Hsu	0.0866	97.95	99.47	99.94
Cpf1	30> GC>70	Hsu agnostic	0.7697	99.39	100	100

Table 4-S1: Reference of the percentiles of On-Target and Off-Target Scores to select Cpf1 and Cas9 sgRNAs from the general database.

Table S2: List of oligos used for plasmid construction

Oligo Name	Sequence (5'-3')	Description
AF276	cctaatagaaatatatcaggatccctcgagatgagcaagctggagaagtttacaaactgctactcgggcccattatg	Fwd oligo of synthetic adaptor fragment to clone LbCpf1 into pFGNr-T7 plasmid
AF277	gggagcatcagtcatggaattcggccggccggcatagtcggggacatcatatgggtatgcataatgggcccgagtagcag	Rv oligo of synthetic adaptor fragment to clone LbCpf1 into pFGNr-T7 plasmid
AF350	cctaatagaaatatatcaggatccctcgagatgagcatctaccaggagttcgtcaacaagtattcgggcccattatg	Fwd oligo of synthetic adaptor fragment to clone FnCpf1 into pFGNr-T7 plasmid
AF351	gggagcatcagtcatggaattcggccggccttaggcatagtcggggacatcatatgggtatgcataatgggcccgaatacttg	Rv oligo of synthetic adaptor fragment to clone FnCpf1 into pFGNr-T7 plasmid
eba-175_LH_Forw	tatagaatactcaagcttgggggatcctctagagtcgacgttatggaactccagataat	Fwd Gibson oligo to amplify and clone EBA-175 left homologous in pUF-1 plasmid [25108687]
eba-175_LH_Rev	cctctaccttcaccactacccatagcaagatgtccataat	Rv Gibson oligo to amplify and clone EBA-175 left homologous in pUF-1 plasmid [25108687]
eba-175_RH_Forw	ttctcaaaaatgaacaataattattcaactaaggcagaaa	Fwd Gibson oligo to amplify and clone EBA-175 right homologous in pUF-1 plasmid [25108687]
eba-175_RH_Rev	agtgtagttaattcatcaaatagcatgcctgcaggtcgacaatattcagcatcacaatta	Rv Gibson oligo to amplify and clone EBA-175 right homologous in pUF-1 plasmid [25108687]
AF299	ggtgacactatagaatactcgcggccgctaatacgactcactatagggtaatttctactaagtgtagat**ccatagcaagatgtccataa**	Fwd klenow/gibson oligo to synthesize and clone EBA-175 T7 prom, DR seq, sgRNA and T7 term
AF300	actttttcttcttcagggtaggcggccgccaaaaaaaccccctcaagacccgtttagaggccccaaggggttatgctattatggacatcttgctatg	Rv klenow/gibson oligo to synthesize and clone EBA-175 T7 prom, DR seq, sgRNA and T7 term
AF236	ttattggttttcaaacttcattgactgtgccggccggccctataattcataataatatgg	Fwd Gibson oligo to amplify and clone PPR chimeric left homologous region into pSN053 plasmid
AF216	ctctttgaagagtttaaacgctttctcataattatctatgtgactataaatttttaac	Rv oligo to amplify native PPR and assemble PCR product with synthetic PPR DNA fragment
AF230	actcacagatcttcttcgctgatgagtttctgttccgtacggtagttctccgtctcagaa	Rv Gibson oligo to amplify and clone PPR chimeric left homologous region into pSN053 plasmid
AF431	ttcaaacttcattgactgtgccggccggccatgaataatccaacaattcag	Fwd Gibson oligo to amplify and clone UNG chimeric left homologous region into pSN053 plasmid
AF432	gataggcacggagaagcaaaggcccatggcttgatcttttgatgataagg	Rv oligo to amplify native UNG and assemble PCR product with synthetic UNG DNA fragment
AF437	acagatcttcttcgctgatgagtttctgttccgtacgttgcggcagttcccacttgattg	Rv Gibson oligo to amplify and clone UNG chimeric left homologous region into pSN053 plasmid
AF217	attatacgaagttatcgtaactataacttattcactgtataaaggaatc	Fwd Gibson oligo to amplify and clone PPR right homologous region into pSN053 plasmid
AF218	gtgagtcgtattattcgctaccttaggaccttacataaggagacagaagc	Rv Gibson oligo to amplify and clone PPR right homologous region into pSN053 plasmid
AF333	ttatgtaaggtcctaaggtagcgaataatacgactcactatagggtaatttctactaagtgtagat**aaagaattaaatatgaaaca**tagc	Fwd klenow/gibson oligo to synthesize and clone PPR sgRNA
AF334	cctaacctagggataacagggtaatcaaaaaaccctcaagacccgtttagaggccccaaggggttatgctatgtttcatatttaatt	Rv klenow/gibson oligo to synthesize and clone PPR sgRNA
AF444	agcatacattatacgaagttatcgtaactataacatatatatatatatatgtcacgtgattc	Fwd Gibson oligo to amplify and clone UNG right homologous region into pSN053 plasmid
AF445	tcgtattattcgctaccttaggaccgtataaatatatatatataacaccttg	Rv Gibson oligo to amplify and clone UNG right homologous region into pSN053 plasmid
AF448	gcgaataatacgactcactatagggtaatttctactaagtgtagattgttcctataggag	Fwd klenow/gibson oligo to synthesize and clone UNG sgRNA
AF449	aagacccgtttagaggccccaaggggttatgctat**ttttactcctataggaaca**atctac	Rv klenow/gibson oligo to synthesize and clone UNG sgRNA

Table 4-S2: List of oligos used for plasmid construction

*Bold font: corresponds to the gene-specific sgRNA sequence

Table S3: List of oligos used to confirm editing

Oligo Name	Sequence (5'-3')	Description
AF324	acagccagtttaactaccaagttcttg	P3 Fwd oligo specific to the yDHODh selective marker to detect integration at the 3' end when using pUF-1 based donor plasmids
AF323	ttaaatgctgttcaacttcccac	P2 Rv oligo specific to the yDHODh selective marker to detect integration at the 5' end when using pUF-1 based donor plasmids
AF346	gaaatatggaaatgttcaaaaaactgataag	P1 oligo to detected PfEBA175 edited event at the 5' end
AF347	aaatcatctaaatcaatatgttcacgcttttcc	P4 oligo to detected PfEBA175 edited event at the 3' end
AF678	cctcgagttgtaaaattgaacagg	P1 oligo to detected PfFC edited event at the 5' end
AF679	atttttaatcatttttttttcgcttatggatttacacc	P4 oligo to detected PfFC edited event at the 3' end
AF680	tttccttctcaaaaaataaagtattatccg	P1 oligo to detected PfHRPII edited event at the 5' end
AF729	acatgttttagctcaaattaatgtgtacagc	P4 oligo to detected PfHRPII edited event at the 3' end
AF285	tcataaggatacgtacgctcacagatc	P2 Rv oligo specific to theTetR/DOZI backbone to detect integration at the 5' end when using pSN053 based donor plasmids
AF266	ggaatactaaatatatatccaatggcccc	P3 Fwd oligo specific to theTetR/DOZI backbone to detect integration at the 3' end when using pSN053 based donor plasmids
AF342	gcaaaaacgggaaatggatataaggc	P1 oligo to detected PfPPR edited event at the 5' end
AF343	gaatgttatttacatttacatgataaaaaaaatatttaattcatcc	P4 oligo to detected PfPPR edited event at the 3' end
AF495	tttataagatattattttaaatatttatacacatcaatctgataaataagtatatgtccc	P1 oligo to detected PfUNG edited event at the 5' end
AF496	gtttgattacgtgtaaggttaattatccg	P4 oligo to detected PfUNG edited event at the 3' end

Table 4-S3: List of oligos used to detect integration events

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

The burden of malaria around the world takes a toll of nearly half million deaths, especially in children and pregnant women. Despite some progress in reducing the number of cases in areas of seasonal or sporadic malaria by the use of first-line antimalarials, the emergence of drug-resistant *P. falciparum* strains has substantially increased the malaria morbidity and mortality. A better understanding of the parasite's biology and the systematic identification of proteins that accomplish essential functions for infection and survival of the parasite represent a promising strategy for the inclusion targets for the development of the novel therapeutics. The work presented here introduces the development of new genetic and molecular tools to manipulate *P. falciparum*. These tools were developed as single modules that could be integrated within a single genetic toolkit to accomplish broader tasks in parallel and effectively overcome the potential limitations of one or another. The genetic toolkit described in this work provides a robust and versatile way to identify and characterize essential and non-essential genes in asexual stages of the parasite. As a demonstration of its utility, I applied this technology to study the biological function of a previously unknown essential gene and characterized the involvement of this gene product in the parasite's life cycle. The principal contributions and findings of this work, as well as the outlook for future research include the following:

Synthetic post-transcriptional regulation for inducible control of gene expression in blood stages of P. falciparum

In this work we developed a RNA-protein synthetic module to inducibly regulate gene expression in *P. falciparum*. Previous work in our lab focused on the engineering of a system for inducible translational control in eukaryotes, including *P. falciparum* (Goldfless et al., 2012; Goldfless et al., 2014). This system demonstrated to be versatile to control the expression of native genes in *P. falciparum* independent of its transcriptional control, protein sequence or localization of the targeted protein in the cell (Goldfless et al., 2014). To improve translational control and enhance the dynamic range to fine-tune regulation of gene expression in the parasite, on Chapter 2 we developed an integrated strategy between native translational machinery and an already established synthetic RNA-protein module. Different translational regulator protein effectors were

fused to TetR and they showed effective and enhanced inducible regulation in an aTc-dependent manner consistently in yeast and *P.falciparum*. We focused our work on the TetR-DOZI fusion to control gene expression in the parasite, which showed steady regulation along the parasite life cycle when a gene of interest is flanked by a dual-aptamer arrangement.

By installing regulatory TetR-DOZI/aptamer machinery at a reporter or at a native locus we achieved tight gene regulation with a substantial increase in the dynamic range and reduced leaky expression compared to what was observed previously without DOZI. Although the TetR-DOZI tool has been proven to tightly regulate gene expression in the parasite, the molecular mechanism of DOZI in blood stage parasite remains elusive. Additional biochemical and biological studies are required to understand the contribution of DOZI in *P. falciparum* post-transcriptional processes in asexual and sexual stages. In addition we showed that when TetR is fused to the translational regulator CITH, the translation regulation of target genes is also improved by employing a dual-aptamer context. This data suggests that our tool could serve as a platform to determine the roles of other translational regulator effectors in blood stage parasites. Some translational regulators have been identified in the *P. falciparum* genome and their potential function in parasite development has been described (Bunnik et al., 2016; Chene et al., 2012; Miao et al., 2013; Reddy et al., 2015; Shrestha et al., 2016; Vembar et al., 2016; Vembar et al., 2015), however their role in blood stage parasites has not been shown and we envision that our synthetic RNA-protein module could be utilized to establish their contribution to translational regulation in asexual stages of *P. falciparum*.

Another important feature of our work is the modularity and the orthogonality of the system. Our gene regulation tool demonstrated its functionality for translational regulation in yeast and in *P. falciparum*. For this reason, our RNA-protein module might hold potential to be implemented in other organisms that lack a system for orthogonal translational regulation. The synthetic module could be adapted to a particular system by

fusing TetR to any endogenous translational regulator effector proteins that could improve the dynamic range of the regulation and move functional genetics forward in other biological systems.

We demonstrated that our strategy is suitable to conditionally regulate essential genes, a task that is of paramount importance when characterizing candidate target genes for antimalarials. We showed the versatility and stability of our system along the erythrocytic cycle. Our tool could be used to regulate the expression of other genes, independent from their transcriptional activity, biochemical properties and localization in the cell and determine their essentiality in asexual stages of the parasite. In addition, this tool could be applied to regulate the expression of genes that are presumed to be the targets of compounds with antimalarial activity. In many instances, the resistance mechanisms to antimalarials are not defined due to the technological limitations to target the potential target genes.

Another application of this technology could be to achieve multiple gene regulation in the parasite, an inexistent feature in other inducible systems described in *P. falciparum*. TetR belongs to a family of translational repressors that have been identified in other bacterial species. These proteins respond to different environmental stimuli and the characterization of their tridimensional structure has helped to identify their specific ligands that modify their response (Ramos et al., 2005). By identifying the interacting RNA molecules for each of these TetR-homologs, we could build a regulatory network to control different loci with diverse synthetic modules in response to different small molecules. This strategy will enable scaling up the number of genes that could be studied in a high-throughput manner. A similar strategy has recently been developed in mammalian cells using repressor components to regulate gene expression at the transcriptional level (Stanton et al., 2014).

Study of an essential RNA-Binding protein using an integrated CRISPR-Cas9 technology

The development of an integrated technology that combines editing technologies with an inducible system enabled the identification of a novel essential RNA-binding protein for parasite survival during the erythrocytic cycle. The strategy we implemented to develop a final integrated tool to target an essential gene required the validation of individual modules we previously developed, which were used first to conditionally regulate the expression of a putative essential gene and then to disrupt the native locus to further investigate its function.

In chapter 3, we presented the applications of previous tools we have developed in our lab to study the biology of an essential gene with unknown function in blood stages of the parasite. PfRRM1 is a protein that belongs to a RRM1 family and it is characterized by the presence of a single RNA-recognition motif (RRM). In *P. falciparum*, 28 genes have been found within this family; however the function of the majority of the proteins those genes encode for is not known (Reddy et al., 2015). Previous attempts in our lab to target this protein using conventional genetic tools failed, suggesting PfRRM1 could be an essential gene. Later on, we presented evidence from different studies that indicated the possible essential role of this protein in apicomplexan parasites. Leveraging effective applications of a translational control module and genome-engineering technologies we determined the essentiality of PfRRM1 and its role in parasite cell cycle progression.

Using the principles learned from the inducible regulatory capabilities from our TetR/DOZI-aptamer system (Ganesan et al., 2016), we implemented the PfCRT module characterized on Chapter 2 to regulate the expression of PfRRM1. We confirmed tight regulation of the protein demonstrating the versatility of our tool using the dual aptamer context to regulate native genes in the parasite. In parallel in our lab, the efficiency of the CRISPR/Cas9 to edit genes in *P. falciparum* was demonstrated on non-essential genes (Wagner et al., 2014). In this work, for the first time we showed the editing activity of Cas9 to target an essential gene in the parasite at a similar efficiency. This data served to

create the foundation of the fundamental principles that we currently use in our lab to target multiple essential genes in *P. falciparum*. The development of this robust platform provides significant tools that can be applied towards the understanding of the biology of the parasite and to the discovery of new therapeutic targets.

The contribution of this work to the general knowledge of RBP biology in *P. falciparum* is highly relevant. We demonstrate that this is the first essential RRM1 protein that showed a role in parasite schizogony. Although the information derived from the biochemical and biological experiments conducted in this work provided fundamental evidence for the assignment of the function of this protein, additional proteomics and RNA sequencing experiments are required to determine the exact molecular mechanism of action and the pathways this protein could be potentially involved with. In addition, our tool could be used to target and regulate other genes that are presumed to be interacting partners of PFRRM1 and that could operate upstream or downstream of this RBP to determine whether PfRRM1 is a master regulator of cell cycle in *P. falciparum*. On a top layer on the evolution of our technology, the development of a multiplex CRISPR/Cas9 system on this parasite would enable a more effective targeting strategy to identify those interacting partners on the PfRRM1 pathway and determine their contribution to post-transcriptional processes in cell cycle regulation of asexual stages of *P. falciparum*.

Our current work with the CRISPR/Cas9 system is limited to the use of a constitutive expression of the effector protein in the parasite. Inducible systems to regulate the expression of Cas9 have been developed in other systems to modulate the expression of this protein spatially and temporary in the cell. This strategy enables this protein to respond to a particular stimuli and, for instance, reprogram cell fate as an outcome (Dai et al., 2017). The ability to induce machinery of editing technologies is lacking in *P. falciparum*. The benefit of having an inducible version of Cas9 in the parasite would serve to expand the control we currently have with our inducible system and will further

control the timing of editing on stage-specific genes. This higher level of temporal control might serve to determine the role of those genes in each phase of the parasite life cycle. For example, the inducible targeting of RBPs that regulate large cascades of post-transcriptional processes involved in parasite cell cycle progression would enable us to directly dissect their biology and to determine the different pathways they control in the cell. If true, RBPs that operate as master regulators could serve as ideal drug targets under the assumption that the disruption of those gene products would also interrupt transmission by inhibiting not only cell cycle progression but also other downstream processes.

A novel genome editing tool to target AT-rich regions

Some of the genes that we attempted to target and regulate using the integrated Cas9 technology developed in chapter 3 were recalcitrant to the Cas9 editing activity. Those genes had an AT-composition of about 80%, which represented a challenge for the design of optimal Cas9 sgRNA and ultimately editing on those genes was precluded. At the time, a novel CRISPR-class II system was characterized with an effector protein capable of recognizing AT-rich sequences using a much simpler crRNA assembly system (Zetsche et al., 2015). In this work, we implemented the editing activity of Cpf1 endonucleases to first assess their efficiency on a non-essential gene using a previously targeted gene with Cas9. In this experiments we showed an efficiency of 100% in the editing of the *PfEBA-175* locus and demonstrated that the Cpf1 activity is observable during the first cycle of the parasite after transfection. These results on the efficiency and dynamics of targeting are similar to what we have observed for Cas9 in *P. falciparum*.

In an important advance, we demonstrated the effective targeting of low complexity regions and the ability to perform inducible regulation at a native locus using the Cpf1 system. We successfully disrupted a target of 80% AT composition and introduced via homologous recombination regulatory elements at the Cpf1-cute site that could be used to regulate this target's expression. In addition, we succeeded in the disruption and

inducible regulation of a target that is of high interest as an antimalarial drug target. This feat was achieved by coupling the Cpf1 editing activity and the regulatory activity of our TetR-DOZI/aptamer system, adopting a similar integrated strategy used in chapter 3 with Cas9. This antimalarial target did not show an essential phenotype when the protein is knockdown at least during the first 3 IDCs that were monitored. Additional drug-sensitivity assays and biochemical experiments are needed to determine a phenotype associated to the knockdown of this protein in blood stage parasites. The incorporation of Cpf1 in our genetic toolkit expands the repertoire of targetable genes in *P. falciparum* to assess their function and dissect more aspects of the biology of this parasite.

Multiplex editing capabilities using Cpf1 have been recently described in mammalian cells (Zetsche et al., 2017), plants (Wang et al., 2017) and yeast (Verwaal et al., 2017). Cpf1 can process its own crRNA and also the Cpf1-crRNA cassette assembly is simpler than the Cas9 one. These features are relevant as they provide a more straightforward strategy to target multiple loci at the same time. Further advances that would enable the construction of donor vectors with more complex designs that include multiple homologous templates will make possible Cpf1-mediated multiplex editing in *P. falciparum*. We showed that almost every gene on the *P. falciparum* genome has a Cpf1-binding site, with this we can envision the establishment of a multiple editing platform to target this parasite's genome using Cpf1 endonucleases.

As a proof of concept, we utilized a co-transfection-based method to express Cpf1 and repair cleavage with a second donor plasmid. The generation of integrated lines with a constitutive expression of Cpf1 and T7RNAP might facilitate single and multiple editing experiments. Our preliminary data on Cpf1 integrated lines indicate they can tolerate constitutive expression of the protein, however additional modifications on the cell lines might be required to achieve consistent gene editing on stable lines. For instance, the development of an inducible system to regulate the expression of Cpf1 and to activate its editing activity at a specific time on stable lines is an alternative we have considered in

our framework. Similarly, our orthogonal T7 promoter expression cassette could be engineered to inducible transcribe the sgRNA.

This work showed the first evidence of the Cpf1 editing activity in *P. falciparum*. This new genetic tool brought an enormous benefit to the *Plasmodium* genetics field by covering the complexity of its AT-rich genome. Cpf1 editing technologies represent a complementary alternative to the existing Cas9 technologies so now a more complete and flexible genetic network can be implemented to expand the number of genes can be targeted on the *P. falciparum* or other Plasmodium species. This will favor a more efficient pathway to identify new essential targets and dissect the parasite biology towards the discovery of new therapies and malaria eradication.

REFERENCES

Bunnik, E.M., Batugedara, G., Saraf, A., Prudhomme, J., Florens, L., and Le Roch, K.G. (2016). The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. Genome Biol *17*, 147.

Chene, A., Vembar, S.S., Riviere, 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*, 3066-3077.

Dai, X., Chen, X., Fang, Q., Li, J., and Bai, Z. (2017). Inducible CRISPR genome-editing tool: classifications and future trends. Crit Rev Biotechnol, 1-14.

Ganesan, S.M., Falla, A., Goldfless, S.J., Nasamu, A.S., and Niles, J.C. (2016). Synthetic RNA-protein modules integrated with native translation mechanisms to control gene expression in malaria parasites. Nat Commun *7*, 10727.

Goldfless, S.J., Belmont, B.J., de Paz, A.M., Liu, J.F., and Niles, J.C. (2012). Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction. Nucleic Acids Res *40*, e64.

Goldfless, S.J., Wagner, J.C., and Niles, J.C. (2014). Versatile control of Plasmodium falciparum gene expression with an inducible protein-RNA interaction. Nat Commun *5*, 5329.

Miao, J., Fan, Q., Parker, D., Li, X., Li, J., and Cui, L. (2013). Puf mediates translation repression of transmission-blocking vaccine candidates in malaria parasites. PLoS Pathog *9*, e1003268.

Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., and Tobes, R. (2005). The TetR family of transcriptional repressors. Microbiol Mol Biol Rev *69*, 326-356.

Reddy, B.P., Shrestha, S., Hart, K.J., Liang, X., Kemirembe, K., Cui, L., and Lindner, S.E. (2015). A bioinformatic survey of RNA-binding proteins in Plasmodium. BMC Genomics *16*, 890.

Shrestha, S., Li, X., Ning, G., Miao, J., and Cui, L. (2016). The RNA-binding protein Puf1 functions in the maintenance of gametocytes in Plasmodium falciparum. J Cell Sci *129*, 3144-3152.

Stanton, B.C., Siciliano, V., Ghodasara, A., Wroblewska, L., Clancy, K., Trefzer, A.C., Chesnut, J.D., Weiss, R., and Voigt, C.A. (2014). Systematic transfer of prokaryotic sensors and circuits to mammalian cells. ACS Synth Biol *3*, 880-891.

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*, 772-792.

Vembar, S.S., Macpherson, C.R., Sismeiro, O., Coppee, 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*, 212.

Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S., and Roubos, J.A. (2017). CRISPR/Cpf1 enables fast and simple genome editing of Saccharomyces cerevisiae. Yeast.

Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014). Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. Nat Methods *11*, 915-918.

Wang, M., Mao, Y., Lu, Y., Tao, X., and Zhu, J.K. (2017). Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System. Mol Plant *10*, 1011-1013.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., *et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., *et al.* (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotechnol *35*, 31-34.