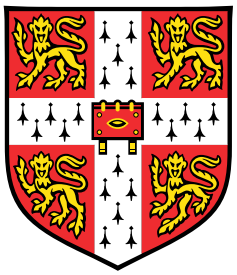
Enabling the *in vitro* study of long noncoding RNAs to understand their role in *Plasmodium falciparum*



Johanna Steffania Hoshizaki

Wellcome Sanger Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Churchill College

March 2023

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text. This dissertation does not exceed the prescribed limit of 60,000 words.

Johanna Steffania Hoshizaki
March 2023

Abstract

Enabling the *in vitro* study of long noncoding RNAs to understand their role in *Plasmodium falciparum* Johanna Steffania Hoshizaki

Long noncoding RNAs (lncRNAs) have been identified in *Plasmodium falciparum*, the parasitic cause for life-threatening malaria, yet their role remains largely undiscovered. Through interactions with nucleic acids and proteins, lncRNAs can modulate gene expression at the transcriptional, post-transcriptional, translational, and post-translational levels. Determining the role of lncRNAs in the regulation of the *P. falciparum* transcriptome and proteome is imperative to further our understanding of gene regulation in the parasite.

The characterisation of *P. falciparum* lncRNAs has been hindered by an incomplete annotation and the absence of disruption methods that together would permit high-throughput systematic knockdown of lncRNAs. During my PhD, I addressed these challenges to enable the study of *P. falciparum* lncRNAs *in vitro*. I generated a high-quality lncRNA annotation using manual curation of sequencing data generated at the Sanger Institute, along with supportive datasets from the literature. I evaluated CRISPR-based approaches for *in vitro* disruption of lncRNAs including gene knockout, knockdown, and interference. CRISPR-associated enzymes were explored including commonly used DNA-cutting enzymes (Cas9), inactivated enzymes to block transcription (dCpf1) and enzymes that target RNA directly (Cas13), the latter of which had not been applied to *Plasmodium*. Furthermore, I implemented these tools to demonstrate the feasibility of lncRNA studies in *P. falciparum*. I interrogated a set of lncRNAs that were selected based on predicted biological significance and targetability using dCpf1. LncRNA-depleted parasites were phenotypically characterised by assessing changes in fitness, drug resistance, gametocytogenesis and expression. I identified potential roles for specific lncRNAs in drug resistance and gametocytogenesis.

By developing bioinformatics and molecular tools, this work enables future studies elucidating the specific roles of lncRNAs in *P. falciparum*. Understanding the transcriptome and gene regulation will inform the development of novel interventions for the control and eradication of malaria, which remains a serious global health concern.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr Marcus Lee, who accepted me into his lab, and without whom, this work would not have been made possible. Thank you for giving me space to explore my ideas and the guidance needed to execute them, especially in times of unprecedented uncertainty, in a welcoming and stimulating environment. I also thank my co-supervisor Dr Matt Berriman for helping devise this study, the precise blend of approaches to parasitology that I was seeking, and my thesis committee for providing grounding feedback early on: Dr Julian Rayner, Dr Gavin Wright and Dr Erik Miska.

Thank you to Dr Sophie Adjalley and Dr Adam Reid for volunteering their time, expertise, and patience to help bring this work to fruition. You have been crucial mentors and teachers to me, from whom I have learned how to conduct science and think critically.

To the members of Team 226, past and present, thank you for making the days in the lab fly by. You are all inspiring scientists; it has been a privilege to have worked alongside you. Special recognition to Emma Carpenter, who has been an invaluable comrade in this endeavour: celebrating the successes, commiserating the failures, and providing motivation to continue on in the most challenging of times. I am grateful to Juli Cudini and Hannah Jagoe for taking us under their wings, bestowing all their wisdom upon us and cheering us on. I also appreciate Mukul Rawat and Heledd Davies for their selfless support and counsel and Liam Prestwood for his instrumental role in running the lab.

I thank the college and institute administrators who supported my studies (Rebecca Sawalmeh, Annabel Smith and Christina Hedberg-Delouka) and the college directors of studies who provided me with an opportunity to teach (Liria Nakagawa and Rita Monson). Thank you to the students that I supervised in Biology of Disease over the years for keeping my parasitology knowledge fresh and preparing me to answer far-out questions.

I would like to express immense gratitude to the friends and family who have held me together during this PhD. To my Cambridge friends: Giacomo Gattoni, Andrea Ferlini, the rest of the One Man's Tree, Leakey (Catherine Leake), Joe Allison, Elizabeth Bearblock, Juli Cudini and Ananth Pallaseni thank you for providing friendship and happiness, especially in the darkest days of lockdown. To my sporting communities (CUIHC and CURUFC): thanks for giving me a home in Cambridge, welcomed distraction, endless joie-de-vivre and another life purpose as a captain. To my wonderful extended family and lifelong friends (Amicha Robertson, Brynley Bertram-Wright and Maki Shimizu), thank you for always reminding me of my identity outside science and who I am as a person. To my greatest friend and partner-in-crime, Kishen Chahwala: thank you for being here every day of this PhD, providing vital encouragement, support, optimism, distraction, sustenance, and perspective.

Finally, I would like to thank those to whom this thesis is dedicated. My parents, who encouraged me to develop hypotheses and opinions from a young age and instilled in me a desire to understand the unknown and improve the lives of others. And my brother Tom, who taught me that perseverance and mindset are essential to overcoming challenges. Although I may not have entered the family business of brain science, the three have provided unwavering love and support.

Table of contents

List of figures

List of tables

Abbreviations

aslncRNA	antisense lncRNA
bp	base pair
BSD	blasticidin S deaminase
cDNA	complementary DNA
CM	complete media
CRISPR	clustered regularly interspaced short palindromic repeats
DE	differentially expressed
DiCre	dimerisable cre recombinase
DNA	deoxyribonucleic acid
DSB	double-stranded breaks
GFP	green fluorescent protein
gRNA	guide RNA
hDHFR	human dihydrofolate reductase
HDR	homology-directed repair
HEPES	4-(2-hydroxyethyl)-1 piperazineethanesulfonic acid

IDC	intraerythrocytic developmental cycle
InDel	insertion or deletion mutation
lncRNA	long noncoding RNA
miRNA	microRNA
mRNA	messenger RNA
ncRNA	noncoding RNA
NHEJ	non-homologous end joining
PAM	protospacer adjacent motif
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFS	protospacer flanking sequence
qPCR	quantitative PCR
RBC	red blood cell
RNAi	RNA interference
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT-qPCR	reverse transcription-quantitative PCR
RT	room temperature
RUF	RNA of unknown function
snoRNA	small nucleolar RNA

SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
TAE	tris-acetate-ethylenademiaminetetraacetic acid
TARE	telomere-associated repetitive elements
tRNA	transfer RNA
TSS	transcription start site
TX	transfection
UTR	untranslated region
WT	wildtype

Chapter 1

Introduction

1.1 Malaria

Malaria is an ancient disease caused by infection with parasites of the *Plasmodium* genus. These eukaryotic parasites have co-opted humans, among other vertebrates, as their hosts and caused deadly infections for millennia. Despite advancements in understanding the disease, implementation of interventions, and substantial investment, malaria has not been eliminated. Elimination is not near, with hundreds of millions of cases occurring annually and billions of people at risk for infection.

1.1.1 Life-threatening infection

Malaria presents as an acute febrile illness with symptoms of high fever, headache, vomiting, diarrhoea, muscle pain, shaking, tiredness and anaemia (WHO, 2022). The fever is notably periodic, occurring every 3-4 days due to the synchronised release of parasites from infected erythrocytes. Depending on the species of the infecting parasite, the host's immune and the treatment, the infection can lead to severe consequences, including kidney failure, persistent liver infection, splenomegaly, cerebral malaria, placental malaria and possibly death if untreated (Milner, 2018). Cerebral malaria presents as acute encephalitis, which can progress to coma. Placental malaria can result in adverse outcomes such as intrauterine growth retardation, premature delivery, low birth weight and fetal loss. Even uncomplicated malaria, albeit not life-threatening, can significantly impact quality of life and the burden on socioeconomic systems (Kayiba et al., 2021).

In endemic regions, the entire population, including healthy individuals, are at risk for *Plasmodium* infection. However, vulnerable populations are at a higher risk for increased disease severity due to suppressed, underdeveloped or absent immunity. These populations include young children, pregnant women, immunocompromised individuals, travellers and migrants.

1.1.2 Historical perspective

References to infections resembling malaria in ancient civilisations appear in numerous sources: Chinese texts, a Mesopotamian clay tablet, an Egyptian papyri, Hindu texts and later in ancient Greek literature (Cox, 2010). These early texts describe the signature periodic fever and enlarged spleen of malaria infection. While the written records date early malaria infections to around 2700 BCE, researchers have detected malaria antigens (aldolase and hrp2) in the remains of Egyptian mummies dating back to 3200 BCE (Cerutti et al., 1999; Al-Khafif et al., 2018). Malaria has inflicted many, including influential people such as monarchs, religious leaders, philosophers, writers and explorers. Although the cumulative deaths are hard to estimate, it is clear that malaria has had a substantial impact on human populations throughout history.

Further, malaria has profoundly impacted human evolution. Mutations that protect against infection are selected for and maintained in endemic populations. For example, changes to erythrocyte shape, such as sickle cell anaemia and ovalocytosis, protect carriers against malaria as the parasite cannot efficiently infect erythrocytes (Williams, 2006). Another example is a variant that emerged 42 000 years ago in sub-Saharan Africa, which disrupts the expression of Duffy antigen, a surface protein involved in erythrocyte invasion (McManus et al., 2017).

Ancient peoples believed that afflictions like malaria had supernatural causes and were unaware of the role of parasites and mosquitoes. Although for thousands of years, people understood that malaria incidence increased near wetlands. At around 400 BCE, the miasma theory arose, primarily advanced by the Greek physician Hippocrates, which claimed that poisonous air caused pestilence (Hempelmann and Krafts, 2013). Thus, people believed that vapours from stagnant bodies of water and decaying matter caused malaria, even the origins of the name, *Mal'aria* translates to bad air in 18th century Italian (Cox, 2010). Although this theory was incorrect, it inspired effective disease management, such as waste removal and moving populations away from wetlands where mosquitoes are prevalent.

Microscopic life was first observed in the 1600s by Robert Hooke and Antonie van Leeuwenhoek; however, it was only in the late 1800s that Louis Pasteur and Robert Koch, respectively, developed and proved germ theory. Thus began the Golden Age of Microbiology, when scientists identified many microbiological agents causing infectious diseases. In 1880, Charles Louis Alphonse Laveran identified the *Plasmodium* parasite in the blood of malaria patients at a military hospital in Algeria (Cox, 2010). Sir Ronald Ross later incriminated mosquitoes as the vectors mediating transmission by observing parasites in the gut of an *Anopheles* mosquito that fed on a patient with malaria. These two discoveries initiated the study of parasite and vector biology to understand malaria.

1.1.3 Current global burden and challenges

Malaria continues to be responsible for human illness and death today. In 2021, there were an estimated 247 million cases and 619 000 deaths (WHO, 2022). Most cases occurred in Africa (234 million), specifically affecting the poorest and most vulnerable communities (Figure 1.1).

Human interventions have aided in controlling malaria since they were implemented in the early 1900s, yet they have failed to eradicate malaria. Mosquito populations are controlled by using insecticides and draining water reservoirs. Houses in endemic areas are fumigated (indoor residual spraying) and outfitted with insecticide-treated nets and metal roofing (Carter and Mendis, 2002). Antimalarials were discovered as early as 1820 and are continuously developed to provide life-saving treatments for the infected (Carter and Mendis, 2002). These therapeutics are also used as malaria prophylaxis for vulnerable populations (pregnant women and children under 5), which are administered seasonally when there is an increased risk of malaria.

One major challenge has been the delay in the development of an effective prophylactic vaccine for malaria. Although the search for a vaccine began in the 1960s, only in October 2021 did the World Health Organization (WHO) recommend the first malaria vaccine for widespread use. This recombinant-protein-based vaccine, RTS,S/AS01 (Mosquirix), combines epitopes of the *P. falciparum* circumsporozoite protein with the surface antigen of the Hepatitis B virus. In clinical trials, after receiving the third dose, the vaccine protected 55% of children aged 5-17 months and 31% of infants 6-12 weeks of age against clinical malaria for 12 months (RTS, 2015). Although the efficacy is modest, the vaccine is currently being

given to children living in regions with moderate to high malaria transmission. There is a continued effort to develop a more efficacious vaccine.

Another challenge is drug resistance, which has arisen after implementing each antimalarial. The oldest treatment for malaria is quinine, a compound extracted from the bark of the cinchona tree in South America by the Quechua people. Since then, many other drugs have been discovered and developed, including chloroquine, sulfadoxine, pyrimethamine, mefloquine, artemisinin, atovaquone, proguanil, lumefantrine and piperaquine (Haldar et al., 2018). Their mechanisms are not all fully-elucidated although the pathways targeted by these therapeutics include heme metabolism, dihydropteroate synthetase, protein synthesis, and cytochrome B. Chloroquine resistance arose ten years after implementation; resistance emerged even quicker in other drugs like sulfadoxine-pyrimethamine, mefloquine and piperaquine, and in 2009, resistance to combination therapies was first observed (Haldar et al., 2018). Tremendous resources are devoted to the implementation of drug intervention programs, surveillance of drug resistance and discovery of novel drugs (Ippolito et al., 2021).

Climate change is another factor that influences malaria incidence both directly and indirectly. Rising temperatures, rainfall, changes in weather patterns and extreme weather incidents like flooding can lead to direct changes to the biology of the parasite or vector, the geographic distribution of mosquitoes or human migration patterns (Nissan et al., 2021). Also, the disruption and instability caused by climate change have consequences for socioeconomic systems and malaria is an infection that disproportionately affects poor populations, and climate change deepens economic inequality and increases poverty (Tusting et al., 2013). Malaria intervention infrastructures such as logistics, supply chains, government programming and funding are vulnerable. The recent COVID-19 pandemic has revealed this vulnerability as disruptions between 2019 and 2021 have been linked to an excess of 13.4 million cases, including 63 000 deaths (WHO, 2022).

1.2 *Plasmodium* parasites: the cause of malaria

Understanding the cause of an illness is essential in developing treatments and prevention strategies; in parasitic infection, the cause is an organism with an intricate relationship with its host. The parasitic cause of malaria is infection with *Plasmodium* spp., which are unicellular eukaryotes (protozoans). They are classified as apicomplexans, possessing the characteristic apical complex structure, and specialised organelle called an apicoplast. *Plasmodium* spp.

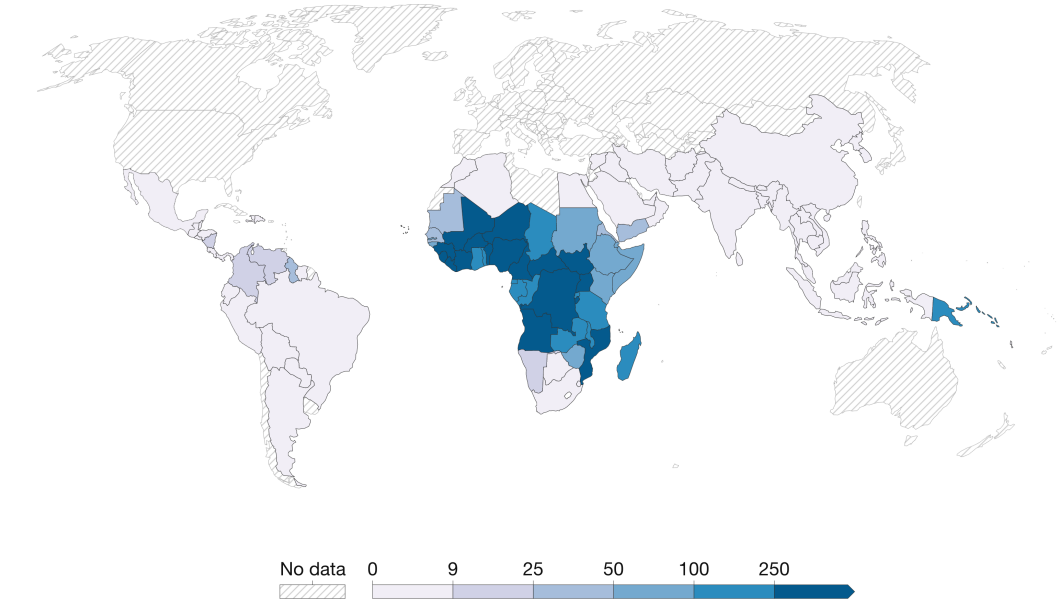


Fig. 1.1 Map showing the global distribution of malaria incidence in 2020. This world map shows the prevalence of malaria measured by the number of new cases of malaria in a year per 1000 people. White regions indicate areas of low incidence, and blue regions indicate areas of high incidence. This map was taken from ourworldindata.org, which was generated from data acquired from the World Health Organization via World Bank. The figure and legend were obtained from Ritchie (2019), which is under a CC-BY license.

are obligate endoparasites that require the infection of both arthropods (mainly *Anopheles* mosquitoes) and vertebrates (such as mammals, birds and reptiles) to complete their life cycle (Molina-Cruz et al., 2016; Galen et al., 2018).

1.2.1 *Plasmodium* species responsible for human disease

Of the hundreds of *Plasmodium* species, only six are known to cause malaria in humans: *P. falciparum*, *P. vivax*, *P. ovale.walliker*, *P. ovale.curtisi*, *P. knowlesi* and *P. malariae* (Figure 1.2). Most of these infect humans exclusively as an intermediate host, while *P. knowlesi* and *P. malariae* also infect other primates (Rutledge et al., 2017). Zoonotic infections from other *Plasmodium* species are rare occurrences.

The species vary in many ways, including in geographical distribution and pathology. *P. falciparum* is the most predominant species with a global distribution and high prevalence in Africa, where 90% of the cases occur (WHO, 2022). *P. falciparum* also causes the most severe form of malaria and is responsible for the most deaths. The remaining species cause less life-threatening malaria (Milner, 2018). However, two of the species, *P. vivax* and *P. ovale*, can be dormant in the liver and reactivate many months or years later. The two species occupy separate niches with *P. ovale* in sub-Saharan Africa and *P. vivax* in Asia and South America. *P. malariae* has a global distribution but it is responsible for few cases. *P. knowlesi* is found in Southeast Asia and primarily infects macaques but can infect humans. From a global perspective, *P. falciparum* is the species of greatest importance as it causes the most harm to humanity. Therefore, most research studying malaria centres on *P. falciparum*.

1.2.2 Life cycle of the *Plasmodium falciparum* parasite

P. falciparum infects *Anopheles* mosquitoes, which transmit the parasite to humans during a blood meal. Mosquitoes are the primary hosts as fertilisation occurs in the mosquito gut, while humans are the secondary hosts; both are necessary for the parasite life cycle. Only female mosquitoes transmit parasites to humans, as males do not feed on humans. Transmission often occurs at night because *Anopheles* mosquitoes are "night-biters", feeding between sunrise and sunset and sensing their hosts through olfaction and carbon dioxide sensing (Konopka et al., 2021).

During a blood meal, haploid *P. falciparum* sporozoites, carried in the mosquito's salivary glands, are injected into the bloodstream (Figure 1.3) (Bousema and Drakeley, 2011). The

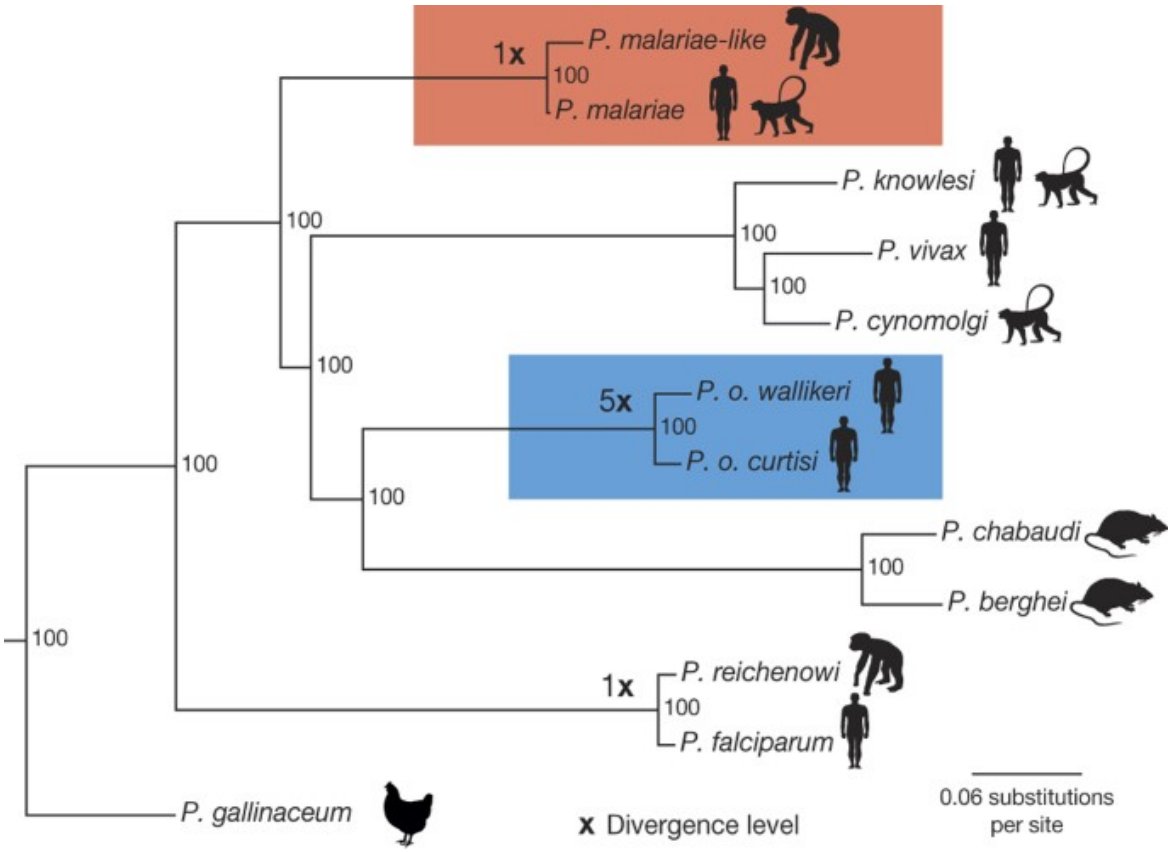


Fig. 1.2 Maximum likelihood phylogenetic tree showing the *Plasmodium* genus. The divergence levels are calculated relative to the *P. falciparum* and *P. reichenowi* split (3–5.5 million years ago). Therefore, the *P. ovale* split is predicted to have occurred 20.3 million years ago, and the *P. malariae* split 3.5 million years ago. Silhouettes show the host specificity of each *Plasmodium* species. Bootstrap values are shown at branching points. The figure and figure legend were obtained and adapted from Rutledge et al. (2017), which is under a CC-BY license.

sporozoites are motile and proceed to invade liver cells, where they replicate asexually for 7-10 days (the exo-erythrocytic cycle). Once released into the bloodstream as merozoites, they replicate asexually in red blood cells (the intraerythrocytic development cycle, IDC). Every 48 hours, parasites cycle through the following stages: merozoites infect erythrocytes becoming ring stages, which mature into trophozoites and then develop into multi-nucleated schizonts. Schizonts rupture and 20-36 daughter merozoites are released that infect new erythrocytes (Rudlaff et al., 2020). The parasites complete the IDC in synchrony, and the simultaneous rupturing of erythrocytes causes periodic high fever. A minority of parasites (fewer than 10%) differentiate into sexual stages called gametocytes (Smith et al., 2002). This process, gametocytogenesis, is formed of five stages, ending with the formation of either female or male gametocytes (Figure 1.3). These sexual stages must be produced and ingested by a mosquito to complete the remaining stages of *P. falciparum* life cycle (the sporogonic cycle).

Mosquitoes ingest gametocytes, which develop into fertile male microgametes and female macrogametes in the mid-gut (Bennink et al., 2016). Microgametes are generated through exflagellation, which enables the gamete to become motile. The gametes fuse to create a zygote (diploid), which then converts into a motile form called an ookinete. Within 20hrs since ingestion, the ookinete exits the gut lumen and settles in the gut epithelium, where it is converted into an oocyst and, over two weeks, undergoes replication. The oocyst ruptures, releasing thousands of sporozoites that travel to the salivary glands, where they restart the parasite life cycle. The average life span of *Anopheles* mosquitoes is only two weeks in nature, which is just enough time to accommodate the *P. falciparum* life cycle.

1.2.3 The *P. falciparum* genome

Plasmodium parasites have three genomes: the nuclear, mitochondrial and apicoplast genomes. The *P. falciparum* nuclear genome is 23.29Mb (megabases) and comprises of 14 chromosomes containing 5280 genes. The remaining genomes are small and circular: the apicoplast genome is 34 kb (kilobases), containing 30 genes, and the mitochondrial genome is 5.9 kb with 3 genes. The *P. falciparum* nuclear genome has strikingly high (A + T) content, which reaches 80% in coding regions and 90% in noncoding regions (Chappell et al., 2020). This is not true of all *Plasmodium* species, *P. vivax* has a comparatively GC-rich genome. AT-rich genomes pose a challenge for many sequencing approaches designed for mammalian species with higher GC content. Despite these challenges, a high-quality *P. falciparum* draft genome was generated in 2002 and has since been improved, including the closure of all gaps

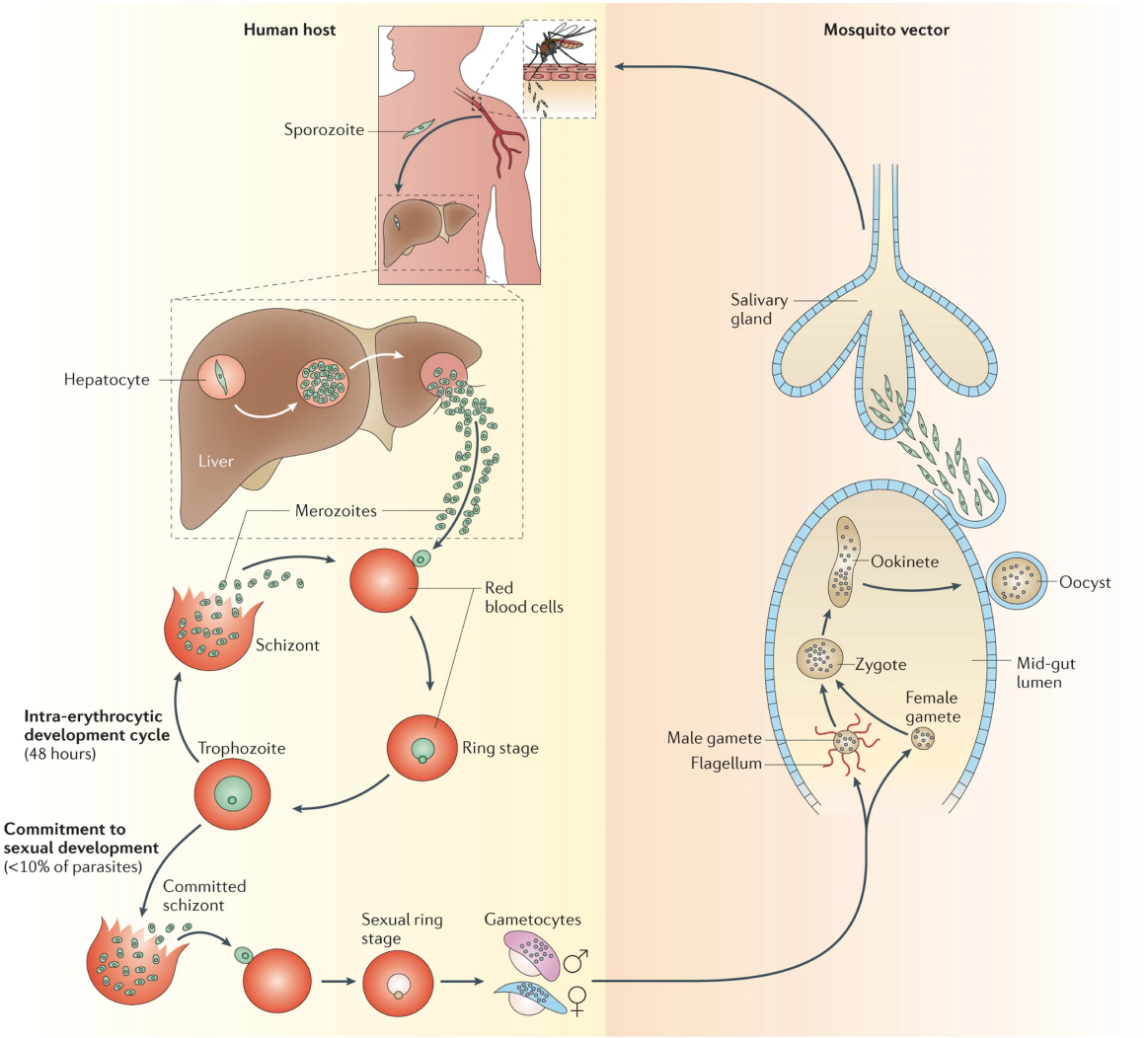


Fig. 1.3 The life cycle of the *Plasmodium falciparum* parasite. Mosquitoes inject sporozoites into the bloodstream, which travel to the liver, where they replicate asexually. After 7-10 days, they are released as merozoites and infect red blood cells. Every 48 hours, they transition through the asexual ring, trophozoite, and schizont stages. Some parasites become sexual stages called gametocytes, which are ingested by mosquitoes. These develop into male microgametes and female macrogametes in the mid-gut, which fuse to create a zygote, which then converts into an ookinete. The ookinete settles in the gut epithelium, becomes an oocyst and replicates. After two weeks, the oocyst releases thousands of sporozoites that travel to the salivary glands. The figure was obtained with permissions from Josling and Llinás (2015).

(Gardner et al., 2002; Böhme et al., 2019). The application and development of additional tools into the parasite such as RNA sequencing (RNA-Seq), immunoprecipitation, chromatin interactions, gene editing, microscopy, and proteomics have also enhanced our understanding of the malaria parasite.

1.3 Gene regulation in *P. falciparum*

The *P. falciparum* life cycle is complex, consisting of many different life stages. The parasite undergoes significant morphological changes, including a 7-fold increase in size in the IDC and transitions between motile and immotile, uninucleate and multinucleate, and sexual and asexual stages (Machado et al., 2021). The parasite must survive in diverse environmental niches in the definitive host and intermediate host, from the midgut and salivary glands of the mosquito to human liver cells and erythrocytes. Furthermore, the parasite transitions through these stages at specific times and sense and respond to external factors. These processes must be underpinned by tight regulation and coordinated programs of gene expression in the *Plasmodium* genome.

1.3.1 The *P. falciparum* transcriptome

Early studies identified that, like most eukaryotes, *P. falciparum* exhibits monocistronic transcription, meaning that each cistron is transcribed as a separate transcriptional unit. In this type of transcription, genes require a proximal promoter and terminator. Transcription in *P. falciparum* was characterised by a "just-in-time" model, where gene expression is tightly regulated and expressed in specific stages (Le Roch et al., 2003; Bozdech et al., 2003; Llinás et al., 2006). This model suggests that transcription is initiated only once per cycle at the required time. However, this model was later challenged when researchers did not find sufficient evidence of transcriptional regulation to support the model in the form of a scarcity of transcription factors. An alternative theory came into favour that suggested that regulation occurs at the post-transcriptional level, which was supported by an observed delay between transcription and translation (Coulson et al., 2004; Le Roch et al., 2004; Hall et al., 2005). This type of regulation was identified in trypanosomes, another protozoan, although these parasites exhibit polycistronic transcription. A tour-de-force search for mediators of post-transcriptional regulation ensued. However, it was non-canonical transcription factors that were discovered and proven to be functional.

1.3.2 Transcriptional and epigenetic regulation

Transcription factors

DNA-binding domains were identified in early genomics studies; however, these proteins were general transcription factors responsible for RNA polymerase recruitment or chromatin structure, not associated with regulating specific genes that could explain the tightly-regulated gene expression. The first specific transcription factors, the apicomplexans AP2s (ApiAP2s), were identified by Balaji et al., who performed a sensitive sequence profile analysis of multiple apicomplexan genomes: *Plasmodium*, *Cryptosporidium*, and *Theileria* (Balaji et al., 2005). ApiAP2s contain a domain similar to Apetala2/ERF (ethylene response factor) integrase DNA-binding domain. AP2s might have entered the genome of ancestral *Plasmodium* via horizontal gene transfer from the red algae endosymbiont; however, AP2s have also been found in bacteria and phages (Wessler, 2005). Although many AP2s are conserved between apicomplexans, there is also evidence of expansion occurring in lineages after the divergence of apicomplexan species. Within *Plasmodium* spp., the AP2 domains are highly conserved, but their putative target genes vary considerably (Campbell et al., 2010).

There are 27 putative ApiAP2s in *P. falciparum*, which each contain one to four AP2 domains, and there is little shared homology outside of these domains. Subsequent analysis revealed that the expression of these transcription factors clustered around specific developmental stages and had different binding motifs, targets, and functionalities (Figure 1.4) (Campbell et al., 2010). AP2-I is associated with erythrocyte invasion, likely inducing chromatin remodelling through its interaction with BDP1 and BDP2. AP2-exp was also associated with sporozoites. Another AP2, PfSIP2, binds to the SPE2 DNA motif found in subtelomeric domains and acts as a DNA-tethering protein in heterochromatin formation and maintenance (Flueck et al., 2010). AP2-tel is a component of the telomere-binding complex, which interacts with BDP1. AP2-G and AP2-G2 play a role in gametocytogenesis (Kafsack et al., 2014; Sinha et al., 2014). AP2-G is essential for sexual commitment and likely targets early gametocyte genes. AP2-O activates gene expression in ookinetes and is synthesised in female gametocytes (Yuda et al., 2009). AP2-Sp contributes to sporozoite formation and is expressed from late oocysts to sporozoites (Yuda et al., 2010). APG-L is expressed in sporozoites and is necessary for parasite development in hepatocytes (Iwanaga et al., 2012).

Non-ApiAP2s exist; there are at least 22 other DNA-binding proteins, some of which contain homology with known transcription factors (Coulson et al., 2004; Toenhake and Bártfai, 2019). Recently, homeodomain-like protein 1 (*hdp1*, PF3D7_1466200) was discovered and

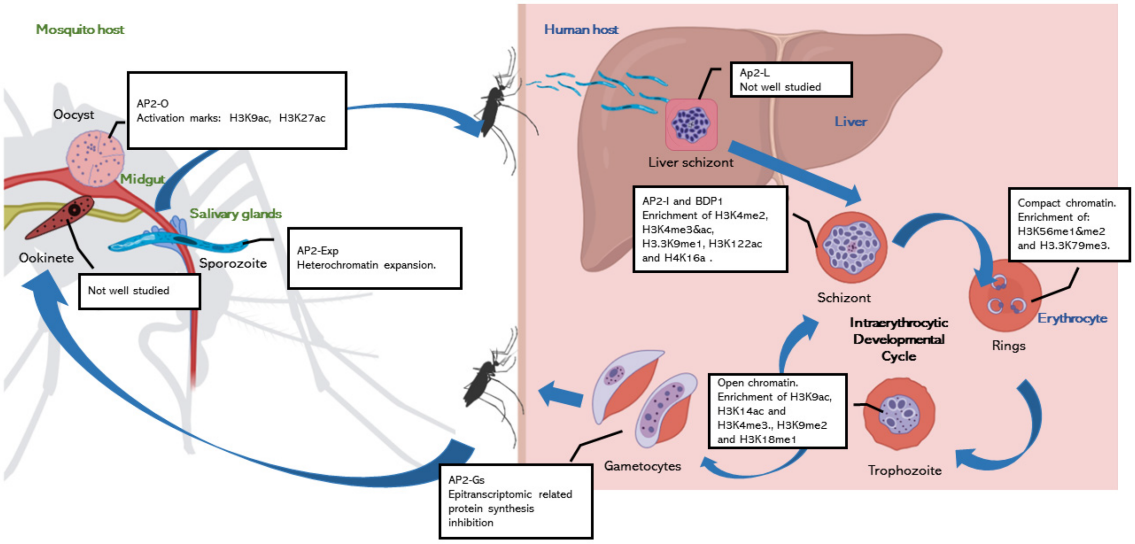


Fig. 1.4 Transcriptional and epigenetic regulation in *P. falciparum*. This figure superimposes elucidated epigenetic pathways for each life stage onto the *P. falciparum* life cycle. Specific AP2 transcription factors regulate transcription at different life stages, such as AP2-G in gametocytogenesis and AP2-L in sporozoites infecting the liver. Chromatin reshaping is integral to the asexual blood-stage stages. Different histone modifications have been identified in different stages. The figure and legend were obtained and adapted from Serrano-Durán et al. (2022), which is under a CC-BY license.

has a role in the regulation of early-stage gametocyte genes (Campelo Morillo et al., 2022). It binds to GC-rich motifs to upregulate gene expression. Although, when disrupted, not all expression was abolished, suggesting that *hdp1* works within a network of transcriptional regulators.

DNA methylation

DNA methylation is a common way that eukaryotes distinguish the regions of the genome that should be expressed. Two bases can be methylated: cytosine and adenine. Early studies thought DNA methylation (5mC) was absent from the parasite, which was supported by the absence of DNA methyltransferase homologues. However, DNA methylation of cytosine residues was identified using BS-Seq (bisulfite conversion of unmethylated cytosines coupled with sequencing) (Ponts et al., 2013). Although, the conclusions from this study are disputed. Hammam et al. detected low levels of 5mC in the parasite and identified a 5hmC-like-modification at slightly higher levels (Hammam et al., 2019). The N6-methyladenosine (m6a) mark has also been identified in the parasite (Luo et al., 2016).

Nucleosome landscape and histone modification

Gene regulation also occurs at the level of histones, which hold 147 base pairs (bp) of DNA in *P. falciparum* and form nucleosomes in octamers. The patterns of nucleosome occupancy are highly variable throughout the parasite life cycle (Ponts et al., 2011). In addition to four core histone types: H2A, H2B, H3 and H4, there are also variant types such as H2A.Z, H2B.Z, H3.3, and H3Cen, the first two preferentially occupy intergenic regions (Hoeijmakers et al., 2019). The parasite lacks a +1 nucleosome that indicates the transcriptional start site (TSS); instead, it is positioned within the coding regions (Ponts et al., 2011; Bunnik et al., 2014).

Histones can be modified to facilitate gene regulation by depositing histone marks, including methylation and acetylation of certain residues. 232 histone post-translation modifications have been identified in *P. falciparum* but only 36 have been characterised (Salcedo-Amaya et al., 2009; Saraf et al., 2016). These activating and repressive marks are deposited by "writers", removed by "erasers", and interpreted by "readers". Histone lysine methyltransferases (PfSET1-10) deposit methyl groups on lysine residues resulting in the H3K4, H3K9, H3K27, H3K36, H3K5/K8/K12 and H4K20 histone marks. Histone demethylases (lysine-specific demethylase 1 and Jumonji C-domain containing histone demethylases) erase these marks. Histone acetylation is facilitated by four histone acetyltransferases (HAT), including GCN5 and MYST, which are specific to H3 and H4, respectively and histone deacetylases (HDACs) such as Sir2a, Sir2b and HDACs (1-3) mediate histone deacetylation. Also, some methyltransferases add methyl groups to protein arginine residues (PRMT 1, 4 and 5). Bromodomain proteins (1-6) read acetylated lysine residues, while other acetylated and methylated histones are read by SET proteins, PhD-domain-containing proteins, MYST and HP1. These marks are found in active promoters but also in silent promoters, the $5'$ end of coding genes and intergenic regions (Cui et al., 2007; Salcedo-Amaya et al., 2009; Bártfai et al., 2010). ChIP-Seq (chromatin immunoprecipitation sequencing) experiments have verified that these marks are stage-specific (Figure 1.4) (Bunnik et al., 2018; Fraschka et al., 2018).

Chromosome conformations

Studies using immunofluorescence identified that regions of *P. falciparum* chromatin where gene expression was known to be repressed, such as subtelomeric regions and antigenic variation genes, were forming clusters around the nucleus (Figueiredo et al., 2002; Horrocks et al., 2009; Epp et al., 2009). This architecture was later confirmed with chromosome

conformation capture, which captures long-range chromatin interactions (Ay et al., 2014; Bunnik et al., 2018; Batugedara and Le Roch, 2019). The *P. falciparum* chromosomes are arranged around the nucleus in folded structures that are parallel to each other, with separate transcriptionally active and repressed compartments, possibly tethered to one area of the nucleus.

Noncoding RNAs and antisense transcription

Noncoding RNAs such as miRNAs and lncRNAs can also modulate gene regulation and participate in regulatory networks. The *P. falciparum* genome does not encode miRNAs. The parasite lacks the RNA interference (RNAi) machinery required for miRNA processing; therefore, it is not a likely mechanism of transcriptional regulation. However, a recent study has proposed that the parasite imports the human miRNA-RISC complex (Dandewad et al., 2019). The parasite encodes thousands of lncRNAs and exhibits extensive antisense transcription. There is also a high proportion of bidirectional promoters in the parasite that drive lncRNA expression (López-Barragán et al., 2011; Adjalley et al., 2016). The role of lncRNAs in *P. falciparum* has not been fully elucidated; however, a few have been implicated in transcriptional regulation, which I will describe in detail in the next section.

1.3.3 Translational regulation

Plasmodium spp. exhibits classic eukaryotic protein translation and, like many other eukaryotes, have mechanisms of translational regulation (Cui et al., 2015; Holmes et al., 2017). Polysomal profiling revealed that more than 30% of genes exhibit a delay between the relative abundance of mRNAs and the polysomal fraction (Bunnik et al., 2014). This delay is particularly prevalent in gametocyte and sporozoite life stages, where some transcripts are produced but not translated until specific times (Hall et al., 2005; Mair et al., 2010; Lindner et al., 2013; Silvie et al., 2014). These transcripts enable the transition stages (gametocytes and sporozoites) to persist and remain ineffective until a blood meal transfers the parasite to the next host.

Translation can be regulated globally in response to conditions of cellular stress. Known as the integrated stress response, phosphorylation of the α subunit of eIF2 leads to global translational repression. Phosphorylated eIF2α acts as an inhibitor of eIF2B, reducing the conversion of eIF2-GDP to eIF2-GTP, which is necessary for forming the 43S pre-initiation complex. *P. falciparum* encodes three eIF2α kinases that facilitate this phosphorylation: IK1,

IK2 and PK4. IK1 phosphorylates eIF2α in response to amino-acid starvation (Fennell et al., 2009). IK2 represses protein synthesis during sporozoite development in the salivary glands to support latency until transmission (Zhang et al., 2010). PK4 inhibits protein synthesis in asexual blood stages to accommodate the energy requirements of schizogony and in mature gametocytes to promote latency similar to sporozoites (Zhang et al., 2012).

Translational repression can also be specific to individual mRNAs. For example, in *P. berghei*, rodent malaria, the Dhh1 RNA helicase DOZI complex and CITH repressors protect mRNAs expressed in gametocytogenesis from degradation (Mair et al., 2006; Braks et al., 2008; Guerreiro et al., 2014). It is unknown if this also occurs in *P. falciparum*. The highly-conserved eukaryotic Puf protein has also been shown to repress the expression of mRNAs in *Plasmodium* spp. in gametocyte and sporozoite stages (Cui et al., 2002; Fan et al., 2004; Müller et al., 2011). It has been proposed that Puf binds the 5′ untranslated region (UTR) of select genes via Puf-binding elements. *P. falciparum* contains long 5′ UTRs, although their role in the translation remains unclear, although there is one example of their role in regulation. The translation of a 360 bp upstream open reading frame in *var2csa* mediates translational repression of *var2csa* mRNA by impacting translation initiation efficiency (Amulic et al., 2009; Bancells and Deitsch, 2013; Fastman et al., 2018).

1.4 Long noncoding RNAs

RNA is a nucleic acid principally known for its role as an intermediate between DNA and protein, carrying the genetic information required for the synthesis of proteins. The secondary function of RNA is to produce noncoding RNAs (ncRNAs) required for housekeeping activity, including mRNA splicing and protein synthesis: small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). RNA also plays a third role as a regulatory molecule. The first regulatory ncRNAs identified were small RNAs known as small interfering RNAs, microRNAs (miRNAs) and small-PIWI interacting RNAs, which interact with Argonaute proteins to facilitate gene regulation. Additional ncRNAs were discovered that did not fit into the categories of ncRNAs, but these were exceptional cases, and noncoding genomic regions were mostly classified as "junk" DNA (Pennisi, 2012). However, after transcriptomic technologies emerged, it became apparent that noncoding elements are transcribed and are a general phenomenon, which required a classification (Berretta and Morillon, 2009). Long noncoding RNAs (lncRNAs) were named based on their structure, longer than well-defined small ncRNAs, and were

classified as at least 200 nucleotides long (Kapranov et al., 2007). Although this classification has been recently questioned due to many ncRNA transcript sizes occurring at this boundary, creating a grey zone in nomenclature (Mattick et al., 2023).

1.4.1 Characteristics

Most lncRNAs resemble mRNAs; they are transcribed by RNA polymerase II and undergo post-transcriptional processing, including splicing, the addition of a 7-methylguanosine cap and polyadenylation. However, some lncRNAs do not undergo any of these processes: they are transcribed by other polymerases, processed from precursors with Rnase P or snoRNA-protein complexes, or form closed circular structures (Mattick et al., 2023). The transcribed lncRNAs form secondary and tertiary structures that could be capable of binding DNA, RNA, proteins or small molecules (Yao et al., 2019). LncRNAs can localise to the nucleus or cytoplasm. Through these interactions, lncRNAs can act as transcriptional regulators. These interactions can be broadly categorised into two classes: *cis*-regulatory mechanisms that act locally or *trans*-regulatory mechanisms that affect distant molecules through diffusion or chromatin conformations (Figure 1.5A).

1.4.2 Mechanisms of action

Anywhere from hundreds to hundreds of thousands of lncRNAs have been identified in eukaryotic organisms (2000 in *Saccharomyces cerevisiae* and 100 000 in humans) (Mattick et al., 2023). Initial theories suggested that most of these lncRNAs were transcriptional noise, byproducts from transcription that served no biological function, which were lowly expressed and exhibited minimal sequence conservation (Mattick et al., 2023). However, examples have emerged describing the vast range of mechanisms by which lncRNAs regulate gene expression in diverse biological contexts. Some mechanisms include recruiting proteins, acting as a scaffold for protein complexes, regulating mRNA stability and sponging miRNAs (Figure 1.5B). Although it remains unclear to what extent these functionalities apply across the entire class of lncRNAs and to what extent they are conserved between species. Most research elucidating lncRNA functionality centres on their role in humans therefore, the examples incorporated in this section are human lncRNAs.

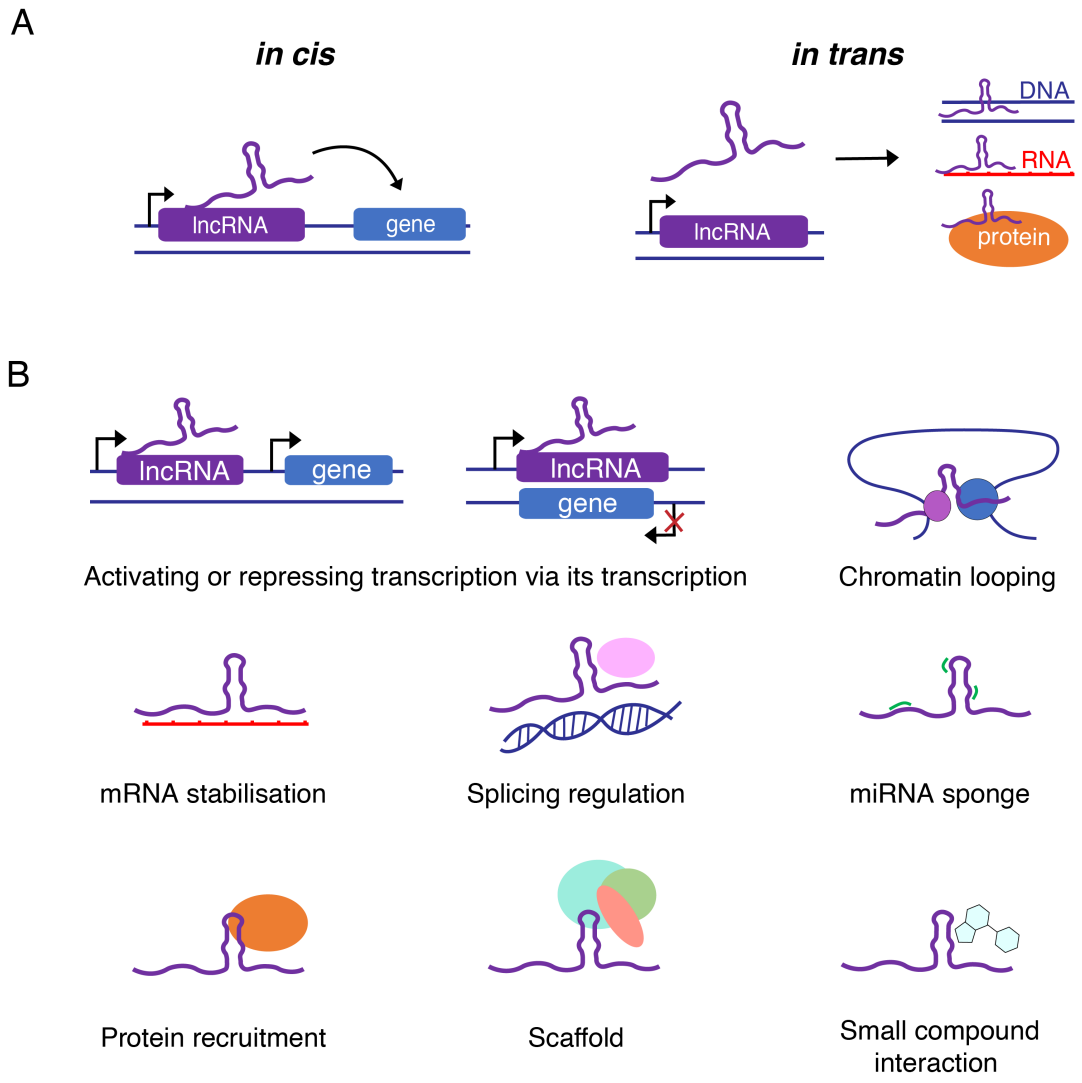


Fig. 1.5 Functionalities of lncRNAs. (**A**) LncRNAs can act locally, affecting the regulation of neighbouring genes (*cis* regulation) or affect distant genes (*trans* regulation) through diffusion or chromatin looping. (**B**) LncRNAs can interact with DNA, RNA, proteins and small molecules to facilitate diverse mechanisms of gene regulation: the transcription of the lncRNA itself can lead to the transcriptional activation or repression of nearby genes, or the lncRNA molecule can be involved in protein recruitment, forming complexes, mRNA stability, splicing, regulating miRNAs and interacting with small molecules.

Role in chromosomal regulation and chromatin remodelling

LncRNAs can regulate chromosome architecture. For example, once the lncRNA *Xist* (X-inactive-specific transcript) is transcribed, it reshapes the architecture of the X chromosome from which it was transcribed (Chen et al., 2016). Through recruitment to the nuclear lamina, *Xist* initialises silencing cascades that lead to chromosome condensation and the formation of the Barr body. Another lncRNA from the X chromosome, *FIRRE* (functional intergenic repeating RNA element), regulates chromosomes by acting as a scaffold that facilitates transchromosomal contacts (interchromosomal interactions) (Hacisuleyman et al., 2014). Some lncRNAs associate with chromatin promote or block the recruitment of chromatin modifiers. These interactions can be *in cis* or *in trans*. For example, *Xist* interacts with SHARP to recruit SMART, which activates HDAC3 to deacetylate histones, silencing transcription on the X chromosome (McHugh et al., 2015). A contrasting example is *lncPRESS1*, which interacts as a decoy with SIRT6, an H3K56 and H3K9 deacetylase, to sequester it away from promoters (Jain et al., 2016).

Role in transcription

LncRNAs can contribute to regulating transcription both as loci and as RNA molecules. The transcription of the loci can influence the transcription of nearby genes. The transcription and splicing of the lncRNA *Blustr* has *cis*-activating effects on the expression of downstream gene *sfmbt2* (Engreitz et al., 2016). Conversely, the transcription of lncRNAs that overlap gene promoters or gene transcripts can silence genes. This phenomenon is often observed with antisense lncRNAs (aslncRNA) such as lncRNA *Air*, which coincides with the *lgf2r* promoter, thereby suppressing the recruitment of RNA Pol II to that gene (Latos et al., 2012).

The loci can also act as enhancer lncRNAs that facilitate chromatin looping through interactions with scaffolding proteins or chromatin modifiers (Yao et al., 2019). For example, lncRNA S*WINGN* interacts with SWI/SNF complex to support transcription of *gas6*, which is contained in the same topological domain (Grossi et al., 2020). A lncRNA loci, *pvti*, competes for enhancer contact *in cis* with other genes (Cho et al., 2018). LncRNAs can form local R loops: triple hybrid RNA:DNA:DNA structures. These R loops tether lncRNAs and recruit transcription cofactors to promoter regions. AslncRNA *KhpsI* forms an R-loop upstream of its antisense gene *sphk1*, which recruits HAT p300 to the *sphk1* promoter, increasing the accessibility of the chromatin to enhance expression (Postepska-Igielska et al., 2015). Similarly, lncRNA *TERRA* (telomeric repeat-containing RNA) forms DNA-RNA hybrids

at telomeres to aid in chromosome maintenance by promoting homologous recombination (Graf et al., 2017). LncRNAs can also interact with transcriptional machinery by acting as a scaffold for initiation machinery or other binding elements. LncRNA *Alu* binds the pre-initiation RNA Pol II transcriptional machinery at promoters to stop transcription (Mariner et al., 2008). Conversely, lncRNA *7SK* suppresses elongation by binding elongation factor B and diminishing its kinase activity (Yang et al., 2001).

Role in nuclear organisation

Some lncRNAs facilitate the formation of RNA-protein complexes called nuclear bodies. For example, paraspeckles are formed and maintained by the lncRNA *NEAT* (Clemson et al., 2009). Paraspeckles are architectural features where specific proteins are sequestered, preventing their activity. Another example is snoRNA-ended lncRNAs (sno-lncRNAs), which form bodies that sequester splicing factors and lncRNA *SatIII*, which creates nuclear stress bodies (Valgardsdottir et al., 2008; Yin et al., 2012).

Role in post-transcription

LncRNAs can affect post-transcriptional mechanisms by sequestering miRNAs, altering mRNA stability, or interacting with splicing machinery. Some lncRNAs act as sponges sequestering miRNAs and abrogating miRNA-mediated RNA silencing, such as *ciRS-7*, a sponge for miR-7 (Hansen et al., 2013). LncRNAs can directly regulate mRNA stability by recruiting degradation proteins or acting as decoys to sequester RNA-binding proteins involved in mRNA decay. Finally, lncRNAs can regulate splicing by interacting with splicing proteins, like *MALAT1*, which interacts with SR proteins and alters their phosphorylation, thereby disrupting their interactions with mRNAs (Tripathi et al., 2010).

Role in translation and post-translation modifications

LncRNAs can be associated with ribosomes, and some lncRNAs can suppress translation by binding mRNAs and recruiting transcriptional activators or repressors. For example, *lincRNA-p21* binds JunB mRNA and recruits translational repressors like Rck (Yoon et al., 2012). There is also evidence that some lncRNAs may be translated into small peptides (Xing et al., 2021). LncRNAs can interfere with phosphorylation, which can impact post-translational modification. For example, *lnc-DC* binds transcription factor STAT3 and promotes its phosphorylation on Tyr705 (Wang et al., 2014). The *lnc-DC* blocks the interaction of STAT3 with SHP1, which usually facilitates its dephosphorylation.

1.4.3 Conservation

Unlike coding regions of the genome, lncRNAs may not exhibit sequence conservation. Instead, they could demonstrate conservation in structure, functionality or syntenic transcription. For example, locally conserved structural motifs have been identified in mammalian lncRNAs; *MALAT1*, *HOTAIR* and *RepA*, which share little sequence similarity between humans and mice (Tavares et al., 2019). Functional conservation has been identified in mammalian lncRNA *JPX*, where neither sequence nor structure is conserved (Karner et al., 2020). Some lncRNAs, referred to as syntologs, are loci of conserved transcription, and their role is based on their expression and not their sequence or structure. This type of conservation has been identified between humans and lancelets using microsynteny analyses (Herrera-Úbeda et al., 2019).

1.4.4 Tools for studying lncRNAs

Expression

Transcriptomic studies provide a practical approach to detecting lncRNAs. Historically, these methods consisted of microarrays and serial analysis of gene expression (SAGE), however now, RNA-Seq is widely-used. Improvements in RNA-Seq technologies have been vital for high-quality lncRNA annotation, such as generating longer reads and full-length transcripts, eliminating the requirement for complementary DNA (cDNA) amplification, or allowing the enrichment for RNAs of interest (rapid amplification of cDNA ends (RACE) or RNA CaptureSeq) (Mercer et al., 2014; Lagarde et al., 2016). Complementary techniques also inform the study of lncRNAs, such as lncRNA transcription dynamics (GRO-Seq, global nuclear run-on sequencing), variation at the single-cell level (scRNA-Seq, single-cell RNA-Seq) and lncRNA sequence modifications (ICE-Seq, inosine chemical erasing coupled with sequencing) (Turner et al., 2019).

There are extensive databases, most notably NONCODE, that are devoted to cataloguing lncRNAs from the literature (in about 40 plant and animal species, not including *Plasmodium*) (Liu et al., 2005). These datasets have enabled the development of bioinformatics tools that can identify lncRNAs from RNA-Seq data, including FEElnc, UClncR, lncScore, LncADeep, lncRNAnet, and longdist (Turner et al., 2019). These tools use statistical approaches or deep learning to analyse various lncRNA features such as neighbouring gene expression, sequence protein-coding potential, association with gene ontology or predicted pathway involvement.

Other tools can support lncRNA annotation by predicting secondary structure (RNAfold and RNAstructure) and evolutionary conservation (EvoFold and RNAz) (Turner et al., 2019).

Interactions

Several high-throughput approaches can capture the interactions of lncRNAs with DNA, RNA or proteins. DNA interactions can be detected using ChIRP-Seq (chromatin isolation by RNA purification), CHART (capture hybridization analysis of RNA targets), GRID (global RNA interactions with DNA by deep sequencing) or MARGI (mapping RNA-genome interactions). ChIRP-Seq and CHART crosslink RNA with chromatin and use biotinylated oligonucleotides to purify RNA-DNA complexes for high-throughput sequencing (Simon et al., 2011; Chu et al., 2012). In comparison, GRID and MARGI use bivalent linkers (ssRNA-dsDNA or ssDNA-dsDNA, respectively) to capture complexes (Li et al., 2017; Sridhar et al., 2017). RNA-RNA interactions can be captured using RNA antisense purification (RAP), which employs probes that tile the target sequence to pulldown binding RNAs (Engreitz et al., 2015). A higher throughput method, LIGR-Seq (ligation of interacting RNA followed by high-throughput sequencing), crosslinks RNA-RNA duplexes using a psoralen derivative and uses RNAse to enrich for duplexes for high-throughput sequencing (Sharma et al., 2016). Finally, an adaptation of the ChIRP protocol, ChIRP-MS, enables the identification of RNA-protein interactions (Chu et al., 2015). RNA-protein complexes are pulled down by the biotinylated oligos, and the proteins are isolated and analysed with mass spectrometry. RNA-dependent protein complexes specifically can be captured with the R-DeeP method, which combines density gradient ultracentrifugation with mass spectrometry (Caudron-Herger et al., 2019). Other RNA immunoprecipitation approaches can be coupled with sequencing to identify RNA-protein interactions, such as RIP-Seq (RNA immunoprecipitation) and CLIP-Seq (cross-inking immunoprecipitation) (Turner et al., 2019).

Localisation

The subcellular localisation of lncRNAs can be determined using RNA fluorescence *in situ* hybridisation (RNA-FISH). Oligonucleotides labelled with fluorophores bind the RNA and give a fluorescent signal. However, traditional RNA-FISH may not be sensitive enough to capture lowly expressed lncRNAs. More sensitive techniques are also used, such as single-molecule RNA FISH or ultra-sensitive RNA-FISH protocols (Turner et al., 2019).

Function

The association of lncRNAs with specific phenotypes or genes can be predicted using bioinformatics. For example, analysing lncRNA expression patterns over differing conditions can reveal an associated phenotype or comparing lncRNA expression patterns with the expression of other genes can show incidences of co-regulation. Genome-wide association studies have also linked lncRNA loci to specific phenotypes; for example, variation in the lncRNA *H19* has been linked to coronary artery disease and ischemic stroke in Chinese populations (Turner et al., 2019). However, these associations require confirmation through *in vitro* or *in vivo* experimentation. The initial approaches to studying lncRNAs experimentally are perturbation studies (the approaches are described in *Chapter 4*). Altering a lncRNA in some way, such as abrogating expression, and observing the impact of that alteration can indicate its features or functions. Furthermore, lncRNA perturbation paired with the broad repertoire of techniques described above (such as sequencing, imaging, and interactome analyses) can further enable the elucidation of lncRNA function.

1.5 LncRNAs in *P. falciparum*

Extensive research elucidating the roles of mammalian lncRNAs has uncovered their contribution to disease and inspired new approaches for therapeutics (López-Urrutia et al., 2019; Rinn and Chang, 2020; Winkle et al., 2021). In parasitic infections, complex host-parasite interactions can facilitate the transfer of regulatory molecules such as miRNAs and lncRNAs between organisms. For example, infection with *Toxoplasma gondii*, *Schistosoma mansoni*, *Cryptosporidium parvum*, and recently, *P. yoelli* has been shown to alter the expression of host lncRNAs (Menard et al., 2018; Amaral et al., 2020; Li et al., 2021a; Chen et al., 2022). Researchers are also endeavouring to identify parasite-derived lncRNAs and define their functions within the parasite and the host. One example from *Cryptosporidium* demonstrates that parasitic lncRNAs can modulate host gene expression, lncRNA Cdg7_Flc_0990 is delivered into host epithelial cells and recruited to specific host promoter regions with a histone methyltransferase (Wang et al., 2017). For malaria, research in parasite-derived lncRNAs has focused on *P. falciparum*, although some studies have identified lncRNAs in *P. vivax* (Boopathi et al., 2013; Kim et al., 2017).

1.5.1 *P. falciparum*-derived lncRNAs

The first evidence of parasite lncRNAs were antisense transcripts that appeared in cDNA analyses (Su et al., 1995; Kyes et al., 2003; Gunasekera et al., 2004; Ralph et al., 2005; Lu et al., 2007). Since then, thousands of putative lncRNAs have been identified in the parasite using many approaches such as serial analysis of gene expression, DNA microarrays, tiling arrays, northern blots and RNA-Seq (Chakrabarti et al., 2007; Raabe et al., 2010; Otto et al., 2010; Sorber et al., 2011; Broadbent et al., 2011; Liao et al., 2014; Siegel et al., 2014; Broadbent et al., 2015; Chappell et al., 2020; Yang et al., 2021). The annotation of lncRNAs in *P. falciparum* is ongoing, no definitive annotation has been established, and few transcripts have been verified *in vitro*.

Although different research groups have implemented differing classification and naming systems (described further in *Chapter 3*), most *P. falciparum* lncRNAs can be categorised by their position relative to genes. LncRNAs can be intergenic, meaning they do not reside within genes (coding or noncoding) on either strand. LncRNAs residing within genes can be within a gene transcript (sense lncRNA), antisense to a gene (aslncRNAs, natural antisense transcripts), transcribed from the intron (intronic lncRNA or intron-derived aslncRNAs) or proximal to a UTR (TSS/5′-associated, 3′-associated). LncRNAs can be generated from the coding sequence but not encode proteins; these lncRNAs are known as sense lncRNAs. Other naming conventions classify lncRNAs for their association with certain positions (lncRNA-TAREs overlapping telomere-associated repetitive elements, bidirectional lncRNAs driven by bidirectional promoters), predicted structure (circular RNAs or dsRNAs), or related to their discovery (nascent-lncRNAs) (Raabe et al., 2010; Yin et al., 2020; Alvarez et al., 2021). ncRNAs that are too short to be categorised under the strict definition of lncRNAs are referred to as ncRNAs, small ncRNAs or intermediate-size lncRNAs (is-lncRNAs) (Chakrabarti et al., 2007; Wei et al., 2014).

1.5.2 Characteristics specific to *P. falciparum* lncRNAs

LncRNAs in *P. falciparum* share many of the features of mammalian lncRNAs. On a genomic level, compared to coding genes, lncRNAs contain more repetitive sequences, decreased GC content and a scarcity of introns (Broadbent et al., 2015). Unlike human studies, lncRNAs appeared to be conserved within the genus (Otto et al., 2010). On a transcriptome level, similarly, the lncRNAs exhibit lower transcript expression and stability compared to coding genes (Broadbent et al., 2015). LncRNAs exhibit tightly-regulated stage-specific expression

across the IDC mediated by RNA Pol II (Siegel et al., 2014; Broadbent et al., 2015). No RNA Pol III-mediated transcription of lncRNAs has been observed, and there has not been an investigation of lncRNAs in mosquito or liver stages. Post-transcriptional processing has been investigated in a few lncRNAs, and 5′ capping and polyadenylation have been observed (Broadbent et al., 2011). Although some lncRNAs are not polyadenylated, such as the lncRNAs that are antisense to *var* genes (Epp et al., 2009). There is evidence that multi-exonic lncRNAs undergo splicing (Siegel et al., 2014; Broadbent et al., 2015). RNA fluorescence *in situ* hybridisation (FISH) has shown that *P. falciparum* lncRNAs can localise to the nucleus or cytoplasm. Some lncRNAs are diffused in the nucleus, while others form perinuclear foci with other proteins (Epp et al., 2009; Raabe et al., 2010; Sierra-Miranda et al., 2012; Heinberg et al., 2022). Global analysis of aslncRNAs revealed no nuclear enrichment suggesting that localisation is inconsistent within this type of lncRNA (Siegel et al., 2014).

1.5.3 LncRNA-mediated gene regulation in *P. falciparum*

Only a minority of lncRNAs have been associated with a potential role in the parasite, and few have been verified as biologically functional. Therefore, our understanding of the functionality of *P. falciparum* lncRNAs is limited and based on few individual lncRNAs.

Regarding *cis* regulation, studies have identified a positive correlation between intergenic lncRNAs and neighbouring genes (Liao et al., 2014; Broadbent et al., 2015). One proposed mechanism to facilitate these interactions is the presence of binding motifs; for example, lncRNA-TAREs contain the SPE2 motif, which interacts with ApiAP2 PfSIP2 (Broadbent et al., 2011). However, this motif was later found not to be essential for their function (Jing et al., 2018). Some lncRNAs are driven by bidirectional TSSs in a head-to-head conformation with a coding gene. Chappell et al. identified 337 mRNA-lncRNA transcript pairs that they classified as co-regulated due to correlated temporal expression (Chappell et al., 2020).

AslncRNAs, however, exhibit a bias towards a tail-to-tail conformation, often with initiation occurring in the 3′ UTR of the gene and termination occurring within the gene body (Broadbent et al., 2015). Tail-tail antisense-sense transcript pairs have been associated with a negative correlation in expression (Broadbent et al., 2015). However, the relationship between antisense-sense pairs remains unclear because some exhibit the opposite relationship. The expression of *var* genes is positively correlated with the expression of their aslncRNAs (Jiang et al., 2013; Amit-Avraham et al., 2015; Jing et al., 2018). To this end, other studies have concluded that there is no global effect for these pairs (Siegel et al., 2014).

Trans gene regulation has not been observed in *Plasmodium* lncRNAs however, it has observed in the RUF6 (GC-rich) family of ncRNAs. Using DNA and RNA FISH, Guizetti et al. determined that these ncRNAs act independently of their gene loci, colocalising to a specific perinuclear domain with the active *var* gene (Guizetti et al., 2016). Barcons-Simon et al. used dCas9-mediated gene knockdown of the RUF6 family to determine that RUF6 ncRNAs are clonally variant and involved in the regulation of *var* gene switching (Barcons-Simon et al., 2020). Fan et al. determined that the accumulation of RUF6 ncRNAs triggers the activation of the adjacent *var* gene via chromatin remodelling with H3K9ac marks (Fan et al., 2020). Furthermore, they demonstrated that RUF6 ncRNAs have global derepression effects on distant heterochromatin, which the Rrp6 (PF3D7_1449700) protein regulated. Diffendall et al. used a ChIRP-mass spectrometry to demonstrate that RUF6 ncRNAs also interact directly with RNA Pol II, nucleosome assembly proteins and DDX5 (PF3D7_1445900) (Diffendall et al., 2023).

Additional lncRNA structures have been predicted or detected. Using computational analysis, Broadbent et al. predicted 1381 circular RNAs in *P. falciparum* and proposed their potential function as a sponge for host miRNAs (Broadbent et al., 2015). Recently using duplex RNA-Seq, Alvarez et al. discovered dsRNA duplexes in the parasite. Some of these dsRNA duplexes contained lncRNAs, although the impact of these structures remains unknown (Alvarez et al., 2021).

1.5.4 Biological processes involving *P. falciparum* lncRNAs

LncRNAs have been associated with various potential biological processes in *P. falciparum* (Figure 1.6). Antisense transcripts were found to play a role in immune evasion via the regulation of *var* gene switching. The *var* multigene family encodes the PfEMP1 surface variant antigen, which is expressed on infected erythrocytes in a mutually-exclusive manner. In addition to the RUF6 ncRNA family described above, there are two other *var*-associated ncRNAs, which are transcribed from the highly-conserved *var* intron, which acts as a bidirectional promoter (Epp et al., 2009; Amit-Avraham et al., 2015). The first is a sense lncRNA, which is associated with *var* gene silencing and may play a role in epigenetic imprinting (Epp et al., 2009). The second is an aslncRNA which is expressed in active *var* genes and has been associated with *cis* activation of *var* genes (Amit-Avraham et al., 2015). The aslncRNA diffuses to a *var*-specific nuclear compartment, incorporates into the chromatin, localises to the *var* gene TSS and forms a DNA-RNA hybrid (Epp et al., 2009; Jiang et al., 2013; Heinberg et al., 2022). Heinberg et al. identified a conserved region

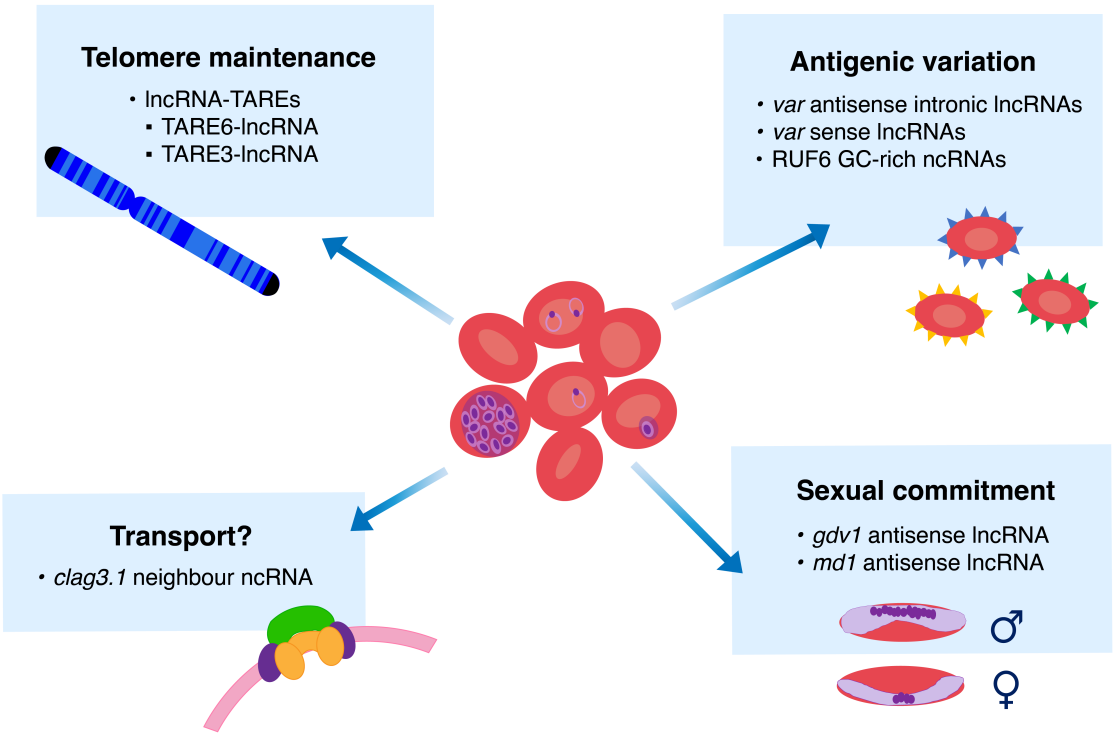


Fig. 1.6 Proposed biological functions of *P. falciparum* ncRNAs. LncRNAs have been associated with various biological processes, including telomere maintenance via the lncRNA-TAREs, antigenic variation via four types of lncRNAs and sexual commitment via two lncRNAs that are antisense to genes mediating gametocytogenesis. In addition, a lncRNA has been proposed to regulate a protein involved in solute transport.

within *var* aslncRNAs that encodes a stem and loop secondary structure, which binds RNA binding proteins that could facilitate *var* switching. Not all *var* genes are regulated in this way because the *var2csa* intron can be deleted and yet, still be activated and silenced (Bryant et al., 2017).

Another example of transcriptional control by lncRNAs is regulating sexual differentiation. *gdv1* (PF3D7_0935400) is an upstream activator of sexual commitment, and when expressed, the GDV1 protein evicts the epigenetic silencer HP1 from its specific loci (Filarsky et al., 2018). The expression of *gdv1* is negatively regulated by an aslncRNA (PF3D7_0935390) during blood stages. When the antisense locus is disrupted, the expression of *gdv1* is increased, leading to increased dissociation of HP1 from heterochromatin, consequently increasing the expression of *ap2-g*, a transcription factor that initiates sexual commitment (Filarsky et al., 2018). Deletions in other lncRNAs at this locus identified in clinical samples have revealed that adaptive selection occurs at this locus (Duffy et al., 2018). Recently, another aslncRNA with a role in determining sexual fate has been discovered in the *pfmd1* (PF3D7_1438800) locus (Gomes et al., 2022). Expression of *md1*-lncRNA is mediated by intron 1; however, the expression of *md1* and *md1*-lncRNA are mutually exclusive. When the authors abrogated *md1*-lncRNA expression by disrupting intron 1, the number of males was increased, suggesting *md1*-lncRNA may play a role in enhancing sex determination towards a male fate.

LncRNAs-TAREs transcribed from 22 chromosome ends have been proposed to be regulators of telomere maintenance and chromatin remodelling in DNA replication. These relatively GC-rich lncRNAs have been shown to exhibit coordinated expression with maximal expression occurring in the schizont stage (Broadbent et al., 2011). Sierra-Miranda et al. determined that there are two types of lncRNA-TARE transcripts: a lncRNA spanning telomere sequences and a TARE-3 element and a lncRNA composed of a TARE6 element (Sierra-Miranda et al., 2012). Among these, the TARE-6 lncRNA has been shown to localise to a perinuclear compartment with some telomeres and bind to histone H3 using a hairpin structure. The authors speculated that the lncRNA could act as a histone chaperone; however, this has not been further elucidated.

A lncRNA has also been identified that is antisense to *clag3.1* (PF3D7_0302500), part of a gene family involved in solute transport. The *clag3.1* and *clag3.2* (PF3D7_0302200) proteins exhibit mutual exclusive expression. It has been proposed that there could be a role for the

lncRNA in regulating their expression (Rovira-Graells et al., 2015). However, this has not yet been investigated.

Although the aforementioned studies provide insights into how lncRNAs may regulate biological processes, many recent reviews highlight that these examples represent only a small subset of the thousands of lncRNAs identified in *P. falciparum* (Yeoh et al., 2019; Li et al., 2020; Lodde et al., 2022; Simantov et al., 2022). Perturbations of more lncRNAs are needed to assess the scope of the role of lncRNAs in the regulation of the transcriptome.

1.6 Gene perturbation in *P. falciparum*

Genetic perturbation has provided major biological insights into gene function in *P. falciparum*. Perturbations such as an alteration of a gene *in vitro* or *in vivo* can allow for the determination of its essentiality and characterisation of its function. The advent of CRISPR-based gene editing and its application into *P. falciparum* has drastically improved this approach. Since then, advances have been made in these existing tools, and additional tools have been applied to the parasite. 30% of coding genes and most predicted ncRNAs still have unknown functions; therefore, the continued development of these tools is needed to facilitate the characterisation of these genomic features (Cárdenas et al., 2022).

1.6.1 Early gene editing approaches

Early approaches to gene manipulation were promising; however, they were limited in their application due to various factors, including extensive molecular cloning, slow efficiency, lack of temporal control, and undesired off-target effects (Meissner et al., 2007). Forward-genetics experiments were completed using transposon-mediated random mutagenesis such as Mini-Tn5 (Sakamoto et al., 2005) and *piggyBac* transposon (Balu et al., 2005). Reverse-genetics experiments were conducted using site-specific recombination to facilitate gene excision or allelic exchange and tetracycline-based regulation systems that enabled inducible gene regulation (Carvalho et al., 2004; Meissner et al., 2005). These techniques were followed by zinc-finger nucleases (Straimer et al., 2012; Singer et al., 2015; Moraes Barros et al., 2015), which could generate double-stranded breaks (DSB) at specific sites in the genome. Donors were provided to facilitate gene editing and support the repair of the DSB. As *Plasmodium* spp. lack non-homologous end joining (NHEJ) DNA repair, these DSBs are repaired by homology-directed repair (HDR) or, more rarely, by microhomology-mediated end joining

(Gardner et al., 2002; Singer et al., 2015). Without NHEJ, which is error-prone and likely to cause indel mutations, DSBs are not expected to generate mutations without a donor repair template.

Zinc-finger nucleases were soon made redundant by CRISPR-based tools, which provided simplified targeting design, reliable cleavage of the target DNA, straightforward off-target prediction and the potential for multiplexing. CRISPR (clustered interspaced palindromic repeat elements) technology was adapted from a rudimentary immune defence mechanism in bacteria and archaea to a powerful gene editing tool. A library of foreign sequences is maintained, and these sequences are transcribed as RNAs, which guide a CRISPR-associated nuclease to their complementary sequences, i.e. bacteriophage DNA or RNA, for targeted degradation. This system was adapted into a three-component molecular tool for gene editing consisting of a Cas9 nuclease that makes DSB, a guide RNA (gRNA) complementary to the target sequence that guides the Cas enzyme and a donor sequence to provide a template for repair and modification. Although many species contain CRISPR-Cas systems and there are many types of systems, most applications of CRISPR harness Cas9 from *Streptococcus pyogenes*, a type II CRISPR system.

1.6.2 CRISPR-based gene editing

CRISPR-Cas9 was first applied to *P. falciparum* and *P. yoelii* (Ghorbal et al., 2014; Wagner et al., 2014; Zhang et al., 2014) in 2014, and has since been applied to *P. knowlesi* and *P. berghei* (Mohring et al., 2019; Shinzawa et al., 2020). The system can facilitate whole gene knockout through excision, the insertion of selection markers, single-base editing, or altering the regulation of expression through promoter targeting (Figure 1.7A). The first tools developed in *P. falciparum* were two-vector systems (Ghorbal et al., 2014; Wagner et al., 2014). The systems differ in how they promote the transcription of the gRNA; Wagner et al. used a T7 phage promoter with a T7 RNA polymerase, and Ghorbal et al. used a parasite U6 snRNA cassette designed to recruit RNA Pol III from the parasite. Other versions of the CRISPR-Cas9 system have since been developed, including a suicide-rescue dual-vector approach (Lu et al., 2016) and a single-vector system (Ng and Fidock, 2019; Lee et al., 2019; Adjalley and Lee, 2022). A plasmid-free delivery system has also been established using ribonucleoproteins (recombinant Cas9 complexed with gRNAs) and single-strand oligodeoxynucleotide donors (Crawford et al., 2017). The need for Cas9 delivery has also been entirely circumvented by integrating the Cas9 sequence into the *P. falciparum* genome. Nishi et al. integrated Cas9 into the *karhp* locus in the 3D7 strain (Nishi et al.,

2021). The integration did not appear to cause genomic instability, and the endogenous Cas9 demonstrated robust gene editing.

Cas9 has been used to target many genes in *P. falciparum* because there are minimal restrictions on the target sequence. Cas9 gRNAs targets must be a 20-nucleotide sequence in the genome, which is followed by a 5′-NGG-3′ protospacer adjacent motif (PAM), where N is any nucleotide followed by two guanine nucleotides. If gene editing is employed, the gRNA should target as close to the mutation site as possible. To avoid off-target effects, the target sequence must be unique and lack significant shared homology with other sites in the genome. gRNAs are designed using predictive software based on gRNA activity in large-scale gene-targeting screens in other organisms. As described above, the *P. falciparum* genome is highly AT-rich compared to mammalian genomes, and the accuracy of these predictions for *P. falciparum* perturbations remain uncertain. Multiple gRNAs need to be designed and tested for certain targets to identify an efficient and specific gRNA. Multiplexing Cas9 gRNAs has been achieved in *P. yoelii* through the uses of a ribozyme cassette for gRNA expression (Walker and Lindner, 2019) and in *P. falciparum* with an integrated system Cas9 expression system (Zhao et al., 2020). Some efforts for *Plasmodium*-specific resources have been made, including a database of gRNAs predicted to be suitable for use *P. falciparum* (Ribeiro et al., 2018) and an *in silico* tool for automated design of donor vectors compatible with pre-specified vectors (Cárdenas et al., 2022). The donor repair templates require at least 250-1000 bp homology to the site of the DSB (Ribeiro et al., 2018; Lee et al., 2019). Silent shield mutations are placed at the seed regions (nucleotides closest to the PAM site) in the target sequence contained in the donor to prevent its excision by the Cas9.

1.6.3 Challenges with CRISPR-Cas9

Firstly, the knockout approach cannot be used to study essential genes whose disruption leads to parasite death. It would be difficult to discern if the gene was successfully knocked out and essential or if the parasites did not recrudesce for other reasons. Therefore, researchers have developed conditional knockout systems using dimerisable cre recombinase Figure 1.7B) (Collins et al., 2013; Jones et al., 2016). The cre recombinase is split into two portions that only dimerise when induced by rapamycin introduced in culture. Once activated, the cre recombinase selectively completes site-specific recombination of sequences flanked with *lox*P sites, resulting in their excision or flipping. Therefore, *lox*P sites can be inserted on either side of a target genomic loci or, if used in conjunction with a CRISPR approach, can be used to conditionally activate the CRISPR system (Knuepfer et al., 2017). Additional

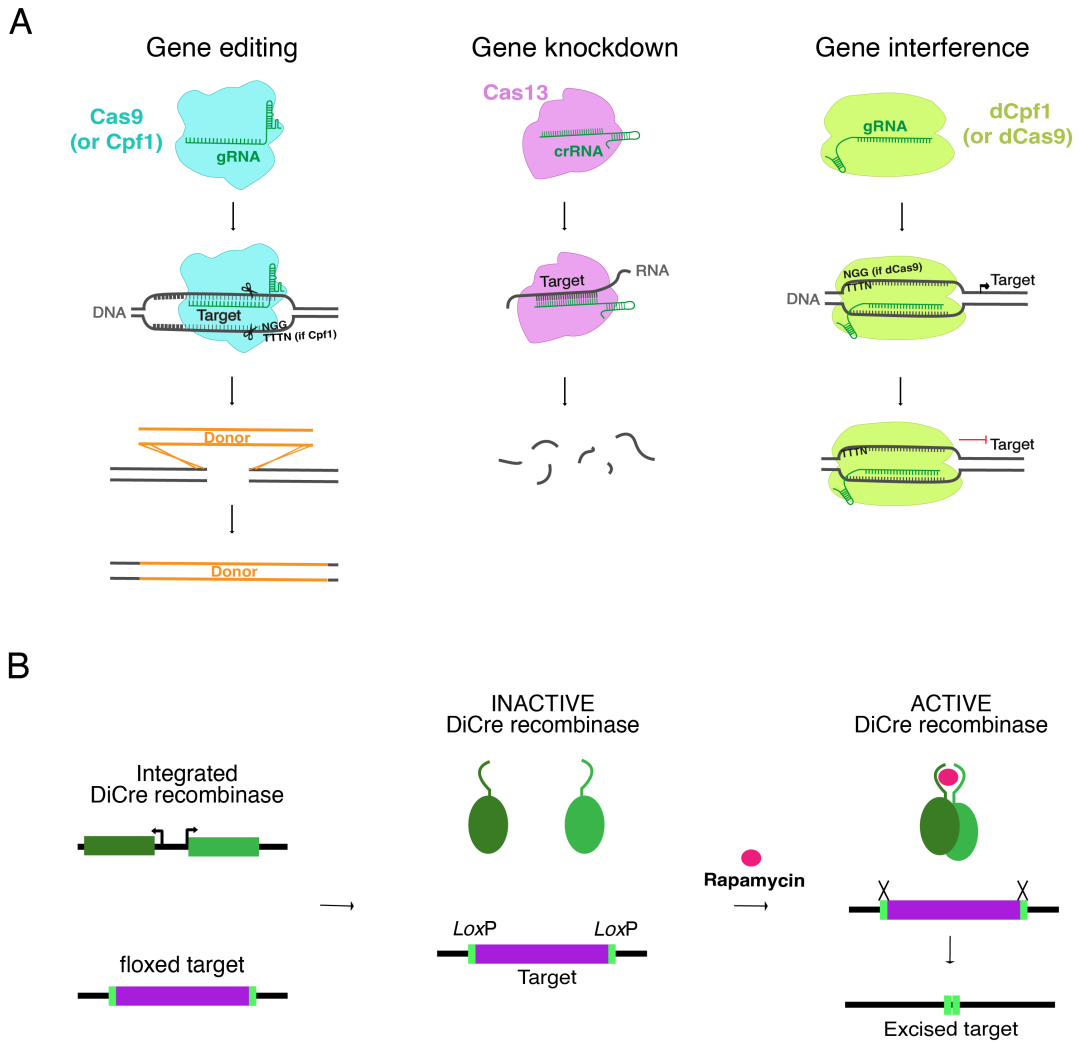


Fig. 1.7 CRISPR approaches to gene perturbation. (**A**) In CRISPR gene editing, a Cas enzyme is guided by an RNA (gRNA) to a complementary target sequence and forms a double-stranded DNA break. To facilitate repair, a donor sequence can be provided with homology regions. In gene knockdown, Cas13 enzymes create breaks in target RNA, leading to their degradation. Gene interference is mediated by dead Cas enzymes, which block the initiation or continuation of transcription. (*B*) Inducible gene editing systems are enabled by the DiCre recombinase system. The system is integrated into the genome, and the cre enzymes are expressed endogenously. *Lox*P sites are flanked on the target, which upon the addition of rapamycin, dimerises and activates the cre enzymes to facilitate excision or flipping of the target.

targeting methods that enable the study of essential genes are discussed below, such as gene interference, mRNA destabilising and protein-targeting.

Secondly, certain sequences are more challenging to target with CRISPR-Cas9, such as repetitive sequences or gene families, where unique gRNAs cannot be identified or super AT-rich regions, which lack suitable PAM sites. More AT-rich PAM sites have been identified in an alternative class II enzyme, Cpf1 (Cas12a) (Zetsche et al., 2015; Chen et al., 2018). The Cpf1 system also differs from Cas9 by making DSBs with sticky ends instead of blunt ends, not requiring trans-activation CRISPR RNA (crRNA), and processing its crRNA array. Cpf1 from *Acidaminococcus* sp., which contains a "TTTV" PAM site (where V is an A, G or C), was adapted for gene editing in *P. falciparum* by two research groups (Zhao et al., 2020; Nessel et al., 2020). The increase in AT content in the PAM site increases the number of possible target sites in the genome by nearly three-fold (Zhao et al., 2020). Nessel et al. also adapted Cpf1 from *Lachnospiraceae bacterium ND2006*, which has a "TTTN" PAM site. Both enzymes demonstrated efficient genome editing in both studies.

Thirdly, there are undesired consequences of CRISPR-Cas editing that can occur. Although off-target effects have not been reported in the literature for CRISPR-Cas9 targeting in *P. falciparum*, the methods for quantifying off-target effects are not sufficient, and currently, only whole-genome sequencing can identify other sites that were also altered. Another adverse effect that can occur when using vector-based delivery to target non-essential genes, is that the entire vector can be inserted into the break site. However, this can be easily checked using PCR amplification of the insertion site. One of the pitfalls of CRISPR-Cas9 is that it can generate unexpected large deletions, sometimes truncating chromosomes. Bryant et al. observed this adverse effect in *P. falciparum* when Cas9 editing the *var2csa* intron, which is located in the subtelomeric region (Bryant et al., 2017). Interestingly, CRISPR-Cas9 has been recently used to split chromosomes to study chromosome spacial arrangement and their role in *P. berghei* (Addo-Gyan et al., 2022).

1.6.4 CRISPR-mediated gene interference and activation

Another way Cas9 has been utilised for gene disruption is for gene interference by using a catalytically inactivated (dead) form of the enzyme (Figure 1.7A) (Qi et al., 2013; Larson et al., 2013). In this case, the CRISPR-Cas9 system localises to the target and binds the DNA but does not induce a DSB: instead, it physically interferes with transcription by sterically blocking the binding or progression of RNA polymerase. Furthermore, the Cas9 can be

fused to other proteins to unlock additional capabilities such as epigenetic repression and activation. CRISPR-dCas9 was adapted for use in *P. falciparum* by Xiao et al. by expressing recombinant dCas9 that harbours two mutations in the catalytic domain (Xiao et al., 2019). This technique has been used to knockdown gene expression in *P. falciparum* and *P. yoeilii* (Baumgarten et al., 2019; Walker and Lindner, 2019) and interfere with features associated with *var* regulation: introns, promoters and GC-rich ncRNAs (Bryant et al., 2020; Barcons-Simon et al., 2020). Xiao et al. also created fusions to the histone deacetylase Sir2A and histone acetyltransferase GCN5, which they used to repress *eba-174* and *pfset1* expression *in vitro* or contrastingly, activate *rh4* expression, respectfully (Xiao et al., 2019). This approach has recently been converted into two inducible systems by integrating the TetR-DOZI or DiCre/*lox*P systems (Liang et al., 2022). Although the TetR-DOZI system demonstrated some leaky expression and therefore, the leak-free DiCre systems (one with GCN5 and one with Sir2a) were used to activate or repress gene expression in eight target genes, including six essential genes. Some CRISPR enzymes, such as Cas13, can be used to target RNA directly for their specific degradation; however, these systems have not yet been applied to *P. falciparum* (Figure 1.7A).

1.6.5 Other approaches to perturbation

Other techniques for perturbation target expression at the transcriptional and translational levels. Most of these strategies predate CRISPR: some have been replaced by CRISPR, and some are still used today, sometimes integrating CRISPR into their systems. For example, the *glmS* ribozyme from bacteria has been co-opted as a method of inducible mRNA degradation in *P. falciparum* (Prommana et al., 2013). The *glmS* ribozyme sequence is inserted in the 3′ UTR of the gene of interest. When glucosamine is added, the *glmS* ribozyme self-cleaves, leading to mRNA degradation and expression knockdown. Prommana et al. knocked down *pfdhfr-ts* expression, and this strategy is still being used, recently applied to *pflpl20* and *pfgm1* (Prommana et al., 2013; Sheokand et al., 2021; Tehlan et al., 2022). Also, numerous different types of small molecules have been designed to bind specific mRNAs directly and promote their degradation, including morpholino oligomers (Augagneur et al., 2012, 2013; Garg et al., 2015), phosphorothioate antisense oligodeoxynucleotides (PS-ODNs) (Rapaport et al., 1992; Barker et al., 1996, 1998; Wanidworanun et al., 1999; Noonpakdee et al., 2003), 2'-OMe modified PS-ODNs (Razavi Vakhshourpour et al., 2022) and peptide nucleic acids (Kolevzon et al., 2014). However, these systems have had limited application due to consequences such

as non-negligible off-target effects, efficiency and challenges with scalability (Stojic et al., 2018).

CRISPR systems can facilitate the generation of dual-aptamer systems that inhibit translation in specific mRNAs. An aptamer array is inserted into a gene after the stop codon, which is transcribed. In the absence of ATc, the aptamer array binds Tet-DOZI, thereby inhibiting the translation of the mRNA. Moreover, this inhibition can be titrated by the addition of ATc. The TET-DOZI system was adapted to *P. falciparum* by (Ganesan et al., 2016), which is widely used today with the support of CRISPR (Spillman et al., 2017; Rajaram et al., 2020; Nessel et al., 2020; Nasamu et al., 2021).

Proteins can also be targeted directly. Protein destabilisation has been achieved in *P. falciparum* using ddFKBP (Armstrong and Goldberg, 2007). The destabilisation can be rescued by adding Shld1 (an analogue of rapamycin). Another strategy for perturbation that has been adapted to *P. falciparum* is knocksideways (anchor away), where proteins functions are disrupted through mislocalisation (Birnbaum et al., 2017). For example, a protein known to localise to the plasma membrane is mislocalised to the nucleus or vice versa. This process is permitted by selection-linked integration, where a plasmid is designed with homology to the target gene and contains FKBP, GFP and a resistance marker separated by a skip peptide. This plasmid lacks a promoter; therefore, it must be integrated for the parasite to become resistant. However, due to the skip peptide, the resistance protein does not attach to the target. The expression of GFP allows protein localisation to be determined, and the dimerisation of FKBP to FRB facilitates the mislocalisation, where the FRB is targeted to specific anchor sites such as the nucleus or parasite plasma membrane.

1.7 Summary of the thesis aims and chapters

Due to limited research and the lack of necessary experimental tools, our understanding of the role of lncRNAs in the malaria-causing *P. falciparum* parasite remains largely unelucidated. In this thesis, I endeavour to improve the understanding of the role of lncRNAs in regulating the *P. falciparum* transcriptome by enabling the *in vitro* study of lncRNAs. In the first research chapter (*Chapter 3*), I generated an improved annotation of lncRNAs in the *P. falciparum* transcriptome using manual curation of long-read sequencing data and existing publicly-available datasets. The updated annotation uncovered novel lncRNAs, validated existing annotations, and identified incorrect annotations. In the second research chapter

(*Chapter 4*), I explored CRISPR-based tools to disrupt lncRNAs in *P. falciparum in vitro*, including applying novel tools to *P. falciparum*. In the final research chapter (*Chapter 5*), I used tools explored in the preceding chapters to disrupt specific lncRNAs *in vitro* and characterise the lncRNA-disrupted *P. falciparum* lines.

Chapter 2

General Methodology

2.1 Parasite lines

The parasite cell lines used in this work are laboratory-adapted (Table 2.1). These strains are widely used for *in vitro* experimentation due to their ability to grow well in a lab environment. Additional lines have been generated from these strains through cloning and genetic modification.

2.2 Parasite handling

All parasite handling was completed in a class II microbiology safety cabinet in a derogated Containment Level 3 laboratory.

Table 2.1 Lab-adapted *P. falciparum* lines used in this work

Line	Region	Origin	Source
NF-54	Africa	Obtained from a patient in the Netherlands	(Ponnudurai et al., 1981)
NF-54^{attB}	Africa	Genetically-modified NF54	(Adjalley et al., 2011)
NF-54^{camEGFP}	Africa	Genetically-modified NF54	Marcus Lee
3D7	Africa	Clone from NF54	(Walliker et al., 1987)
3D7^{attB}	Africa	Genetically-modified 3D7	(Nkrumah et al., 2006)
3D7 DiCre	Africa	Genetically-modified 3D7	(Collins et al., 2013)
3D7 DiCre-attB	Africa	Genetically-modified 3D7	(Knuepfer et al., 2017)
Dd2	Southeast Asia	Clone from W2-MEF	(Oduola et al., 1988) (Guinet et al., 1996)
Dd2^{bipGFP}	Southeast Asia	Genetically-modified from Dd2	(Baragaña et al., 2015)

2.2.1 Cell culture

Asexual blood-stage *P. falciparum* parasites were cultivated in complete media (CM, recipe in Table 2.2) with 2.5-3% human red blood cells (RBCs) (Trager and Jensen, 1976). Leukocyte-depleted RBCs were obtained weekly from anonymous donors from National Health Services Blood and Transplant. Their use was in accordance with relevant guidelines and regulations, with approval from the NHS Cambridgeshire Research Ethics Committee and the Wellcome Sanger Institute Human Materials and Data Management Committee. O+ blood was used when available because it is compatible with any serum type and some *P. falciparum* parasites show a preference for this blood type in culture (Read and Hyde, 1993; Theron et al., 2018). Before use, blood was washed twice with CM and diluted to 50% haematocrit (the proportion of red blood cells in a volume). An anticoagulant, citrate-phosphate-dextrose with adenine (Sigma-Aldrich), was added at 10% to prolong RBC integrity. Parasites were maintained in continuous culture with 2-3% haematocrit and incubated in a gaseous environment of 3% CO_2, 1% O_2 and 96% N_2 at 37°C.

2.2.2 Enumeration

Light microscopy

The quantification of *P. falciparum* in culture is described by the term "parasitaemia", which is the percentage of RBCs that are parasitised. The parasitaemia and distribution of life stages in cultures were monitored regularly using light microscopy, and cultures were maintained at 0.5-5% parasitaemia unless stated otherwise. For light microscopy, a thin blood smear was prepared with 2μL of infected RBCs. Smears were fixed in 100% methanol and stained in 10% Giemsa (Sigma-Aldrich) for 10 minutes, which stains DNA and, therefore, only parasites (free-living and intraerythrocytic) as RBCs are anucleate. Slides were visualised with 100x objective on a compound microscope with oil immersion. At least 1000 cells were counted, and parasitaemia was calculated by dividing infected RBCs by total RBCs and multiplying by 100. The distribution of *P. falciparum* intraerythrocytic stages was recorded (i.e. rings, trophozoites and schizonts).

Flow cytometry

When the precise measurement was required, flow cytometry was used to quantify parasitaemia by staining with SYBR® Green I (Thermo Fisher), a green-fluorescent dye that stains double-stranded DNA and MitoTracker^{TM} Deep Red FM (MitoDR, Thermo Fisher),

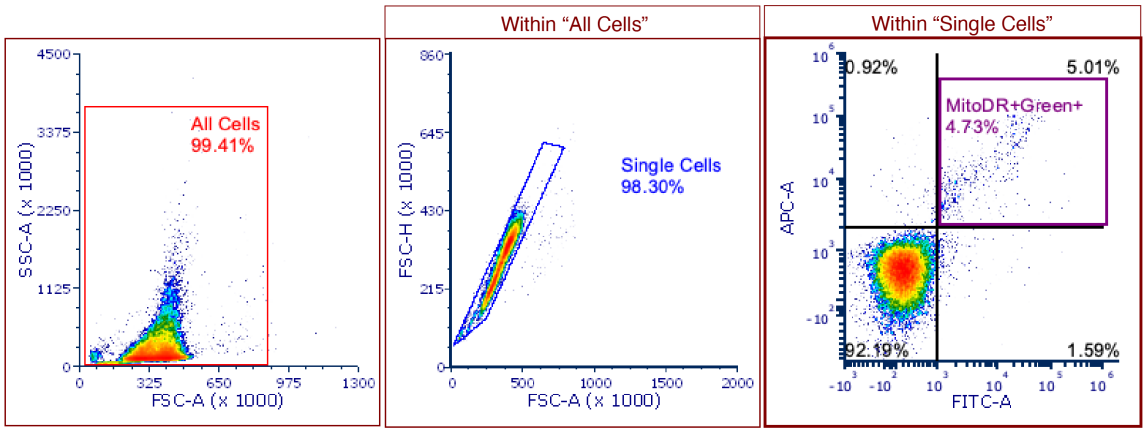


Fig. 2.1 Schematic of gating for two-colour flow cytometry. The flow cytometry events were first gated using the forward scatter area vs. side scatter area. Single cells were then obtained from this population using forward scatter area vs. forward height. Fluorescence within this single cell population was then quantified with FITC area (representing green fluorescence or SYBR® green) vs. APC area (representing MitoTracker^{TM} Deep Red).

a red-fluorescent dye that stains mitochondria. In cases where parasites expressing green fluorescence were quantified, only MitoDR^{TM} was used. 3-5μL of culture was added to 200nM MitoDR^{TM} solution (in NaCl 0.9%/Dextrose 0.2%), and SYBR® Green (1X, Thermo Fisher) prepared in phosphate-buffered saline (PBS) and incubated for 30 minutes in the dark at 37°C. After a 1/40 dilution in PBS, parasites were analysed with a flow cytometer (CytoFLEX S, Beckman Coulter), and at least 20 000 events were recorded for each experiment. The blue 488nm laser was used to detect green fluorescence, and the red 638nm laser was used to detect MitoDR. Quantification of green-fluorescent+ (GFP or mNeonGreen) and MitoDR+ cells was performed using FlowJo (v10) or FCS Express (v7). The gating strategy is described in Figure 2.1.

2.2.3 Synchronisation

In synchronous cultures, parasites transition through the stages of IDC at the same time. Both synchronous and asynchronous infections have been observed in *P. falciparum*-infected patients (Hawking et al., 1968). In addition to mimicking *in vivo* infection, inducing synchronicity can also facilitate experiments in which specific stages are required. Synchronisation was completed using two methods: sorbitol ring enrichment to isolate early stages and Percoll gradient separation to enrich for late stages.

Ring enrichment

Sorbitol bursts RBCs containing trophozoites and schizonts, thereby enriching for RBCs containing ring-stages (Lambros and Vanderberg, 1979). Ring-staged cultures were centrifuged (2000RPM, 5 minutes), resuspended using 0.5X culture volume of 5% D-sorbitol (Sigma-Aldrich, filtered with a 0.2μM filter) and incubated for 10 minutes at 37°C. The pellet was centrifuged, washed with 10mL of CM and centrifuged again. The pellet was resuspended in CM and RBCs returning the culture to 2-3% haematocrit.

Schizont isolation

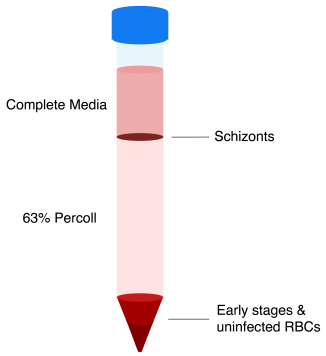


Fig. 2.2 Schematic of 63% Percoll gradient. Three layers form on top of a pellet containing early-stage parasites and uninfected RBCs. The top layer is CM, the middle layer is purified schizonts, and the bottom layer is Percoll®.

Percoll® can separate RBCs containing schizonts from other stages via a density gradient (Kramer et al., 1982). 63% Percoll® (Cytiva) was prepared in RPMI with 7% 10X PBS and filtered. Cultures containing schizonts were centrifuged (1600g, 3 minutes) and resuspended in RPMI to 40% haematocrit. The culture was carefully layered on 7.5mL of pre-warmed 63% Percoll® in a 15mL centrifuge tube and centrifuged (1300g, 11 minutes, break 0). The top layer was aspirated then the middle layer was transferred into a 50mL centrifuge tube with 20mL of pre-warmed CM (Figure 2.2). The resulting culture was centrifuged (600g, 3 minutes) and resuspended with CM and RBCs to return the culture to 2-3% haematocrit.

2.2.4 Parasite storage

P. falciparum cultures can be frozen and preserved for long-term cold storage (Radfar et al., 2009; Scherf et al., 2013). To be suitable for freezing, cultures had to be at least 3% parasitaemia in 5-10mL and mostly ring-staged. Cultures were centrifuged (1600RPM, 3 minutes) and resuspended dropwise in 2X pellet volume of filter-sterilised glycerolyte solution 2.2 and transferred into a cryovial and stored in a -80°C freezer. Cryovials were thawed at room temperature (RT) and transferred into a 50mL centrifuge tube. 500μL of 12% NaCl (Sigma-Aldrich) solution followed by 5mL of 1.6% NaCl solution was slowly added drop-wise while swirling the tube gently. The tube was incubated at RT for 5 minutes and centrifuged (1600RPM, 3 minutes). The supernatant was

discarded, the pellet disturbed, and 5mL of 0.9%NaCl/0.2%dextrose solution was added dropwise while swirling. The tube was incubated at RT for 5 minutes and centrifuged (1600 RPM, 3 minutes). The tube was washed with CM, centrifuged (1600RPM, 3 minutes) and resuspended in CM and fresh blood to re-establish 2-3% haematocrit.

2.2.5 Parasite cloning

Cloning is a technique used to obtain homogeneous (clonal) cultures from a mixed culture. Cultures are grown from a single parasite. Cultures were diluted and seeded into a 96-well plate to obtain 0.5-0.8 parasites per well in 200μL at 3% haematocrit. Cultures were fed once a week with CM and fresh RBCs at 0.4% haematocrit. Parasite growth was monitored every 2 days beginning on day 14. 30μL of SYBR® Green I (2X, Thermo Fisher) was mixed with 30μL culture and incubated at 37°C for 30 minutes, then read with a microplate reader using a 485nm excitation filter and 523nm emission filter.

2.2.6 Transfection

Transfection is a method to transfer nucleic acids into *P. falciparum* parasites. By passing an electric current through the culture (electroporation), the plasma membranes of the RBC, parasitophorous vacuole and parasite become porous, which leads to the uptake of materials in the media. This protocol is modified from a previously described method (Wu et al., 1995; Fidock and Wellems, 1997), which utilises a buffer with a pH 7.6 called cytomix (Table 2.2) (van den Hoff et al., 1992).

Cultures containing 5% ring-stage parasites were harvested in 15mL centrifuge tubes (3-5mL of culture per transfection), spun (1500rpm, 3 minutes, RT), washed with cytomix and spun again (1500rpm, 3 minutes, RT). 100μL of the pelleted cells was mixed with 50-100 μg of plasmid and cytomix to a final volume of 500μL and transferred to a 2mm cuvette. The cuvettes were electroporated (Gene Pulser Xcell, Bio-Rad) at 0.31kV for 10 milliseconds. Immediately, 1mL of pre-warmed CM was added to the cuvette. Then to aid in recovery, the contents were transferred to a 15mL tube, an additional 4mL of CM was added, and the culture was incubated (37°C, 1 hour). To remove debris, cultures were spun (1500rpm, 3 minutes, RT) and resuspended in 5mL of CM and 150μL of RBCs. Post-transfection cultures were fed fresh media every day for 7 days, with drug selection (if applicable to the construct) beginning on the first day following transfection and continuing as required. On day 5 or 6, 4/5ths of the culture was discarded and replaced with fresh media and RBCs, as this has

been observed to improve recrudescence in the lab. Cultures were then fed every second or third day, and once a week, 1/5th of the culture was discarded and replaced. Cultures were monitored with microscopy at least once a week to check for recrudesced parasites.

2.3 Molecular cloning

2.3.1 Plasmid propagation and storage

Plasmids were propagated in competent *E. coli* (XL-10 Gold Ultra-competent cells, NEB) using bacterial transformation. 2-50ng of plasmid was incubated with 50μL of competent cells for 20 minutes on ice and then heat-shocked for 30 seconds at 42°C. The cells were cooled on ice for 1 minute and recovered in 450μL of SOC media for 30 minutes at 37°C with shaking. 100μL of the transformation was plated and spread on LB agar plates containing ampicillin (100μg/mL) and incubated overnight at 37°C. The following day, individual colonies were inoculated into 2-200mL of LB broth supplemented with ampicillin (100μg/mL) and incubated overnight (37°C with shaking). Bacterial cultures were harvested the next day using centrifugation (30 minutes, 4000-10000g, 4°C). Plasmids were isolated from bacterial cultures using mini or midi preparation (prep) kits (Macherey-Nagel). When small quantities of plasmid were required, mini preps were used, while when large quantities were required, midi preps were completed. For molecular cloning purposes, preps were resuspended in elution buffer or water, although for transfections, elutions were resuspended in cytomix. Plasmid concentrations were measured with a NanoDrop spectrophotometer (Thermo Fisher). Plasmids were stored at -20°C. For long-term storage, glycerol stocks were made from bacterial cultures. 700μL of culture was combined with 300μL of glycerol (50%) in a cryovial and stored at -80°C.

2.3.2 Polymerase chain reaction

Polymerase chain reaction (PCR) is a means of amplifying fragments of DNA using sequence-specific primers, DNA polymerase, and thermocycling. DNA fragments can be used for molecular cloning, such as for ligation reactions and to confirm sequence composition by sequencing or length of the sequence by separation on agarose. PCR invokes thermocycling to denature DNA (95°C), anneal primers to specific loci (55-70°C depending on the primers' T_m), and extend the sequence with polymerases (72°C) to generate complementary sequences which amplify the sequences in subsequent cycles. For most reactions, to accommodate the AT-rich *P. falciparum* genome, the extension temperature was lowered to 62 °C. Polymerases

have varying levels of fidelity and proofreading activity, which were appropriate depending on the requirement of the PCR: GoTaq® Green (Promega) was used to check plasmids propagated in bacterial colonies, CloneAmp^{TM} HiFi (Takara) was used to check parasite pellets genomic DNA, and Kapa HiFi^{TM} (2X, Roche) was used for high-quality PCR for cloning. The temperature cycling timing varied depending on the enzyme used and the length of the desired product.

2.3.3 PCR purification

PCR products were purified with PCR purification kits (Macherey-Nagel) or when needed, separated with gel electrophoresis. Gel electrophoresis enabled the visualisation of DNA bands and the separation of these bands by size. 1-2% agarose gels were prepared by dissolving agarose (Sigma-Aldrich) into TAE (tris-acetate-ethylenediaminetetraacetic acid) buffer, adding either ethidium bromide or SYBR^{TM} Safe gel stain to stain the DNA and pouring the solution into a tray with a ladder to set. The trays were submerged in 1X TAE buffer in an electrophoresis chamber, and the DNA and 1kB DNA ladder (HyperLadder, Meridian Bioscience) was loaded into the wells (5-30μL). Gels were run at 210V for 0.5-2 hours and observed under UV light using a gel reader (iBright 1500, Thermo Fisher). When necessary, bands were excised, dissolved and purified using a gel purification kit (Macherey-Nagel).

2.3.4 Ligation

To modify cloning vectors, fragments of DNA were prepared using restriction digests and PCR, and then combined using ligation. Vectors were digested with the appropriate restriction enzymes and, if required, isolated using gel electrophoresis and gel extraction. Vectors were also dephosphorylated by adding 2μL shrimp alkaline phosphatase (rSAP) and incubating at 37°C for 30 minutes. T4 DNA ligase (Promega) with 10X ligase buffer (Promega) was used to ligate insert and vector DNA fragments together. The proportion of insert to vector DNA was determined from the NEB website (https://nebiocalculator.neb.com/#!/ligation). The reaction was incubated for 10 minutes - 1 hour and then transformed into competent cells and incubated overnight.

Annealing oligonucleotides

Before inserting oligonucleotides (oligos) like gRNAs into a plasmid, the forward and reverse oligos must be annealed for efficient ligation. Both oligos (100μM) were annealed with T4 Polynucleotide Kinase (NEB) in 10X ligation buffer and water. The reaction was incubated in a thermocycler at 37°C for 30 minutes, then 94°C for 5 minutes and ramped down to 25°C at 5°C/min.

Gibson assembly

Gibson ligation was used when multiple fragments needed to be ligated or when basic ligation failed. Pre-digested vector and the insert(s) were combined with 2X NEBuilder® HiFi mix (NEB) containing T5 exonuclease, Phusion DNA polymerase and Taq DNA ligase. The NEB calculator was used to calculate the amount of vector and inserts using a 2:1 insert:vector mass ratio for 2-3 fragments and 1:1 for 4-6 fragments. The reaction was incubated for 50 °C for 15 minutes or 1 hour if >3 fragments and transformed using bacterial transformation.

2.3.5 Sequence verification

Gel electrophoresis can be used to verify the size of a sequence fragment and, when completed in combination with restriction digest, can confirm the composition of the plasmid. Restriction enzyme reactions were prepared in a volume of 10μL with CutSmart® buffer (NEB), enzymes, plasmid and nuclease-free water. The quantities and conditions of restriction digests were determined using NEB (https://nebcloner.neb.com/#!/redigest), although often a 37°C incubation for 1 hour to overnight. To verify sequences at the base level, DNA fragments or plasmids were sent to GATC for Sanger sequencing with primers. Sequences were analysed using Lasergene (DNASTAR), Benchling or SnapGene (Benchling, 2020; SnapGene, 2022).

2.4 Genotyping

2.4.1 Genomic DNA extraction

0.1% saponin was added to 10mL of culture in a 15mL centrifuge tube, incubated for 5-10 minutes at RT and spun (3000g, 3 minutes). The supernatant was removed and the pellet was washed with 10mL of PBS and then 1.5mL of PBS and transferred to a microcentrifuge tube. The solution was spun (6000g, 3 minutes) and the supernatant was removed and stored

at -20°C. If purified DNA was required, a DNAeasy kit (Qiagen) was used. Briefly, the pellet was resuspended in 200μL of PBS and mixed by inversion with 20μL Proteinase K and 200μL of buffer AL. The mixture was incubated at 70°C for 10 minutes then mixed 200μL of 100% EtOH and transferred to a spin column. Spun (8000rpm, 1 min), washed with 500μL of buffer AW1, and 500μL of buffer AW2, transferred to a tube and eluted with 50μL AE buffer. Sequence verification was completed as described above.

2.5 Phenotyping

2.5.1 Expression

RNA isolation and purification

10mL of culture at 5-15% parasitemia was centrifuged (1800rpm, 4 minutes). The pellet was resuspended in 10-20X pellet volume of pre-warmed TRIzol^{TM} reagent (Thermo Fisher), shaken vigorously and incubated for 5 minutes at 37°C to facilitate cell lysis. The mixture was transferred into phase-lock gel tubes (Quantabio). To promote phase separation, chloroform (Sigma-Aldrich) was added at a 1:5 chloroform:TRIzol^{TM} ratio and tubes were mixed by inverting and incubated at RT for 3 minutes. The phases were separated by centrifugation (12000g, 15 minutes, 4°C, brake=0). The tubes were placed on ice and the upper aqueous layer was transferred into microcentrifuge tubes. An RNeasy mini kit with on-column DNase treatment (Qiagen) was used to further purify the RNA. Briefly, 1 volume of 100% ethanol was added to the isolated RNA, mixed gently, and then loaded into a spin column. The column was spun (>8000g, 15 seconds) and flow-through discarded, followed by washes and spins with 700μL Buffer RW1 and twice with 500μL Buffer RPE. An additional spin (>8000g, 2 minutes) was completed to dry the column, and the column was eluted into a microcentrifuge tube with 30-50μL RNase-free water incubated at RT for 1-2 minutes and then spun (>8000g, 1 minute). RNA quantity and quality were measured using a NanoDrop spectrophotometer (Thermo Fisher).

RT-qPCR

RNA can be quantified by using reverse transcription to generate complementary DNA (cDNA), which can be measured using quantitative PCR (qPCR). Isolated RNA (with DNase treatment) was transformed into cDNA using the SuperScript^{TM} IV First-Strand kit (Thermo Fisher). RNA (up to 5μg) was combined with 1μL of oligo-dTs (0.5μg/μL), 0.2μL of

random primers/hexamers (3μg/μL), 1μL dNTPs (10mM) and water, then incubated at 65°C for 5 minutes and cooled on ice for 2-3 minutes. Meanwhile, a reaction was prepared to contain 4μL of First-Strand buffer (5X), 1μL of DTT (0.1M), 1μL of ribonuclease inhibitor and 1μL of Superscript IV RT (reverse transcriptase or nuclease-free water for -RT (controls lacking reverse transcriptase). The reverse transcriptase reaction was combined with annealed RNA and ran on a thermocycler (MJ Research) for 10 minutes at 23°C, 1 minute at 50-55°C and 10 minutes at 80°C. To remove any residual RNA, 1μL RNase H was added and incubated for 20 minutes at 37°C. The resulting cDNA was aliquoted and stored at -20°C.

DNA was diluted to 3.5ng in 1μL and combined with 0.6μL of primer 1, 0.6μL of primer 2, 10μL of PowerUp^{TM} SYBR^{TM} Green Master Mix (2X, Thermo Fisher) and 7.8μL water. Samples were prepared in optical plates (Thermo Fisher), sealed with optical adhesive tape (Thermo Fisher) and centrifuged. qPCRs were run using a real-time PCR system (QuantStudio5, Thermo Fisher) with the following program: 50°C for 2 minutes, 95°C for 2 minutes, and then 40 cycles of 95°C for 15 seconds, annealing at 55-65°C (depending on the primer melting temperature) for 20 seconds and extending at 60°C for 1 minute. For each plate, samples were run alongside two control samples (3D7 and a control line with ineffective dCpf1 repression) using 3 sets of primers: 1 for the lncRNA of interest and 2 for housekeeping genes (cyclophilin and 18S rRNA). All samples were run in triplicate, -RT controls were included for each reaction, and a no-template control was included for each primer.

Primers (18-30 bp in length) were designed to target sequences and generate 80-150 bp products. Primer efficiencies were determined by running a qPCR with a 10-fold serial dilution of a control sample (3D7) in triplicate. Average Ct was calculated from triplicates, and the slope of the regression between average Ct values and log(Sample dilution values) was determined. Primer efficiency was calculated using Equation 2.1, an efficiency of 90-110 was deemed suitable. For the qPCR assays, average Ct values from technical replicates were obtained from the qPCR machine software, Design and Analysis 2 (v.2.6.0, AppliedBiosytems, Thermo Fisher) and relative gene expression was calculated using the Pfaffl method (Pfaffl, 2001). This method uses Equation 2.2, where E stands for primer efficiency, GOI stands for gene of interest, and HKG stands for housekeeping gene. ΔCt values are calculated by subtracting avgCt values from the control average (average obtained from biological controls). Both housekeeping genes were incorporated into this equation by obtaining a geometric mean for both denominator calculations. Standard deviation and statistical significance (two-tailed t-tests) were also determined.

$$\text{Primer efficiency (E)} = (10^{\frac{-1}{slope}} - 1) * 100 \tag{2.1}$$

$$\text{Gene expression ratio} = (\frac{(E_{GOI})^{\Delta}CtGOI}{(E_{HKG})^{\Delta}CtHKG} \tag{2.2}$$

RNA-Seq

DNase-treated RNA was sent to Sanger DNA Pipelines for library prep and RNA sequencing. Further sequencing and library prep details are outlined in *Chapter 3* and *Chapter 5*. For annotation, sequences were mapped with minimap2 against Pf3D7 (v3) reference or for *P. knowlesi*, also mapped with HISAT2 against the PKNH (v2) reference (Li, 2018; Kim et al., 2019; Amos et al., 2021). Sequences were visualised using Artemis (Carver et al., 2012).

For expression profiling, reads were processed using samtools (Li et al., 2009), trimmed with cutadapt (Martin, 2011) and mapped with STAR against the *P. falciparum* 3D7 reference genome (Dobin et al., 2013; Amos et al., 2021). Quality control was completed using fastqc, and multiqc (Andrews, 2010; Ewels et al., 2016). Feature counts were obtained using featurecounts (Liao et al., 2013). Differential expression analysis was completed with DEseq2 was used to determine differential expression using a negative binomial general linear model (Love et al., 2014). Shrunken log fold changes were estimated using the ashr program (Stephens, 2016). Genes with a log 2-fold change greater than 0.5, and an FDR (false discovery rate) less than 5% were considered deferentially expressed.

2.5.2 Fitness

Fitness was measured in two ways: growing parasites in parallel cultures in a growth assay or competing parasites against fluorescent parasites in the same culture in a competition assay. The flow cytometry data was analysed in FlowJo (v10) or FCS Express (v7).

Growth assay

1mL of each culture was seeded into 24-well plates at 0.5% parasitemia in triplicate and confirmed by flow cytometry. Every following day or other days for 2 weeks, parasitaemia was measured by flow cytometry, and cultures were cut to maintain parasitaemia between 0.25-5%. Growth curves were generated using mean cumulative parasitaemia. Cumulative

parasitaemia was calculated by multiplying measured parasitaemia by the fractions by which cultures were cut and adding to the previous day's parasitaemia.

Competition assay

0.5mL of each culture was seeded into 24-well plates at 0.5% parasitemia and combined with an equivalent volume of fluorescent parasites (NF-54^{camGFP}, Dd2^{bipGFP}, Dd2^{pareNeon} or 3D7^{pareNeon}) at 0.5% parasitaemia. Immediately after and every following day or other days for 3 weeks, the proportion of fluorescent parasites was measured by flow cytometry and cultures were cut to maintain parasitaemia between 0.25-5%. Competition assay graphs were generated using averages between replicates.

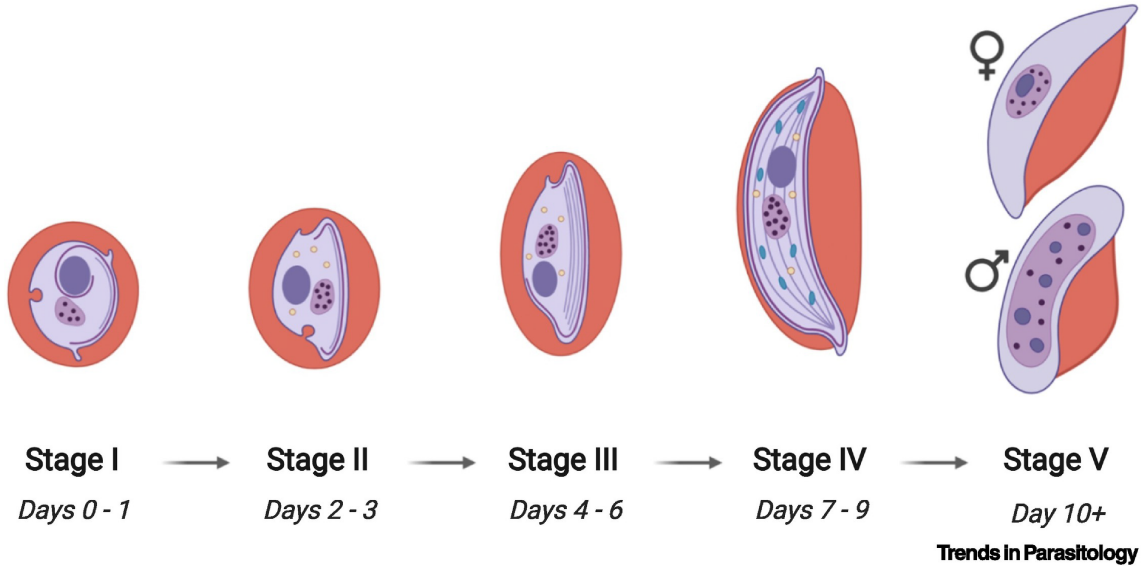


Fig. 2.3 *P. falciparum* gametocyte stages. Gametocytogenesis is formed of five stages (stages I to V), which over 10 days, transform asexual parasites into male and female gametocytes. The figure was obtained with permissions from McLean and Niles (2022).

2.5.3 Gametocytogenesis

Gametocytes can be obtained by inducing sexual commitment of asexual blood stages using stress (Carter and Miller, 1979; Ogwan'g et al., 1993). Trophozoite-staged cultures with high parasitemia (5-8%) were fed 2/3 of fresh media and 1/3 of conditioned media (day 0). Starting 24 hours later (day 1) and every following day, cultures were fed with pre-warmed media supplemented with 5% inactivated horse serum (Gibco), which supports gametocyte growth and 1μL/mL heparin (Sigma-Aldrich, 20U/mL in PBS, 0.2μM filtered), which blocks

new invasions (Fivelman et al., 2007). Gametocytes were visible from day 2. To assess gametocyte conversion rate, cultures were prepared in triplicate in a 24-well plate, rings were counted on day 1 and gametocytes (stage III) were counted on day 4. Gametocyte stages are shown in Figure 2.3. Conversion rate = (gametocytes on day 4)/over rings on day 1). To assess the proportions of each sex: male and female gametocytes were counted on days 10 and 13. The reported gametocyte sex ratio is the proportion of males to all gametocytes.

2.5.4 Drug sensitivity

The susceptibility of parasites to antimalarial drugs was quantified with a drug dose-response assay, where the read-out is the half-maximal effective concentration (EC_{50}, referred to as IC_{50} in the malaria field) can be determined, which is the drug concentration required for 50% of the maximal response (where the maximal response is killing the parasites). 96-well plates were prepared with two-fold serial dilutions of the drug from columns 1-10 and designed to reach the IC_{50} around wells 4-6 (Figure 2.4). The drugs were dispensed either manually or using a digital dispenser (D300e, Tecan). CM was dispensed into column 11 as a no-drug negative control and into the top and bottom rows to counter edge effects. Cultures were seeded into columns at 1% haematocrit and 1% parasitaemia, excluding column 12, which was seeded with only RBCs. Cultures were maintained at normal conditions and then lysed and stained three days later with an equal volume of 2X SYBR lysis buffer (Table 2.2). The plate was incubated for 30 minutes in the dark and measured with a microplate reader (FLUOstar® Omega, BMG Labtech) using a 485/535 nm excitation and emission filter. The fluorescence data were analysed in GraphPad Prism 8 where parasite survival was assessed by comparison between the fluorescence of drug and no-drug treatment groups. Background fluorescence was also subtracted using the RBCs-only wells. A non-linear regression model for log(inhibitor) and variable(slope) was used to calculate IC_{50} values, and standard deviation and statistical significance (two-tailed t-tests) were also determined.

2.5.5 Fluorescence

Parasites engineered to express fluorescent markers were visualised using fluorescent microscopy. 200μL of parasite culture was harvested at 5% parasitemia, spun at 3000rpm for 30 seconds in a centrifuge tube and washed with 0.5mL of PBS. Parasites were fixed in 4% (v/v) paraformaldehyde + 0.0075% glutaraldehyde, incubated for 30 minutes and washed twice with 1mL of PBS. Parasite DNA was stained with 1mL of Hoechst stain (10μg/mL, Thermo Fisher) for 5 minutes and resuspended in 1mL of PBS. The stained parasites were transferred

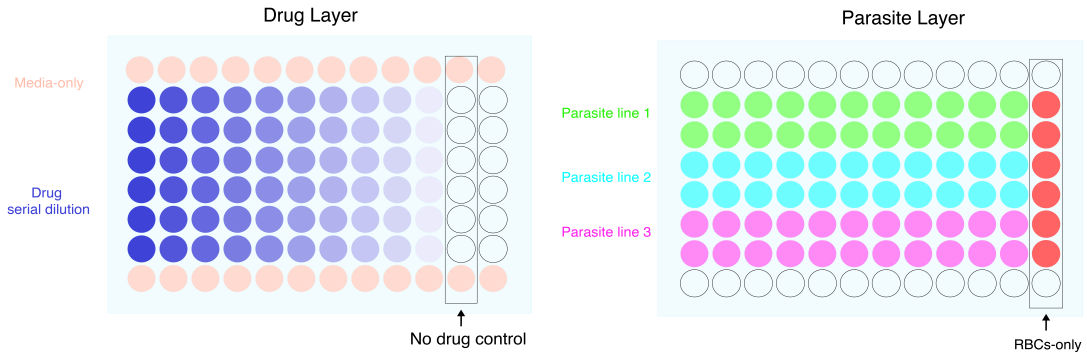


Fig. 2.4 Schematic of drug dose-response assay set-up. In a 96-well plate, a two-fold serial dilution of a drug was set up across columns 1-10 of the plate, excluding the top and bottom rows. No drug was added to columns 11 and 12 as these were the no-drug and RBC-only controls. Row A and H were given media instead of the drug. Cultures containing parasites were also seeded into the plate in columns 1-11 from rows B-G, while RBCs-only were seeded into column 12. Final volumes were 100μL at 1% haematocrit and 1% parasitemia.

to a 4-well chambered coverglass (M-Slide, Ibidi) that was pre-treated with 0.25mL of poly L-lysine (0.2mg/ml, Sigma-Aldrich) for 10 minutes. The coverglass was transferred to an inverted fluorescence microscope (DMi8, Leica). Blue and green-filtered fluorescence microscopy as well as bright-field microscopy with a 100x objective (total 1000x magnification) was used to image infected erythrocytes and fluorescent parasites. Image processing and analysis were completed with Leica Application Suite X.

Table 2.2 Solutions

Solution	Components
Complete media (CM)	RPMI 1640 with serum substitute AlbuMAX® (Gibco) and supplemented with L-glutamine substitute GlutaMax® (Gibco), antibiotic gentamicin (Gibco) and buffering agent HEPES (20mM, equilibrated to pH 7.0)
Glycerolyte	576mL 99% glycerol, 16g L-sodium lactate, 300mg KCl, 1g NaH_2PO_4, 2.366g Na_2HPO_4, 424mL deionised water, equilibrated to pH 6.8
Cytomix	120mM KCl, 0.15mM $CaCl_2$, 10mM K_2HPO_4/KH_2PO_4, 25mM Hepes, 2mM EDTA, 5mM $MgCl_2$, equilibrated to pH 7.6
2X SYBR lysis buffer	2X SYBR® Green, 20mM Tris-HCl, 5 mM EDTA, 0.1% w/v saponin, and 1% v/v Triton X-100

Chapter 3

Generating a manually-curated lncRNA annotation in *Plasmodium falciparum*

3.1 Overview

A substantial limitation to studying lncRNAs in the *Plasmodium falciparum* parasite is the lack of a comprehensive and updated annotation. An annotation provides the locations of lncRNAs in the genome and the sequences of the transcribed RNAs, which is valuable information for experimental and computational research. In this chapter, I generated an improved lncRNA annotation for *P. falciparum*. By generating long-read RNA sequencing data and integrating information extracted and curated from multiple sources, I annotated lncRNAs, evaluated supporting evidence, and characterised their features. The content of this chapter has been published in *BMC Genomics* as *A manually curated annotation characterises genomic features of P. falciparum lncRNAs* (Hoshizaki et al., 2022a). The sequencing data used to generate annotation was generated and processed by Dr Sophie Adjalley, Sanger DNA pipelines and Dr Adam Reid; these colleagues are cited in the relevant sections for their contributions. Additional *P. falciparum* and *P. knowlesi* sequencing datasets were obtained from Dr Vandana Thathy and Dr Chris Newbold (University of Oxford), and Dr Lia Chappell (Wellcome Sanger Institute). Proteomics analysis was completed by collaborators Dr Scott Chisholm and Dr Ross Waller (University of Cambridge). Supplemental data and acknowledgements are in Appendix A.

3.2 Introduction

Modern research in malaria studying the biology of the parasite relies heavily on the genome assembly and its annotation (Böhme et al., 2019). For example, molecular approaches such as PCR amplification, gene disruption, and gene tagging require precise start and stop locations of gene features. Similarly, bioinformatics analyses such as multiple gene alignment, structure prediction, and coding potential prediction analyses require accurate sequence information. While certain genomic features, such as protein-coding genes, are relatively well-annotated, the role of non-coding transcription in the parasite's biology remains poorly understood.

3.2.1 Current *P. falciparum* lncRNA annotations

There is no definitive lncRNA annotation for *P. falciparum*, instead, annotations or transcript lists must be extracted from various publications. LncRNAs have been largely excluded from the *P. falciparum* 3D7 genome annotation (Böhme et al., 2019). The only ncRNA annotations included are snRNAs, snoRNAs, spliceosomal RNA, the RUFs (RNAs of Unknown Function) and one lncRNA, GDV1-aslncRNA, which has been experimentally validated and demonstrates biological function (Filarsky et al., 2018). Boehme et al. state that previously annotated ncRNAs were discarded from the annotation due to the lack of supportive evidence as these annotations have not been verified *in vitro* as *bona fide* transcripts or associated with biological function.

LncRNAs were initially identified through early transcriptomic techniques such as serial analysis of gene expression, DNA tiling arrays, microarrays and northern blotting (Su et al., 1995; Patankar et al., 2001; Kyes et al., 2003; Gunasekera et al., 2004; Ralph et al., 2005; Lu et al., 2007; Li et al., 2008; Epp et al., 2009; Broadbent et al., 2011). Additionally, computational techniques were used to predict putative lncRNAs such as comparative genomics (Upadhyay et al., 2005; Chakrabarti et al., 2007), genomic-wide secondary structure prediction (Mourier et al., 2008) and through the analysis of existing cDNA and expression sequencing tag datasets (Raabe et al., 2010). However, the development of RNA sequencing (RNA-Seq) revolutionised genome-wide annotation of lncRNAs. Otto et al. applied RNA-Seq to *P. falciparum* and identified 107 novel ncRNA transcripts and additional antisense transcripts, which were supported by subsequent RNA-Seq studies (Otto et al., 2010; Sorber et al., 2011; López-Barragán et al., 2011; Liao et al., 2014). However, early RNA-Seq methods did not facilitate strand-specific resolution. Therefore, the next advancement in annotation arrived with strand-specific RNA-Seq and when PCR-amplification was improved

Table 3.1 RNA sequencing-based approaches capturing ncRNAs in *P. falciparum*

Sequencing	Methodology	Output	Reference
Short-read RNA-Seq	Illumina RNA-Seq	310 AS events	(Sorber et al., 2011)
	Strand-specific RNA-Seq	1204 AS events	(Siegel et al., 2014)
		1134 lncRNAs & 1381 circRNAs annotations	(Broadbent et al., 2015)
		2252 nascent-ncRNAs	(Yin et al., 2020)
	Bespoke RNA-Seq (DAFTseq)	5000 ncRNA events	(Chappell et al., 2020)
Long-read RNA-Seq	ONT direct RNA-Seq	1112 AS events	(Lee et al., 2021)
	PacBio SMRT RNA-Seq	3623 ncRNA events	(Yang et al., 2021)
	ONT & PacBio	491 AS and 167 intergenic isoforms	(Shaw et al., 2022)

Table modified from (Simantov et al., 2022).

by using polymerases with less sequence bias that better captured AT-rich ncRNAs. Two studies applied this technology to *P. falciparum* across numerous timepoints: Siegel et al. identified 1247 genes with evidence of antisense transcription, and Broadbent et al. generated an annotation of 660 intergenic lncRNAs, 474 antisense lncRNAs (220 of which are novel) and 1381 circRNAs (Siegel et al., 2014; Broadbent et al., 2015). Recently, Yin et al. explored the dynamics of nascent transcription by using EU labelling with RNA-Seq and identified 2252 putative nascent-transcribed ncRNAs (Yin et al., 2020). Other recent transcriptomic studies using improved methodologies, such as long-read sequencing, have captured more ncRNA transcripts (Table 3.1). However, these datasets have not been used to generate new lncRNA annotations, i.e. with collapsed reads and consensus start and stop coordinates or compared with existing annotations (Chappell et al., 2020; Lee et al., 2021; Yang et al., 2021; Shaw et al., 2022). Collating and repurposing existing RNA-Seq datasets for the annotation of lncRNAs has recently proven effective in another apicomplexan, *Cryptosporidium* sp., where 396 novel lncRNAs were identified (Li et al., 2021b).

3.2.2 Challenges in annotating lncRNAs in *P. falciparum*

The annotation of *P. falciparum* lncRNAs, which have reduced G + C content and increased repetitive sequences, have been made difficult by low coverage of short reads that map to low-complexity regions. Long-read RNA sequencing technology is more suitable for detecting and annotating *P. falciparum* lncRNAs. Pacific Biosciences (PacBio) and Oxford Nanopore

Technologies (ONT) have extended RNA sequencing lengths while maintaining high-fidelity sequencing. Even further, in the case of ONT, RNA sequencing can be completed directly (direct RNA-Seq) without the need for cDNA generation, amplification, or fragmentation. Recent studies have applied long-read RNA-Seq to *P. falciparum*, which has proven effective for refining the annotation of untranslated regions (UTRs) and analysing transcript isoforms (Chappell et al., 2020; Lee et al., 2021; Shaw et al., 2022). Additionally, when generating or re-purposing sequence data for lncRNA annotation, another consideration is library preparation. As some *P. falciparum* lncRNAs are not poly-adenylated, such as var-aslncRNAs, poly-A enrichment may not capture all lncRNAs (Amit-Avraham et al., 2015).

Due to pervasive transcription in *P. falciparum*, another challenge in lncRNA annotation is the complexity of resolving overlapping expression from neighbouring transcriptional units. The boundaries of gene features must be well-defined. Since existing annotations were developed, the UTRs have extended considerably for many genes across the *P. falciparum* 3D7 genome (Chappell et al., 2020). Previous annotations must be revisited to determine the impact of the changed UTR boundaries. Distal UTR reads may have been interpreted as separate transcriptional units and labelled intergenic lncRNAs.

The current annotation pipeline is often limited to transcript assembly and basic filtering. However, other types of datasets could inform and curate a lncRNA annotation. For example, powerful evidence for the presence of a transcribed sequence is an associated transcriptional start site (TSS). Studies have comprehensively mapped out TSSs across the genome using specialised approaches that improve the capture of intact $5'$ ends (Adjalley et al., 2016; Chappell et al., 2020). Additionally, epigenetic data sets describing the accessibility of the chromatin (ATAC-Seq) have been generated, which could also provide supportive evidence of transcription at putative lncRNA loci (Ruiz et al., 2018; Toenhake et al., 2018). A collation and curation of publicly available datasets can provide context and evidence for lncRNAs.

In this chapter, I created an updated and improved lncRNA annotation for blood-stage *P. falciparum* 3D7. I developed a high-quality annotation by generating long-read direct RNA sequencing and integrating the updated UTR annotations. Uniquely, I manually curated collated datasets from the literature to provide supportive evidence for each lncRNA annotation. I compared the annotation to existing annotations to verify existing lncRNAs and identify novel lncRNAs and characterised the features of *P. falciparum* lncRNAs.

3.3 Objective and Aims

To generate a high-quality, updated annotation of lncRNAs in *P. falciparum* using long-read sequencing and manual curation of existing datasets.

1. Verify previously-annotated *P. falciparum* lncRNAs
2. Identify novel lncRNAs in *P. falciparum*
3. Characterise the genomic and sequence features of annotated lncRNAs

3.4 Results

3.4.1 The generation of a manually curated lncRNA annotation

Long-read RNA sequencing of asexual intraerythrocytic *P. falciparum*

The annotation was developed using a two-pronged approach of generating new RNA sequencing and harnessing existing datasets. Asexual intraerythrocytic-staged *P. falciparum* 3D7 parasites were sequenced using long-read ONT direct RNA sequencing (Table 3.2). Mixed stages were sequenced to capture the broad scope of lncRNA expression in the intraerythrocytic development cycle. ONT was chosen over other long-read sequencing technology for the unique advantages of direct-RNA-Seq, specifically the ability to generate full-length RNA transcripts without any cDNA generation, amplification, or fragmentation step, making this particularly suited for the capture of full-length lncRNAs. Libraries were prepared with and without TEX to obtain a library enriched with primary transcripts. TEX (5′-monophosphate-dependent exonuclease) treatment degrades rRNAs and other processed RNA transcripts, thereby enriching primary (unprocessed) transcripts, which can improve the capture of transcriptional start sites. The RNA was extracted by Dr Sophie Adjalley, the library prep and sequencing were completed by Sanger DNA pipelines, and sequence QC and mapping were conducted by Dr Adam Reid (sequencing information in Appendix: Table A.1).

Assembling and curating existing datasets from the literature

Preliminary observations revealed low read depth at lncRNA loci, likely due to their diminished expression. Therefore to improve read depth and gain additional confidence in

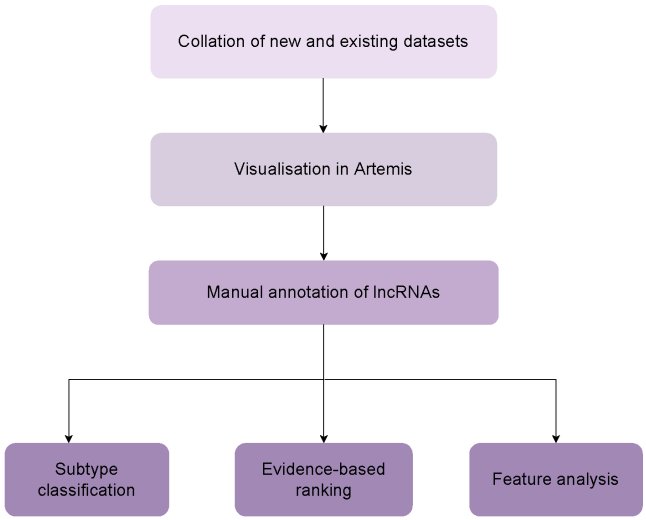


Fig. 3.1 Flow diagram depicting the approach for manual curation and annotation of lncRNAs

detecting lncRNA transcripts, long-read RNA-Seq data from the present study was collated with data from Lee et al. (Table 3.2) (Lee et al., 2021). Unlike the other existing *P. falciparum* 3D7 long RNA-Seq datasets (from Yang et al. and Chappell et al.), the Lee et al. dataset was generated using the same sequencing approach (Chappell et al., 2020; Yang et al., 2021; Lee et al., 2021). I also obtained stage-specific short-read RNA sequencing of asexual *P. falciparum* 3D7 parasites to support the annotations made from the long-read data (provided by Dr Vandana Thathy and Dr Chris Newbold, University of Oxford). Furthermore, to provide contextual support, datasets from TSS and chromatin accessibility studies as well as existing lncRNA annotations were extracted and compiled from numerous sources (Table 3.2) (Broadbent et al., 2015; Kensche et al., 2016; Adjalley et al., 2016; Ruiz et al., 2018; Chappell et al., 2020). These sources can be subdivided into two groups: those that were used for contextual support during annotation and those that were later used for comparative analysis.

Manual annotation

Annotation was completed by visualising and evaluating the datasets in a genome browser (Figure 3.1). LncRNAs were defined as noncoding RNAs of at least 200 nucleotides in length that were not otherwise annotated as another type of noncoding RNA (rRNAs, tRNAs, snRNAs and snoRNAs). One exception was lncRNAs that contained other ncRNAs; however, these transcripts had to differ distinctly from the annotated ncRNAs. The lncRNA boundaries

Table 3.2 Datasets used for manual curation of *P. falciparum* lncRNA annotation

Use	Dataset	Type	Reference
Annotation	Pf nanopore 1	Nanopore long read	**This work**
	Pf nanopore 2	Nanopore long read	Lee et al. (Lee et al., 2021)
Contextual support	Pf short read	Illumina short read	**This work**
	Pf TSS-seq 1	Illumina short read	(Chappell et al., 2020)
	Pf TSS-seq 2	Illumina short read	(Kensche et al., 2016)
	Pf TSS-seq 3	Illumina short read	(Adjalley et al., 2016)
	Pf ATAC-seq	Illumina short read	(Ruiz et al., 2018)
	Pf ncRNA calls	Illumina short read	(Chappell et al., 2020)
	Pf lncRNA annotation	Annotation	(Broadbent et al., 2015)
	Pf ncRNA annotation	Annotation	PlasmoDB (Amos et al., 2021)
Comparative analysis	Pf lncRNA annotation	Annotation	(Liao et al., 2014)
	Pf antisense transcripts	Gene list	(Siegel et al., 2014)
	Pf ncRNA calls	Predicted transcripts	(Chappell et al., 2020)
	Pf lncRNA calls	Predicted transcripts	(Yang et al., 2021)

were defined as the outermost read in the collated long-read sequencing. An evidence-based ranking was assigned to each annotation, with a 1 signifying the most supportive evidence and 9 the least (Figure 3.2, supplemental data). LncRNAs were scored based on three criteria: the presence of the lncRNA in the long-read RNA-Seq datasets (one or both), the number of reads (single vs multiple) and finally, evidence of a distinct TSS in the TSS datasets (none, single or multiple datasets). The rankings reflect the level of confidence for each annotation and can be used to subset the dataset.

3.4.2 The *P. falciparum* transcriptome contains over two thousand lncRNAs

I identified a total of 2369 lncRNAs in *P. falciparum*, of which 1119 were novel to this study. The remaining 1250 were previously annotated by Broadbent et al. or Liao et al., listed on PlasmoDB (from various sources) or were predicted by Siegel et al., Chappell et al. or Yang et al. (Figure 3.3A, Table 3.2, supplemental data) (Liao et al., 2014; Siegel et al., 2014; Broadbent et al., 2015; Chappell et al., 2020; Yang et al., 2021; Amos et al., 2021). Previous annotations were updated; for example, long-read sequencing enabled the extension of lncRNAs that were previously partially annotated and the fusion of those previously annotated as multiple lncRNAs (Figure 3.3B). The long-read sequencing also expanded most lncRNA boundaries (start and stop positions), although some boundaries were reduced (Table 3.3). The new UTR boundaries provided by the current 3D7 genome

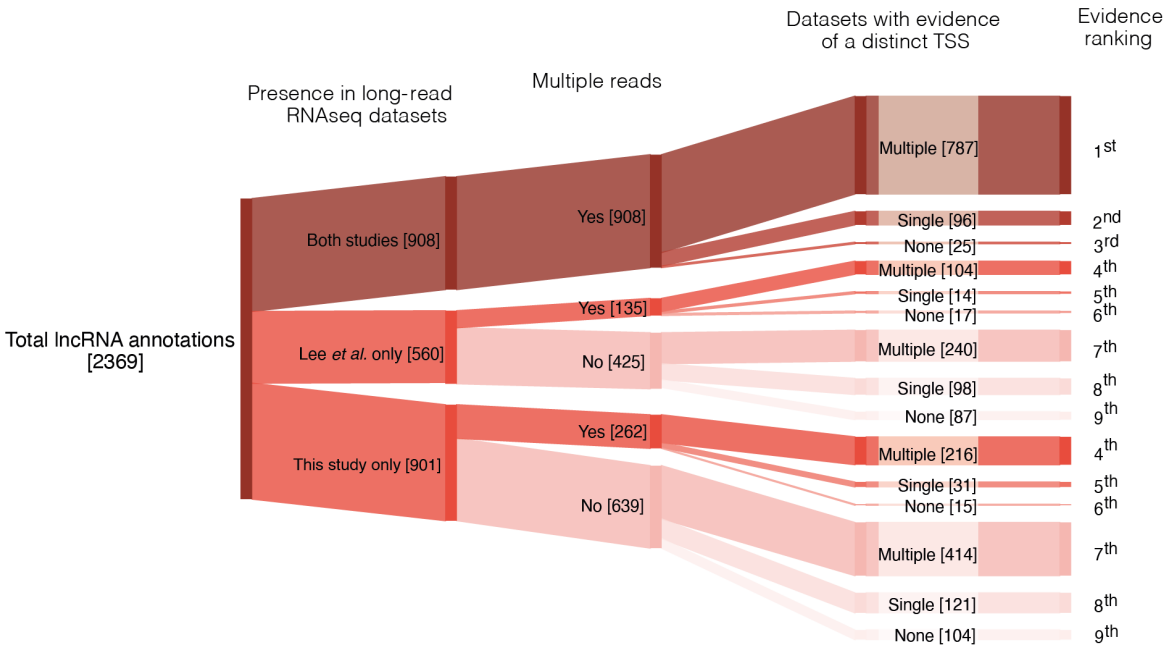


Fig. 3.2 Evidence ranking based on supportive evidence of lncRNA annotations. LncRNAs were ranked from 1-9, 1st representing lncRNAs with the most evidence and 9th representing lncRNAs with the least evidence. Three criteria were used to define the ranking: 1- Presence in both vs one of the long-read RNA-Seq datasets (this study and Lee et al.) (Lee et al., 2021); 2- the presence of multiple reads vs a single read; 3- evidence of a distinct TSS from multiple datasets vs a single dataset vs no evidence of a TSS (Adjalley et al., Kensche et al. and Chappell et al. datasets) (Adjalley et al., 2016; Kensche et al., 2016; Chappell et al., 2020). A Sankey plot shows how these criteria were used to determine rank.

annotation led to significant changes to previously annotated lncRNAs. These lncRNAs were either misannotations or wrongly categorised but are encompassed by genes or UTR sequences (Figure 3.3B, Table 3.3). Also, some previously annotated lncRNAs were not captured in this study's collated long-read RNA sequencing.

3.4.3 LncRNAs are produced from distinct genomic contexts

The genomic context was determined based on the presence of overlapping (on the same strand), antisense (on the opposing strand) or nearby (associated) genomic features. LncRNAs were classified into eight subtypes: *intergenic*, *antisense* (to genes, UTRs, introns, or lncRNAs), *UTR-associated*, *intronic* and *sense* (within an exon) (Figure 3.4, supplemental data). The most common subtypes were *antisense-to-gene* (44%), followed by *antisense-to-UTR* (24.7%) and *intergenic* (11.9%) (Figure 3.5A). The remaining subtypes were less common: *UTR-associated* (8.9%), *antisense-to-lncRNA* (6.4%), *antisense-to-intron* (2.5%),

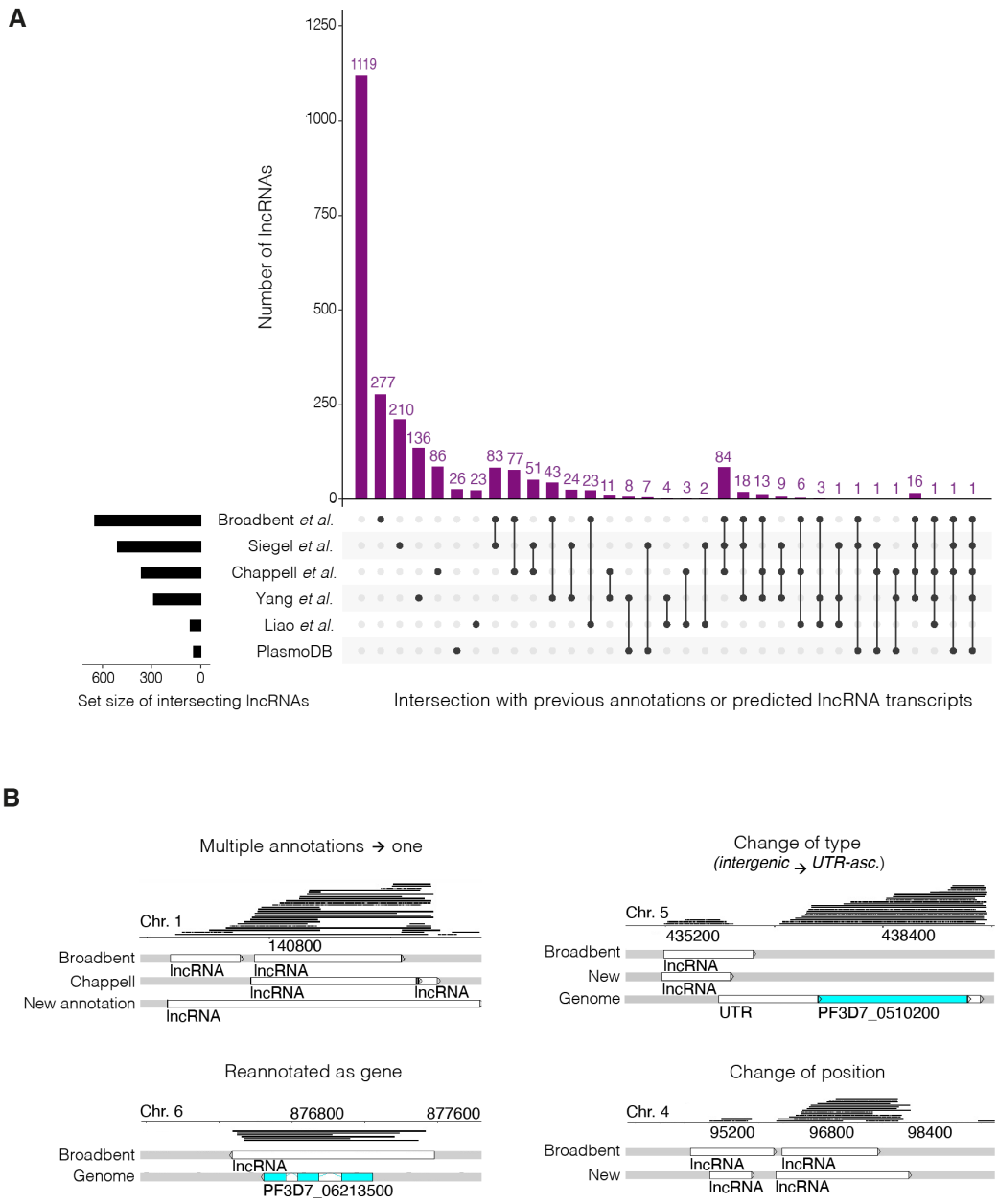


Fig. 3.3 Verification of previous *P. falciparum* lncRNA annotations in the literature. Of the 2369 lncRNAs, 1119 were unique to this study, and 1250 were previously annotated by Broadbent et al. or Liao et al. or predicted by Siegel et al., Chappell et al., Yang et al., and/or listed by PlasmoDB from various published studies (Broadbent et al., 2011; Liao et al., 2014; Siegel et al., 2014; Chappell et al., 2020; Yang et al., 2021; Amos et al., 2021). (**A**) An upset plot shows the number and membership of the previously annotated lncRNAs as well as the size of the set. Gene IDs from the Siegel et al. dataset were intersected with gene IDs of genes antisense to lncRNAs in this work (Siegel et al., 2014). (**B**) Snapshots demonstrate examples of changes to previous annotations.

Table 3.3 Comparison of previous annotations to new *P. falciparum* lncRNA annotation

Difference with new annotation		Number of previous annotations		
		Broadbent	Chappell	PlasmoDB
Change of subtype	*intergenic→UTR-asc.*	64	N/A	N/A
	intergenic→antisense-to-gene	42		
	intergenic→antisense-to-UTR	108		
	antisense-to-gene→UTR-asc.	38		
	antisense-to-gene→intergenic	1		
Change of position	start site expanded	433	327	17
	start site reduced	211	33	7
	stop site expanded	315	310	20
	stop site reduced	332	49	4
Change in number of annotations	single→multiple	1	0	0
	multiple→single	49	543	4
Misannotated	reannotated as gene	3	0	0
	reannotated as UTR	180	39	4
	less than 200 bp	15	4	3
Not observed		240	4493*	14

The new annotations were compared against the Broadbent et al. lncRNA annotations (Broadbent et al., 2015), Chappell et al. ncRNA calls (Chappell et al., 2020), and PlasmoDB ncRNA annotations (Amos et al., 2021). *Represents the number of ncRNA calls in place of annotated lncRNAs.

sense (1.5%) and *intronic* (0.04%). While the other subtypes required a direct overlap of the lncRNA with the genomic feature, the *UTR-associated* lncRNAs were evaluated from a 150 bp proximity. One hundred fifty was selected based on previous methods suggesting some UTRs may extend 100 nucleotides or more beyond the position predicted by sequence coverage (Chappell et al., 2020). Of the *UTR-associated* lncRNAs, 65% were associated with a 5′ UTR, 32% were associated with a 3′ UTR, and 3% were spanning two genes, associated with a 5′ and 3′ UTR (Figure 3.5B).

3.4.4 LncRNA loci are ubiquitous in the *P. falciparum* genome

The lncRNAs were distributed ubiquitously throughout the 14 *P. falciparum* chromosomes. A cluster analysis with Clustor Locator identified that lncRNA subtypes occasionally formed location-based clusters of 3-5 lncRNAs (Figure 3.5C, Table 3.4). There was no apparent strand preference for the production of lncRNAs, with 49% on the negative strand and the remaining 51% on the positive strand (Figure 3.5D).

Previous research has suggested that the vast majority of promoters in *P. falciparum* are bidirectional, suggesting that the majority of lncRNAs are driven by gene promoters (Adjalley

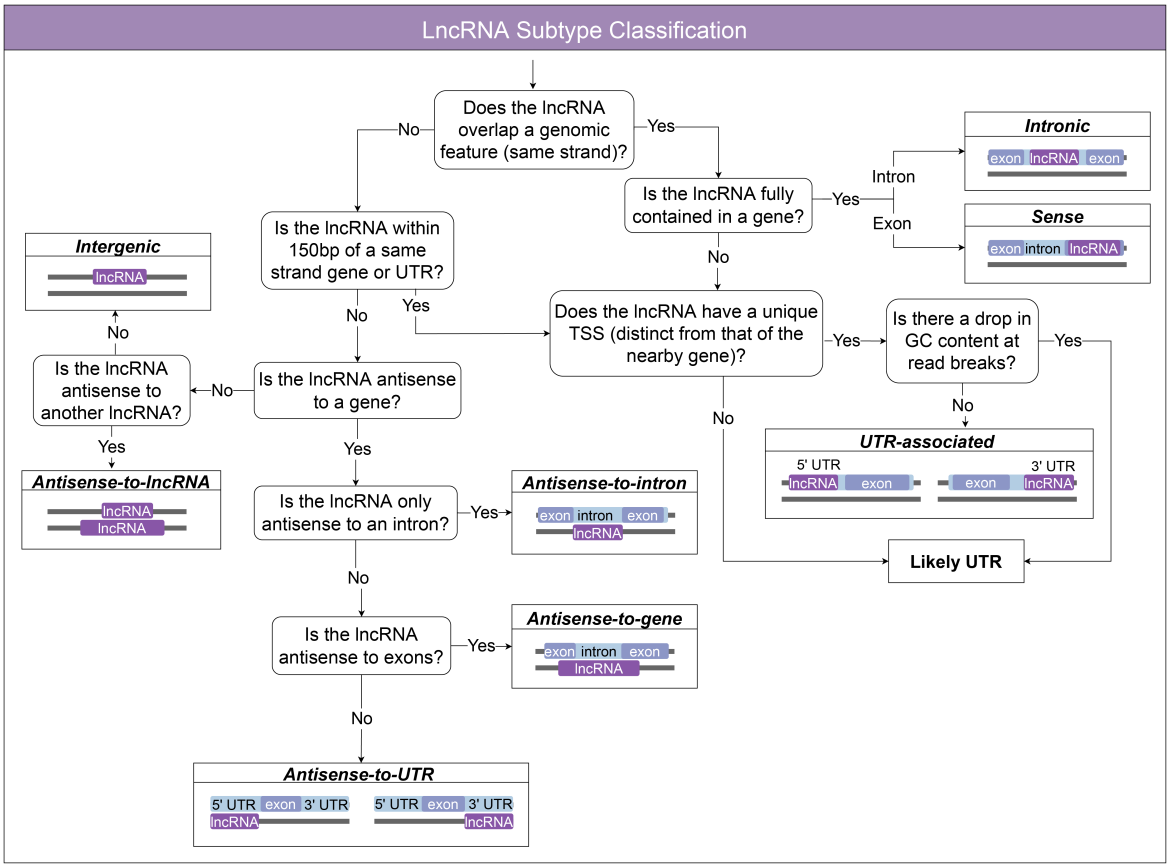


Fig. 3.4 Schematic representation of the classification of lncRNA into genome context-based subtypes. Annotations were categorised by genomic context using a decision tree. LncRNAs overlapping a gene on the same strand were classified as either *intronic* if contained within the intron or *sense* if contained within a single exon. No lncRNAs were annotated that spanned multiple exons in a gene. LncRNAs that overlapped a UTR and lncRNAs nearby genes (within 150 bp of an annotated UTR or exon or read from the gene) were flagged as potential UTR-associated lncRNAs. A careful examination of collated data was performed to delineate UTR-associated lncRNAs from UTR transcripts (that could be fragmented due to drops in GC content or alternative start sites). This included an analysis of the level of overlap between reads from the putative lncRNA and gene/UTR, the presence of a unique transcriptional start site (distinct from the gene) and the lack of evidence of a drop in GC content. LncRNAs that were antisense (opposite strand) to genomic features were classified based on the type of antisense genomic feature: *antisense-to-gene, antisense-to-intron, antisense-to-UTR* and *antisense-to-lncRNA*. The *antisense-to-intron* lncRNAs were contained within the intron boundaries (with little to no overlap with the exon). The *antisense-to-UTR* lncRNAs only overlapped the UTR, not the exons, and the level of overlap varied. Some lncRNAs could be classified as more than one subtype if overlapping multiple features–the classification has a hierarchy starting with: *intronic, sense, UTR-associated, antisense-to-intron, antisense-to-gene, antisense-to-UTR* and *antisense-to-lncRNA*. LncRNAs that were not overlapping, antisense to, or nearby (150 bp) any feature were classified as *intergenic*.

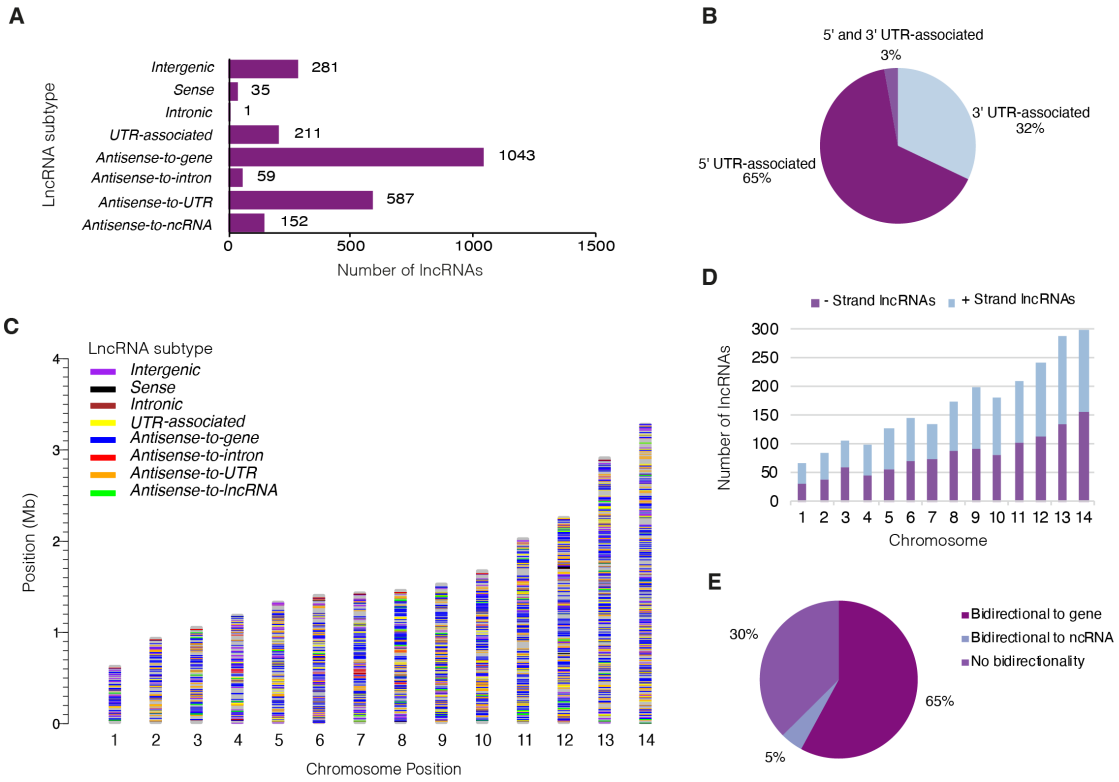


Fig. 3.5 Genomic features of *P. falciparum* lncRNAs. (**A**) The majority of lncRNAs were *antisense-to-gene* lncRNAs or *antisense-to-UTR* lncRNAs, followed by *intergenic* lncRNAs, *UTR-associated* lncRNAs and *antisense-to-lncRNA* lncRNAs. A minority were *antisense-to-intron, intronic* or *sense* lncRNAs. (**B**) *Antisense-to-UTR* lncRNAs were most commonly (65%) associated with a 5′ UTR, followed by 32% associated with a 3′ UTR and a small subset (3%) were nestled between 5′ and 3′ UTRs. (**C**) LncRNAs and their genome-context subtypes were distributed throughout the genome. (**D**) LncRNAs were equally distributed between positive and negative strands, and their abundance in chromosomes was relative to chromosome size. (**E**) For lncRNAs with an associated transcriptional start site, 70% had evidence of bidirectionality (transcription in both directions in the same location and timepoint).

Table 3.4 Genomic location-based clustering of *P. falciparum* lncRNAs by subtype

Subtype	Cluster size	Number of clusters	Number of lncRNAs	Percentage of lncRNAs		p-value
				Per cluster (%)	Total (%)	
Intergenic	3	7	21	7.42	34.28	3.09e-07
	4	2	8	2.83		
Antisense-to-gene	3	47	141	13.43	87.90	2.99e-02
	4	23	92	8.76		
	5	2	10	0.95		
Antisense-to-intron	3	1	3	6.13	18.37	3.73e-05
	4	2	8	2.83		
Antisense-to-UTR	3	30	90	15.31	39.80	9.07e-02
	4	3	12	2.04		
	5	2	10	1.70		
Antisense-to-lncRNA	3	5	15	9.93	63.55	1.00e-10
	4	1	4	2.65		
	5	1	5	3.31		

Cluster Locator was used to identify positional-based clusters of lncRNAs based on genome position (Pazos Obregón et al., 2018). A max gap, the number of genes between cluster elements, was assigned a value of 2. Clusters of two lncRNAs were excluded. The *intronic* subtype was not included as it only contains one lncRNA. The p-value was calculated in comparison to random clustering.

et al., 2016). For the 2199 lncRNAs with evidence of an associated TSS, 70% had evidence of bidirectionality with 65% potentially sharing a promoter with genes, and 5% with other lncRNAs (Figure 3.5E). I determined if a promoter was bidirectional by assessing if a TSS was on the opposite strand at the exact location and same stage (timepoint) using the Chappell et al. dataset (Chappell et al., 2020).

3.4.5 LncRNA subtypes are associated with specific gene ontology terms

I observed that the *sense* and *antisense-to-intron* lncRNAs were almost exclusive to *var* genes, where this configuration has been previously shown to be functionally relevant (Jing et al., 2018; Amit-Avraham et al., 2015). I, therefore, completed a Gene Ontology (GO) term-enrichment analysis to investigate functional similarity between the genes contextually associated with the lncRNAs (genes that overlapped for *sense* and *UTR-associated* lncRNAs, and antisense genes for *antisense-to-gene/UTR/intron* lncRNAs). GO enrichment analysis calculates statistical significance based on the frequency of a GO term in a sample compared to the frequency of the term, in the full set of genes, in this case the *P. falciparum* genome. The generated p-value represents the likelihood of obtaining a particular GO term in the set of genes as a result of chance. For each subtype, significant enrichment ($p<0.01$) of

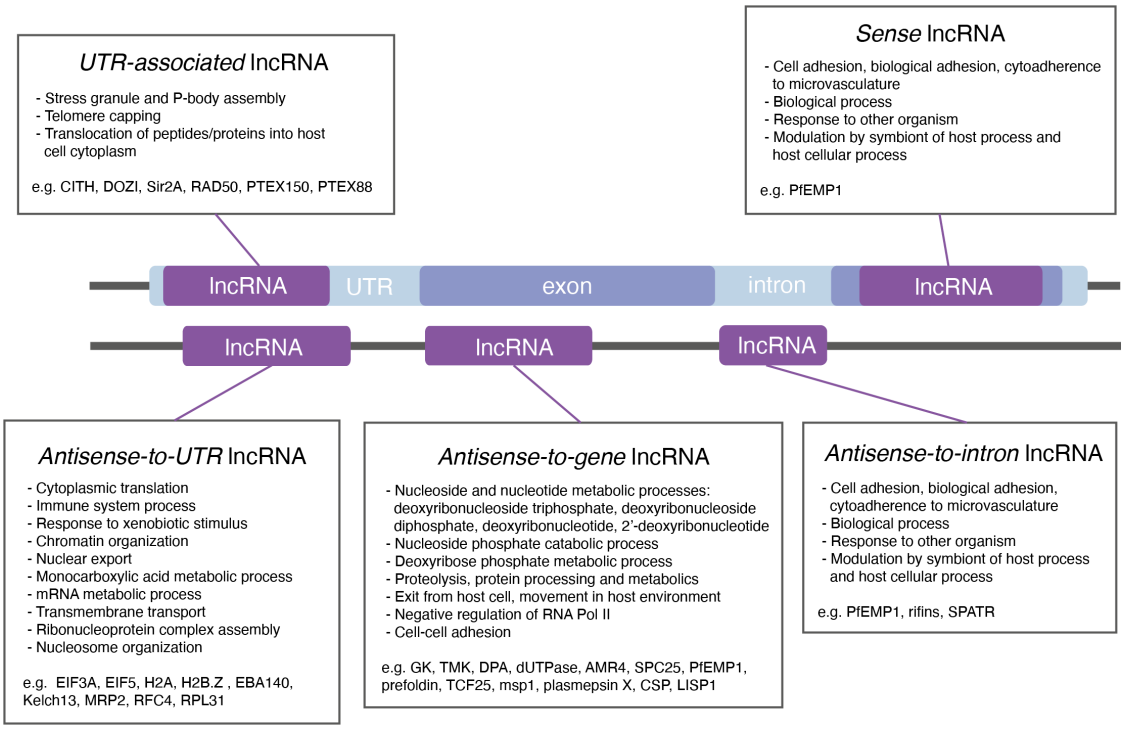


Fig. 3.6 GO term enrichment of genes contextually-associated with *P. falciparum* lncRNA subtypes. Gene ontology (GO) term enrichment was completed on PlasmoDB using biological process ontology for genes associated with each lncRNA subtype: for *sense* and *UTR-associated* lncRNAs, these were the overlapping genes (same strand), and for lncRNAs antisense to genomic features, these were the antisense genes (Amos et al., 2021). Enrichment was determined based on the fold change and odds ratio calculated from the occurrence of GO terms in the set compared to the background. The p-value cut-off selected was 0.01. The GO terms were reduced using REVIGO (Supek et al., 2011).

multiple GO terms was observed (Figure 3.6). Matching my observations, genes associated with the terms *adhesion*, *response to other organisms*, and *modulation by symbiont of host process* (mainly *var* genes as well as other genes encoding surface-exposed proteins) were enriched in *sense* lncRNAs and *antisense-to-intron* lncRNAs. *Antisense-to-gene* lncRNAs (the most abundant class) were enriched for genes involved in nucleoside and nucleotide metabolic and catabolic pathways along with protein metabolism, adhesion and movement in the host environment. *Antisense-to-UTR* lncRNAs were enriched for genes associated with chromatin organisation and translation machinery. *UTR-associated* lncRNAs were enriched for genes relating to stress granule and P-body assembly, telomere capping and translocation of proteins in the cytoplasm.

3.4.6 Some lncRNAs contain structural RNA sequences

Searches against the RNA families database (Rfam) revealed that 19 lncRNAs contained sequences associated with 22 unique described RNA families (Figure 3.7A, Table 3.5), including those encoding known structural RNAs such as the signal recognition particle RNA, the ribozyme ribonuclease P and several RNAs of unknown function (RUFs) (Chakrabarti et al., 2007). Additionally, some lncRNAs contained sequences corresponding to smaller RNAs (usually shorter than 200 nucleotides), including 13 snoRNAs, four tRNAs and one snRNA that I describe respectively as sno-lncRNAs, tRNA-lncRNAs and sn-lncRNAs (Figure 3.7B). LncRNAs containing structural RNA sequences have been previously identified in other organisms, including humans (sno-lncRNAs) and plants (lncRNA containing a tRNA-like molecule) (Yin et al., 2012; Wu et al., 2016; Plewka et al., 2018). I also identified examples of more than one structural RNA sequence within a single lncRNA. There were three examples where two snoRNAs flanked the ends of a single lncRNA, which resembles the structure of sno-lncRNAs in humans (Figure 3.7B). There was also one example of multiple snoRNAs, a RUF and ncRNA forming a single RNA product (PF3D7lncRNA_2170) (Figure 3.7B). Co-transcription of snoRNAs at this locus has been previously suggested by Chakrabarti et al. (Chakrabarti et al., 2007).

3.4.7 Several lncRNAs may code for small proteins

P. falciparum lncRNAs have an average length of 1146 bp (ranging from 200 bp to 7452 bp) and average GC content of 16%. The GC content is lower than that of the overall *P. falciparum* genome, which is 19.4% but less than that of all non-coding regions, which approaches 10% (Hamid et al., 2014). Transcript lengths and GC content vary between subtypes (Figure 3.7C, D). *Sense* lncRNAs, i.e. located within gene exons, displayed a clear bias towards higher length distributions and greater average GC content. Similarly, *antisense-to-gene* lncRNAs had a higher average GC content than those lncRNAs found in non-coding regions of the genome. Like the Broadbent et al. study, I noted that it is uncommon for *P. falciparum* lncRNAs to contain introns, with only 5% detected in the analysis (Broadbent et al., 2015). Of these lncRNAs, most were *antisense-to-gene* lncRNAs (59%), and the rest consisted of other subtypes. Broadbent et al. previously highlighted lncRNAs that contain multiple introns as notable due to the rareness of this property (Broadbent et al., 2015). In addition to the three examples that they highlighted (lncRNAs associated with *gdv1*, *etramp9* and rRNA methyltransferase), I identified a further 25 lncRNAs that share these features

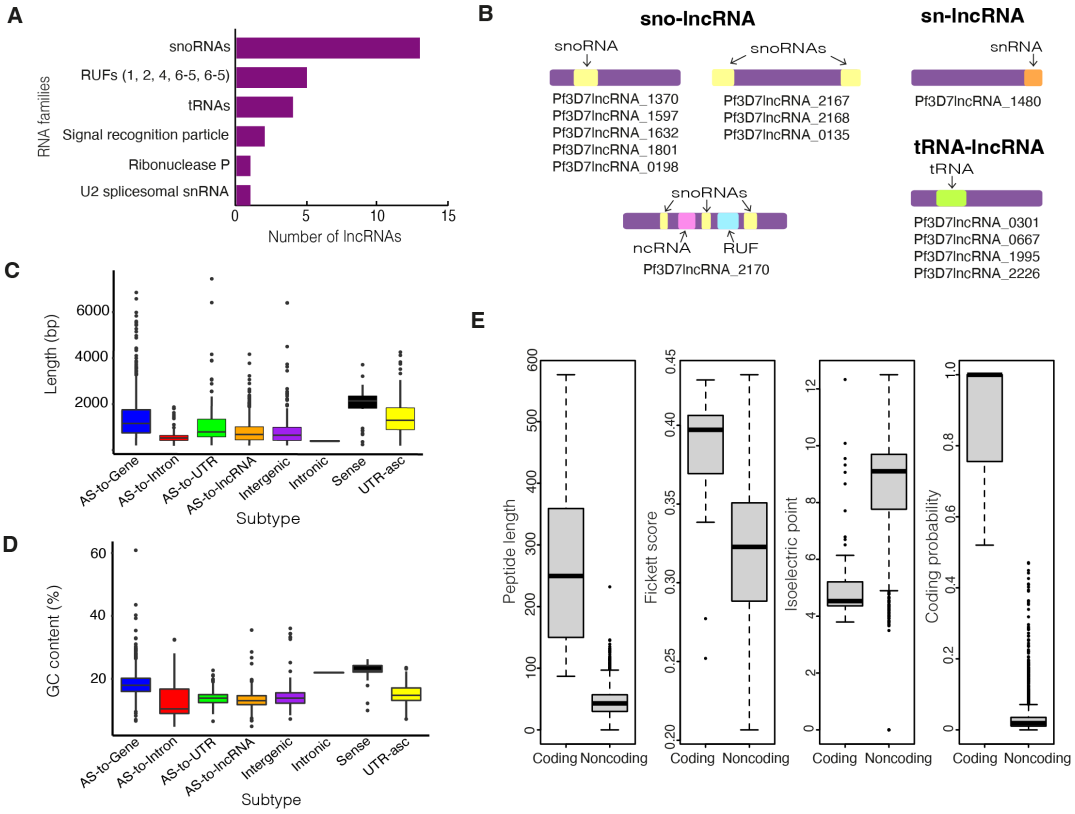


Fig. 3.7 Sequence features of *P. falciparum* lncRNAs. (**A**) 26 RNA families aligned to annotated lncRNA sequences: 13 snoRNAs, 5 RUFs, 4 tRNAs, ribonuclease P, signal recognition peptide and U6 spliceosomal snRNA using Rfam (Kalvari et al., 2018). (**B**) Visual representation and IDs of lncRNAs containing snoRNAs, snRNAs and tRNAs. (**C**) LncRNA lengths ranged from 200-7452 bp, and the distribution of lengths differed between subtypes (AS:antisense to). (**D**) LncRNAs were AT-rich with an average GC content of 15.97%, and the distribution of GC content differed between subtypes (AS:antisense to). (**E**) 16 lncRNAs were identified as putatively coding by sequence feature-dependent coding potential analysis with CPC2 (Kang et al., 2017). The distributions of the measures used to determine coding potential (peptide length, isoelectric point and Fickett score) are presented for putative coding and noncoding lncRNA annotations.

Table 3.5 Noncoding motif analysis of *P. falciparum* lncRNAs

Feature	Family ID	Family	Total Hits	LncRNAs
Signal recognition peptide	RF00017	Metazoa_SRP	1	Pf3D7lncRNA_2143
	RF01856	Protozoa_SRP	1	Pf3D7lncRNA_2143
Ribonuclease P	RF01577	RNase_P	1	Pf3D7lncRNA_0134
RNA of unknown function	RF01578	RUF1	1	Pf3D7lncRNA_1888
	RF01579	RUF2	1	Pf3D7lncRNA_0135
	RF01582	RUF4	1	Pf3D7lncRNA_2170
	RF01581	RUF6-5	2	Pf3D7lncRNA_0063 Pf3D7lncRNA_0501
U2 snRNA	RF00004	U2	1	Pf3D7lncRNA_1480
snoRNAs	RF01583	snoR01	1	Pf3D7lncRNA_0135
	RF01585	snoR07	1	Pf3D7lncRNA_1370
	RF01590	snoR14	1	Pf3D7lncRNA_1597
	RF01591	snoR15	1	Pf3D7lncRNA_1632
	RF01593	snoR16	1	Pf3D7lncRNA_1801
	RF01598	snoR23	1	Pf3D7lncRNA_2167
	RF01599	snoR24	1	Pf3D7lncRNA_2167
	RF01602	snoR27	1	Pf3D7lncRNA_2168
	RF01604	snoR28	1	Pf3D7lncRNA_2168
	RF01603	snoR29	1	Pf3D7lncRNA_2170
	RF01605	snoR30	1	Pf3D7lncRNA_2170
	RF01606	snoR31	1	Pf3D7lncRNA_2170
	RF00133	SNORD33	1	Pf3D7lncRNA_0198
tRNA	RF00005	tRNA	4	Pf3D7lncRNA_0301 Pf3D7lncRNA_0667 Pf3D7lncRNA_1995 Pf3D7lncRNA_2226

LncRNAs containing known RNA families were identified using Rfam batch search (Kalvari et al., 2018).

(supplemental data). Two examples, which are antisense to PF3D7_1115200 (set7) and conserved protein PF3D7_0918400 (unknown function), are shown here (Figure 3.8).

Some apparent lncRNAs might be protein-coding genes missed in previous annotations. In particular, open reading frames (ORFs) encoding small proteins or peptides are hard to identify (Ruiz-Orera and Albà, 2019). Therefore, I calculated the coding potential of each lncRNA using the coding potential calculator algorithm CPC2, which can be used for non-model organisms without retraining the model (Kang et al., 2017). CPC2 uses four sequence-intrinsic features to predict the coding probability of RNA transcripts: Fickett score, ORF length, ORF integrity and isoelectric point. As expected, most lncRNAs were predicted to be noncoding transcripts. However, 16 lncRNAs were determined to have the potential to encode proteins and warrant further investigation (Figure 3.7E, Table 3.6). Most of these putative proteins were 100–150 amino acids in length, and when queried in the Caro et al. *P. falciparum* ribosomal profiling dataset, most had some modest evidence of ribosomal footprints, although often not spanning the length of the lncRNA and would require further experimental validation (Table 3.6) (Caro et al., 2014). Only two shared similarities with other proteins: PF3D7lncRNA_1391, a lncRNA antisense to PF3D7_1116500 (folate transporter 2), and PF3D7lncRNA_0624, an *intergenic* lncRNA, shared similarity with predicted proteins in other *P. falciparum* strains like Dd2 (Figure 3.9). To further evaluate their potential coding capacity, the 16 putative peptide sequences were sent to collaborators (Dr Scott Chisholm and Dr Ross Waller, University of Cambridge) with expertise in *Plasmodium* proteomics. They ran an analysis of human and Pf3D7 predicted proteomes, with and without the putative sequences, against mass-spectrometry data. No peptide matches were identified against our sequences.

3.4.8 LncRNAs secondary structures can be predicted

The secondary structures of the lncRNAs were predicted using RNAfold, a thermodynamic-based RNA prediction method, which demonstrates strong (50-70%) prediction performance for lncRNAs (Gruber et al., 2008; Bugnon et al., 2022). The predicted secondary structures contained an average of 16.7 hairpins, 31.8 inner loops and 6.5 multi-loops, although this varied between subtypes (Figure 3.10A). The secondary structures had an average minimum free energy (MFE) calculation of -188.84 kcal/mol (Figure 3.10B). It is well-reported in the literature that MFE is negatively correlated with RNA length and, to a lesser extent, GC content, which was recapitulated in this dataset (Figure 3.10C, D) (Trotta, 2014). An example

Table 3.6 Sequence intrinsic features of *P. falciparum* lncRNAs determined to have coding potential

LncRNA	Peptide length (AA)	Fickett score	Isoelectric point	ORF integrity	Coding probability	Ribosomal footprints
Pf3D7lncRNA_0624	576	0.28	9.32	1	1.00	No
Pf3D7lncRNA_0937	255	0.35	5.21	1	1.00	Yes
Pf3D7lncRNA_0817	150	0.40	6.14	1	0.93	Yes
Pf3D7lncRNA_0229	174	0.34	6.70	1	0.92	Yes
Pf3D7lncRNA_0892	134	0.40	7.71	1	0.76	Yes
Pf3D7lncRNA_1445	167	0.34	9.06	1	0.74	Yes
Pf3D7lncRNA_1714	116	0.35	4.58	1	0.69	Yes
Pf3D7lncRNA_0100	103	0.40	4.52	1	0.68	Yes
Pf3D7lncRNA_1391	95	0.39	3.79	1	0.66	Yes
Pf3D7lncRNA_2030	143	0.37	10.08	1	0.62	Yes
Pf3D7lncRNA_0682	87	0.25	3.91	1	0.58	Yes
Pf3D7lncRNA_0260	140	0.35	8.66	1	0.55	Yes
Pf3D7lncRNA_0275	155	0.39	12.33	1	0.55	Yes
Pf3D7lncRNA_1822	121	0.37	6.79	1	0.54	Yes
Pf3D7lncRNA_1890	132	0.37	9.55	1	0.53	Yes
Pf3D7lncRNA_0262	119	0.37	6.51	1	0.52	Yes

LncRNAs with coding potential were identified using Coding Potential Calculator (CPC2) (Kang et al., 2017). The features used to make the prediction: peptide length, Fickett score, isoelectric point, ORF length and coding probability, are listed here for the 16 lncRNA determined to have coding potential. Ribosomal footprints from Caro et al. were viewed in PlasmoDB and MochiView (Homann and Johnson, 2010; Caro et al., 2014; Amos et al., 2021). A result of "yes" was reported if a mark was present for any timepoint; otherwise, a result of "no" was reported. AA: amino acids.

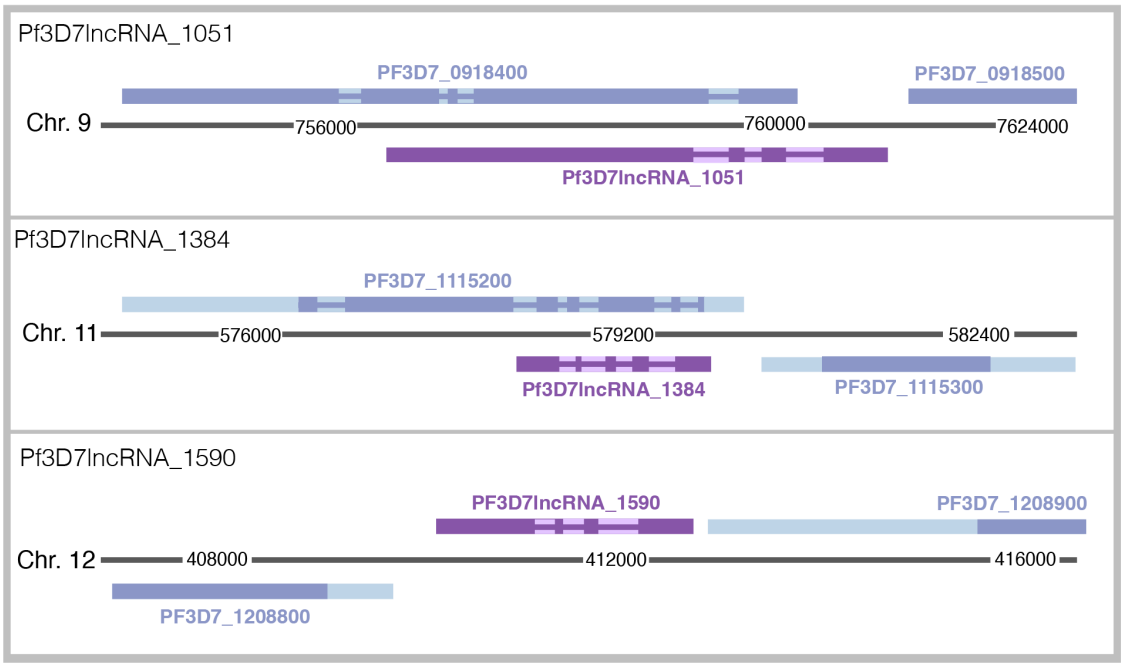


Fig. 3.8 Examples of lncRNAs that contain multiple introns. Schematic of the loci of three lncRNAs that contain multiple introns: PF3D7lncRNA_1051, PF3D7lncRNA_1384 and PF3D7lncRNA_1590.

of a predicted lncRNA secondary structure for PF3D7lncRNA_0120 is shown here (Figure 3.10E).

3.4.9 A subset of lncRNAs may be essential

Intersecting the lncRNA sequencing dataset with that of the *piggyBac* transposon mutagenesis study from Zhang et al. determined that 68% (1602) of lncRNAs are predicted to be non-essential in asexual blood stages (Figure 3.11) (Zhang et al., 2018). In contrast, no *piggyBac* insertions were found in the remaining 32% (767) lncRNA sequences, suggesting these lncRNAs may be essential. Among these, 433 are antisense to or overlapping genes deemed essential, meaning that the essentiality of the protein-coding gene and the lncRNA cannot be disentangled. The remaining 335 lncRNAs are not associated with essential genes and thus may have potentially critical functions for parasite growth and viability. However, the absence of insertions does not definitively demonstrate essentiality.

Fig. 3.9 Putative proteins from lncRNAs with predicted coding potential share sequence similarity with hypothetical proteins from other *P. falciparum* strains. Sequences were obtained from an NCBI BLASTp search and were aligned using Clustal Omega (ClustalW with character counts) (Johnson et al., 2008; Camacho et al., 2009; Sievers et al., 2011). (**A**) Pf3D7lncRNA_1391 shared sequence similarity with hypothetical proteins PFLG_00935 from *P. falciparum* RAJ116 (e-value: 8e-35) and PFDG_04709 from *P. falciparum* Dd2 (e-value: 3e-28). (**B**) Pf3D7lncRNA_0624 shared sequence similarity with many hypothetical proteins, a few of which are shown here: PFDG_04022 and PF3D7_01724 from *P. falciparum* Dd2 (e-value: 0), Keratin KB40 from *P. falciparum* RAJ116 (e-value: 0) and Thr-Ser protein *P. falciparum* IGH-CR14 (e-value: 0). The NCBI IDs for these sequences are KNC35879.1, KOB89161.1, KOB87748.1, KOB86616.1, KNC36782.1 and KNG78363.1.

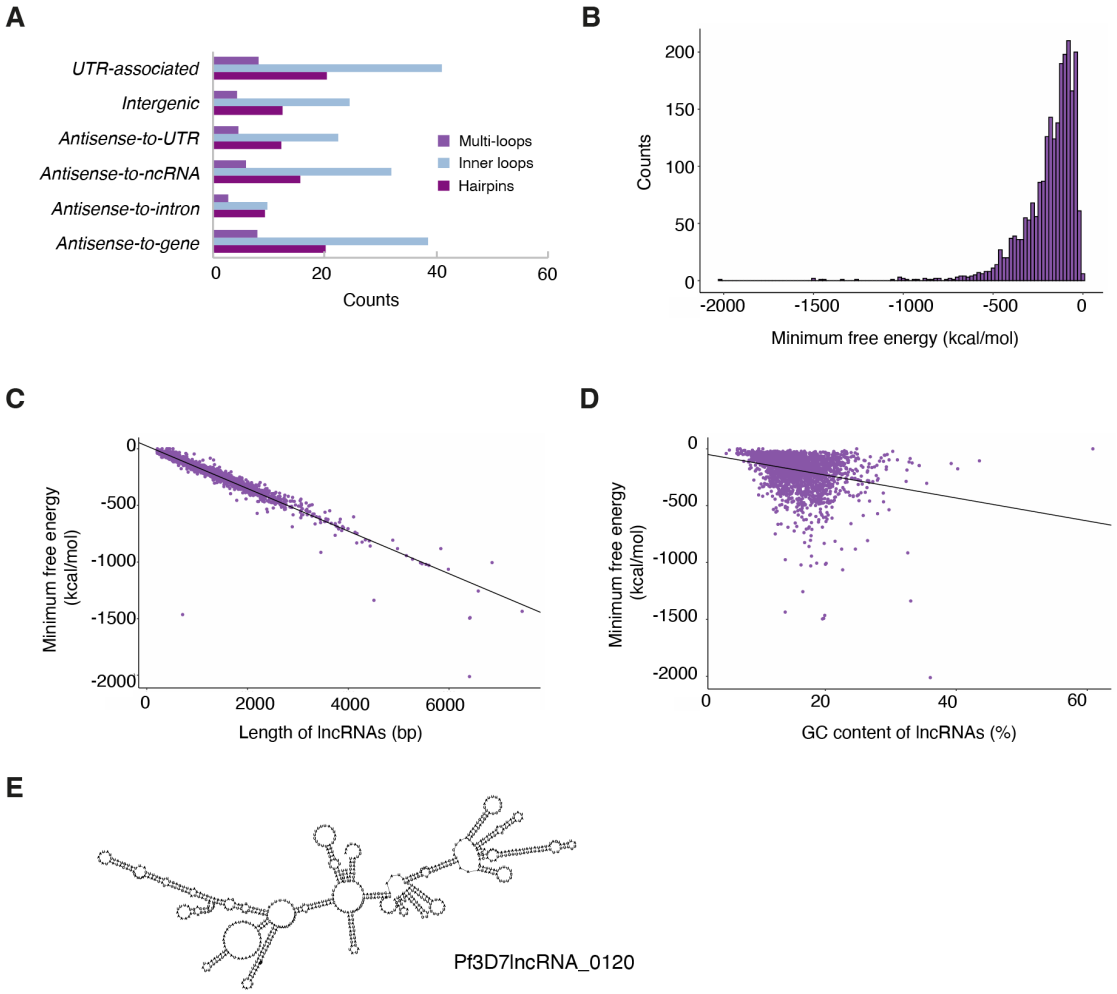


Fig. 3.10 Secondary structure prediction of *P. falciparum* lncRNAs. (**A**) The distribution of minimum free energy calculations for predicted lncRNA secondary structures. (**B**) The average number of multi-loops, inner loops and hairpins predicted for each lncRNA genomic context subtype. The distribution of minimum free energy calculations for predicted lncRNA secondary structures by (**C**) lncRNA length and (**D**) GC content. (**E**) An example of a predicted secondary structure for PF3D7lncRNA_0120.

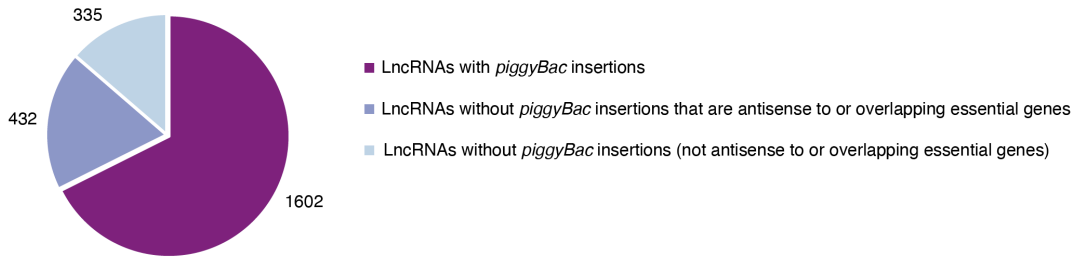


Fig. 3.11 The majority of lncRNAs can be disrupted by the *piggyBac* transposon system. The lncRNA annotations were intersected with insertion site positions from surviving mutants in the Zhang et al. *piggyBac* transposon mutagenesis screen (Zhang et al., 2018). 1602 lncRNAs overlapped with insertion positions suggesting that these lncRNAs are not essential, while the remaining 767 lncRNAs did not have any insertions. Of these 767 lncRNAs, 432 are antisense to or overlapping with genes identified as essential in the Zhang et al. study and, therefore, cannot be discerned for essentiality. The remaining 335 are potentially essential lncRNAs.

3.4.10 The level of conservation of *P. falciparum* lncRNAs remains unclear

An aspect of lncRNAs that can indicate biological relevance is conservation within strains or species. To assess the conservation of *P. falciparum* lncRNAs, a preliminary investigation was completed exploring approaches using whole genome alignment tools (Goujon et al., 2010; Hickey, 2021). A pilot investigation using BLAST was completed exploring the conservation of the 335 putatively essential lncRNAs identified in the *piggyBac* analysis in seven strains (3D7, 7G8, Dd2, GB4, HB3, KH01 and CN01). 293 lncRNAs had hits in all strains, 36 lncRNAs were not conserved in all strains, and 7 lncRNAs were not conserved in any other strain. The conserved lncRNAs demonstrated a high level of sequence conservation, with 99.3% of lncRNAs having e-values approaching zero ($<10^{-100}$). It became evident that an analysis of lncRNA conservation would require a more complex approach beyond the time scale of this thesis. For instance, the sequence conservation of lncRNAs needs to be compared with the sequence conservation of non-coding sequences that do not encode ncRNA molecules.

I also endeavoured to create a *P. knowlesi* lncRNA annotation from PacBio ISO-Seq RNA sequencing using a congruent manual annotation approach supported by short-read RNA sequencing and updated UTR annotations (obtained from Dr Adam Reid and Dr Lia Chappell, unpublished). However, the sequencing could not be reliably mapped to the correct strand.

The strand mapping was improved by trialling different mappers (minimap2 and HISAT2) but was insufficient to enable lncRNA annotation (HISAT2 mapping was completed by Dr Adam Reid) (Li, 2018; Kim et al., 2019). In addition, the sequencing was more fragmented than the ONT sequencing used for the *P. falciparum* annotation, which also posed a challenge to the annotation of longer RNAs.

3.4.11 Two novel lncRNAs associated with *var* genes

Three types of *var*-associated ncRNAs have been described in *P. falciparum*: an *antisense-to-intron* lncRNA, a *sense* lncRNA (overlapping exon 2) and a GC-rich RUF6 ncRNA (usually in a head-to-head configuration and 135 bp in length) (Figure 3.12A) (Epp et al., 2009; Amit-Avraham et al., 2015; Barcons-Simon et al., 2020). Previous research has suggested that these ncRNAs are widespread in *var* genes. To understand if they are expressed at all *var* loci in mixed asexual blood stages, I analysed each *var* locus. *Antisense-to-intron* lncRNAs were identified in 51 *var* genes, while *sense* lncRNAs were found in only 36 *var* genes. Only 2 GC-rich RUF6 ncRNAs were detected (Figure 3.13A). Through CRISPR-interference knockdown, these ncRNAs have been shown to activate the expression of 15 *var* genes *in trans* through predominant transcription of a single member adjacent to the active *var* gene (Barcons-Simon et al., 2020; Fan et al., 2020). GC-rich RUF6 ncRNAs PF3D7_071270 and PF3D7_1240800 were detected in the Lee et al. ONT long-read sequencing dataset and the latter was adjacent to the single active *var* gene (PF3D7_1240900) (Lee et al., 2021). No RUF6 ncRNAs were detected in the sequencing data generated in this study. However, the active *var* gene (PF3D7_1200600, also known as *var2csa*) is not proximal to a RUF6 ncRNA (Guizetti et al., 2016). I also identified two additional lncRNAs at *var* loci that had not been previously described. A downstream *intergenic* lncRNA was detected close to 31 *var* genes, and an *antisense-to-gene* lncRNA (antisense to exon 2) was detected in 28 *var* genes (Figure 3.12A). Examples of lncRNAs at specific *var* gene loci are shown here (Figure 3.12B).

3.4.12 Most lncRNA-TAREs were not detected in this dataset

Other studies have found lncRNAs that are transcribed from subtelomeric regions of the *P. falciparum* genome. LncRNAs-TAREs have been detected on 22 chromosome ends in schizont-staged parasites and associated with telomere maintenance during replication (Broadbent et al., 2011; Sierra-Miranda et al., 2012). LncRNAs were not fully captured by the long-read RNA sequencing generated in this study and obtained from Lee et al. (Lee et al., 2021). Only three reads overlapped with three different lncRNA-TAREs: lncRNA-TARE-4L,

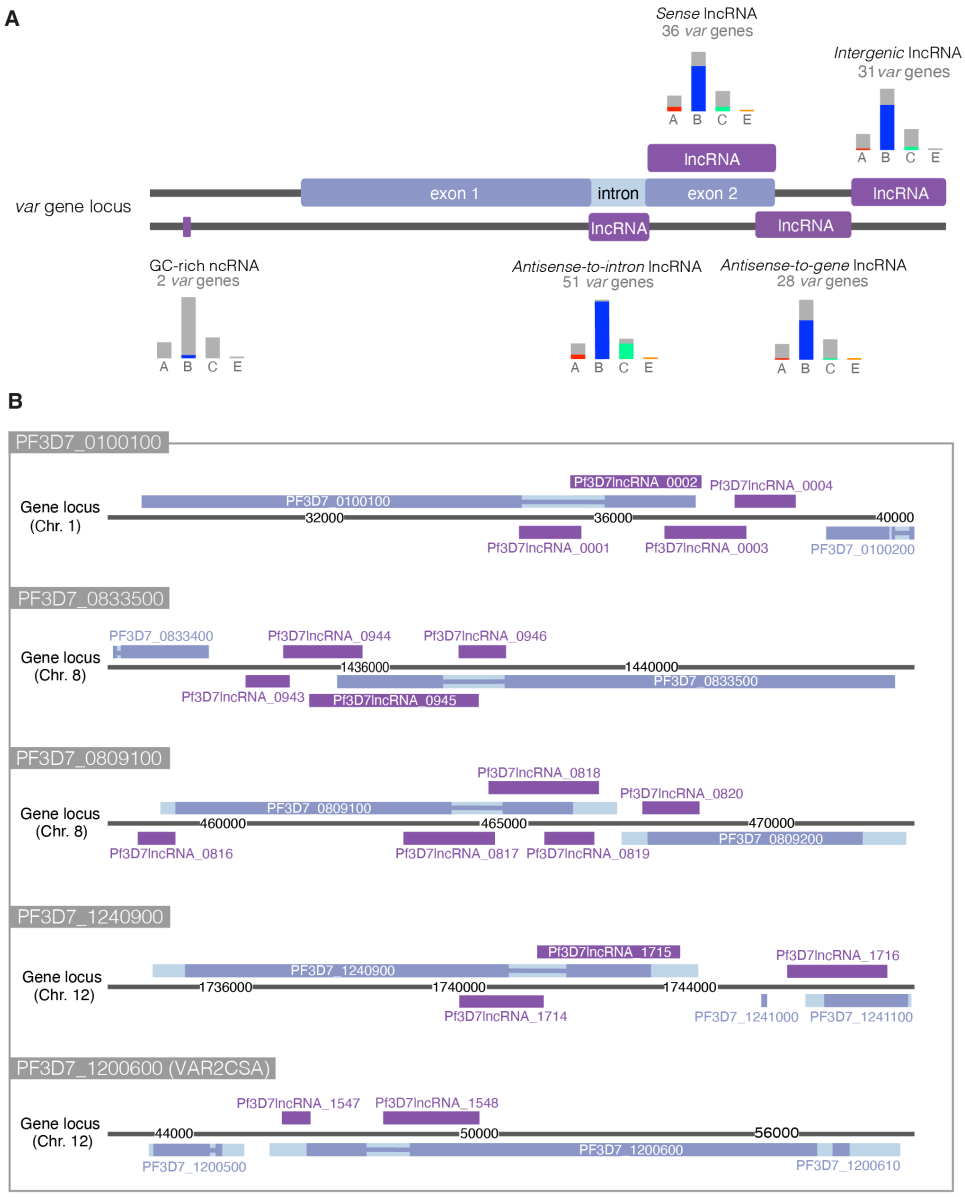


Fig. 3.12 LncRNAs present at *var* gene loci. (**A**) Schematic of lncRNAs found at *var* gene loci and the number of loci containing each subtype in this study. Bar charts denote the distribution of *var* gene subtypes for these loci by colour (A-red, B-blue, C-green and E-orange). The GC-rich (RUF6) family, *antisense-to-intron*, and *sense* lncRNAs are well-established *var*-associated lncRNAs. This study observed two additional lncRNAs, a downstream *intergenic* lncRNA and an *antisense-to-gene* lncRNA. (**B**) Examples of lncRNAs at five *var* gene loci.

lncRNA-TARE-5L and lncRNA-TARE-13L (Figure 3.13B). As this group of lncRNAs are most highly expressed during schizogony, I investigated if the RNA sequencing used in this study adequately captured lncRNAs expressed in schizont stages. I obtained the stage-specific expression data from Broadbent et al., which assigned a max-expression timepoint for each lncRNA. I then determined for each timepoint the proportions of stage-specific lncRNAs that were captured in this study (Broadbent et al., 2015) (Figure 3.13C). LncRNAs from each stage were represented in similar proportions, verifying that the parasite cultures used to generate the RNA sequencing datasets included all stages in the intraerythrocytic development cycle and were not biased towards specific stages.

3.5 Discussion and future outlook

Long noncoding RNAs are involved in regulating developmental pathways and immune evasion strategies in the malaria parasite *Plasmodium falciparum* (Amit-Avraham et al., 2015; Filarsky et al., 2018; Gomes et al., 2022). Evidence suggests that thousands of genes encode lncRNAs, but due to limited research and the lack of necessary experimental tools, our understanding of their broader role remains poor (Broadbent et al., 2015; Yang et al., 2021). In this study, I sought to provide a basis for future research into these elements by improving their annotation in the *P. falciparum* genome. I employed manual curation, an approach not previously used for *P. falciparum* lncRNAs, in combination with long-read sequencing, to generate a more comprehensive annotation. Long-read sequencing provided a clear improvement in capturing full-length lncRNAs, by enabling the accurate determination of lncRNA boundaries and providing sequence coverage for lncRNAs not captured previously by short-read sequencing. For instance, I identified several lncRNAs that had previously been annotated as multiple lncRNA units. I also expanded the annotation significantly, suggesting the total number of lncRNAs in *P. falciparum* is over two thousand, which aligns with recent transcriptomic studies that have predicted thousands of ncRNA transcripts (Chappell et al., 2020; Yang et al., 2021). Furthermore, manual curation allowed me to harness the plethora of publicly available datasets to create high-quality genome annotations. Context and supportive evidence were investigated to facilitate each annotation and reduce errors common to automated annotation.

The characterisation of the genomic and sequence features of lncRNAs in this work predominantly validates the findings in the literature. LncRNAs are widespread throughout the genome, often found at sites with bidirectional promoters, and their sequences are AT-rich,

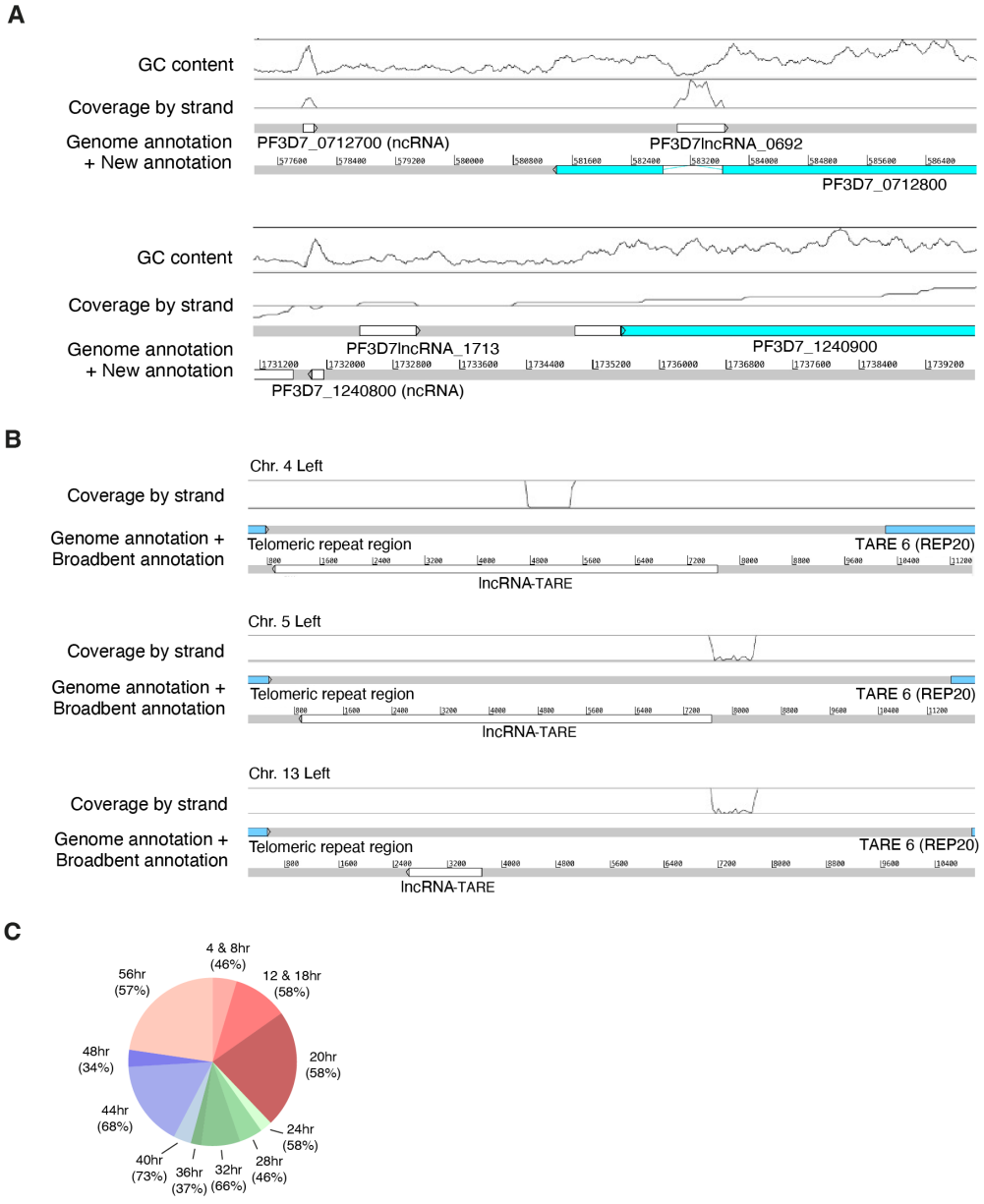


Fig. 3.13 GC-rich *var*-associated ncRNAs and lncRNA-TAREs were not fully captured by the long-read sequencing (**A**) Two members of the GC-rich *var*-associated ncRNA family were captured in the long-read sequencing, snapshots from Artemis show these ncRNAs (Carver et al., 2012; Amos et al., 2021). (**B**) The long-read sequencing compiled in this study did not fully capture lncRNA-TAREs. Three lncRNAs were found in the subtelomeric region (antisense to the TAREs); however, their length and location differed from the Broadbent annotation. Snapshots from Artemis show these lncRNAs (Carver et al., 2012; Broadbent et al., 2015). (**C**) The proportion of Broadbent et al. lncRNAs that were captured in this study for each max-expression timepoint (Broadbent et al., 2015).

vary in length and contain few known RNA motifs. I introduced a genome context-specific classification system instead of the simplified intergenic and antisense lncRNA system. This allows rich information on the genome context of each lncRNA, which provides a helpful tool for wet lab applications. For instance, off-target effects are of utmost concern in experiments targeting lncRNAs for genetic modification. Presenting contextual subtypes may enable differing approaches to be refined for *in vitro* study. The subtype classification also allowed genes associated with specific lncRNA subtypes to be identified using gene ontology. It is evident that *sense* and *antisense-to-intron* lncRNAs are subtypes almost exclusive to *var* genes, barring nine additional genes with *antisense-to-intron* lncRNAs, three of which are rifins and one is a *var* pseudogene. The other subtypes are associated with various genes but are enriched for certain biological processes. Interestingly, the *antisense-to-gene* lncRNA subtype was enriched for genes involved in multiple nucleoside and nucleotide processes, with almost all genes labelled with these and related GO terms contextually associated with a lncRNA of this subtype. Genes involved in protein processes and cell-cell adhesion were also enriched in the *antisense-to-gene* lncRNA subtype, such as the new lncRNAs I identified at *var* loci. Genes enriched in the *antisense-to-UTR* lncRNA subtype were involved in cytoplasmic translation, immune system processes, chromatin organisation and transport, which included most proteins involved in translation, such as the elongation initiation factor genes, ribonucleoproteins and epigenetic proteins like histones. LncRNAs could regulate these biological processes and others, although due to the nature of significance testing in gene ontology enrichment, poorly represented GO terms can lead to spurious associations if they are represented in a sample. Further studies on transcriptional expression and biological interactions are required to define these possible roles.

Most of the well-studied lncRNAs from the literature were verified in this study, although I did not fully capture the lncRNA-TAREs. These lncRNAs could have been absent due to the stage-specificity of their expression. LncRNA-TARE expression peaks during parasite invasion, and therefore, a mixed culture would not be expected to contain large numbers of these parasites (Broadbent et al., 2011, 2015). There could also have been challenges in mapping their highly repetitive sequences. The three lncRNA-TAREs observed were much shorter than expected (Figure 3.13). However, alternative transcription or post-transcriptional processing could explain these short transcripts, which has been observed in lncRNA-TAREs (Broadbent et al., 2011; Sierra-Miranda et al., 2012). Sequencing more deeply with long reads and from a wide range of life stages would likely capture these lncRNAs and improve the annotation further. Moreover, high-quality datasets should continue to be mined as

a resource for lncRNA annotation or supportive evidence. For example, since this work was completed, a long-read RNA-Seq dataset has become available that combines both ONT and PacBio sequencing (Shaw et al., 2022). Another group of lncRNAs that I did not verify in this study is the circRNAs that were previously predicted by Broadbent et al. (Broadbent et al., 2011). However, in this case, no circRNA prediction was completed. Moving forward, bioinformatics tools could be used to examine if the annotated lncRNAs form circular structures (López-Jiménez et al., 2018).

It has been suggested that some *P. falciparum* genes resembling lncRNAs could encode short polypeptides (Yang et al., 2021). But it is unclear how effective coding potential tools developed using mammalian or bacterial systems are for the *P. falciparum* transcriptome. I identified a small subset of 16 lncRNAs with a high coding probability however, these peptides cannot be found in *P. falciparum* mass spectrometry data and most lacked any sequence similarity to any known proteins. This suggests that these sequences are not likely to be coding and that protein-coding prediction tools used in isolation are not sufficient to predict the coding potential of *P. falciparum* lncRNAs.

I also identified lncRNAs that could be classified based on containing shorter structural ncRNAs, such as snoRNAs, snRNAs and tRNAs. Although these structural ncRNA-lncRNAs hybrids have not been previously reported in *P. falciparum*, they have been observed in other species. In humans, snoRNAs at the Prader-Willi Syndrome locus have been shown to exist as sno-lncRNAs (lncRNA flanked by two snoRNAs) and SPA-lncRNAs ($5'$ snoRNA capped and $3'$ polyadenylated lncRNA), which play a role in post-transcriptional processing of snoRNAs and regulate mRNA metabolism through association with RNA-binding proteins (Yin et al., 2012; Wu et al., 2016), respectively. The lncRNAs identified in this study could play a similar role in regulating these structural ncRNAs involved in mRNA metabolism and protein synthesis. Or even more simply, these lncRNAs could be processed into snoRNAs in a way similar to genes that contain snoRNAs that splice out and process the snoRNAs from pre-snoRNAs. Further investigation is needed to determine if there is a role for these lncRNAs.

To further the impact of the annotation, transcriptomic data from other stages in the lifecycle and other *P. falciparum* strains and *Plasmodium* species should be analysed. Like coding genes, lncRNAs display dynamic regulation across asexual blood stages, but little is known about their regulation across other stages of the parasite lifecycle (Broadbent et al., 2015). Studies examining lncRNA expression in other stages, such as gametocyte and liver stages,

are necessary to provide a complete lncRNA annotation and better understand their regulation and potential functional roles. A recent preprint has indicated that more lncRNAs with roles in gametocytogenesis are being discovered (Batugedara et al., 2022). It is more difficult to predict the conservation of noncoding RNAs than that of coding RNAs, where the amino acid code restricts sequence variation. I was unable to sufficiently analyse the inter and intra-species conservation of *P. falciparum* lncRNAs. Further investigation is needed to enable a comprehensive analysis of lncRNA conservation in *P. falciparum*.

I examined many characteristics that may indicate that a lncRNA is biologically relevant or has a biological function such as essentiality, secondary structure and GO term analysis. However, these data are not sufficient to define a role for lncRNAs in the regulation of the *P. falciparum* transcriptome. Extensive *in vitro* studies are required to validate the presence of these lncRNAs and, subsequently, characterise their features and elucidate their functions. LncRNAs have many possible mechanisms to regulate gene expression, and new advances in CRISPR technology may enable the deciphering of the specific functions of *P. falciparum* lncRNAs. This lncRNA annotation will support future studies by providing high-quality sequence annotations that can facilitate functional characterisation, which is demonstrated in *Chapter 5*.

3.6 Additional methodology

3.6.1 Long-read RNA-Seq

Two long-read libraries (with and without exonuclease treatment) were prepared by running the Pf3D7 RNA samples (prepared by Dr Sophie Adjalley) on the Oxford Nanopore GridION using the direct RNA-Seq protocol, avoiding PCR amplification. Exonuclease treatment (TEX) with exonuclease 2 spiked in was used to enrich primary transcripts as sequencing from both libraries was later combined. Raw data from exonuclease treated and untreated RNA samples in .fast5 files were converted into fastq files of reads using the base caller Guppy v3.1.5. The exonuclease-treated (TEX plus) sample yielded 55130 reads. The untreated (TEX minus) sample yielded 377999 reads. The reads were then mapped against Pf3D7 v3 reference (plus the enolase 2 gene sequence, which is spiked into samples as a control) using minimap2 (Li, 2018; Amos et al., 2021) (completed by Dr Adam Reid). The two sets of reads were then merged and used for annotation. The median length of the combined read set was 852 bp, with the longest read being 12084 bp.

3.6.2 Short-read RNA-Seq

Unpublished mapped short-read RNA sequencing was obtained from Dr Vandana Thathy and Dr Chris Newbold (University of Oxford). The short-read libraries had been prepared using the Illumina TruSeq kit. They were sequenced using Illumina HiSeq (ENA project ERP104547) as 150 bp paired-end reads and mapped to the Pf3D7 (v3) reference genome using HISAT2 (v2.0.0).

3.6.3 Data collection, curation and visualisation

Previous lncRNA annotations were obtained from Liao et al., Broadbent et al. and PlasmoDB (Liao et al., 2014; Broadbent et al., 2015; Amos et al., 2021). The predicted transcripts from the Chappell et al. and Yang et al. studies were obtained from the supplemental material and the authors, respectively (Chappell et al., 2020; Yang et al., 2021). The gene IDs of genes with natural antisense transcripts were derived from the Siegel et al. publication due to the absence of antisense transcript coordinate information (Siegel et al., 2014). Additional RNA sequencing datasets were downloaded from PlasmoDB (Amos et al., 2021). Sequences were viewed using Artemis, with separate windows created for GC content, long-read and short-read sequencing datasets, TSS datasets, and genome annotations (Carver et al., 2012). The plot for supportive evidence was generated with webserver sankeyMATIC, and the comparative annotation analysis plot was generated with UpSetR (v1.4.0) (Conway et al., 2017; Bogart, 2022).

3.6.4 Sequence, structure and coding potential analyses

The comparative analyses with other annotations were completed using Bedtools (Quinlan and Hall, 2010). Location-based clustering of lncRNAs by subtype was completed using Cluster Locator (v2) with max-gap = 2 (Pazos Obregón et al., 2018). Gene ontology (GO) enrichment analyses for antisense and overlapping genes were completed in PlasmoDB (v56) and visualised using REVIGO (v1) (Supek et al., 2011; Amos et al., 2021). A motif and RNA families search was completed using Rfam (v14.7) batch search (Kalvari et al., 2018). Seqkit was used to obtain AT content information and length, and the presence of exons was determined during annotation (Shen et al., 2016). Coding-potential, based on intrinsic sequence features, was analysed using Coding Potential Calculator (v2) and putative proteins were queried in BLAST against all other proteins (blastp and tblastn, v2.12.0) and Pfam (v35.0) and aligned using Clustal Omega (Camacho et al., 2009; Sievers et al., 2011; Kang

et al., 2017; Mistry et al., 2020). RNAfold was used to predict RNA sequence structures and obtain dot-bracket notations and MFE calculations (Gruber et al., 2008). A python script was used to count hairpins, loops and multi-loops (Schudoma, 2018). Ribosomal footprints were observed in MochiView (v1.46) (Homann and Johnson, 2010). Plots were created in R using ggplot2 (v3.3.5) and idiogramFISH (v1.16.1) packages (Lex et al., 2014).

3.6.5 Conservation analysis

Sequences for *P. falciparum* strains 3D7, 7G8, Dd2, GB4, HB3, KH01 and CN01 were obtained from PlasmoDB (Amos et al., 2021). BLAST multiple sequence alignments were run for each putatively essential lncRNA from the command line (Goujon et al., 2010).

3.6.6 *P. knowlesi* sequencing

RNA was extracted from asexual mixed intraerythrocytic stages of *P. knowlesi* by Dr Sophie Adjalley. The samples were sequenced with PacBio Iso-Seq sequencing and mapped with minimap2 and HISAT2 (by Dr Adam Reid and myself) to the PKNH v2 reference (Li, 2018; Kim et al., 2019; Amos et al., 2021). The sequences, along with short-read RNA-Seq and UTR annotations, were viewed in Artemis (Carver et al., 2012).

Chapter 4

Developing molecular tools to study *Plasmodium falciparum* lncRNAs *in vitro*

4.1 Overview

Molecular tools are needed to enable the efficient and scalable study of lncRNAs in *Plasmodium falciparum*. In this chapter, I discuss the development of two tools: one for perturbation and one to facilitate phenotypic assays. I explored using CRISPR-based approaches to disrupt lncRNA expression *in vitro* by attempting to adapt Cas13b and Cas13d for use in *P. falciparum*. To support the characterisation of lncRNA-disrupted mutants, I developed a system to rapidly generate fluorescent *P. falciparum* lines to facilitate competition assays with strain-matched fluorescent parasites.

Portions of this chapter have been published in *Frontiers in Cellular and Infection Microbiology* as *Efficient generation of mNeonGreen Plasmodium falciparum reporter lines enables quantitative fitness analysis* (Hoshizaki et al., 2022b). This work contains data obtained from plasmids created by colleagues: CRISPR-Cas9 (Adjalley and Lee, 2022), CRISPR-Cas13b (Dr Marcus Lee, unpublished), CRISPR-dCas9-loxP-attP plasmid (Dr Sophie Adjalley, unpublished), and *pfpare*-targeting CRISPR-Cas9 (Dr Hannah Jagoe, unpublished). These colleagues are cited in the relevant sections for their contributions. Supplemental data and acknowledgements are in Appendix B.

4.2 Introduction

Reverse genetics entails targeting specific genes for modification and then observing the phenotypic consequences. Reverse genetics has allowed scientists to elucidate the biological functions of particular lncRNAs and comment on their role more broadly in organisms (Gao et al., 2020). Few *P. falciparum* lncRNAs have been studied experimentally due to the absence of an efficient and scalable reverse genetics approach. Two tools that would facilitate this approach but require further development are tools for lncRNA disruption and the rapid generation of fluorescent lines that would enable the phenotyping of lncRNA-disrupted mutants.

4.2.1 Challenges with developing molecular tools in *P. falciparum*

Plasmodium species vary biologically from other eukaryotes, particularly from the species considered in the design and testing of molecular tools. Firstly, *Plasmodium* parasites are deficient in RNA interference machinery, eliminating the possibility of co-opting an endogenous RNAi pathway for RNA targeting (Baum et al., 2009). Secondly, the parasites lack a major DNA repair pathway used by eukaryotes, canonical non-homologous end joining (NHEJ), which facilitates the repair of double-stranded DNA breaks in the absence of a template (Kirkman et al., 2014). Instead, the parasites primarily use the other major DNA repair mechanism: homology-directed repair (HDR). Thirdly, the time to recrudescence after transfection is longer than other protozoans, approximately 4-6 weeks compared to 1 week in other protozoans (Beneke et al., 2017; Zhang et al., 2017; Bryant et al., 2019). Furthermore, to study *P. falciparum* specifically, there are additional challenges unique to this species. *P. falciparum* exhibits lower transfection efficiency compared to other *Plasmodium* species (Skinner-Adams et al., 2003; Caro et al., 2012; Moon et al., 2013; Bushell et al., 2017). Also, the extremely AT-rich *P. falciparum* genome limits the selection of targets, impacts the design, specificity and efficacy of oligonucleotide primers, guide RNAs and donor sequences and the stability of transfection plasmids. For these reasons, the application of molecular tools into *P. falciparum* has been challenging.

4.2.2 Current approaches for targeting *P. falciparum* lncRNAs *in vitro*

The literature has four examples of lncRNA-targeting in *P. falciparum*, all of which use a CRISPR-Cas9 deletion approach. The deletion is either in the part of the lncRNA that does not overlap an antisense gene or, in the case of intron-derived lncRNAs, in the intron that acts

as a promoter or TSS to the antisense lncRNA. In 2018, Filarsky et al. disrupted the *gdv1*-lncRNA through CRISPR-Cas9 insertion of the human dihydrofolate reductase (hDHFR) or blasticidin S deaminase (BSD) resistance cassettes in the F12 strain and 3D7 conditional AP2-G mutants (Filarsky et al., 2018). *gdv1*-lncRNA transcripts were undetectable in these mutants. A recent preprint describes the disruption of a lncRNA on chromosome 14 by inserting a BSD resistance cassette (Batugedara et al., 2022). Bryant et al. disrupted the intron-derived antisense lncRNA of *var2csa* using CRISPR-Cas9 in the 3D7 strain (Bryant et al., 2017). Complete deletion of the *var2csa* intron, the lncRNA's promoter, resulted in the abrogation of transcription. In 2022, Gomes et al. also disrupted an intron-derived antisense lncRNA using complete intron deletion using Cas9 but in the NF54 strain (Gomes et al., 2022). Removing the *md1*-lncRNA transcriptional start site within intron 1 of *md1* resulted in a 5-fold reduction in the expression of the lncRNA.

It is necessary that current *P. falciparum* CRISPR tools are evaluated based on their potential application to the study of lncRNAs, and new alternative CRISPR tools are explored. The Cas9-deletion approach has limited application as it is only suitable for the minority of lncRNAs that are intergenic or lncRNAs that have promoters in non-coding sequences, where a perturbation at the genome level will be less likely to impact neighbouring genes. The deletion of promoters, UTRs or a change in chromosome conformation could affect the expression of neighbouring genes. The Cas9 approach also often requires the design of donor sequences, which adds a layer of complication to genome-wide disruption screens. Another major issue with only using gene knockout approaches for lncRNA disruption is that the impact of transcribing the lncRNA and the functionality of the lncRNA molecule itself cannot be distinguished. CRISPR interference approaches that use catalytically-inactived enzymes like dCas9 (dead Cas9) provide suitable replacements for RNAi and have been used for studying lncRNAs (Stojic et al., 2018).

dCas9 has been used to interfere with the transcription of the GC-rich ncRNA family in the 3D7 strain (Barcons-Simon et al., 2020). Although these ncRNAs are too short (135 bp) to be classified as lncRNAs, they exemplify another disruption approach. Barcons-Simon et al. targeted a homologous region common to all 15 members of the RUF6 family. They demonstrated enrichment of dCas9 at all the target loci and at least 3-fold downregulation of ncRNA expression. In the Gomes et al. study, they also used dCas9 to target the promoter of *md1* but not the lncRNA (Gomes et al., 2022). They do not report if this gene interference approach interfered with the expression of the lncRNA.

Recent reviews have suggested that new CRISPR tools like Cpf1 and Cas13 should be investigated as strategies for CRISPR disruption in *Plasmodium* (Lee et al., 2019; Bryant et al., 2019; Quansah et al., 2023). Cpf1 (Cas12) was recently adopted to *P. falciparum* for diagnostics. It functions like Cas9, although has the advantage of a "TTTN" PAM site that can enable the targeting of extremely AT-rich sequences, such as lncRNAs (Zhao et al., 2020; Nessel et al., 2020). Meanwhile, Cas13 enzymes are a type VI CRISPR system that contains a programmable single-effector RNA-guided ribonuclease (Abudayyeh et al., 2017). Cas13 enzymes have recently been applied to lncRNA disruption to target *HOTTIP*, *MALAT*, *GACAT3* and the new class of very long (>50 kb) *intergenic* ncRNAs (vlincRNAs) (Konermann et al., 2018; Xu et al., 2020; Zhang et al., 2021). There are multiple classes of Cas13 enzymes; Cas13b was initially identified as the most efficient Cas13 ortholog for RNA knockdown with high specificity and no detectable off-target interactions (Smargon et al., 2017; Cox et al., 2017). Whereas Cas13d (also known as CasRx) was discovered later and demonstrated some advantages over Cas13b, such as a more compact structure while maintaining high efficiency and specificity in RNA targeting in human cells (Konermann et al., 2018). In this chapter, I explored the application of novel RNA-targeting CRISPR enzymes, Cas13b and CasRx, into *P. falciparum* while also evaluating current CRISPR approaches such as Cas9 and catalytically-inactive Cpf1 (dCas12a).

4.2.3 Developing tools for phenotyping lncRNA-disrupted mutants

The generation of fluorescent *Plasmodium* lines has been instrumental in interrogating gene function and providing biological insights into drug activity, life cycle, and host-parasite interactions (Frischknecht et al., 2006; Talman et al., 2010; Wilson et al., 2010; Portugaliza et al., 2019; Voorberg-van der Wel et al., 2020; Thommen et al., 2022). An important use is for quantifying parasite fitness via competitive head-to-head growth assays. A line containing an engineered mutation of interest, such as a lncRNA knockout or knockdown, is queried against a fluorescent isogenic wild-type parasite. The change in relative abundance over time between the queried and fluorescent lines provides a measure of the fitness impact of the mutation (Baragaña et al., 2015; Gabryszewski et al., 2016; Ross et al., 2018; Stokes et al., 2021).

The first *Plasmodium* reporter lines expressed the exogenous proteins firefly luciferase and chloramphenicol using episomes (Goonewardene et al., 1993; Wu et al., 1995; Horrocks and Kilbey, 1996). Shortly after, green fluorescent protein (GFP) from *Aequorea victoria* was developed as a reporter and adapted into *P. falciparum*. GFP became widely used and is

still the predominant fluorescent marker today as it provides stable and strong fluorescence without requiring a cofactor (Chalfie et al., 1994; VanWye and Haldar, 1997). The reporter lines enabled functional and genetic analyses, particularly the bioluminescent and fluorescent reporters, which supported imaging. The application of other reporters such as mCherry and yellow fluorescent protein and the transition to integrated reporters were gradual due to hindrances with homologous recombination-based integration (Armstrong and Goldberg, 2007; Engelmann et al., 2009).

The advent of CRISPR-Cas9 genome editing and its application in *P. falciparum* made the development of reporter lines straightforward, allowing for the selective integration of fluorescent markers into specific genomic sites (Mogollon et al., 2016; Kuang et al., 2017; Miyazaki et al., 2020). Favourable integration sites could be selected based on the essentiality and phenotype of the gene and its expression profile, allowing for the development of reporter lines with stage-specific or multi-stage expression (Marin-Mogollon et al., 2019; Miyazaki et al., 2020). Furthermore, reporters in different strains would support phenotypic analyses where the reporter line must be the same strain or parental line as the queried strain. Especially in the use case of competition assays, where different *P. falciparum* strains or lab-adapted lines can vary substantially in fitness. Although, the diversity of the current repertoire of *Plasmodium* fluorescent reporters is limited to the Dd2 and NF54 strains.

In this chapter, I developed an efficient CRISPR-Cas9 approach for the rapid generation of fluorescent *P. falciparum* lines and explored the application of mNeonGreen, a newly developed fluorescent protein, in *P. falciparum*. Recently developed fluorescent proteins like mNeonGreen and mGreenLantern could generate better performing *Plasmodium* reporter lines as these proteins have a 3- and 6-fold increase in brightness compared to GFP, respectively (Shaner et al., 2013; Campbell et al., 2020). Compared to standard fluorescent proteins, these fluorescent reporters exhibit faster maturation, improved acid tolerance, increased photostability, and enhanced thermostability. mNeonGreen has been successfully applied to protozoans, including apicomplexans, trypanosomes and *Toxoplasma* (Beneke et al., 2017; Glushakova et al., 2018; Markus et al., 2019).

4.3 Objective and Aims

To develop molecular tools to facilitate the study of *P. falciparum* lncRNAs *in vitro*.

1. Generate a Cas13 tool for RNA knockdown in *P. falciparum*

2. Apply existing CRISPR-based approaches to lncRNA disruption in *P. falciparum*

3. Develop a tool for the rapid generation of mNeonGreen-fluorescent lines

4.4 Results

4.4.1 Evaluation of CRISPR approaches for targeting *P. falciparum* lncRNAs

The CRISPR toolkit has unlocked many approaches to gene disruption: gene editing, gene knockout, gene interference and gene knockdown; however, these tools are not equally suited to lncRNA targeting (Figure 4.1). To evaluate CRISPR strategies, lncRNA features such as sequence features, genomic context, and potential functionality must be considered. Firstly, *P. falciparum* lncRNAs are highly AT-rich and contain regions of low complexity, features that are not amenable to all Cas enzymes. Some Cas nucleases require specific sequences in certain positions to facilitate their activity. For example, Cas9 requires the protospacer adjacent motif (PAM) NGG, which could limit target selection more than enzymes like Cpf1 with an AT-rich motif (TTTN) or those that lack any required motif like Cas13. Secondly, lncRNAs with many repetitive sequences could cause challenges in designing gRNAs or primers. These sequences could provide few gRNA options, require longer gRNAs to diminish off-target hits or be non-targetable. In terms of functionality, gene editing is not an appropriate general strategy to disrupt lncRNAs because precise nucleotide edits may not invoke a change in function. The alteration of a limited number of nucleotides in a long transcript may not dramatically affect the secondary structure or activity of the lncRNA. Gene knockout, by contrast, would likely lead to a phenotype. However, *P. falciparum* lncRNAs often are found in complex genomic contexts leading to only a minority that can be knocked out without affecting other nearby gene features. In these contexts where other genes or ncRNAs could be impacted, it would be difficult to decipher which disruption led to a phenotype. Furthermore, donor templates for targeting lncRNAs could be highly AT-rich, which leads to plasmid instability. AT-rich sequences are susceptible to local melting in supercoiled plasmids, which reduces torsional strain (Benham, 1979).

Gene interference could also provide a solution: by disrupting the transcription of lncRNAs, their function in the parasite would be disrupted without altering the genome. Like gene knockout, this approach does not target the lncRNA directly. Gene knockdown or direct RNA

Table 4.1 Features of CRISPR-Cas enzymes relevant to lncRNA disruption

Feature	**Cas9**	**Cpf1**	**dCas9**	**dCpf1**	**Cas13**
Disruption strategy	knockout	knockout	interference	interference	knockdown
Molecular target	DNA	DNA	DNA	DNA	RNA
PAM motif	NGG	TTTN	NGG	TTTN	none
Target essential sequences	x	x	✓	✓	✓
Target antisense lncRNAs	x	x	✓	✓	✓✓✓
Established in *Pf*	✓	✓	✓	✓(unpublished)	x

targeting would be the ideal approach to target at the RNA level. When used in parallel with gene interference, it could delineate the effects of transcription of the lncRNA versus the action of the lncRNA itself. None of the CRISPR tools available for *P. falciparum*, including Cas9, dCas9 and the latest addition, Cpf1 (Cas12a), can facilitate RNA knockdown. One RNA-targeting CRISPR enzyme, LwCas13a, has been recently applied to malaria research, although for diagnostic use, for the detection of *P. falciparum* and *P. vivax* 18S rRNA in clinical isolates (Cunningham et al., 2021). Therefore, I prioritised the development of a Cas13 RNA-targeting CRISPR enzyme for *P. falciparum* while also exploring the application of Cas9 and dCpf1Sir2a, newly developed by our lab, to lncRNA targeting.

4.4.2 CRISPR-Cas: a single-vector plasmid system

Each CRISPR experiment in this chapter uses a plasmid-based expression system derived from the single-vector CRISPR-Cas9 plasmid developed for *P. falciparum* gene disruption (Figure 4.1) (Adjalley and Lee, 2022). Although the Cas enzymes and U6 cassettes are unique in each plasmid, many principal components remain consistent. Selection markers included an ampicillin resistance cassette for plasmid propagation in bacterial cultures and an hDHFR or blasticidin S resistance cassette to select for plasmid uptake in transfected parasites. For the plasmid features translated in the parasite, namely the Cas enzyme and the selection marker, 5′ UTRs or promoters and 3′ UTRs from *Plasmodium* genes are added to improve their expression. The Cas enzyme was driven by a calmodulin (*cam*) promoter (PF3D7_1434200) positioned upstream and followed by the 3′ UTR from *pfhsp86* (PF3D7_0708400). The 5′ UTR of *pcdt* (PCHAS_0728300) and 3′ UTR of *pfhrp2* (PF3D7_0831800) were positioned before and after the selectable marker. The Cas enzymes obtained from bacterial systems can have different amino acid usage and frequencies compared to *P. falciparum*. Recodonisation that mirrors *P. falciparum* amino acids frequencies can support the expression of a bacterial protein in the parasite, and therefore, Cas enzymes were recodonised using a Python script from Dr Adam Reid. A nuclear localisation signal was also added to ensure the localisation

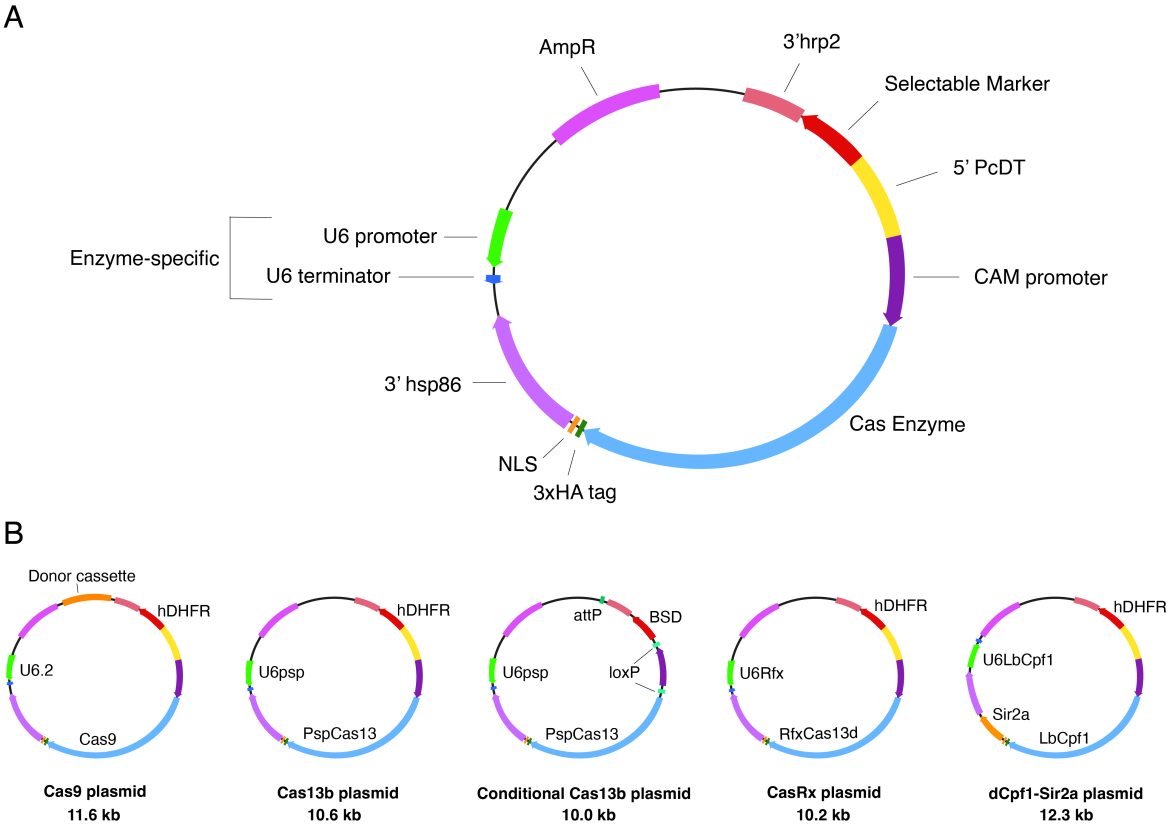


Fig. 4.1 CRISPR-Cas delivery system in *P. falciparum*. (**A**) This vector was developed by Dr Marcus Lee (Adjalley and Lee, 2022). The Cas enzyme includes a nuclear localisation signal and a 3X HA tag and is often recodonised for *P. falciparum*. The enzyme is driven by a calmodulin (CAM) promoter (PF3D7_1434200 and followed by the 3′ UTR of *pfhsp86*. A U6 snRNA cassette was included for the insertion and expression of gRNAs. A selectable marker for selection in the parasite is included and flanked by the *P. chabaudi* DHFR-TS promoter and 3′ UTR of *pfhrp2*. An ampicillin resistance cassette was included for propagation in bacterial culture. (**B**) The plasmids used in this work with their unique features noted.

of the protein to the nucleus, where it is active. Hemagglutinin (HA) tags have 9 amino acids to enable the isolation of the protein with an anti-HA antibody. The U6 cassette included in the plasmid is specific to the species of the Cas enzyme, as it drives the expression of small RNAs such as gRNAs. The U6 cassette also contains a BbsI restriction site for gRNA insertion.

4.4.3 RNA knockdown in *P. falciparum* using CRISPR-Cas13

I created expression systems for two different Cas13 enzymes: Cas13b from *Prevotella* sp. and Cas13d from *Ruminococcus flavifaciens*.

Expression of Cas13b in *P. falciparum* and knockdown of endogenous GFP

To determine if a new Cas enzyme could be expressed in *P. falciparum* using a plasmid system and that this enzyme was capable of RNA knockdown in the parasite, a target was selected with an easily quantifiable phenotype when knocked down. The target chosen was GFP, which is expressed endogenously in the NF54$^{\text{camEGFP}}$ line from the *cg6* (PF3D7_0709200) locus, a nonessential gene in the intraerythrocytic cycle. The level of green fluorescence, which is representative of GFP expression, is quantifiable by flow cytometry. Furthermore, the *gfp* sequence is more GC-rich than *P. falciparum* sequences, resembling more closely the *Prevotella* sp. and the mammalian applications of the technology.

To target GFP RNA, I used a PspCas13b expression system (pDC2-cam-PspCas13-U6psp-hDHFR, designed and generated by Dr Marcus Lee). The 4.4kb Cas9 sequence was replaced by a smaller 3.27 kb PspCas13b with AvrII and XhoI digestion. Site-directed mutagenesis was used to remove an internal BbsI site from PspCas13b to enable BbsI-mediated gRNA insertion into the U6 cassette. PspCas13b was also recodonised, and a nuclear localisation signal, HIV nuclear export signal and 3xHA tag were added. The U6 was replaced with a U6psp cassette which contained a direct repeat specific to PspCas13 downstream of the cloning site to facilitate the expression of gRNAs. Although Cas13 guides are also known as crRNAs (as they do not require tracrRNA that is usually included in the gRNAs), in this work, I will refer to them as gRNAs. Nine 30 bp gRNAs (8 GFP-targeting and 1 control non-targeting) were designed using Benchling and DNAstar, and BLAST was used to check for off-target hits in the *P. falciparum* 3D7 reference genome (Johnson et al., 2008; Benchling, 2020). Given that Cas13 requires single-stranded substrates, RNAfold was used to predict RNA folding and regions with predicted secondary structures were avoided

(Gruber et al., 2008). Although there are a few constraints in the protospacer flanking site requirements, some preferences were identified in bacteria assays (Cox et al., 2017). For example, a slight preference for GG at the 5′ protospacer flanking sequence (PFS) in bacterial assays and an association of a 3′ G with increased off-target motifs in HEK93FT cells (Cox et al., 2017). Each gRNA was ligated into a separate pDC2-cam-PspCas13-U6psp-hDHFR plasmid. The control gRNA targeted foreign DNA that did not map to any region in the *P. falciparum* genome to assess non-specific targeting. Plasmids were transfected individually into NF54^{camEGFP} and WR99210 (1nM) drug pressure was applied on the following day and maintained continuously to select for plasmid uptake and maintenance (Figure 4.3A).

The first round of transfections did not result in recrudescence after 40 days, and the cultures were not maintained longer due to contamination (Table 4.2). However, due to molecular cloning of the PspCas13b expression plasmid for concurrent experiments, I discovered that the Cas13 sequence in this plasmid was truncated. Upon XhoI and AvrII digestion of the plasmid and gel electrophoresis, I did not observe the expected 3456 bp band; instead, a band reflecting a lighter fragment was present (Figure 4.2A). Single-enzyme digestions confirmed that both restriction enzymes were functional, and further double-enzyme digests confirmed that the plasmids with other gRNAs also were truncated. Returning to the glycerol stocks of the two original clones of the plasmid (before gRNA insertion), 1 of the 2 clones had truncations (digest, gel and sequencing completed by Dr Sophie Adjalley), including the clone that was used to create the GFP-targeting plasmids (Figure 4.2B). However, an intact sequence had been observed in the CRISPR-Cas13 plasmid clones prior to midi-prep by PCR and sequencing (Dr Sophie Adjalley). Therefore, I hypothesised that the truncation might have occurred during the propagation in *E. coli*.

Therefore, the GFP gRNAs were re-ligated into the correct clone (clone 2). I was concerned about the potential instability of the plasmid in bacterial culture; therefore, I ensured that overnight incubation times were dropped to 10 hours and that restriction digests were completed at every step. 6 gRNAs (all but gRNA 2 and 6) were confirmed correct with sequencing and could be midi prepped. In this round of transfection, both a negative control (non-targeting gRNA) and positive transfection control (CRISPR-Cpf1 plasmid) were included. One GFP transfection (labelled Tx6-1) re-emerged after 43 days (Table 4.2). This transfection was expanded, and GFP fluorescence was quantified using flow cytometry. No change in the GFP+ parasite population was observed between TX6-1 and NF54^{camEGFP} line (Figure 4.3B). To check if the Cas13 plasmid was present in the parasite and WR99210 resistance was not acquired through an unusual event such as recombination, DNA was

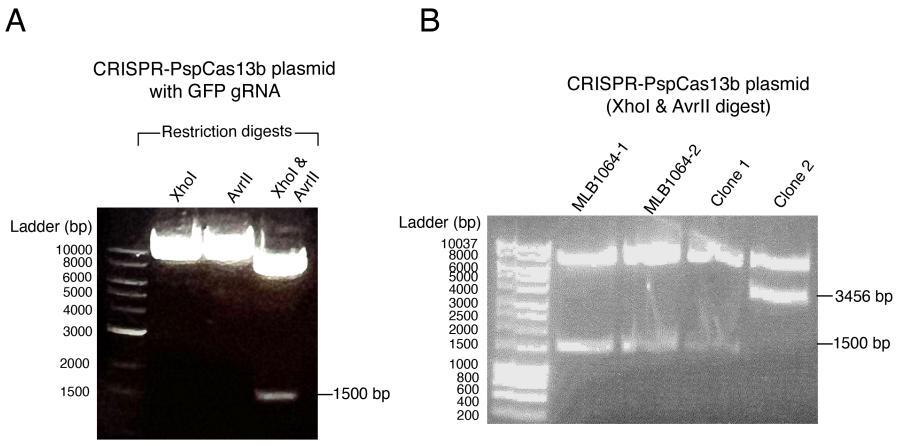


Fig. 4.2 Initial PspCas13b plasmid contained a truncation in PspCas13b. (**A**) Gel electrophoresis of XhoI and AvrII single and double-enzyme digestions of the GFP-targeting PspCas13b plasmid (pDC2-Cas13-U6-hDHFR-GFPgRNA). 1kB Plus ladder (NEB) was used instead of HyperLadder. (**B**) Gel electrophoresis of XhoI and AvrII double-enzyme digestions of two original clones from the generation of the PspCas13b plasmid (clone 1 and 2) and two preps (MLB1064-1 and MLB1064-2) from a glycerol stock of clone 1. This gel was completed by Dr Sophie Adjalley.

extracted from the parasite, and primers were used to amplify the PspCas13b. PspCas13b DNA was detected in the parasite (Figure 4.3C). Due to the HA tag, the PspCas13b protein in the parasite can be quantified from a parasite pellet. However, the protein was undetectable by western blot with an anti-HA antibody (completed by Mukul Rawat). The transfections were repeated with the same gRNAs, to see if any transfectants expressing PspCas13b could be obtained; however, no transfections recrudesced (Table 4.2).

Conditional expression of Cas13b in *P. falciparum*

Given the poor recovery rate following transfection with the PspCas13b expression plasmids, including transfections containing non-targeting gRNAs, I hypothesised that Cas13 could be toxic to the parasite. Thus, I designed a strategy to investigate and potentially circumvent any toxicity of Cas13 using conditional expression. The rationale was to establish an inducible system with Cas13b coding sequence integrated into the genome whose expression could be regulated using the dimerised cre (DiCre) recombinase-based system. By integrating Cas13 into the genome, this line would only require the delivery of gRNAs and rapamycin to induce cre dimerisation to knockdown any target.

To generate this line, I exploited the 3D7 DiCre-attB line, which expresses the DiCre recombinase in the *P230* (PF3D7_0209000) locus and contains an attB site for site-specific

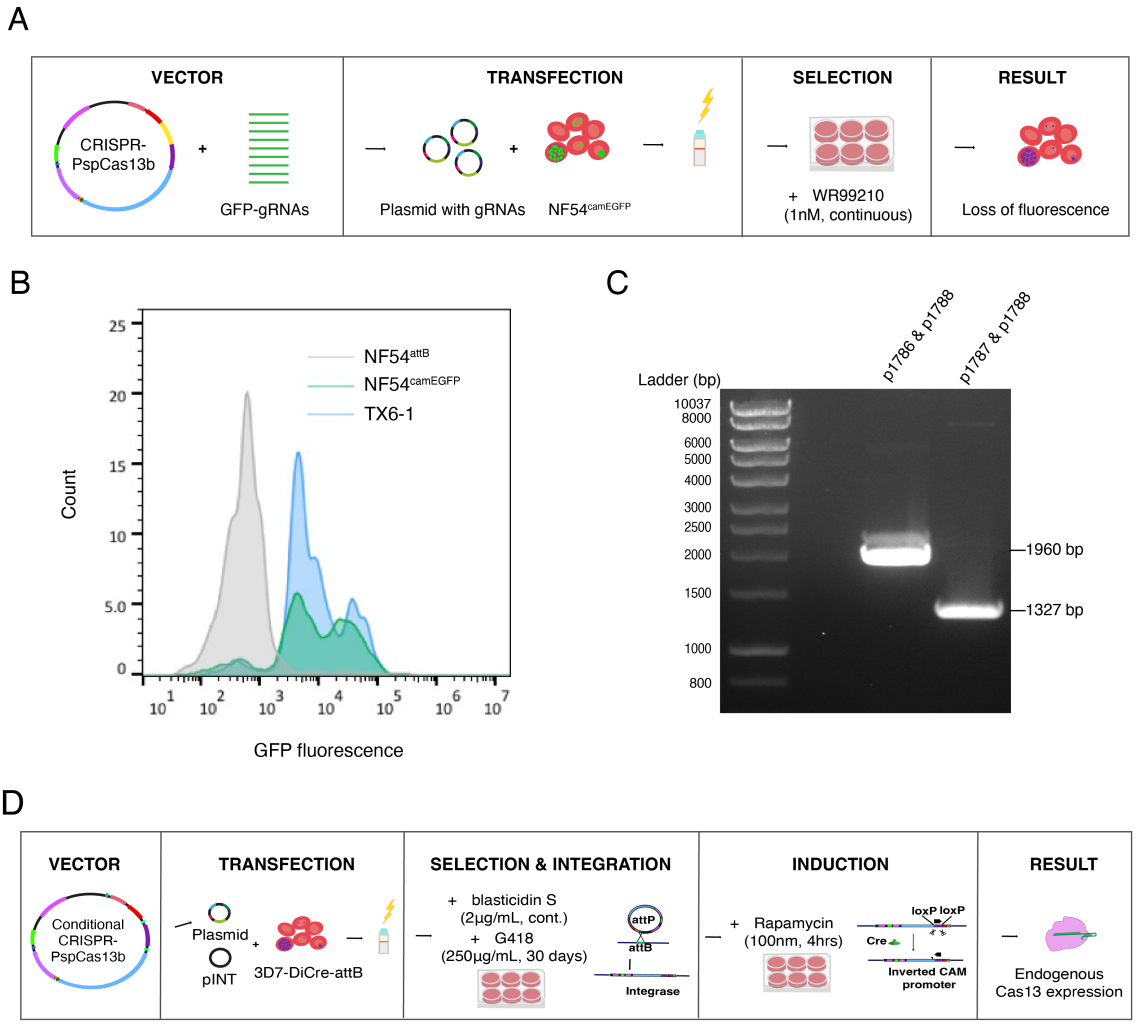


Fig. 4.3 GFP-targeted knockdown using CRISPR-PspCas13b does not enable result in decreased GFP expression. (**A**) Transfectant 6-1 (TX6-1) was obtained by transfecting NF54^{camEGFP} with a PspCas13b-expression plasmid containing a GFP-targeting gRNA and selecting for the plasmid with WR99210 (1nM, continuous). To assess Cas13b-mediated knockdown of endogenous GFP in NF54^{camEGFP}, flow cytometry was completed with TX6-1, NF54^{camEGFP}, the parent line and non-fluorescent control NF54^{attB} and a control containing RBCs-only. (**B**) The distribution of GFP fluorescence revealed no difference when comparing TX6-1 (blue) with NF54^{camEGFP} (green). Parasite DNA was extracted, and plasmid DNA was amplified using the p1786 and p1787 forward Cas13b primers with the p1788 reverse Cas13b primer. (**C**) Gel electrophoresis of the amplified fragments confirmed the presence of the CRISPR-PspCas13b plasmid. (**D**) The strategy for conditional CRISPR-Cas13b expression in *P. falciparum*. The construct was co-transfected with the pINT plasmid into 3D7-DiCre-attB. Selections were completed with blasticidin S and G418. The construct would integrate into the attB site. If transfectants returned, rapamycin would be added to induce the inversion of the CAM promoter and Cas13 expression.

Table 4.2 Summary of PspCas13b and RfxCas13d transfections in *P. falciparum*

Plasmid	Line	Selection	Total TX	TX recru.	Pheno-type
trCRISPR-PspCas13b-GFPgRNA1	NF54^{camEGFP}	WR99210 1nM cont.	1	0	N/A
trCRISPR-PspCas13b-GFPgRNA2			1	0	N/A
trCRISPR-PspCas13b-GFPgRNA3			1	0	N/A
trCRISPR-PspCas13b-GFPgRNA4			1	0	N/A
trCRISPR-PspCas13b-GFPgRNA5			1	0	N/A
trCRISPR-PspCas13b-GFPgRNA6			1	0	N/A
trCRISPR-PspCas13b-GFPgRNA7			1	0	N/A
trCRISPR-PspCas13b-GFPgRNA8			1	0	N/A
trCRISPR-PspCas13b-NTgRNA			1	0	N/A
CRISPR-PspCas13b-GFPgRNA1	NF54^{camEGFP}	WR99210 1nM cont.	2	1	No Kd
CRISPR-PspCas13b-GFPgRNA3			2	0	N/A
CRISPR-PspCas13b-GFPgRNA4			2	0	N/A
CRISPR-PspCas13b-GFPgRNA5			2	0	N/A
CRISPR-PspCas13b-GFPgRNA7			2	0	N/A
CRISPR-PspCas13b-GFPgRNA8			2	0	N/A
CRISPR-PspCas13b-NTgRNA			2	0	N/A
Conditional-CRISPR-PspCas13b + pINT	3D7 DiCre-attB	Blasticidin S 2μg/mL cont. G418 250μg/mL 30d	4	0	N/A
CRISPR-RfxCas13d-GFPgRNA1	NF54^{camEGFP}	WR99210 1nM cont.	3	0	N/A
CRISPR-RfxCas13d-GFPgRNA2			3	0	N/A
CRISPR-RfxCas13d-GFPgRNA3			3	0	N/A
CRISPR-RfxCas13d-GFPgRNA4			3	0	N/A
CRISPR-RfxCas13d-GFPgRNA5			3	0	N/A
CRISPR-RfxCas13d-GFPgRNA6			3	0	N/A
CRISPR-RfxCas13d-GFPgRNA7			3	0	N/A
CRISPR-RfxCas13d-GFPgRNA8			3	0	N/A
CRISPR-RfxCas13d-NTgRNA			3	0	N/A
Pool 1 (gRNAs 1+2+3)			2	0	N/A
Pool 2 (gRNAs 4+5+6)			2	0	N/A
CRISPR-RfxCas13d-pfparegRNA1	Dd2	WR99210 5nM cont.	2	0	N/A
CRISPR-RfxCas13d-pfparegRNA2			3	0	N/A
CRISPR-RfxCas13d-pfparegRNA3			3	0	N/A
CRISPR-RfxCas13d-pfparegRNA4			3	0	N/A
CRISPR-RfxCas13d-pfparegRNA5			3	0	N/A
CRISPR-RfxCas13d-pfparegRNA6			3	0	N/A
CRISPR-RfxCas13d-NTgRNA			3	0	N/A
Pool 1 (gRNAs 1+2+3)			2	0	N/A
Pool 2 (gRNAs 4+5+6)			2	0	N/A

TX: transfection; recru.: recrudesced; trCRISPR-PspCas13b: truncated CRISPR-PspCas13b plasmid; NT: non-targeting gRNA; cont.:continuous; d:days; Kd: knockdown; N/A: not applicable. The CRISPR-Cpf1 plasmid was used as a positive transfection control in WR99210 selections, and the mRFP plasmid was used as a transfection control in blasticidin S selections.

integration of transgenes at the *cg6* locus. While the addition of rapamycin activates the DiCre recombinase to mediate recombination between *lox*P sites, co-transfection with the pINT plasmid (Addgene) that encodes the BxB1 phage integrase enables genome-integration of an attP plasmid via recombination between the attP and attB sites (Nkrumah et al., 2006; Knuepfer et al., 2017). Therefore, I designed a PspCas13b plasmid containing an attP site for genome integration and a promoter in reverse orientation with *lox*P sites so that DiCre recombination induced a promoter flip and, thereby, the transcription of Cas13b. I inserted PspCas13b into a plasmid designed to integrate inducible dCas9 (generated by Dr Sophie Adjalley) using XhoI and AvrII digestion and Gibson assembly. The conditional PspCas13b plasmid was co-transfected into 3D7 DiCre-attB with pINT, and dual-drug pressure was added. G418 (250 μg/ml) was maintained for 30 days to select for pINT uptake, and blasticidin S (2 μg/ml) was maintained continuously to select for uptake and maintenance of the conditional PspCas13b plasmid (Figure 4.3D). In four transfections, I could not recover any transfectants after 50 days (Table 4.2). A positive transfection control (mRFP plasmid) was included for each transfection, 50% recrudesced. Four additional transfections were also completed without the inclusion of the pINT plasmid. One culture had evidence of surviving parasites at 44 days post-transfection; however, the parasites did not expand, and the culture did not recrudesce.

In parallel, I also generated a version of this plasmid that contained gRNAs, instead of relying on external gRNA delivery. A U6psp cassette was synthesised and inserted into the plasmid using BamHI digestion and T4 ligation. Site-directed mutagenesis was used to remove a bbsI site from the BSD sequence, as this restriction enzyme is required to insert gRNAs into the U6 cassette. I designed Cas13 gRNAs to target an endogenous gene, *pfpare* encoding a prodrug activation and resistance esterase for RNA knockdown. Disruption of *pfpare* has no apparent fitness cost; however, it confers resistance to the antimalarial compound MMV011438, which can pulse for low-abundance mutants and is used to confirm a disruption phenotype. Due to the exploration and prioritisation of a different enzyme, Cas13d, the transfection of this plasmid was not attempted.

Expression of RfxCas13d in *P. falciparum* and knockdown of endogenous GFP

Due to the lack of success with PspCas13b, I explored the application of another type of Cas13 enzyme, RfxCas13d (CasRx), into *P. falciparum*, which was shown to exhibit efficient RNA knockdown but with a smaller-sized enzyme (Konermann et al., 2018). The plasmid was generated similarly to the CRISPR-PspCas13b plasmid by completing a Gibson ligation

into the CRISPR-Cas9 plasmid backbone. The RfxCas13d was recodonised, and a nuclear localisation signal, and 3xHA tag were added to the end of the RfxCas13d sequence. The normal U6 was replaced with the U6 from *Ruminococcus flavefaciens*. The RNA cleavage facilitated by RfxCas13d has no dependence on the PFS; therefore, no additional gRNA considerations were required (Konermann et al., 2018). The gRNAs sequences from the PspCas13b were also used for RfxCas13b, although the gRNAs required different BbsI overhangs for insertion into the U6 cassette (from (5′-CACC-forward/5′-AAAC-reverse) to (5′-CAAC-forward/5′-AAAA-reverse)).

Similarly, I looked to GFP knockdown as a proof-of-concept experiment. 8 GFP-targeting gRNAs and 1 non-targeting gRNA were designed and inserted into the CRISPR-RfxCas13d plasmid using BbsI. Plasmids were transfected individually and in pools of three gRNAs into NF54$^{\text{camEGFP}}$ and selected with continuous WR99210 (1nM) (Figure 4.4A). No parasites recrudesced in these initial transfections, although the cultures had to be discarded early (Day 24) due to yeast contamination. However, the set of transfections was completed twice more, and still, no parasites recrudesced except the positive control (CRISPR-Cpf1 plasmid) (Table 4.2).

In parallel, I also targeted an endogenous target from *P. falciparum* genome, *pfpare* (PF3D7_0-709700). *pfpare* encodes a prodrug activation and resistance esterase that is not essential (Istvan et al., 2017). In addition to the dispensable nature of *pfpare*, loss-of-function mutations in *pfpare* confer resistance to the antimalarial compound MMV011438, which requires *pfpare* for its activation. Drug resistance can be used to quantify the RNA knockdown of *pfpare*. 6 *pfpare*-targeting gRNAs were designed and inserted into the CRISPR-RfxCas13d plasmid using BbsI. These plasmids and the plasmid containing a non-targeting gRNA were transfected individually and in pools of three gRNAs into Dd2 and selected with continuous WR99210 (5nM) (Figure 4.4B). Because the cultures from the first round of transfections were kept in proximity to the GFP experiment, they also succumbed to contamination. These transfections were repeated twice, and no parasites recrudesced Table 4.2. Except discordantly, the positive control did not recrudesce.

4.4.4 Knockout of lncRNAs using Cas9

Because new tools are often low or difficult to develop in *Plasmodium*, I explored using the readily-available Cas9 enzyme for lncRNA disruption in *P. falciparum*. Although Cas9 is

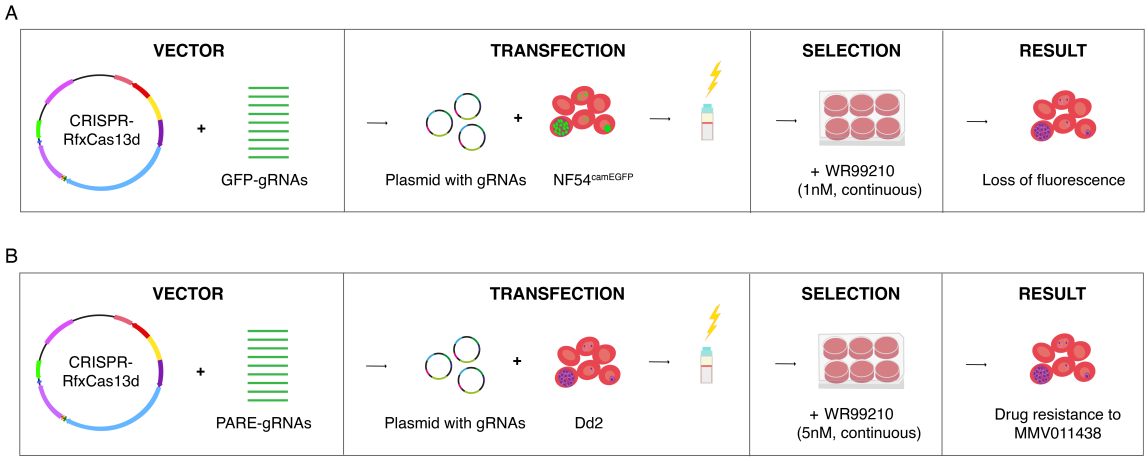


Fig. 4.4 No transfectants could be obtained from GFP- or *pfpare*-targeted knockdown using CRISPR-RfxCas13d. (**A**) GFP was targeted with CRISPR-RfxCas13d in NF54 parasites endogenously expressing GFP. The knockdown of GFP expression can be quantified by measuring parasite fluorescence. (**B**) *pfpare* was targeted with CRISPR-RfxCas13d in Dd2 parasites. The knockdown of *pfpare* is anticipated to affect drug sensitivity to MMV011436, which can be quantified with drug assays. The CRISPR-RfxCas13d requires continuous selection with WR99210 (1nM in NF54-background parasites and 5nM in Dd2).

unsuitable for most lncRNA targets, it could provide an option for the small minority that are truly intergenic or studied on a small scale, where knockouts can be carefully designed.

Selection of lncRNA targets

Three lncRNAs were selected for Cas9 targeting from the Broadbent et al. lncRNA annotation because the annotation in *Chapter 3* was not yet completed (Broadbent et al., 2015). The annotation was intersected with a list of intergenic mutations identified in whole genome sequencing for drug-resistant lines generated by labs in the Malaria Drug Accelerator (MalDA) consortium (SNP and InDel list obtained from Dr Madeline Luth (University of California San Diego). 16 mutations were identified in lncRNAs. From this list, three lncRNAs identified as targets by considering the distance from expressed genes (ideally > 1 kb), the feasibility with considerations for gRNA and donor designs and associated drugs in terms of their known mechanisms and availability. Although the mutations may not necessarily be responsible for the drug resistance as they were not found in isolation, drug resistance could still be a potential consequence of lncRNA knockout and phenotype. The three lncRNAs selected for this pilot study were PF3D7lncRNA_0019, PF3D7lncRNA_0337 and PF3D7lncRNA_0567.

The first lncRNA target, PF3D7lncRNA_0019, is located in a cluster of three lncRNAs nestled between PF3D7_0104200 (STAR-related lipid transfer protein) and PF3D7_0104300 (ubiquitin carboxyl-terminal hydrolase 1), where a switch of transcriptional direction occurs. The mutation in PF3D7lncRNA_0019 was an 8 bp deletion mutation (TTTTTATTA →T) at position 373, which was identified in a parasite resistant to MMV668399. The target of MMV668399 is an FHA domain-containing protein (PF3D7_1307800) (Cowell et al., 2018). However, mutations in other proteins such as amino acid transporter *pfaat1* (PF3D7_0629500 and *pffhd* (PF3D7_0909700) and *pfsulp* (PF3D7_1471200) have been associated with MMV668399 resistance (Cowell et al., 2018). The neighbouring protein of PF3D7lncRNA_0019 is involved in ubiquitin processing, a pathway implicated in artemisinin tolerance in *P. chabaudi* (Hunt et al., 2007).

PF3D7lncRNA_0337, the second lncRNA target, is located between PF3D7_0420200 (holo-[acyl-carrier-protein] synthase) and PF3D7_0420300 (putative AP2). Both genes are expressed on the forward strand, while the lncRNA is expressed on the reverse strand. Two SNPs were identified in the lncRNA, both adenine to thymine changes that occurred in close proximity (5 bp apart at positions 382 and 387). These SNPs were identified in atovaquone-resistant lines, a drug known to target the CytBC1 (Siregar et al., 2015). The neighbouring AP2 gene was implicated in the regulation of *pfcrt* in chloroquine-resistant recombinant progeny (Siwo et al., 2015). The other neighbouring gene was identified as a gene undergoing balancing selection in *P. falciparum* in Mali, Western Africa (Coulibaly et al., 2022).

The mutation in the third lncRNA target, PF3D7lncRNA_0567, was also identified in atovaquone-resistant lines and was an adenine to thymine change at position 1159. The lncRNA is located between PF3D7_0617000 (mitochondrial import receptor TOM40) and PF3D7_0617100 (AP-2 complex subunit alpha). Like the first lncRNA target, this lncRNA exists at a switch of transcriptional direction. Of the three targets, this lncRNA has the greatest distances (1500+ bp) from other gene features. SNPs associated with artemisinin resistance have been identified in the neighbouring AP-2 complex gene (Rocamora et al., 2018).

Disruption of lncRNA targets

To disrupt lncRNAs, I completed a knockout approach by inserting a selectable marker. This approach is necessary when targeting newly identified lncRNAs that may not have a known or observable phenotype to ensure complete disruption of the lncRNA and selection

of the mutants. To form the hDHFR-containing donor sequences: the 5′ and 3′ HRs for each lncRNA were amplified using PCR and inserted into an EcoRI/AatI and ApaI-digested CRISPR-Cas9 plasmid (pDC-cam-Co.Cas9-hdhr) using Gibson ligation. The plasmid was digested with BbsI and gRNAs were inserted. Three to four gRNAs were designed for each lncRNA and ligated into separate plasmids. The plasmids were transfected in pools of 3 different gRNAs (at 30-50μg each) into NF54 and 3D7 (Figure 4.5A). Plasmid uptake and insertion were selected using WR99210 (1nM) for 7 days. Of the three targets, only one could be knocked out, PF3D7lncRNA_0567 (Table 4.3). Another CRISPR-Cas9 plasmid targeting the *MD1* gene (PF3D7_1438800) was included as a positive control; however, it did not recrudesce. Single gRNA transfections were then completed in case one of the pooled gRNAs was toxic and prevented parasite recrudescence. However, even with the individual gRNA approach, only PF3D7lncRNA_0567 could be disrupted (Table 4.3). To confirm the disruption of PF3D7lncRNA_0567, gDNA was extracted from both transfectants (Δ567.1 and Δ567.2), and the donor region and lncRNA loci were amplified using primers and visualised using gel electrophoresis. Initially, primers amplified regions within the donor and confirmed their presence in the parasite: PcDT & 3′ homologous region (p675 & p2015), HRP2 to the 5′ homologous region (p282 & p2012) and the hDHFR fragment (p261 & p365) (Figure 4.5B). However, to confirm that donor sequences were inserted in the genome, a primer pair from inside (p675) and outside the donor sequence (the 3′ UTR of the lncRNA, p2173) was used. The expected 1105 bp fragment was observed, and the knockout was further confirmed with Sanger sequencing (Figure 4.5C).

I then assessed the susceptibility of the Δ567 mutants to atovaquone, the antimalarial drug that the drug-resistant lines containing the mutations were selected against. Dose-response curves were generated for every replicate, representative curves are presented in Figure 4.5D, and IC_{50} values for each replicate in Figure 4.5E. A significant increase in IC_{50} ($P<0.01$) values was observed in both Δ567.1 and Δ567.2. Drug assays were also completed for six other drugs (chloroquine, dihydroartemisinin, mefloquine, KDU691, KAE609, MMV008434 and MMV020746), but no significant changes in IC_{50} were observed.

4.4.5 Interference of lncRNAs using dCpf1Sir2a

Additionally, I explored using another CRISPR enzyme that is functional in *P. falciparum*, LbCpf1 (Cas12a). LbCpf1 has been used in diagnostic testing (Lee et al., 2020) and for gene interference by our lab (unpublished). Like Cas9, when Cpf1 is catalytically inactivated, it can disrupt transcription at target sites, although, unlike dCas9, dCpf1 has an AT-rich

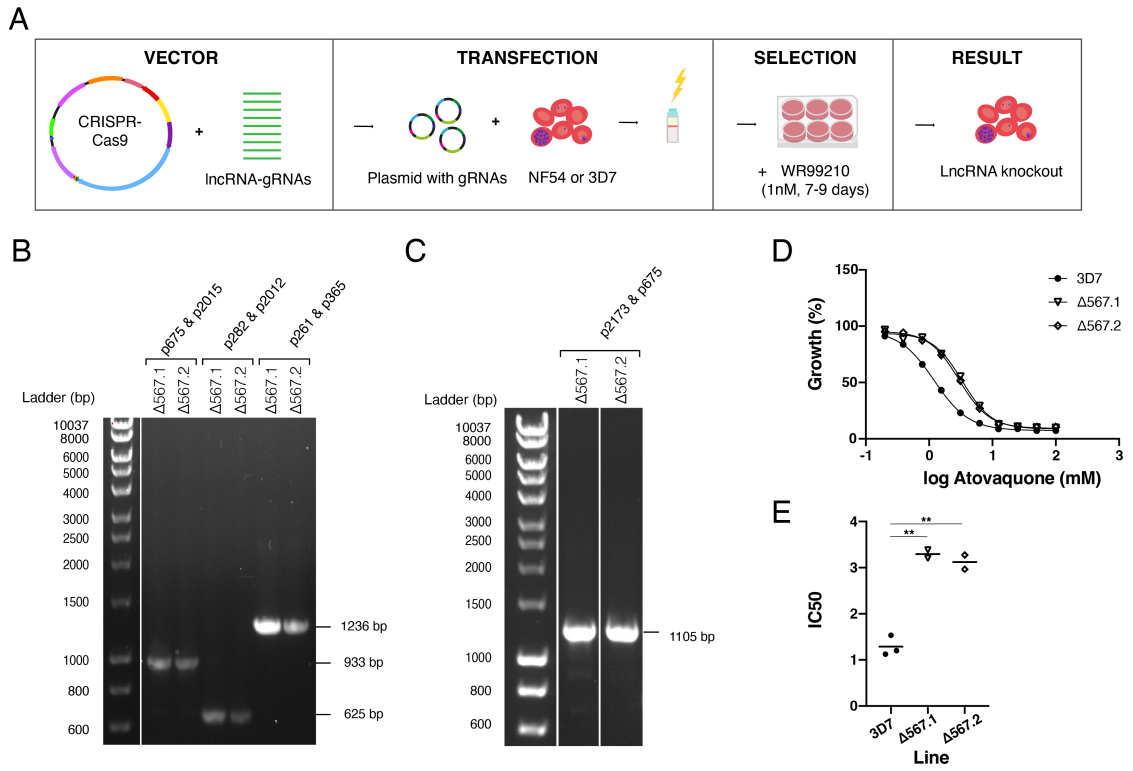


Fig. 4.5 CRISPR-Cas9-mediated knockout of *P. falciparum* lncRNAs. (**A**) LncRNAs were targeted with CRISPR-Cas9 in Dd2 parasites. Donor sequences contained the hDHFR gene, which enabled the selection of knockouts with WR99210 (1nM continuous). (**B**) Gel electrophoresis demonstrates that the donor fragments are present in the transfectants Δ567.1 and Δ567.2. PcDT & 3′ homologous region (p675 & p2015), HRP2 to the 5′ homologous region (p282 & p2012) and the hDHFR fragment (p261 & p365). (**C**) Using a primer (p2173) in the 3′ UTR outside the donor sequence with one in PcDT (p675), it was determined that the donor sequence had been inserted into the target sites and that the lncRNA was disrupted. (**D**) Drug assays for Δ567.1 and Δ567.2 were completed in duplicate against atovaquone. A non-linear regression model for log(inhibitor) and variable(slope) was used to generate a dose-response curve from technical replicates, a representative dose-response curve is shown here. (**E**) IC_{50} values were calculated from dose-response curves, a line representing the mean IC_{50}. Two-tailed t-tests were completed for every line versus the control (3D7). Significant p-values are denoted with representative asterisks: **(0.001 to 0.01).

Table 4.3 Summary of Cas9 lncRNA-targeting transfections in *P. falciparum*

Plasmid	Line	Selection	Total TX	TX recru.	Pheno-type
CRISPR-Cas9-lncRNA_0019(gRNA 1-3)			2	0	N/A
CRISPR-Cas9-lncRNA_0337(gRNA 1-3)	NF54	WR99210 1nM 7d	2	0	N/A
CRISPR-Cas9-lncRNA_0567(gRNA 1-3)			1	0	N/A
CRISPR-Cas9-lncRNA_0019(gRNA 1-3)			1	0	N/A
CRISPR-Cas9-lncRNA_0337(gRNA 1-3)			1	0	N/A
CRISPR-Cas9-lncRNA_0567(gRNA 1-3)	3D7	WR99210 1nM 7d	1	1	Edited
CRISPR-Cas9-lncRNA_0337gRNA1			1	0	N/A
CRISPR-Cas9-lncRNA_0567gRNA3			1	1	Edited
CRISPR-Cas9-lncRNA_0019gRNA1			1	0	N/A
CRISPR-Cas9-lncRNA_0019gRNA2			1	0	N/A
CRISPR-Cas9-lncRNA_0019gRNA3			1	0	N/A
CRISPR-Cas9-lncRNA_0337gRNA1	NF54	WR99210 1nM 7d	1	0	N/A
CRISPR-Cas9-lncRNA_0337gRNA2			1	0	N/A
CRISPR-Cas9-lncRNA_0337gRNA3			1	0	N/A
CRISPR-Cas9-lncRNA_0337gRNA4			1	0	N/A

TX: transfection; recru.: recrudesced; d:days; N/A: not applicable.

PAM site (TTTN). Previous lab members generated a CRISPR-dCpf1 plasmid (Dr Sophie Adjalley) with and without the fusion of dCpf1 to the epigenetic modifier, Sir2a deacetylase. Preliminary evidence showed that the plasmid facilitated gene interference: a 40-60% decrease in GFP fluorescence was achieved when targeting GFP in NF54^{camEGFP} with CRISPR-dCpf1Sir2a (unpublished, Dr Sophie Adjalley). Instead of completing a pilot study, I apply this approach directly on a more extensive lncRNA target set, described in *Chapter 5*.

4.4.6 Rapid generation of mNeonGreen *P. falciparum* reporter lines

Regardless of the method of perturbation, one of the key questions will be whether lncRNA-disrupted lines will have an observable phenotype. Parasite fitness is an important characteristic to quantify because the disruption of a molecule of biological significance often impacts parasite survival and reproduction, manifesting as a slow-growth phenotype. Competition growth assays offer a more sensitive readout than measuring parasitemia. In these assays, parasite lines are competed against a fluorescent line, and the proportion of fluorescent parasites is quantified using flow cytometry. However, due to substantial variances in growth rates between strains, the fluorescent line used should be strain-matched. Thus, I developed a vector for the rapid generation of mNeonGreen-expressing reporter lines and generated Dd2 and 3D7 mNeonGreen-expressing lines.

Selection of the *pfpare* locus for fluorescent markers integration in *P. falciparum*

The *pfpare* locus was identified as an optimal safe-harbour site for integrating the mNeonGreen fluorescent marker. In addition to the dispensable nature of *pfpare*, loss-of-function mutations in *pfpare* confer resistance to the antimalarial compound MMV011438, which requires *pfpare* for its activation. *pfpare* is expressed during blood stages; therefore, its promoter would facilitate the expression of a fluorescent marker throughout the intraerythrocytic life cycle. In addition, I included a BSD resistance cassette in the inserted sequence, which would only be expressed from the *pfpare* promoter once integrated. Thus my strategy would permit both positive and negative selection options for the efficient isolation of a homogeneous culture of mNeonGreen-tagged parasites without cloning. To identify if *pfpare* is suitable as an integration site across multiple *P. falciparum* strains, a multiple sequence alignment of the *pfpare* locus for eight geographically diverse strains (3D7, HB3, 7G8, GB4, CD01, GA01, IT and Dd2) was completed. A highly conserved gRNA target site was identified that had no mutations in the guide RNA or PAM sequences (Figure 4.6A). Examination of the flanking upstream and downstream sequences that would constitute the donor homology regions revealed no mutations in the 5′ homology region and 2-3 single nucleotide polymorphisms in the 3′ homology region of the donor sequence of the plasmid (Figure B.1). This level of sequence diversity would not be expected to strongly impact editing. I validated this prediction below using a 3D7-based donor sequence to edit Dd2, which has three polymorphisms relative to 3D7.

Construct design, generation, and transfection of Dd2$^{\text{pareNG}}$

To design a construct to integrate mNeonGreen into *pfpare*, I co-opted a *pfpare*-targeting CRISPR-Cas9 plasmid (generated by Dr Hannah Jagoe), which expresses Cas9 driven by the calmodulin promoter and transcribes a gRNA targeting *pfpare*. I subcloned a codon-optimised mNeonGreen into the donor region, which also contained a BSD resistance cassette and flanking 5′ and 3′ *pfpare* (3D7) homology regions of 255 bp and 466 bp respectively (Figure 4.6B). Plasmids were transfected into three independent *P. falciparum* Dd2 parasite cultures and continuous selection with blasticidin S selected for plasmid uptake and the insertion of the donor into *pfpare* by HDR (Figure 4.6C). WR99210 could also be used to select for plasmid uptake alone because an hDHFR resistance cassette is in the CRISPR-Cas9 plasmid but not in the donor region. This was not required in my experiments as blasticidin S selection alone was sufficient. Edited parasites were obtained from two of three transfections at around three weeks post-transfection, referred to as Dd2$^{\text{pareNG.1}}$ and

Dd2$^{pareNG.2}$. PCR-amplification of the *pfpare* locus using primers p2174 and p2175 (Figure 4.6C) and gel electrophoresis demonstrated successful donor integration with no wild-type locus detectable (Figure 4.6D). This indicated that positive selection was sufficient to deplete any unedited parasites, obviating the need for clonal isolation. Sanger sequencing confirmed the exact sequence and correct insertion. To further confirm the applicability of this construct to other strains, I also transfected the construct into *P. falciparum* 3D7 and both transfections, 3D7$^{pareNG.1}$ and 3D7$^{pareNG.2}$, recrudesced and were resistant to blasticidin S.

4.4.7 Dd2^{pareNG} fluorescence is comparable to existing GFP lines

To determine if the integrated mNeonGreen yields fluorescent parasites, I first examined the transgenic lines by fluorescence microscopy. Infected erythrocytes were detected using Hoechst DNA stain, which does not stain uninfected erythrocytes because they are anucleate, unlike *Plasmodium* parasites. Erythrocytes infected with either of the Dd2^{pareNG} lines showed strong green fluorescence, unlike the parental Dd2 line (Figure 4.7A). Flow cytometry was used to quantify the level of fluorescence and distribution within the population of a mixed-stage culture. The fluorescence profiles of Dd2$^{pareNG.1}$ and Dd2$^{pareNG.2}$ were highly similar and are comparable to the profiles of other green-fluorescing lines used for competition assays, including Dd2^{bipEGFP} and NF54^{camEGFP}, which express GFP from the strong constitutive promoters of ER-Hsp70 (BiP, PF3D7_0917900) and calmodulin (PF3D7_1434200) respectively (Adjalley et al., 2011; Baragaña et al., 2015). The peaks of the Dd2^{pareNG} lines were modestly shifted left in comparison to the GFP lines, which means that the bulk of Dd2^{pareNG} parasites in mixed culture are less fluorescent than the bulk of GFP-expressing parasites but still readily distinguishable from non-fluorescent parasites (Figure 4.7B). The bimodal peaks suggested that subpopulations expressed different levels of fluorescence. Fluorescence was also observed in the 3D7$^{pareNG.1}$ and 3D7$^{pareNG.1}$ lines (Figure 4.7C).

4.4.8 Dd2^{pareNG} fluorescence varies between asexual blood stages

In a healthy asexual intraerythrocytic *P. falciparum* culture, the sub-populations include briefly free-roaming merozoites and intraerythrocytic ring, trophozoite and schizont stages. To assess the level of fluorescence in specific stages, I enriched Dd2^{pareNG} cultures for different stages using synchronisation. Stage-specific levels of fluorescence were observed, with schizonts producing the highest fluorescence, followed by trophozoites and, lastly, ring stages (Figures 4.8A, C). This pattern was consistent with transcriptomic studies that

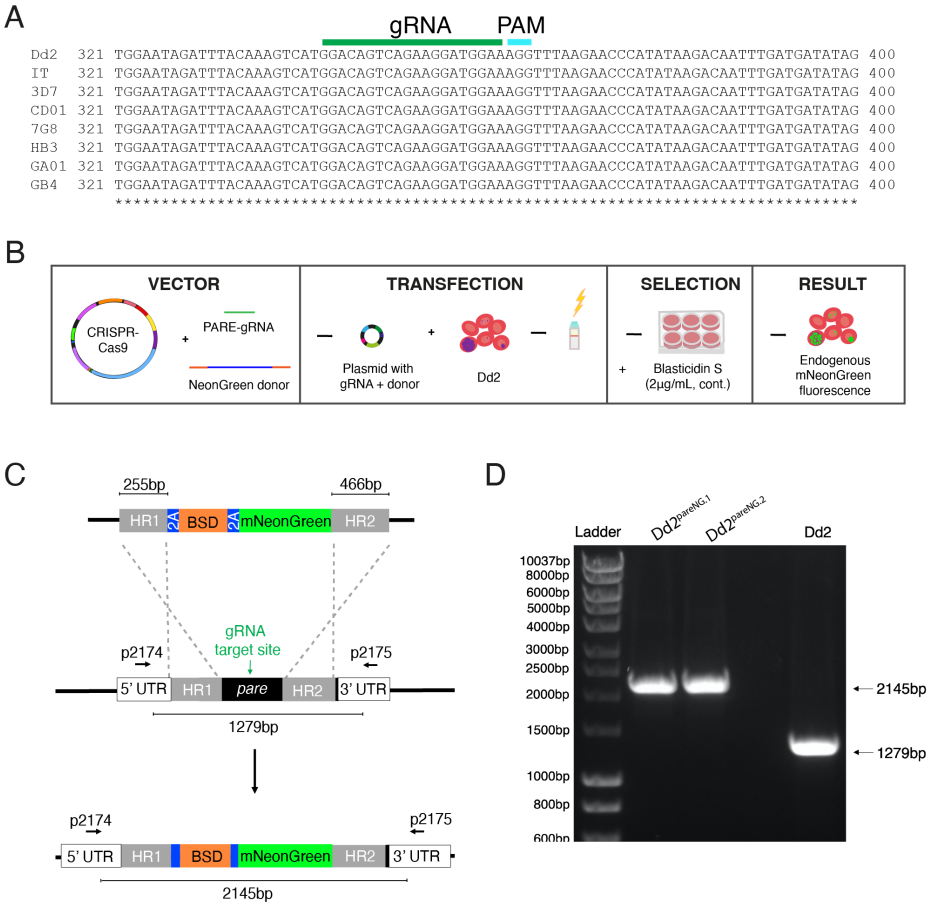


Fig. 4.6 Generation of a *P. falciparum* Dd2 reporter line expressing mNeonGreen under the control of the highly conserved *pfpare* locus. (**A**) DNA sequences for *pfpare* (PF3D7_0709700) in *P. falciparum* 3D7, HB3, 7G8, GB4, CD01, GA01, IT and Dd2 strains were obtained from PlasmoDB, and Clustal Omega was used to create a multiple sequence alignment (Goujon et al., 2010; Sievers et al., 2011; Amos et al., 2021). (**B**) Schematic of the CRISPR-Cas9 *pfpare*-targeting approach. The plasmid encodes a guide RNA cassette containing a guide targeting *pfpare* and a donor region containing a BSD selectable marker flanked by 2A linkers with mNeonGreen downstream. The payload is flanked by two homology regions for the *pfpare* locus. The construct was transfected into Dd2 and 3D7, and blasticidin S was used to select for plasmid uptake. Successful transfectants fluoresced. (**C**) Upon transfection of the pDC2-coCas9-pare-BSD-mNeonGreen plasmid into *P. falciparum* Dd2, the *pfpare*-targeting gRNA directs Cas9 to make a site-directed cut of *pfpare*. The donor region facilitates homology-directed repair and the insertion of BSD and mNeonGreen. (**D**) The *pfpare* locus of blasticidin-resistant recrudesced parasites ($Dd2^{pareNG.1}$ and $Dd2^{pareNG.2}$) was PCR-amplified using 5′ and 3′ UTR primers (p2174 and p2175) to check for correct insertion of mNeonGreen. Lengths of PCR products were quantified using gel electrophoresis. No detectable wild-type product was observed in the transfected bulk culture.

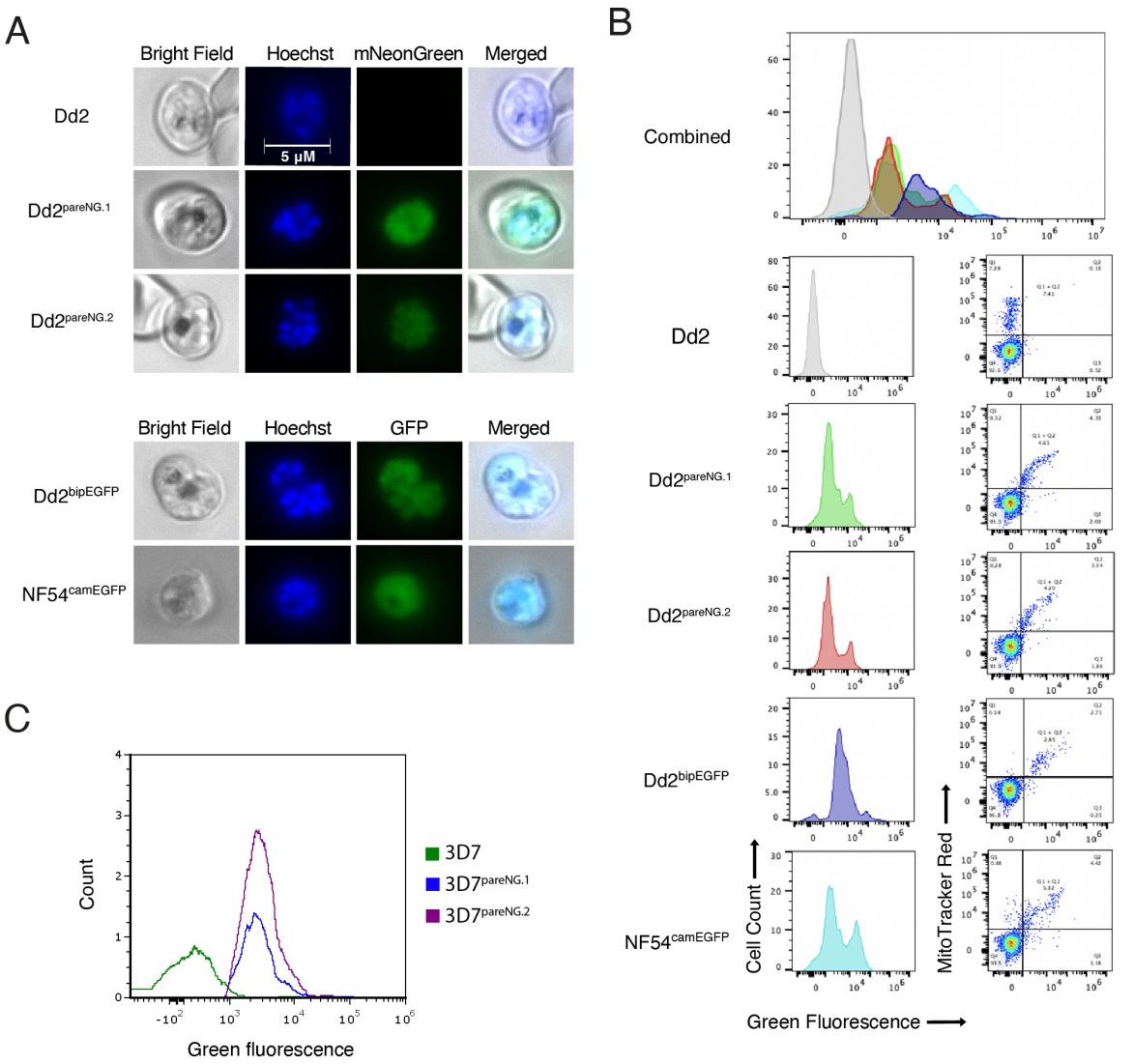


Fig. 4.7 mNeonGreen fluorescence in *P. falciparum* is comparable to GFP fluorescent lines. (**A**) Mixed-staged parasites (Dd2, Dd2$^{pareNG.1}$, Dd2$^{pareNG.2}$, Dd2^{bipEGFP}, NF54^{camEGFP}) were fixed, stained with Hoechst DNA stain, and visualised using bright-field and fluorescence microscopy at 1000x magnification. Flow cytometry of mixed-stage parasites (**B**) (Dd2, Dd2$^{pareNG.1}$, Dd2$^{pareNG.2}$, Dd2^{bipEGFP}, NF54^{camEGFP}) and (**C**) (3D7, 3D7$^{pareNG.1}$, 3D7$^{pareNG.2}$). Parasites were stained with MitoTracker DeepRed and analysed on a flow cytometer. Quantification of green fluorescent (GFP and mNeonGreen) and MitoTracker+ cells was performed using FlowJo. All single-cell, parasitised RBCs (MitoTracker+) were gated, and histograms of green fluorescence were generated (GFP+ or mNeonGreen+).

demonstrate that the expression of *pfpare* peaks at 32-40 hours post-erythrocyte-invasion, which would suggest that the expression of mNeonGreen from the *pfpare* locus would also peak during late trophozoite and schizont stages (Figure 4.8B) (Chappell et al., 2020; Amos et al., 2021). Further, I induced gametocytogenesis and observed fluorescence in all the gametocyte stages (I-V) (Figure 4.8C).

4.4.9 Dd2$^{\text{pareNG}}$ demonstrates robust fitness

To assess the fitness of the Dd2$^{\text{pareNG}}$ lines, competition assays were performed against Dd2 and 3D7. The two populations were seeded at a 1:1 ratio of fluorescent to test line and maintained for three weeks. The ratio of fluorescent (Dd2$^{\text{pareNG}}$) to non-fluorescent (Dd2 or 3D7) populations was measured every 2-3 days using flow cytometry to quantify the competition between the two populations. Dd2$^{\text{bipEGFP}}$ was also included as a control, which has a slight fitness defect from the GFP integration (Baragaña et al., 2015). The Dd2$^{\text{pareNG}}$ lines showed nearly comparable fitness to their parent line, Dd2, over three weeks (Figure 4.9A). Contrastingly when competed against 3D7, a slower-growing lab line, the Dd2$^{\text{pareNG}}$ lines outcompeted 3D7, which demonstrates the value of strain-matched competitor lines (Figure 4.9B).

4.5 Discussion and future outlook

4.5.1 Development of CRISPR-Cas13 RNA knockdown tools

I was unsuccessful in developing a functional CRISPR-Cas13 system in *P. falciparum* with either of the PspCas13b or RfxCas13d enzymes. Recent reviews have proposed Cas13 as a promising strategy, although there is no evidence of their successful application in *P. falciparum* for RNA knockdown (Bryant et al., 2019; Lee et al., 2019; Quansah et al., 2023). Multiple challenges have been associated with adopting CRISPR-Cas9-based tools in *Plasmodium*; therefore, it is unsurprising that there would be similar difficulty adopting CRISPR-Cas13 (Bryant et al., 2019; Lee et al., 2019).

There are many potential reasons why the transfections with CRISPR-Cas13 plasmids did not result in recrudesced parasites. In order to continue pursuing the application of PspCas13b and RfxCas13d into *P. falciparum* each should be explored. Firstly, it is possible that the Cas13 enzyme was toxic to the parasite. I tried to explore this possibility by designing an inducible system to control the expression of Cas13. However, no recrudesced parasites

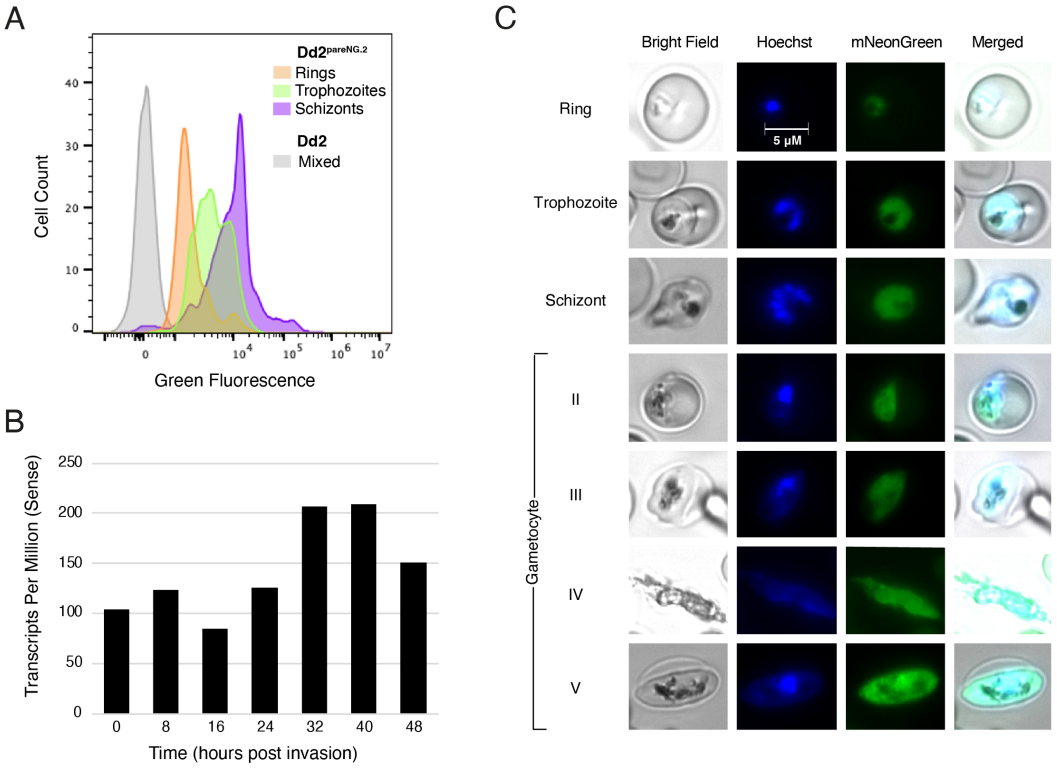


Fig. 4.8 Dd2^{pareNG} fluorescence varies between different asexual blood stages. (**A**) Dd2$^{pareNG.2}$ parasites were sorbitol-synchronised, and microscopy was used to confirm stages. Flow cytometry with MitoTracker DeepRed was used to enumerate parasites expressing green fluorescence in ring, trophozoite and schizont-staged cultures. FlowJo was used to gate single-cell, parasitised RBCs, quantify green fluorescence, and generate histograms. Dd2^{pareNG} schizonts demonstrated the greatest fluorescence, followed by trophozoites and rings. (**B**) Stage-specific RNA sequencing obtained from PlasmoDB shows that *pfpare* is expressed throughout the entire 48hr intraerythrocytic life cycle, and expression peaks at late trophozoite and schizont stages (Chappell et al., 2020; Amos et al., 2021). (**C**) Fluorescence microscopy of synchronised Dd2^{pareNG} parasites at the asexual ring, trophozoite and schizont stages, and gametocyte stages II-V (obtained through gametocytogenesis induction). Cultures were fixed, stained with Hoechst DNA stain, and visualised using bright-field and fluorescence microscopy at 1000x magnification.

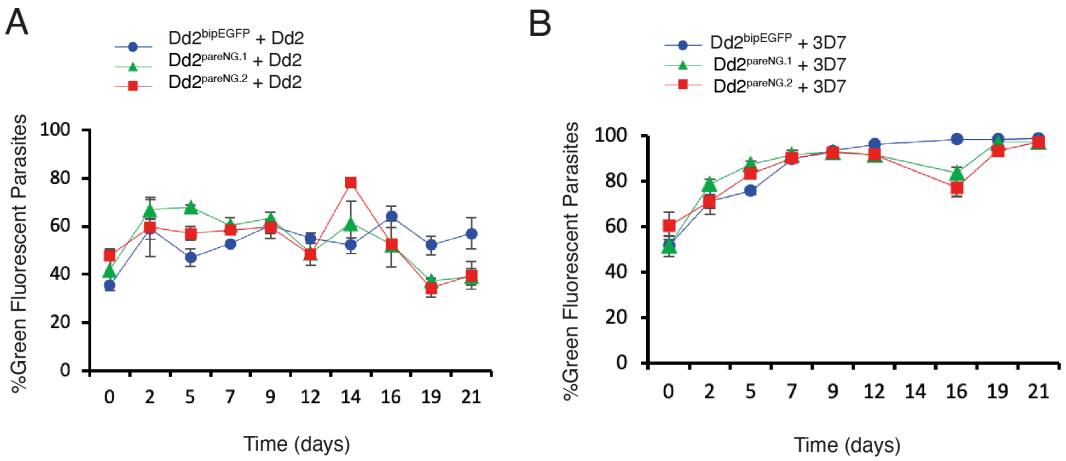


Fig. 4.9 Dd2^{pareNG} lines demonstrate comparable fitness to Dd2^{bipEGFP} when competed against Dd2 or 3D7. Fluorescent lines (Dd2$^{pareNG.1}$, Dd2$^{pareNG.2}$ and Dd2^{bipEGFP}) were competed against (**A**) Dd2 or (**B**) 3D7 in triplicate biological replicates. The two populations were seeded at a 1:1 ratio at 1% parasitaemia and maintained for 3 weeks. The parasitaemia and fluorescence were measured using flow cytometry every 2nd or 3rd day from day 0 to day 21. The percentage of green fluorescent parasites over total parasites was averaged between the triplicates and graphed over time for each line. The error bars represent the standard deviation between triplicates. The three fluorescent lines did not outcompete Dd2 and remained close to the seeded percentages over three weeks. However, the fluorescent lines quickly outcompeted 3D7.

could be generated, which, if Cas13 is toxic, could indicate that the DiCre system was leaky. To assess the toxicity of the Cas13 protein, an experiment delivering the Cas13 protein directly into the parasite (as is or enzymatically-inactivated) would be essential. I was unable to demonstrate that the PspCas13b protein was expressed in TX6-1, the single recrudesced PspCas13b transfection. This could have been due to challenges with capturing the large protein (124 kd) on the western blot. In future, western blot protocols specific to large proteins should be trialled, and also, RNA could be extracted from TX6-1 to determine if Cas13b RNA is transcribed from the CRISPR-Cas13b plasmid. TX6-1 may have a significant fitness defect, encumbering recrudescence; thus, a fitness assay should be completed to assess the costs of carrying the plasmid.

It may be worth exploring the application of LwaCas13a to RNA targeting as Cunningham et al. have already demonstrated that it can be used *in vitro* with *P. falciparum* (Cunningham et al., 2021). This enzyme was not explored due to its inferior RNA targeting efficiency and increased collateral activity, which is useful in ultrasensitive nucleic acid detection but is detrimental to specific RNA knockdown. Although if it is tolerated by the parasite, it may help inform our understanding of Cas13 toxicity in the parasite.

Secondly, if the Cas13 protein is tolerated by the parasite, the next reason to consider is an issue with the plasmid-based delivery system. Early molecular cloning efforts revealed that instability within the CRISPR-Cas13 plasmid when it is propagated in bacterial culture could be causing susceptibility to truncation. It is unlikely that the rest of the plasmid would be a source of toxicity as this system works well with other Cas enzymes, excluding notably the added U6 cassettes (U6psp and U6Rfx). Nasamu et al. have shown that bacterial artificial chromosomes can be used to express unstable sequences in *P. falciparum* episomally and still permit functional CRISPR-Cas9 activity (Nasamu et al., 2021). This system could permit the expression of this plasmid as long as the instability is not occurring in the parasite. RNPs could also be explored as they have been used for CRISPR-Cas13 knockdown in human T cells (Méndez-Mancilla et al., 2022) and for Cas9 systems in *P. falciparum* (Crawford et al., 2017). Other delivery methods might also help with dosage moderation. Dosage effects were observed in studies integrating Cas13 into the *Drosophila* genome, where homozygous transgenes exhibited toxicity while heterozygous transgenes did not (Huynh et al., 2020). On a related note, the transfection technique used in this study is inefficient with a transfection efficiency of approximately 10^{-5}-10^{-6} (Meissner et al., 2007). In cases of toxicity, insufficient uptake of the plasmid would further contribute to the lack of parasite

recrudescence. Alternative transfection protocols like pre-loading or schizont transfection may improve plasmid delivery (Carrasquilla et al., 2020).

Thirdly, Cas13 is capable of off-target RNA cleavage, which can be toxic to eukaryotes (Bot et al., 2022). Although minimal collateral activity was observed in bacterial systems with PspCas13b and CasRx, the extent of their collateral activity in eukaryotes has yet to be completely elucidated (Bot et al., 2022). For Cas13 enzymes to be adopted more broadly in eukaryotes, more sensitive tools must be developed to assess collateral damage. Once Cas13 expression is established in the parasite, an exploration of gRNA constraints and preferences should be completed. Models that predict optimal gRNA design have been developed from Cas13d knockdown screens of GFP transgenes using thousands of gRNAs (Wessels et al., 2020). Although there are no PFS requirements, these other studies have suggested that Cas13 gRNAs have sequence preferences largely based on structure-based targeting. It is vital that the target region is single-stranded to allow Cas13-mediated cleavage. Even though prediction software can predict RNA structure, comprehensive structure-seq data has recently been completed in *P. falciparum* by Alvarez et al., which may better inform the targeting of the transcriptome (Alvarez et al., 2021). Identifying optimal targets and efficient gRNAs will be crucial for ncRNAs targeting, where duplex RNA structures are abundant.

This study has demonstrated the application of Cas13 into *P. falciparum* will be challenging and may require significant time and resources to resolve or may not be possible. However, the potential pay-off of achieving efficient programmable RNA-targeting in *Plasmodium* is considerable and requires further exploration. In addition to investigating why the approaches I used have not worked, exploring other delivery methods and improving gRNA design, newly-discovered classes of Cas13 enzymes should also be explored. Ultra-compact Cas13 enzymes such as Cas13X.1 (445aa) and Cas13-bt (775-804aa) have demonstrated comparable RNA knockdown efficiency to current Cas13 enzymes at a fraction of the size (Xu et al., 2021; Kannan et al., 2022).

4.5.2 Disruption of lncRNAs *in vitro* using existing tools

I successfully disrupted a lncRNA (PF3D7lncRNA_0567) in *P. falciparum* using CRISPR-Cas9-mediated insertion of a selectable marker, confirming that it is an approach that can be used to study *intergenic* lncRNAs *in vitro*. This agrees with the literature, where two studies successfully disrupted two other lncRNAs using the same method (Filarsky et al., 2018; Batugedara et al., 2022). Although, this technique is most suited to intergenic targets

as targets in more complicated circumstances would require careful knockout strategies that avoid disrupting nearby features if knockout is even feasible. Furthermore, both Δ567 mutants had a drug-resistant phenotype to atovaquone. This result provides evidence that the mutation seen in this lncRNA in the drug-resistant lines, adenine to thymine SNP, may contribute to atovaquone resistance.

The other targets PF3D7lncRNA_0019 and PF3D7lncRNA_0337 could not be disrupted, which could suggest that they are essential in asexual blood stage development. Neither of which were identified as putatively essential in *Chapter 3*. However, it is also possible that an issue with the knockout occurred, such as ineffective or toxic gRNAs or unsuccessful DNA repair. Moreover, the insertion of an 1800 bp sequence into the genome sequence in place of 3070 bp and 705 bp lncRNAs could disrupt chromosomal conformation impacting other genes that may be essential. A conditional knockout and gene interference would aid in confirming essentiality. This highlights a major drawback to knockouts: essential lncRNAs cannot be targeted. In high-throughput experiments such as a reverse genetics screen with thousands of targets, some of which would be essential lncRNAs, other techniques would be required to accompany or replace Cas9. A strategy used for essential genes that may be applicable to essential lncRNAs is conditional gene editing using *lox*P sites contained in a synthetic intron (Jones et al., 2016). The intron is introduced into essential genes but is actively spliced out so does not compromise gene expression. When the *lox*P sites are conditionally activated to induce alter the gene, such as removing localisation signals or binding domains.

Without a functional RNA knockdown method, gene interference is a suitable alternative technique that can disrupt essential genes. Therefore, in the following chapter, I employed dCpf1-mediated gene interference for the functional characterisation of a set of prioritised lncRNAs. To ensure that transfection conditions can be appropriately monitored in larger-scaled disruption experiments, robust transfection controls must be used. Therefore, in addition to changing the Cas enzyme, I also changed the positive transfection controls because the controls used in the Cas13 and Cas9 experiments failed to generate mutants reliably.

In this work, I only explored CRISPR-based approaches for lncRNA-targeting, but there are non-CRISPR-based approaches. For example, locked nucleic acids (LNAs) are widely used in human lncRNA studies because they have a special modification that triggers RNAse-H-mediated degradation when the oligo forms an RNA:DNA hybrid with the target RNA

(Sarma et al., 2010). Similar methods have been used in *P. falciparum* including morpholino oligomers (Augagneur et al., 2012, 2013; Garg et al., 2015), phosphorothioate antisense oligodeoxynucleotides (PS-ODNs) (Rapaport et al., 1992; Barker et al., 1996, 1998; Wanidworanun et al., 1999; Noonpakdee et al., 2003), 2'-OMe modified PS-ODNs (Razavi Vakhshourpour et al., 2022), peptide nucleic acids (Kolevzon et al., 2014), and ribozymes (Flores et al., 1997; Prommana et al., 2013). A peptide nucleic acid was even used to target a *var* antisense lncRNA (antisense to PF3D7_421300) (Amit-Avraham et al., 2015). Many large-scale mammalian lncRNA interrogations use RNAi-based approaches; however, there is a novel approach that may warrant investigation. XNAzymes, which are DNAenzymes that have been modified using synthetic alternative nucleic acids to act as site-specific RNA endonucleases (Donde et al., 2022; Gerber et al., 2022). Although these approaches are potentially functional alternatives, they lack the efficiency and scalability that CRISPR-based approaches offer.

4.5.3 Rapid generation of fluorescent parasite lines

Genetic engineering of fluorescent proteins has improved their functionality as tools in molecular biology, and the integration of these enhanced proteins into existing applications in malaria research should be explored. In this work, I developed an efficient and facile approach for generating new fluorescent reporter lines in *P. falciparum* using CRISPR-based integration of mNeonGreen. mNeonGreen was first applied to *P. falciparum* to generate a tagged Exp2 protein; however, here, I assess its broader applicability in developing reporter lines for competitive fitness assays and visualisation (Glushakova et al., 2018). Therefore, I designed a CRISPR/Cas9 construct to integrate mNeonGreen into *pfpare*, a highly conserved, nonessential gene that would endogenously drive expression throughout the intraerythrocytic cycle.

I used the construct to fluorescently tag the Dd2 and 3D7 *P. falciparum* strains. I then further assessed Dd2$^{\text{pareNG}}$ for features suitable in a reporter line, i.e. the strength of fluorescence, localisation, stage-specificity, and impact on parasite fitness. Dd2$^{\text{pareNG}}$ demonstrated strong fluorescence diffused throughout the parasite. The fluorescence was expressed throughout asexual blood stage development and during gametocytogenesis, suggesting that these tagged lines are a valuable tool for *in vitro* research studying the biology of *P. falciparum* in these stages. Gametocyte-competent lines that express fluorescent and luciferase reporters help evaluate antimalarial activity against sexual stages; however, before advances in genome editing, these lines were laborious to construct (Adjalley et al., 2011). The editing strategy

would allow mNeonGreen and potentially other reporters to be inserted into different strains under investigation. The ability to efficiently tag Dd2 indicates that minor differences in the homology regions between the donor, based on the 3D7 sequence, and the target region are tolerated. The absence of wild-type locus in bulk transfections reflects the effect of positive selection resulting from *bsd* expression from the endogenous *pfpare* promoter. Although disruption of *pfpare* also affords the possibility of negative selection using the commercially available compound MMV011438, this was not required in practice due to the stringency of the blasticidin S positive selection. Integrating mNeonGreen into *pfpare* caused a minor fitness defect leading to slightly slower growth compared to Dd2, similar to other GFP-based reporters (Baragaña et al., 2015). However, this defect was relatively minimal, as the resulting line outcompeted 3D7 in a competition assay. These findings support that the Dd2$^{\text{pareNG}}$ is a suitable reporter line and is comparable with the standard *P. falciparum* GFP lines that are currently used.

The Dd2 mNeonGreen reporter line and the construct, which can generate new reporter lines, will support malaria research. Dd2$^{\text{pareNG}}$ is a valuable addition to the repertoire of reporter lines in *P. falciparum* that facilitate experiments involving visualising, tracking and counting parasites. The construct will enable the rapid generation of other fluorescently-tagged *P. falciparum* parasites from different strains, such as field isolates, which can facilitate their study *in vitro*. I exemplified this in the *Chapter 5* by using the 3D7$^{\text{pareNG}}$ to facilitate competition assays of lncRNA-disrupted mutants.

4.6 Additional methodology

See relevant *General Methodology* sections for molecular cloning, culturing, transfection, flow cytometry, fluorescent imaging and competition assays.

4.6.1 Guide RNA and primer sequences

The gRNAs and primers used in this chapter are listed in Tables B.1 and B.2 in Appendix B.

4.6.2 Plasmid maps

Maps for the plasmids used in this chapter are in Appendix B.

4.6.3 Site-directed mutagenesis

Oligonucleotides were synthesised with the desired mutation and flanked with unmodified sequence (p1795 and p1796). 125ng of each primer was combined with 5-50ng of the construct, 1μL dNTP mix, 5μL 10X reaction buffer and 1μL PfuUltra HF DNA polymerase (2.5U/μL). The mixture was incubated at 95°C for 30 seconds and for 12-18 cycles of 95°C for 30 seconds, 55°C for 1 minute, 68°C for 1 minute per kb. It was then cooled to RT, digested with 1μL DpnI (10U/μL) and incubated for 37°C for 1 hour to digest other amplified products. 5μL was run on a gel, and if the expected band was visible, competent cells were transformed with 1μL of the reaction mixture, following normal molecular cloning transformation conditions.

4.6.4 *pfpare* sequence alignment and expression

DNA sequences for *pfpare* locus (PF3D7_0709700) in *P. falciparum* 3D7, HB3, 7G8, GB4, CD01, GA01, IT and Dd2 strains were obtained from PlasmoDB and alignment figures were generated using Clustal Omega (Goujon et al., 2010; Sievers et al., 2011; Amos et al., 2021). Transcriptomic data of 3D7 *pfpare* expression in the intraerythrocytic life cycle was obtained from PlasmoDB; specifically, RNA-sequencing data from Chappell et al. (2020) was used.

Chapter 5

Characterising *Plasmodium falciparum* lncRNAs using CRISPR interference

5.1 Overview

Few of the thousands of lncRNAs annotated in *P. falciparum* have been experimentally studied. Therefore, most lncRNAs have no predicted or known function. In this chapter, I endeavoured to demonstrate the feasibility of lncRNA disruption screens in *P. falciparum* and characterise a small set of lncRNAs. Uncovering specific functions of lncRNAs will aid in understanding their role as regulatory molecules in the *P. falciparum* parasite.

This work contains data and tools generated by colleagues. The identification of stage-specific lncRNAs from single-cell RNA-Seq was completed by Dr Juliana Cudini as part of her thesis (Cudini, 2021). LncRNA-disruption was completed using a plasmid (CRISPR-dCpf1Sir2a) created by Dr Sophie Adjalley (unpublished). These colleagues are cited in the relevant sections for their contributions. Supplemental data and acknowledgements are in Appendix C.

5.2 Introduction

A recent review has claimed that the role of lncRNAs has been overstated and is not sufficiently supported by functional experimentation (Ponting and Haerty, 2022). Although

this review describes lncRNAs implicated in human disease, these questions are pertinent to malaria, where only a couple of lncRNAs have been functionally characterised. There is evidence that lncRNAs are actively involved in the regulation of the *P. falciparum* transcriptome; however, it is unclear to what extent this role is shared by the thousands of lncRNAs identified in the parasite. CRISPR-based screens can enable the systematic characterisation of lncRNAs transcriptome-wide; however, they have not yet been employed to study lncRNAs in *P. falciparum*.

5.2.1 CRISPR-based reverse genetic screens for mammalian lncRNAs

Mammalian lncRNAs have been disrupted using various CRISPR-based approaches for reverse genetic screening. Cas9 has been used to delete lncRNA loci (Zhu et al., 2016) and lncRNA splice sites (Liu et al., 2018). CRISPR activation with dCas9-VP64 has been used to overexpress lncRNAs (Bester et al., 2018) while CRISPR interference with dCas9-KRAB has been used to repress lncRNAs expression (Liu et al., 2017; Cai et al., 2020). LncRNAs have also been knocked down directly at the RNA level with Cas13 (Xu et al., 2020). In these screens, the disrupted cells are predominantly characterised through gene expression profiling with additional phenotyping based on the cell type or tissue, such as assessing morphology, essentiality and growth assays. Many lncRNA functions have been elucidated from this approach, such as *PRANCER*, which plays a role in epidermal homeostasis, *GAS6-AS2 lncRNA*, which activates a chemotherapy resistance mechanism in several cancers and *FOXD3-AS1*, which is essential for haematopoietic stem cell pluripotency and differentiation (Bester et al., 2018; Cai et al., 2020; Haswell et al., 2021).

5.2.2 Considerations in lncRNA-targeting reverse genetic screens

In addition to the technical challenges outlined in *Chapter 4*, there are other obstacles to targeting and studying lncRNAs *in vitro*. Firstly, when lncRNAs are disrupted, there may not be any consequential phenotype. LncRNAs can be involved in complex regulatory networks, where redundancies or compensatory mechanisms could void the impact of a disruption to a regulatory lncRNA. For example, a recent study in zebrafish disrupted a selection of 25 lncRNAs and found no impact on embryogenesis, viability and fertility (Goudarzi et al., 2019). Even lncRNAs with known roles like *MALAT1* and *HOTAIR* report no observable phenotypes and appear dispensable when knocked out (Nakagawa et al., 2012; Amandio et al., 2016). This example reflects the second challenge, which is that a phenotype may not be observable or measurable using the assays available. Taking *P. falciparum* culturing as

an example, there are a limited number of measurable phenotypes that can be quantified at scale. These include parasite fitness, drug sensitivity, observable morphology and expression profiling, but the impact of the loss of regulation from the disruption of a lncRNA may not be evident in these assays.

5.2.3 Reverse genetic screens in *P. falciparum*

Reverse genetic screens have been completed in *Plasmodium*, made possible by the large-insert genomic libraries created by the *Plasmodium* genetic modification (PlasmoGEM) project (Ishizaki et al., 2022). The project developed a scalable pipeline that enables the generation of *P. berghei* knockout vectors and overcomes the challenges of amplifying AT-rich long homology arms in *E. coli* (Pfander et al., 2011). Over 2900 *P. berghei* knockout vectors have been generated and exploited for reverse genetics approaches. Furthermore, the vectors also contain barcodes so they can be transfected and quantified using next-generation sequencing. This approach, in combination with phenotypic assays, has characterised over 3000 genes in *P. berghei* across different stages of the life cycle (Gomes et al., 2015; Bushell et al., 2017; Stanway et al., 2019; Russell et al., 2023). Major advances have been made in CRISPR-based gene editing in *Plasmodium* sp.; however, these tools have not yet permitted large-scale reverse genetic screens (Ishizaki et al., 2022).

CRISPR interference is an attractive avenue for reverse genetics screens in *Plasmodium* because the design and generation of donor sequences are not required. CRISPR interference induces transcriptional repression instead of DNA breaks; therefore, it does not require repair templates to facilitate homology-directed repair or rely on low-frequency microhomology-mediated end joining. The organisation for a screen is then simplified, only requiring gRNAs. CRISPR interference tools are functional in *P. falciparum* such as dCas9, dCas9Sir2a, and dCas9GCN5; however, their applications to large-scale genomic screens have been slowed by insufficient efficacy (Xiao et al., 2019).

In this chapter, I aimed to demonstrate the feasibility of a transcriptome-wide lncRNA interference screen using a relatively small number of lncRNAs. To enrich the target set for lncRNAs with potential biological significance and measurable phenotypes, I identified *P. falciparum* lncRNAs associated with drug resistance, stage specificity and intergenicity. I disrupted the targetable lncRNAs using CRISPR-dCpf1 fused to Sir2a, a transcriptional repressor. I then characterised a subset of the lncRNA-disrupted lines using expression

profiling and various phenotypic assays quantifying fitness, gametocytogenesis, and drug sensitivity.

5.3 Objective and Aims

To characterise *P. falciparum* lncRNAs using CRISPR interference.

1. Identify lncRNAs with potential biological significance
2. Generate lncRNA-disrupted *P. falciparum* mutants using CRISPR-dCpf1Sir2a
3. Characterise lncRNA-disrupted mutants

5.4 Results

A CRISPR interference approach was designed for targeting *P. falciparum* lncRNAs, with the knowledge acquired and tools developed in the preceding chapters. dCpf1 was determined to be the most suitable and functional Cas enzyme at my disposal, and the 3D7^{pareNG} line would be useful in the phenotyping of lncRNA-disrupted mutants. The approach consisted of (1) generating and transfecting 3-4 CRISPR-dCpf1Sir2a plasmids containing different gRNAs for each target into 3D7; (2) selecting for plasmid uptake; (3) confirming target interference with RT-qPCR, and (4) phenotyping lncRNA-disrupted mutants with expression profiling and assays (Figure 5.1). Targets were selected taking the approach into consideration.

5.4.1 Selection of *P. falciparum* lncRNA targets

To enrich the target set for measurable phenotypes and explore a variety of target prioritisation approaches, I identified 54 targets based on features that may indicate potential biological significance. Five interest groups were identified based on four attributes: association with drug resistance mutations, stage-specificity (asexual blood stages or gametocyte stages), long-length and intergenicity (Figure 5.2).

The drug resistance and stage-specific target sets were prioritised as they provide measurable phenotypes when disrupted. Drug sensitivity can be quantified with drug assays, and involvement in a specific life cycle stage can be assessed through microscopy, fitness and gametocyte induction. For selecting lncRNAs associated with drug resistance, a similar approach was

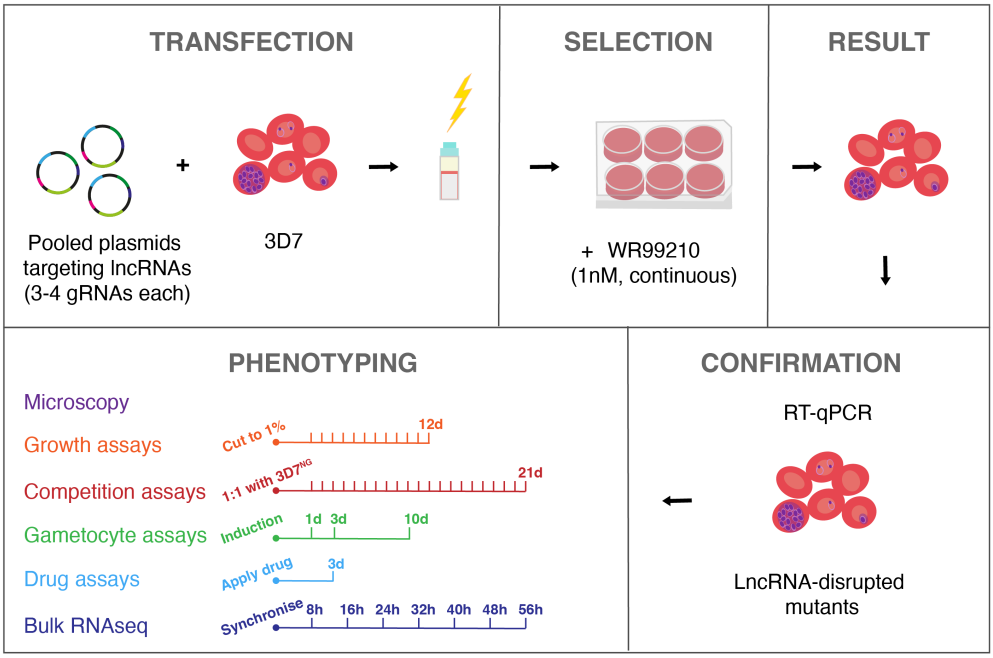


Fig. 5.1 Schematic of approach to lncRNA disruption and characterisation. 3-4 gRNAs were designed for each target and ligated into separate CRISPR-dCpf1Sir2a plasmids. Plasmids for the same target were pooled at 30μg each and transfected into 3D7. Plasmid uptake was selected for using WR99210 (1nM). Target interference was confirmed with RT-qPCR, and mutants were phenotypes with expression profiling and various assays. Timelines are drawn denoting when measurements were taken. h: hour(s); d: day(s).

employed to that used to select lncRNA targets in *Chapter 4*. Although instead of using the Broadbent annotation (Broadbent et al., 2015), the annotation generated in *Chapter 3* was intersected with the list of intergenic mutations obtained from whole genome sequencing of drug-resistant lines. This SNP and InDel list was generated by labs in the Malaria Drug Accelerator (MalDA) consortium and obtained from Dr Madeline Luth (University of California San Diego). 10 lncRNAs were identified that contained mutations associated with drug resistance in two different compound selections or that contained multiple mutations associated with resistance to the same compound (Table 5.1 and C.1). LncRNAs associated with stage-specificity were identified by Dr Juliana Cudini (Wellcome Sanger Institute). As part of her PhD thesis, she used single-cell RNA sequencing to analyse the stage-specificity of lncRNA expression in *P. falciparum* using in part the annotation generated in this thesis (Cudini, 2021). 24 highly stage-specific lncRNAs were identified: 11 were gametocyte-stage-specific, and 13 were specific to asexual blood stages (Table 5.1 and C.1). Two of these targets contain sense-antisense pairs of lncRNAs (1342 & 1343 and 0787 & 0788).

Research has suggested that lncRNAs with unique features such as long lengths, high GC content or multi-exonic sequences may be notable (Broadbent et al., 2015). I decided to explore a small group of seven long-length lncRNAs. Six were selected with lengths above 5000 bp, and one additional target was selected with a shorter length of 1170 bp but is a potential example of a cluster of two lncRNAs (Table 5.1 and C.1).

The final interest group of *intergenic* lncRNAs was selected for two reasons. Firstly, *intergenic* lncRNAs are independently regulated as they do not tend to share bidirectional promoters, which could indicate biological significance. Secondly, they are robust to varying targeting strategies. If the CRISPR-dCpf1Sir2a transfections were unsuccessful, a contingency approach such as a CRISPR-Cas9 knockout strategy would be feasible. Of the 281 *P. falciparum* lncRNAs classified as *intergenic* (see *Chapter 3*), 67 have a high ranking in supportive evidence (1st rank), meaning that they were captured by both long-read RNAseq datasets and had multiple sources of TSS evidence. To further narrow the target set to 14 independently regulated lncRNAs, I looked for evidence of stage-specific TSSs and lack of bi-directional promoters (Table 5.1 and C.1).

5.4.2 Disruption of targets with CRISPR-dCpf1Sir2a

Multiple gRNAs can increase the efficiency of CRISPR interference; therefore, I designed four (23 bp) gRNAs for each target (McCarty et al., 2020). The gRNAs were designed to tile

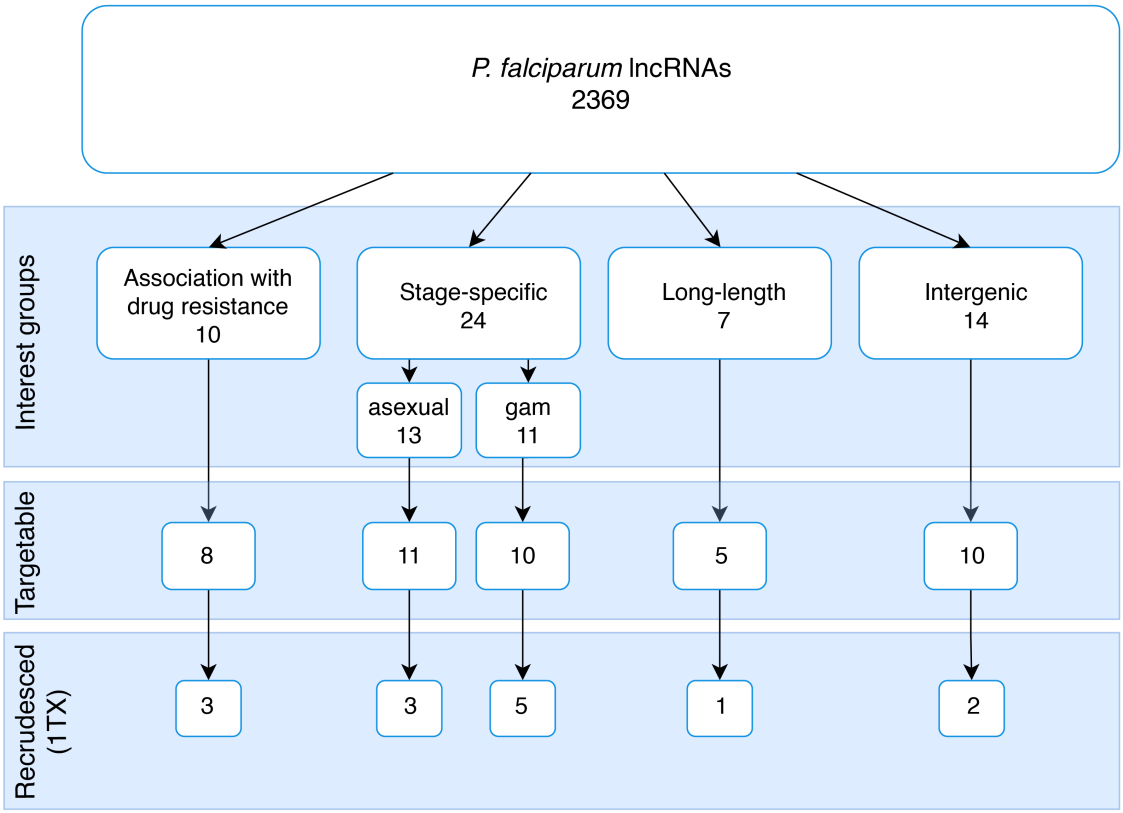


Fig. 5.2 Target selection: identification of five key interest groups. Four criteria were used to identify lncRNAs in five interest groups: association with drug resistance, stage-specificity for asexual blood stages or gametocyte stages, long-length and intergenicity, which contain a total of 55 targets. gRNA design eliminated 11 other targets. The remaining 44 lncRNAs were disrupted using CRISPR-dCpf1Sir2a, and 14 recrudesced. blood: asexual blood stage-specific gam: gametocyte-stage-specific; TX: transfection

the region near the transcriptional start site as previous work in our lab using CRISPR-dCpf1 (with and without Sir2a) to interfere with GFP expression identified that gRNAs targeting the $55'$ end were most effective. The gRNAs could be on either strand, although gRNAs targeting the same strand were prioritised. gRNAs were designed using Benchling and DNAstar, and BLAST was used to check for off-target hits in the *P. falciparum* 3D7 reference genome (Johnson et al., 2008; Benchling, 2020). 11 lncRNA targets were eliminated because highly specific gRNAs could not be designed (Table C.1, Figure 5.2). The lncRNAs were situated near gene families like *var* and *rif* genes or ribosomal RNA, sequences with multiple copies in the genome. A total of 44 targets remained, including four where only three gRNAs could be designed (Table 5.1).

Each gRNA was ligated into separate CRISPR-dCpf1Sir2a plasmids (Figure C.1). Plasmids containing gRNAs for the same target were pooled at 30μg each and transfected into wildtype (WT) 3D7 (Table 5.1). Three controls were included in every batch: a positive transfection control (pDC2-ef1a-Vps4wt-mRFP-hDHFR), a non-targeting gRNA control and a gene-targeting gRNA control (targeting *mfr3*, (PF3D7_0312500)). WR99210 (1nM) drug pressure was applied the following day and maintained continuously to select for plasmid uptake and maintenance. The transfections were only completed once for each target. 14 transfections recrudesced, the disrupted targets span all five target sets (Table 5.1, Figure 5.2). The positive transfection control recrudesced in every batch, and the gene-targeting gRNA recrudesced in one batch. The non-targeting guide was attempted five times with no successful recrudescence, although in two transfections, parasites appeared transiently.

5.4.3 CRISPR-dCpf1Sir2a interferes with the transcription of *P. falciparum* lncRNAs

To assess the effectiveness of the gene interference in the 14 recrudesced lines, the relative expression of each target was determined using RT-qPCR. qPCR primers were designed to amplify *mfr3*, the lncRNA targets and two housekeeping genes (cyclophilin (PF3D7_0510200) and 18S rRNA (PF3D7_0725600)) (Table C.3). Effective primers could not be generated for lncRNA targets PF3D7lncRNA_0342 and PF3D7lncRNA_1342. Three primer sets were generated for both targets and attempted at varying annealing temperatures; however, sufficient amplification could not be obtained. For the 12 remaining lines, RNA was extracted from mixed stages cultures, and RT-qPCR was completed in triplicate using 3D7 as the control line. The relative change in expression was calculated using the $\Delta\Delta$CT method

Table 5.1 Summary of CRISPR-dCpf1Sir2a lncRNA-targeting transfections in *P. falciparum*

Interest group	LncRNA number	Phenotype	Number of gRNAs	TX recru.
Association with drug resistance	0282	BI-2536 & Suloctidil	4	0
	0335	MMV006901 & MMV0407834	4	0
	0675	MMV007181 & OSM-S-106	4	0
	0736	BI-2536 & Suloctidil	4	1
	1098	MMV024114 & MMV028038	4	0
	1690	MMV020746 & MMV665794	4	1
	1967	BI-2536 & OSM-S-106	4	0
	2149	MMV020746 & MMV008434	4	1
Asexual blood-stage-specific	0127	Late schizonts	4	0
	0337	Late schizonts	4	0
	0456	Late schizonts & males	4	1
	0553	Late trophozoites & early schizonts	4	0
	0801	Late schizonts	4	0
	1120	Late schizonts	4	0
	1342 & 1343	Rings	3	1
	1471	Early schizonts	4	0
	1478	Late schizonts	3	1
	1696	Late schizonts	3	0
	2342	Rings	4	0
Gametocyte-stage-specific	0055	Gametocytes & early trophozoites	4	1
	0364	Late males	4	1
	0503	Males	4	0
	0787 & 0788	Early gametocytes	4	1
	0979	Late gametocytes & trophozoites	3	0
	1129	Gametocytes	4	0
	1191	Late gametocytes	4	1
	1259	Late gametocytes	4	1
	1611	Males	4	0
	2244	Early gametocytes	4	0
Long-length	1132	7452 bp	4	0
	1890	5836 bp	4	0
	2003	6854 bp	4	0
	2027	5987 bp	4	0
	2353	Cluster	4	1
Intergenic	0342	Intergenic	4	1
	0370	Intergenic	4	0
	0614	Intergenic	4	1
	0740	Intergenic	4	0
	0785	Intergenic	4	0
	0981	Intergenic	4	0
	1210	Intergenic	4	0
	2184	Intergenic	4	0
	2323	Intergenic	4	0
	2361	Intergenic	4	0

TX: transfection; recru.: recrudesced

(described in *General Methodology*) (Pfaffl, 2001). All lncRNA lines bar one demonstrated a decreased relative expression of the target lncRNA in mixed cultures (Figure 5.3A). The level of decreased relative expression varied substantially between the lines, with mean fold changes ranging from 0.27 to 0.87. These measures of gene interference are of bulk, mixed-stage cultures, which are likely heterogeneous and are not indicative of the exact level of repression that may be discerned by using clonal and synchronised cultures.

The one line that did not have decreased relative expression, Δ0055, had a 2-fold increase in PF3D7lncRNA_0055 expression. The gene-targeting line, Δ*mfr3* also had increased relative expression of its target gene, *mfr3* (Figure 5.3B). To investigate if dCpf1Sir2a was expressed in these two lines, qPCR with primers for dCpf1Sir2a was completed using the generated cDNA. Δ0055 had comparable dCpf1Sir2a expression to lines with evidence of target repression (Δ2353) unlike Δ*mfr3*, which had significantly depleted dCpf1Sir2a expression approaching zero (Figure 5.3C).

5.4.4 Selection of six lncRNA-disrupted mutants for further characterisation

Of the 11 lines with evidence of lncRNA repression by CRISPR-dCpf1Sir2a, six were selected for further investigation and phenotyping. The selection enriched for lines with greater repression of the target lncRNA to improve the possibility of identifying consequential phenotypes. The RT-qPCR was repeated for the selected lines, using three biological replicates, and maintained in culture for varying times (all under two weeks). For five of the six lines (all except Δ2149), the biological replicates demonstrated higher mean relative expressions. However, these lines also showed a high degree of variability between biological replicates (Figure 5.3D).

The lncRNA-disrupted mutants selected for characterisation span four interest groups: drug resistance, asexual blood-stage specificity, gametocyte-stage specificity and intergenicity.

> Δ2149 interferes with an *antisense-to-gene* lncRNA situated on chromosome 14, which is expressed in rings and trophozoite stages of the IDC (Broadbent et al., 2015). It is antisense to PF3D7_1420100, a conserved protein of unknown function, and shares a bidirectional promoter with PF3D7_1420000 (splicing factor 3B subunit 4). Lab-generated parasite lines with drug resistance to two drugs with similar structures, MMV020746 and MMV008434, contained a 7 bp deletion (TATATTC→T) in this

lncRNA at position 728. The mutation is within the part of the lncRNA that is antisense to the 3′ UTR of PF3D7_1420100.

Δ0456 and Δ1478 targeted lncRNAs associated with late schizonts, although the Δ0456 target is also associated with male gametocytes (Cudini, 2021). The first lncRNA (PF3D7lncRNA_0456) is intergenic and appears to share a bidirectional promoter with PF3D7_0525400, 7-helix-1 protein, a protein involved in stress granule formation and associated with the formation of female gametocytes (Bennink et al., 2018). While PF3D7lncRNA_1478 is an *antisense-to-UTR* lncRNA on chromosome 11 that is antisense to the 5′ UTR of PF3D7_1136900, subtilisin-like protease 2 (*SUB2*), a protein shown to be required for the sealing of host RBCs during merozoite invasion (Collins et al., 2020).

Δ0364 and Δ1191 targeted lncRNAs associated with gametocytes: PF3D7lncRNA_0364 with male gametocytes and PF3D7lncRNA_1191 with late gametocytes. The first lncRNA (PF3D7lncRNA_0364) is an *antisense-to-gene* lncRNA on chromosome 5, antisense to PF3D7_0500900 (serine/threonine protein kinase) and shares a promoter with PF3D7_0500800 (mature parasite-infected erythrocyte surface antigen). The second (PF3D7lncRNA_1191) is an *antisense-to-UTR* lncRNA on chromosome 10, antisense to PF3D7_1008700 (tubulin beta chain) to which it shares a bidirectional promoter.

Finally, Δ0614 targets an *intergenic* lncRNA on chromosome 6, which is maximally expressed in schizont stages of the IDC (Broadbent et al., 2015).

5.4.5 Fitness is diminished in all lncRNA-disrupted mutants

Two of the six lncRNA-disrupted lines selected for characterisation had targets with associations with asexual blood stages, specifically late schizonts (PF3D7lncRNA_0456 and PF3D7lncRNA_1478). To understand the impact of the disruption of these lncRNAs on the IDC: growth and competition assays were completed to quantify fitness, and microscopy was completed to capture any morphological changes. For the growth assay, lines were seeded in triplicate at 1% parasitaemia, maintained at 0.5-3% parasitaemia for 12 days, and cumulative parasitaemia was calculated (see *General Methodology*). The vector control (the attempted Δ*mfr3*), which carries the CRISPR-dCpf1Sir2a plasmid but does not express dCpf1Sir2a, had no fitness defect compared to the parental 3D7 line (Figure 5.4A). However, without

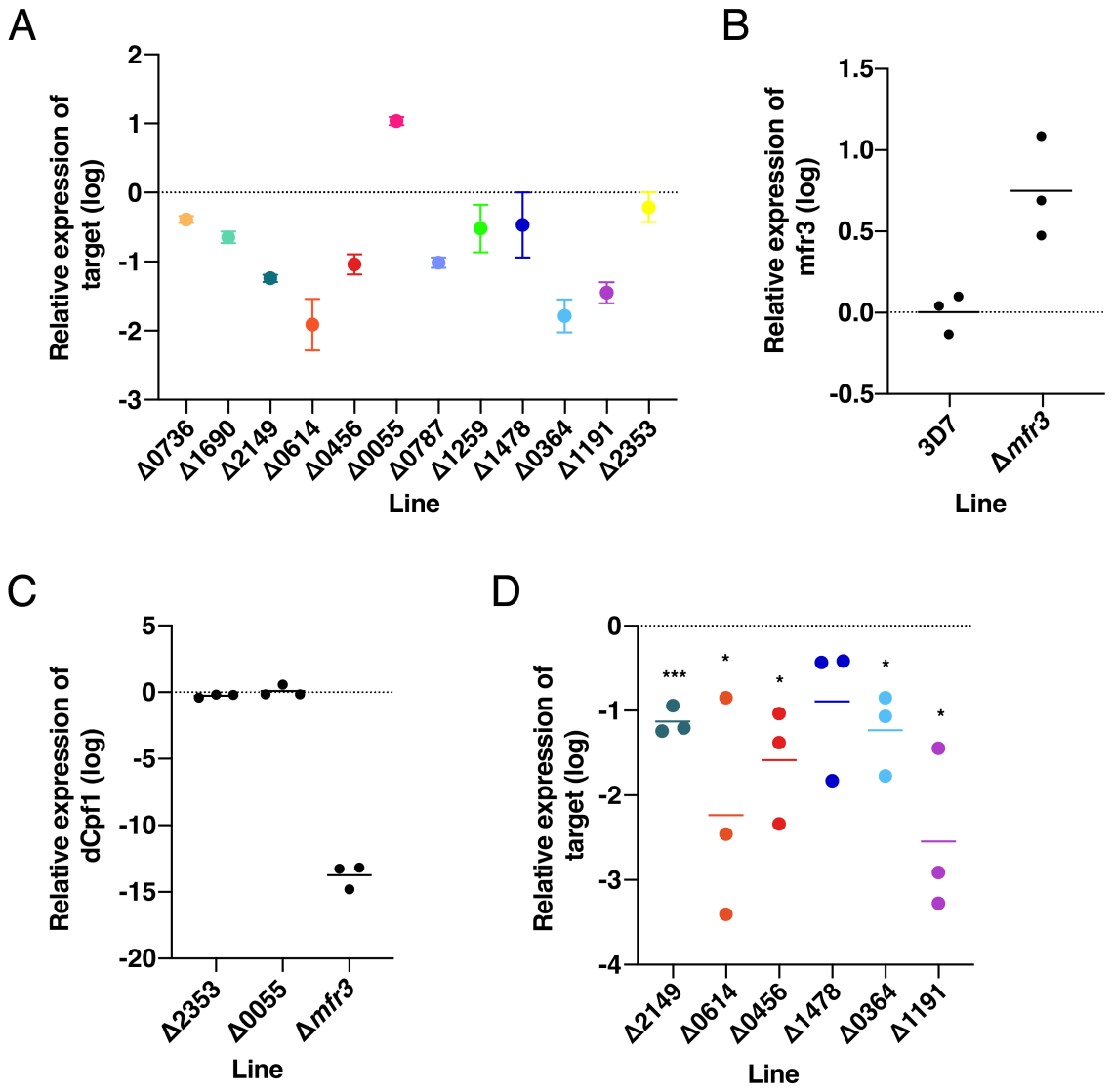


Fig. 5.3 dCpf1Sir2a-based targeting interferes with the transcription of *P. falciparum* lncRNAs. RT-qPCR was completed on RNA extracted from mutant lines and 3D7 using primers for the target and housekeeping genes (cyclophilin and 18S). Relative expression of the target lncRNA was calculated for each mutant (**A**) and the gene-targeted control (**B**) using the ΔΔCT method and 3D7 as a control. Error bars represent standard deviation between technical replicates. (**C**) The relative expression of dCpf1Sir2a was determined in the *mfr3*-targeted line compared to two lncRNA-targeted lines (Δ2353 and Δ0055). (**D**) Additional biological replicates for relative target expression were completed for six lines selected for further characterisation. The coloured lines represent the mean of the biological triplicates. Two-tailed t-tests were completed between every line and the control (3D7). Significant p-values are denoted with representative asterisks: *(0.01 to 0.05), and ***(0.0001 to 0.001).

dCpf1Sir2a expression, this result does not serve as an appropriate measure of the impact that containing a functional dCpf1Sir2a in *P. falciparum* has on fitness. The lncRNA-disrupted lines demonstrated slower growth compared to the parent line 3D7 and the vector control. To further investigate fitness, I completed competition assays, competing the lines head-to-head with the fluorescent line 3D7^{pareNG} over three weeks for a more sensitive readout of parasite fitness (Figure 5.4B). The measurements from days 20 and 21 were discarded due to technical challenges with the flow cytometer. Most of the lines outcompeted the strain-matched fluorescent line. Δ2149 and Δ0614 were initially outcompeted by 3D7^{pareNG} for the first 10 days, and then they began to outcompete 3D7^{pareNG}. By the end of the assay, Δ0614 was still outcompeting the fluorescent line while Δ2149 was level with the fluorescent line. To assess potential morphological phenotypes, the lines were visualised using light microscopy. There were no observable changes in morphology in the lncRNA-disrupted lines (Figure 5.4C).

5.4.6 Gametocyte conversion rate is increased in mutants with disrupted lncRNAs associated with gametocytes

Three of the six lncRNA-disrupted lines selected for phenotyping had targets with predicted associations with gametocytes: PF3D7lncRNA_0456 and PF3D7lncRNA_0364 with specificity to males and PF3D7lncRNA_1191 to late gametocytes. Therefore, I employed a gametocyte assay to determine if the disruption of these lncRNAs had an impact on gametocytogenesis. Sexual commitment was induced in triplicate cultures of lncRNA-disrupted lines using the gametocyte induction protocol (see *General Methodology* and Figure 5.1). Upon induction (Day 0), gametocytes are visible from Day 2 and progress through the five stages of gametocytogenesis (see stages in Figure 2.3). I quantified the rate at which asexual stages convert to sexual stages using the gametocyte conversion rate. The conversion rate is the quotient between the rings counted on Day 1 and the stage III gametocytes counted on Day 4. All three lines with predicted associations with gametocytogenesis had significant increases in gametocyte conversion (Figure 5.5A). Δ0456 had a mean conversion rate of 0.63 ($p<0.05$), Δ0364 0.38 ($p<0.01$), and Δ1191 0.61 ($p<0.05$). Another lncRNA-disrupted line, Δ1478, associated with schizonts, also had a significantly increased gametocyte conversion rate of 0.63.

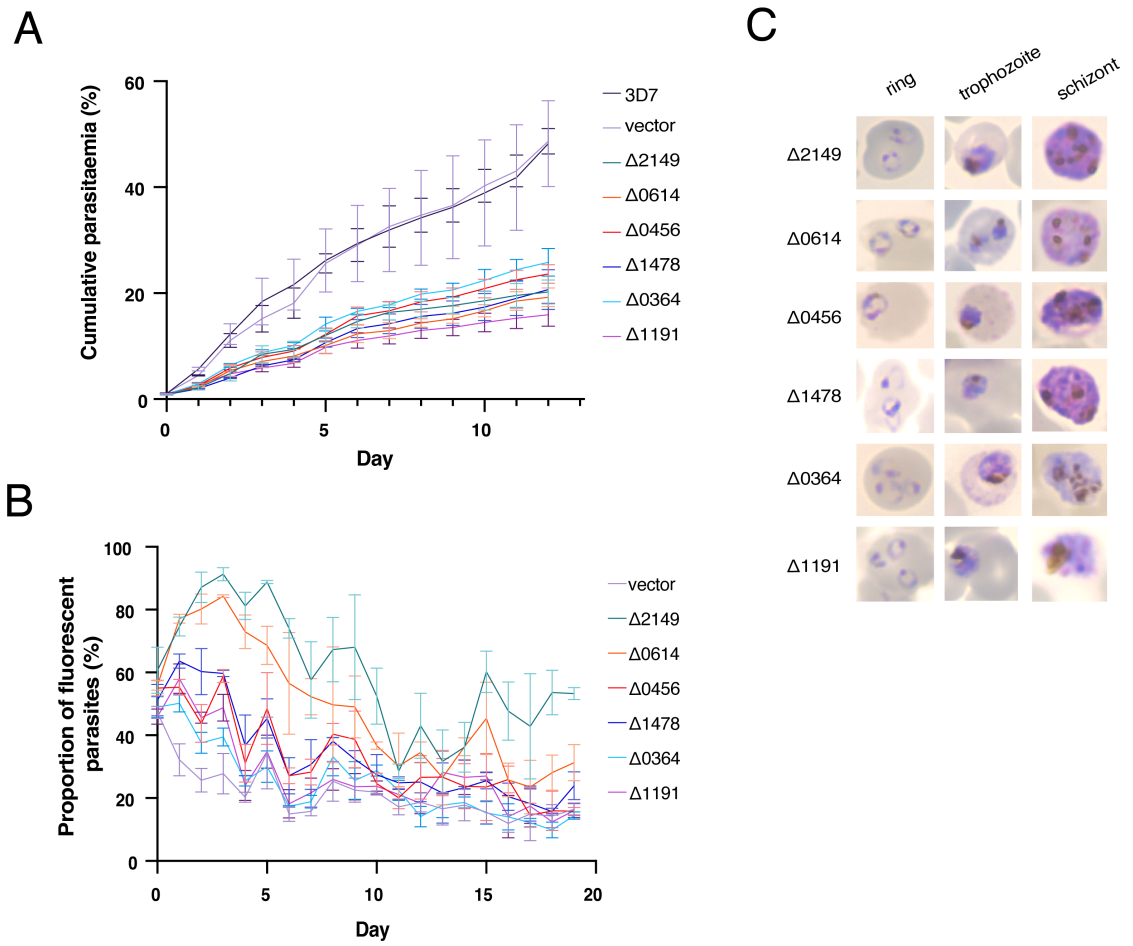


Fig. 5.4 Fitness is marginally diminished in all lncRNA-disrupted mutants. (**A**) Growth curves were completed in triplicate over 12 days. Cumulative parasitaemia was calculated by multiplying measured parasitaemia by culture cuts and adding to the previous day's parasitaemia. Mean cumulative parasitaemia is presented in this figure with error bars representing standard deviation. (**B**) Competition assays were completed against fluorescent 3D7^{pareNG}. The mean proportions of fluorescent parasites are presented in this figure, with error bars representing standard deviation. The downward slope indicates that the query line is outcompeting the fluorescent line. (**C**) Morphology was visualised by observing blood smears under light microscopy with a 100X oil immersion lens.

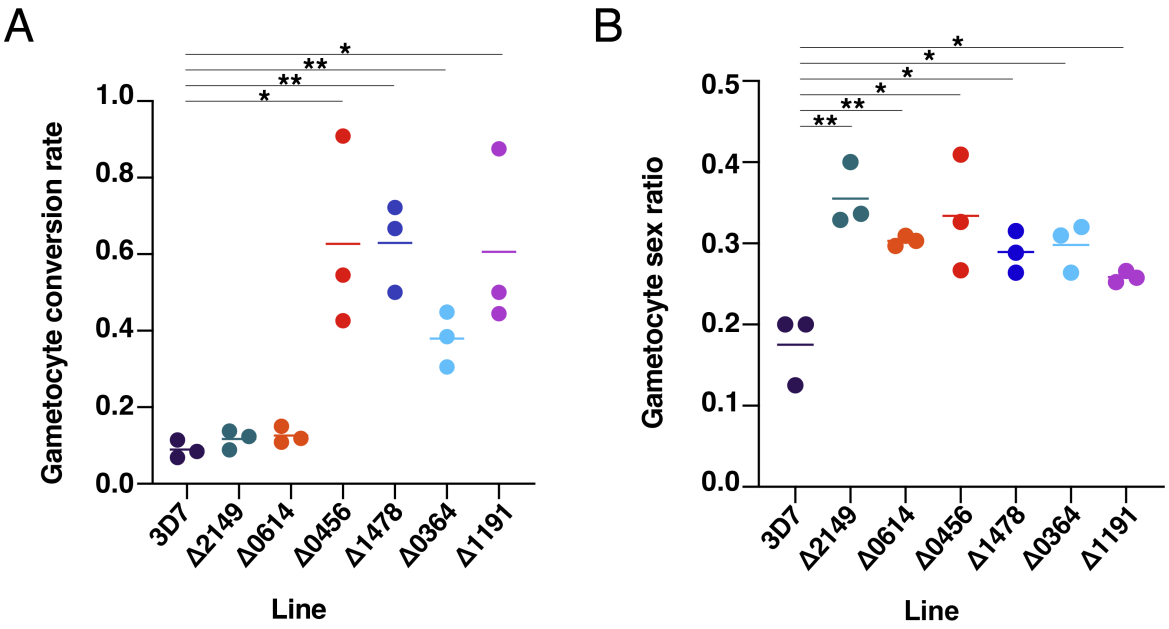


Fig. 5.5 Gametocyte conversion rate and the proportion of male gametocytes are increased in lncRNA-disrupted mutants. Gametocytogenesis was induced in triplicate cultures, and microscopy was used to enumerate the number of ring-stage parasites on Day 1, stage III gametocytes on Day 4 and stage 5 gametocytes (males and females) on Day 10 and 13. Conversion rate (Day 4 stage III gametocytes/Day 1 rings) (**A**) and sex ratio (number of males/the total of gametocytes) (**B**) were calculated. The coloured lines represent the mean between triplicates. Two-tailed t-tests were completed between every line and the control (3D7). Significant p-values are denoted with representative asterisks: *(0.01 to 0.05) and **(0.001 to 0.01).

5.4.7 The proportion of male gametocytes is increased in all lncRNA-disrupted mutants

As Δ0456 and Δ0364 were associated with male gametocytes in particular, I investigated the ratios of the male and female sexes. I maintained the previous gametocyte assay, counted male and female gametocytes on Day 10 and determined the ratio of males to all gametocytes. On Day 10, a proportion of stage 4 gametocytes were still observed in lncRNA-disrupted lines, likely due to their slower-growing phenotype. Therefore, I counted male and female gametocytes on Day 13 and averaged the ratios between the days. The Δ0456 and Δ0364 demonstrated significantly ($p<0.05$) increased gametocyte sex ratios of 0.33 and 0.30, respectively, compared to 3D7, suggesting an increased differentiation towards the male fate (Figure 5.5B). However, this phenotype is not unique as the other four lncRNA-disrupted lines demonstrated similar significant increases in gametocyte sex ratios.

5.4.8 Drug resistance is increased in lncRNAs-disrupted mutants for certain drugs

The final phenotype that I assessed was the susceptibility of parasites to antimalarial drugs because one mutant line, Δ2149, targeted a lncRNA associated with mutations in drug-resistant lines (generated against MMV008434 and MMV020746). Drug assays were completed for all six lncRNA-disrupted lines in triplicate against a panel of eight drugs that cover a broad spectrum of mechanisms, including atovaquone, chloroquine, dihydroartemisinin, mefloquine, KDU691 and KAE609 and the drugs of interest: MMV008434 and MMV020746. Dose-response curves were generated for every replicate, and representative curves are presented in Figure 5.6. The IC_{50} values for each biological replicate are presented in the panel below each curve. The Δ2149 line had a significant ($p<0.05$) increase in IC_{50} for MMV020746 but not for the other associated drug MMV008434. The Δ2149 line also demonstrated a significant ($p<0.01$) increase in IC_{50} for atovaquone. In terms of the other lncRNA-disrupted lines, all had significant increases in IC_{50} for atovaquone. Δ0364 and Δ1191 had significant ($p<0.05$) increases in IC_{50} for mefloquine, and Δ1478 and Δ1191 had significant ($p<0.05$ and $p<0.01$, respectively) increases in IC_{50} for MMV020746. No differences were seen in any of the lines for the remaining drugs.

5.4.9 Repression of lncRNA loci elicits changes in the transcriptome

To assess the impact of lncRNA repression on the transcriptome and further investigate potential roles in the parasite, I analysed gene expression profiles for four mutant lines (Δ2149, Δ1478, Δ0364, and Δ1191) in a timecourse encompassing all of the asexual stages of the IDC. The remaining two lines were omitted (Δ0614 and Δ0456) due to time constraints in sample prep. RNA was extracted and sequenced (Illumina short-read bulk RNA-Seq) from synchronised, triplicate cultures at 8, 16, 24, 32, 40, 48 and 56 hours post-ring-stage synchronisation. The RNAseq analysis was completed collectively on all the samples using a design of *condition + timepoint + condition:timepoint*, where the condition is the parasite line. 3D7 was included as the control condition. Significantly differentially expressed (DE) genes across the IDC were determined for each mutant line (Figure 5.7).

The fitness defect of the lncRNA mutant lines was evident in the PCA plot of the samples (Figure C.2 in Appendix C). Genes that were differentially expressed due to the shift in later timepoints were noted as "shifted" in Figure 5.7 and often involved in erythrocyte egress or invasion and cell division. All mutants exhibited differences in the expression of multigene

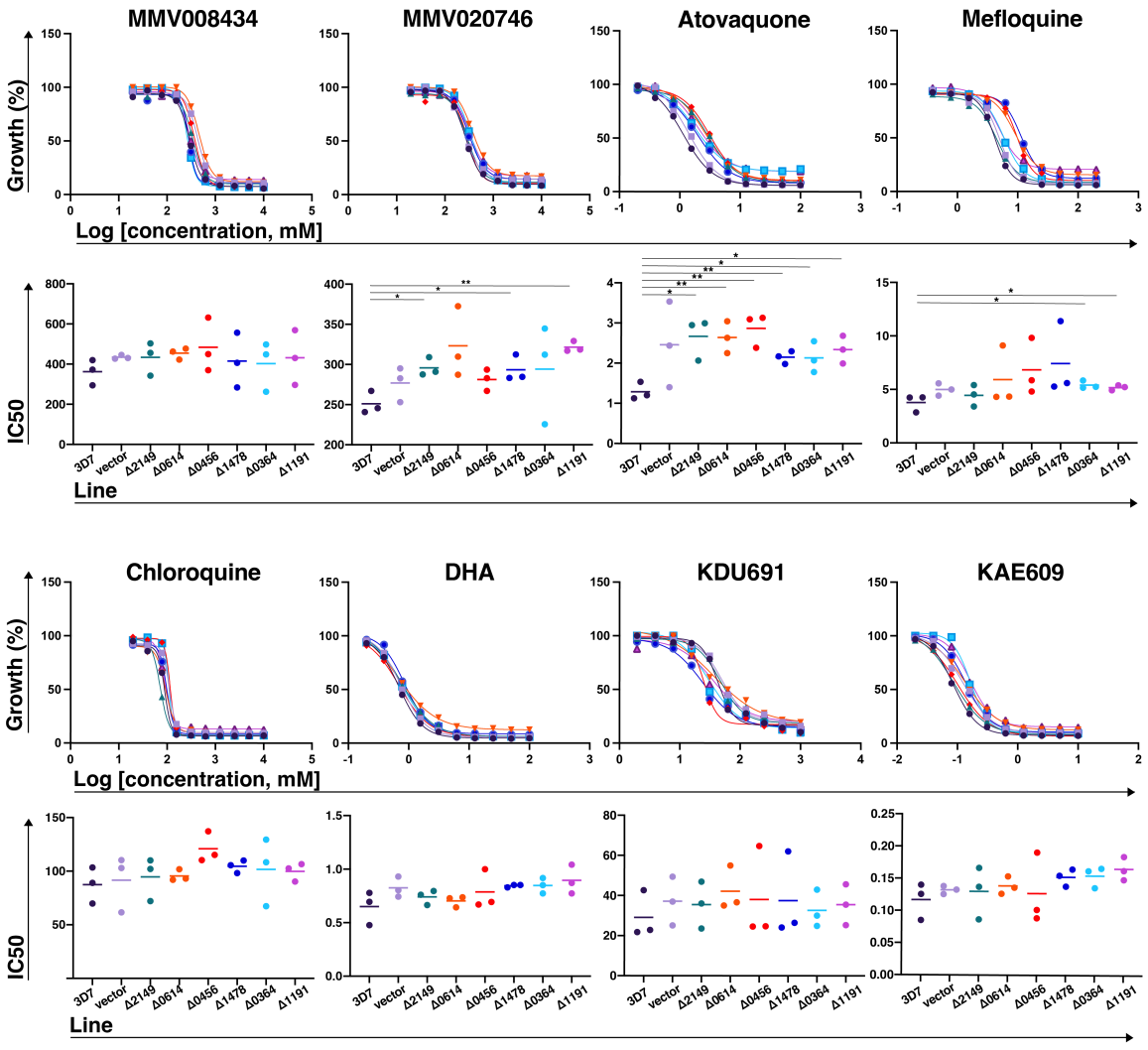


Fig. 5.6 Drug resistance is increased in lncRNA-disrupted lines for certain drugs. Drug assays were completed for three biological replicates against all eight drugs: atovaquone, chloroquine, dihydroartemisinin, mefloquine, KDU691, KAE609, MMV008434 and MMV020746. A non-linear regression model for log(inhibitor) and variable(slope) was used to generate a dose-response curve from technical replicates and calculate IC_{50} values. The upper panel shows a representative dose-response curve, and the lower panel presents all three IC_{50} values with a line representing the mean IC_{50}. Two-tailed t-tests were completed for every line versus the control (3D7). Significant p-values are denoted with representative asterisks: *(0.01 to 0.05) and **(0.001 to 0.01).

families such as surface antigens (such as *var*, *rif* and *stevor*), and Maurer's clefts two transmembrane proteins (*mc-2tm*). While the *MC-2TM* genes were upregulated, the surface antigen genes demonstrated nonspecific trends in expression. These genes are frequently variable in *P. falciparum* RNA-Seq experiments; therefore, I will focus on the remaining DE genes. A heatmap shows DE genes (FDR>0.01) (Figure 5.7), and an extended list (FDR>0.05) can be found in Appendix C, Table C.4). GO term enrichment was completed on differentially expressed genes to identify associated biological processes (Table C.5). Each lncRNA mutant also exhibited specific changes in the transcriptome. Although, all lncRNA mutants downregulated the expression of *ptef* (PF3D7_0202400), a translation elongation factor.

Δ02149, the mutant selected based on drug resistance, upregulated 78 genes and downregulated 8 genes. Among the upregulated genes, there was significant enrichment (p<0.01) of GO terms relating to cell motility and invasion, particularly relating to actin, cytoskeleton, and vesicular-based processes. These genes included *ama1*, *myoJ*, glideosome proteins (*phil1*, *pic2*, *pic3*, *pic4*, *pic5*, *gac* and *GAP45*), inner membrane complex proteins (*imc1g*, *imc1c*, *imc1m* and *imc1f*), microneme (PF3D7_0316000) and rhoptry (PF3D7_0414900) proteins. Also upregulated were merozoite surface proteins such as *msp3*, *msp5*, *msp6*, *msp9*, *rh4*, *rh5*, *eba140*, *eba165* and *msa180*. Differential expression was not observed in the antisense gene, PF3D7_1420100, or the gene which may share a bidirectional promoter, PF3D7_1420000 (Figure C.3).

Δ1478 upregulated 107 genes and downregulated 13 genes. Δ1478 upregulated nine gametocyte-specific genes: *mdv1*, *g27/25*, *p48/45*, *p230*, *pfg27*, *gep*, *pf11-1*, radial spoke protein (PF3D7_1403000) and flagellar outer arm dynein-associated protein (PF3D7_1020100). Furthermore, other genes associated with sexual stages were also upregulated, including plasmepsin VI, ookinete protein *p25* and two exported proteins, *epf1* and *gexp13*. Two putative RNA-binding proteins (PF3D7_0414500 and PF3D7_1438800) and the NOT2 subunit of the CCR4-NOT transcription complex (PF3D7_1128600) were also upregulated. Although not statistically significant after multiple testing corrections, there is a 2.3-fold increase in expression in *sub2*, the UTR of which is antisense to PF3D7lncRNA_1478 (Figure C.3).

Unlike the three other mutants, which predominantly upregulated gene expression, genes in Δ0364 were upregulated and downregulated in near-equal proportions (14 and 15 genes, respectively). Δ0364 upregulated genes encoding PHIST proteins (PHISTa,

PHISTb, and LyMP), although not the PHIST proteins associated with gametocytogenesis. Two RNA-binding proteins were downregulated, including a pre-mRNA splicing factor ATP-dependent RNA helicase (PF3D7_1030100) and PF3D7_1028500, a protein of unknown function, which contains a CSTF RNA-binding domain. No change was observed in the gene, PF3D7_0500900 or the gene that shares a bidirectional promoter (PF3D7_0500800) (Figure C.3).

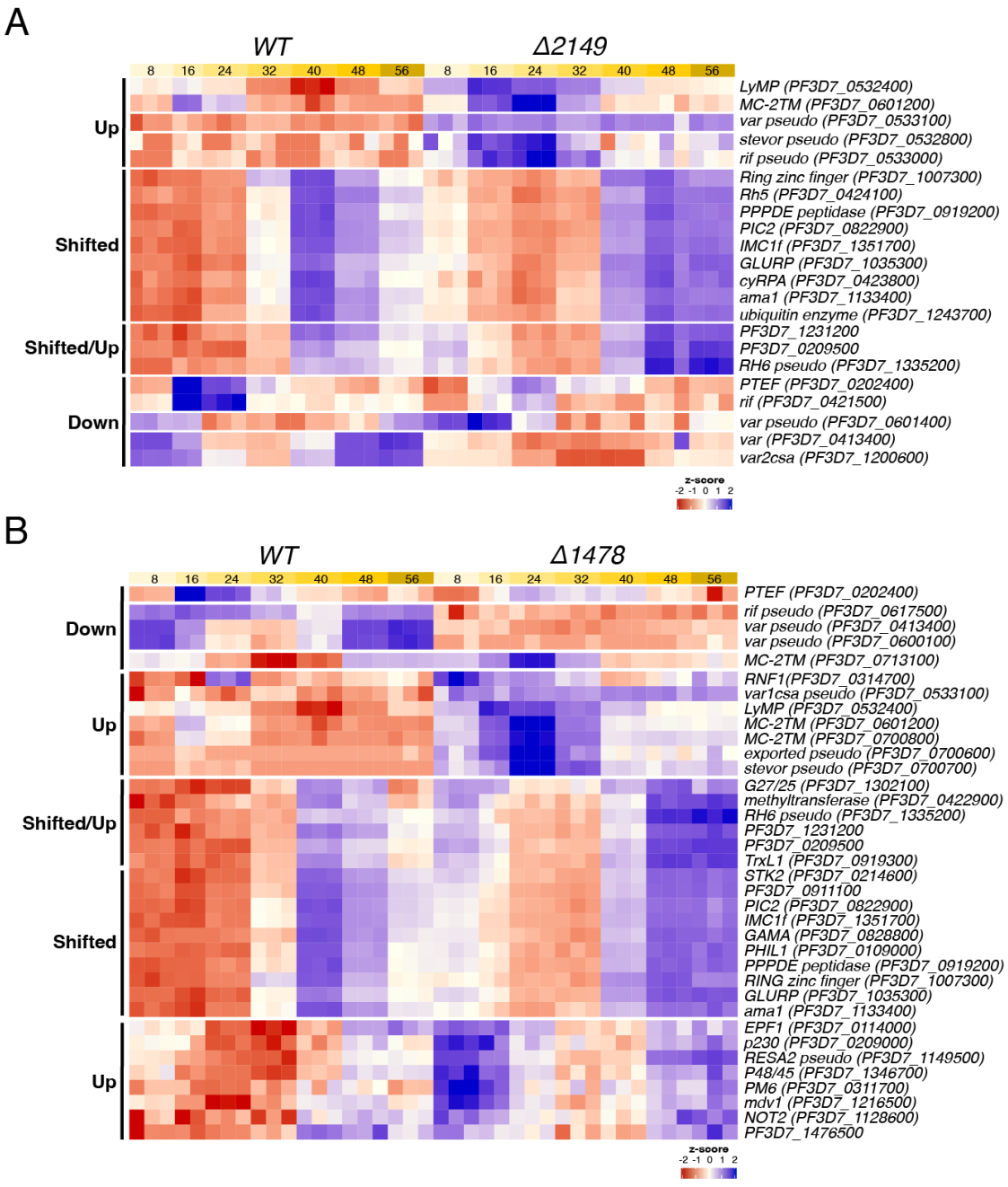
Δ1191 upregulated 32 genes and downregulated 9 genes. Like Δ1478, Δ1191 upregulated gametocyte-specific and associated genes, including *mdv1*, *g27/25*, *p48/45*, *pfg27*, radial spoke protein and plasmepsin VI. RNA-binding proteins were also upregulated, including histone RNA hairpin-binding protein (PF3D7_0413500) and both putative RNA-binding proteins upregulated in Δ1478: PF3D7_0414500 and PF3D7_1438800. Other genes that were upregulated were enzymes such as adenylate cyclase alpha (PF3D7_1404600), PPPDE peptidase (PF3D7_0919200), phosphoglucomutase-2 (PF3D7_0413500) and serine/threonine kinase (PF3D7_0214600). Differential expression was not observed in the gene, PF3D7_1008700, the UTR of which is antisense to this lncRNA (Figure C.3).

5.5 Discussion and future outlook

The thousands of lncRNAs in *P. falciparum* need to be experimentally validated and functionally characterised to understand their role in the parasite. Large-scale disruption screens are a practical approach to characterising thousands of targets, which have been successfully applied to lncRNA characterisation in other organisms. In this chapter, I endeavoured to provide proof-of-concept for the application disruption screens to the characterisation of *P. falciparum* lncRNAs. I applied CRISPR-dCpf1Sir2a interference to a small set of *P. falciparum* lncRNA targets enriched for potential biological significance and observable phenotypes. I then functionally characterised six lncRNA-disrupted mutants using phenotypic assays and expression profiling.

5.5.1 A proof-of-concept for utilising CRISPR interference in *P. falciparum* lncRNA screens

I successfully generated lncRNA-disrupted lines with CRISPR-dCpf1Sir2a and provided an approach that serves as a model for future experiments. One limitation of the approach is


A
WT
Δ2149
8 16 24 32 40 48 56 8 16 24 32 40 48 56
Up
Shifted
Shifted/Up
Down
LyMP (PF3D7_0532400)
MC-2TM (PF3D7_0601200)
var pseudo (PF3D7_0533100)
stevor pseudo (PF3D7_0532800)
rif pseudo (PF3D7_0533000)
Ring zinc finger (PF3D7_1007300)
Rh5 (PF3D7_0424100)
PPPDE peptidase (PF3D7_0919200)
PIC2 (PF3D7_0822900)
IMC1f (PF3D7_1351700)
GLURP (PF3D7_1035300)
cyRPA (PF3D7_0423800)
ama1 (PF3D7_1133400)
ubiquitin enzyme (PF3D7_1243700)
PF3D7_1231200
PF3D7_0209500
RH6 pseudo (PF3D7_1335200)
PTEF (PF3D7_0202400)
rif (PF3D7_0421500)
var pseudo (PF3D7_0601400)
var (PF3D7_0413400)
var2csa (PF3D7_1200600)
z-score
-2 -1 0 1 2
B
WT
Δ1478
8 16 24 32 40 48 56 8 16 24 32 40 48 56
Down
Up
Shifted/Up
Shifted
Up
PTEF (PF3D7_0202400)
rif pseudo (PF3D7_0617500)
var pseudo (PF3D7_0413400)
var pseudo (PF3D7_0600100)
MC-2TM (PF3D7_0713100)
RNF1(PF3D7_0314700)
var1csa pseudo (PF3D7_0533100)
LyMP (PF3D7_0532400)
MC-2TM (PF3D7_0601200)
MC-2TM (PF3D7_0700800)
exported pseudo (PF3D7_0700600)
stevor pseudo (PF3D7_0700700)
G27/25 (PF3D7_1302100)
methyltransferase (PF3D7_0422900)
RH6 pseudo (PF3D7_1335200)
PF3D7_1231200
PF3D7_0209500
TrxL1 (PF3D7_0919300)
STK2 (PF3D7_0214600)
PF3D7_0911100
PIC2 (PF3D7_0822900)
IMC1f (PF3D7_1351700)
GAMA (PF3D7_0828800)
PHIL1 (PF3D7_0109000)
PPPDE peptidase (PF3D7_0919200)
RING zinc finger (PF3D7_1007300)
GLURP (PF3D7_1035300)
ama1 (PF3D7_1133400)
EPF1 (PF3D7_0114000)
p230 (PF3D7_0209000)
RESA2 pseudo (PF3D7_1149500)
P48/45 (PF3D7_1346700)
PM6 (PF3D7_0311700)
mdv1 (PF3D7_1216500)
NOT2 (PF3D7_1128600)
PF3D7_1476500
z-score
-2 -1 0 1 2

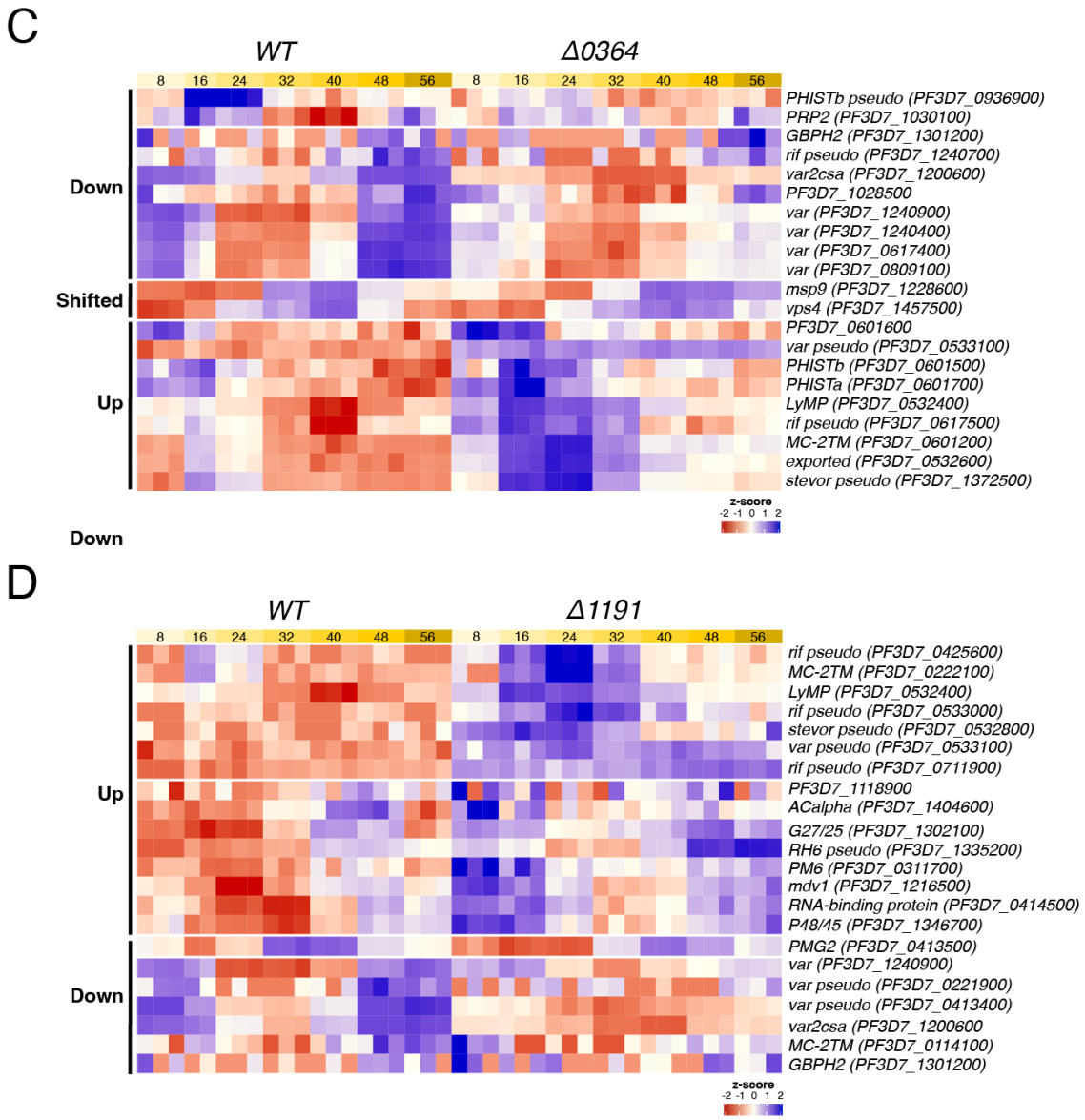


Fig. 5.7 Genes differentially expressed between lncRNA mutant lines across the intraerythrocytic development cycle. Differentially expressed genes are displayed in both WT (3D7) and lncRNA mutant lines: (**A**) Δ2149, (**B**) Δ1478, (**C**) Δ0364, and (**D**) Δ1191. Cells are ordered by timepoint (8, 16, 24, 32, 40, 48 and 56 hours) and show biological triplicates. All genes shown are significantly differentially expressed with an FDR<0.01 (an extended list of all DE genes FDR < 0.05 is in Appendix C) and a log2FC superior of 0.50. Z-scores represent gene expression levels, and labels have been included (left-side) denoting upregulated and downregulated genes and shifted (delayed) gene expression. pseudo: pseudogene.

that not all lncRNAs were targetable, even with the advantages of using dCpf1Sir2a, which uses an AT-rich PAM site. 20% of the initial 55 targets were eliminated because sufficiently specific gRNAs could not be designed. One reason for this is that some of these lncRNAs resided near the UTRs of gene families or rRNAs. These lncRNAs may not targetable, at least not specifically. I also generated two recrudesced lines that were potentially lncRNA-disrupted mutants, but I could not generate effective qPCR primers for the targeted lncRNAs. Therefore, in future, I would design qPCR primers during the target selection process to avoid encountering this issue after the lines are generated. To salvage these lines, RNA sequencing could be used to assess if lncRNA expression is repressed.

The generation of gRNAs, without the need for donors, streamlined the design and molecular cloning required to generate disruption plasmids. One way that the methodology could further simplify the approach and support scalability is with multiplexing. Cpf1 gRNAs can be multiplexed into a CRISPR array (Zetsche et al., 2017), which would reduce molecular cloning and plasmid propagation considerably in experiments where multiple gRNAs are required for each target. Moreover, multiplexed disruption plasmids would ensure that every cell uptakes all the gRNAs.

Recrudesced parasites were obtained from CRISPR-dCpf1Sir2a transfections, with 32% of targets recovered in a single transfection attempt. Considering the nature of transfection in *P. falciparum* in terms of transfection efficiency, this is reasonably robust. The set of transfections should be retried to determine how many additional lncRNA-disrupted mutants can be obtained. Furthermore, transfections with different gRNAs combinations or single gRNAs could circumvent any toxicity that a single gRNA causes. Due to the partial knockdown that the gene interference approach provides, it is less likely, compared to knockout approaches, that the disruption of essential lncRNAs would lead to parasite death. However, with certain levels of reaching a 90% reduction in expression, it is possible that parasites cannot survive with that level of diminished expression of essential lncRNAs. Modulating the level of CRISPR-dCpf1Sir2a interference would be an ideal solution, such as integrating a conditional regulatory element, although these tools have not yet been engineered. Of the 30 targeted lncRNAs that did not recrudesce, five were predicted putatively essential (in *Chapter 3*). Contrastingly, two lncRNAs predicted as putatively essential, PF3D7lncRNA_0364 and PF3D7lncRNA_2353, were able to be disrupted.

The expression of lncRNA targets was disrupted by CRISPR-dCpf1Sir2a, although the level of reduction varied considerably, and the expression levels were only determined

for bulk mixed-stage cultures. The level of reduction in target gene expression ranged from 90% to none at all, including one outlier (Δ0055) with a 2-fold increase in lncRNA expression. Excluding this outlier, the average level of lncRNA repression was 45%. This level of repression is comparable to the 40-60% repression observed in the GFP repression experiments previously undertaken by our lab. This repression level is also similar to the inducible CRISPR-dCas9Sir2a system developed by Liang et al. in *P. falciparum*, which for most genes ranged from 1.5-3-fold (Liang et al., 2022). For some targets, there was variability in the level of interference by CRISPR-dCpf1Sir2a. Ideally, the level of interference should be consistent to minimise variation in subsequent phenotypic assays. As gRNAs are expressed on different plasmids, which are pooled, the variability may be caused by differences in plasmid uptake. As it has not yet be published for its application in *P. falciparum*, it is important that the efficacy of the CRISPR-dCpf1Sir2a system is confirmed and further quantified in stage-specific, clonal cultures.

In the Liang et al. study, CRISPR interference varied significantly depending on gRNA design (Liang et al., 2022). As discussed above, multiplexing could be used to control this and ensure all parasites receive all target gRNAs. Moreover, further work may be required to optimise gRNA design, such as to enable the knockdown of PF3D7lncRNA_0055. gRNA optimisation studies could determine the optimal number of gRNAs for efficient lncRNA disruption in *P. falciparum* and elucidate the impact of targeting the opposing versus the same strand as lncRNA transcription and upstream or downstream of the transcriptional start site.

Furthermore, the specificity of the CRISPR-dCpf1Sir2a system requires additional investigation as I was unable to obtain the controls necessary to evaluate the impact of the enzyme on the parasite. I attempted to obtain controls that could be used to determine any toxicity caused by the CRISPR-dCpf1Sir2a vector, such as off-target effects or impact on parasite fitness or other phenotypes. I could not obtain a control line with a non-targeting gRNA, and the gene-targeting control that I did generate did not express dCpf1Sir2a. These controls are imperative for understanding the specificity of this gene interference approach and the general applicability of CRISPR-dCpf1Sir2a in *P. falciparum*. As sir2a is a general histone deacetylase, there is the possibility that if the fusion enzyme is diffused in the nucleus, other sites in the genome could become deacetylated non-specifically. Validation of the specific action of the dCpf1Sir2a enzyme on its target binding sites should be investigated as has been necessary for the development of similar tools such as dCas9Sir2a (Xiao et al., 2019). For example, a whole-genome analysis of dCpf1Sir2a binding specificity could be completed

by using ChIP-Seq on dCpf1 via its HA tag, to assess any off-target binding sites. Also, the generation of a conditional dCpf1Sir2a system would allow the effects of repression to be triggered. A conditional line expressing dCpf1Sir2a with a non-targeting gRNA could be used to examine any effects of dCpf1 expression on growth and other phenotypes.

To further investigate the variability in interference, the relative expression of a target gene should be evaluated over short and long time frames. If the variation occurs across the stages of the IDC, using synchronised cultures over mixed cultures may be better. I selected lines for phenotyping based on the level of interference observed in initial recrudesced parasites. However, as substantial variability in repression as observed between replicates in later experiments, the other six lines should be revisited as cultures may exhibit increased target interference that would facilitate phenotyping. Furthermore, cloning could enable the isolation of clones that have increased levels of repression, such as those that have taken up all four gRNAs.

To enrich my dataset for lncRNAs with a higher likelihood of having observable phenotypes, I selected lncRNAs with features that had the potential to be linked with biological significance. This approach to target prioritisation was successful, as differences in predicted phenotypes were observed in lncRNAs-disrupted lines. When designing experiments with a more extensive set of targets, careful consideration will be needed to determine appropriate phenotypic assays that will capture the phenotypes of lncRNA-disrupted lines.

5.5.2 Characterising lncRNAs in *P. falciparum*

From the proof-of-concept screen, six lncRNA-disrupted mutants were further characterised to identify the potential functionality and biological significance of their target lncRNAs. Mutants exhibited phenotypes that related to their selection features and predicted roles. For example, PF3D7lncRNA_2149 contained a 7 bp deletion (TATATTC →T) in lab-generated parasite lines with drug resistance to two drugs with similar structures, MMV020746 and MMV008434. The Δ2149 mutant line, with 50-60% knockdown of PF3D7lncRNA_2149 expression, demonstrated a significant change in drug sensitivity for one of the aforementioned drugs (MMV020746), along with a growth defect greater than the other lncRNA-disrupted lines. Three mutants targeting lncRNAs linked with gametocyte-stage specificity and a schizont-associated mutant (Δ1478) had significantly increased gametocyte conversion rates and gametocyte sex ratios. These phenotypes were further supported by expression profiling, which showed an upregulation of gametocyte-specific genes in Δ1478 and Δ1191. However,

it is not clear whether this reflects an increased number of gametocytes in the culture due to the increased conversion rate or if the lncRNAs are upregulating gametocyte-specific genes directly. Overall, the observed phenotypes contribute evidence to the association of these lncRNAs with their potential roles; however, these phenotypes need to be confirmed, in particular given the variability in experimental gametocyte induction, and the functionalities of the lncRNAs need to be elucidated.

Moreover, some phenotypes do not reflect the predicted roles for the lncRNAs. For example, PF3D7lncRNA_0456, which is associated with male gametocytes, may act as a transcriptional switch as it shares a promoter with a gene associated with an opposing phenotype, female gametocytogenesis. However, when this lncRNA was disrupted, an increased ratio of male to female gametocytes was observed. In contrast, if it acted as a switch, then when disrupted, an increase in female gametocytes would have been expected. Due to potential inaccuracies in counting stage V male and female gametocytes, exflagellation assays should also be completed in future for lncRNA-disrupted lines.

LncRNA disruption, even with a diverse set of phenotypic assays, is insufficient for elucidating lncRNA function. Other strategies should be employed to study these lncRNAs, such as alternative CRISPR-based disruption methods or overexpression. PF3D7lncRNA_0614 is an *intergenic* lncRNA, so CRISPR-based whole-lncRNA knockout could also be used to disrupt the lncRNA. Targets in more complex contexts, such as PF3D7lncRNA_2149, an antisense-to-gene lncRNA, would require careful altering of the transcriptional start site to obtain a knockout, as removal of the entire lncRNA would disrupt the antisense gene. Also, the deletion found in PF3D7lncRNA_2149 in the drug-resistance studies could be recapitulated using CRISPR gene editing and further provide evidence of the impact of this mutation on drug resistance. Episomal lncRNA overexpression has been used previously to study lncRNA function in *P. falciparum* and can provide an opportunity to observe an opposite phenotype (Amit-Avraham et al., 2015). Another strategy that is helpful for elucidating lncRNA function is determining the interacting partners of the lncRNA. ChIRP-Seq can identify DNA binding sites, RNA antisense purification can identify binding RNAs, and RNA immunoprecipitation pulldown protein binding partners. For example, PF3D7lncRNA_2149 may be involved in drug resistance to MMV020746 and the mechanism of action of this drug is not known. The characterisation of the interactions of PF3D7lncRNA_2149 and its role in the parasite may aid in elucidating the mechanism of action of MMV020746.

Additionally, there were potentially smaller numbers of significantly differentially expressed genes than might be expected in mutants with observable altered phenotypes (ranging from 29 to 120 genes). This may be because the RNA-Seq analysis was collective across all time points. The RNA-Seq dataset should be further analysed by completing individual timepoint-comparisons to capture any timepoint-specific differentially expressed genes that were not deemed significant by the model generated across all timepoints.

In this study, I was able to capture lncRNA-associated phenotypes primarily because I enriched the target set for lncRNAs containing features potentially related to biological significance, which also had measurable phenotypes. For future large-scale experiments to be effective and meaningful, further development of the approach to phenotyping is required. As phenotypes may not be observable using traditional *P. falciparum* phenotypic assays, other approaches to phenotyping should be used in parallel, such as capturing lncRNA interactions such as ChIRP-Seq, LIGR-Seq and ChIRP-MS or localisation with RNA-FISH. The vector or transfected line could be modified to facilitate these methods. For example, unique gRNA barcodes have been incorporated into dCas9 experiments to enable pooling en masse and quantitative CRISPR screening (McGlincy et al., 2021) although other studies have been able to achieve the same end without barcoding by amplifying gRNA sequences directly (Wong et al., 2022).

5.6 Additional methodology

See relevant *General Methodology* sections for molecular cloning, parasite handling, transfection, and phenotyping.

5.6.1 Target selection

Targets were prioritised based on one of the following features: stage-specificity, genomic context (*intergenic* lncRNAs), length, or association with drug resistance. The stage-specificity targets were selected by Dr Juliana Cudini using her single-cell sequencing of blood-staged parasites, which identified lncRNAs with a strong association with specific stages. The genomic context and length targets were identified from the supplementary data of *Chapter 3*. To identify lncRNAs associated with drug resistance, Bedtools (v2.30.0) intersect (option -wa) was used to intersect the lncRNA annotation with a list of SNPs and InDels identified in

whole-genome sequencing of lab-generated drug-resistant strains (provided by the MalDA consortium) (Quinlan and Hall, 2010).

5.6.2 Guide RNA and primer sequences

The gRNAs and qPCR primers used in this chapter are listed in Tables C.2 and C.3 in Appendix C. The cyclophilin qPCR primers were designed by Dr Sophie Adjalley, and the 18S rRNA primers were obtained from (Usui et al., 2019). The p36 primer (GTTGTGTG-GAATTGTGAGCGG) was used to amplify the U6 cassette and check for correct gRNA insertion.

5.6.3 Plasmid map

The map of the plasmid used in this chapter is in Appendix C.

5.6.4 RNA timecourse preparation and sequencing

30mL cultures at 3-6% parasitaemia were synchronised to ring stages. Cultures were then expanded and aliquoted at 0.75% into 21 flasks (3 x 7 timepoints). The 8hr timepoint flasks received 10mL of culture, while all other flasks received 5mL of culture. Flasks were gassed and incubated at 37°C. Flasks were removed from the incubator, and RNA was extracted at 8, 16, 24, 32, 40, 48 and 56 hours post-synchronisation. Flow cytometry enumeration and blood smears were completed for each sample. RNA was purified using both phenol-chloroform and an RNA purification kit and treated with DNase. Isolated RNA was sent to Sanger DNA Pipelines for library prep and sequencing. Stranded libraries with oligo-dT pulldown were prepared. Samples were multiplexed (groups of 42), and sequencing was completed on NovaSeq with 100 bp paired-end reads. Expression profiling was completed as described in *General Methodology*. Gene ontology (GO) term enrichment was completed on PlasmoDB using biological process ontology for genes associated ($p<0.01$).

Chapter 6

Conclusions and Outlook

How can a disease that has afflicted humans for millennia, has been studied by scientists for over a century, and receives global investment and attention still evade eradication? Malaria is an enigma in many ways, as the conventional toolkit for eradicating infectious diseases has not been sufficient. Interventions have effectively controlled the parasite, significantly reducing mortality and transmission. However, these interventions are threatened by the emergence and spread of insecticide and antimalarial resistance, the limited efficacy and slow development of vaccines, and the lack of access to health services in endemic areas. Recent years have revealed unpredictable threats such as climate change, conflict and pandemics. New and sustainable interventions are needed; their development will rely on our understanding of parasite biology.

The underlying mechanisms of gene regulation in *Plasmodium* are poorly understood. While specific regulatory molecules such as transcription factors and epigenetic modulators are relatively well-studied, others are not. In this thesis, I highlight lncRNAs as potential contributors to the regulation of the *P. falciparum* transcriptome. While many studies have predicted an abundance of parasite-derived lncRNAs, only a few have been functionally characterised, partially due to challenges with existing methodology. Through my chapters, I addressed this shortcoming and developed the tools necessary to enable the *in vitro* study of lncRNAs in *Plasmodium falciparum*. To demonstrate the applicability of these tools, I disrupted and characterised a modest set of lncRNAs.

6.1 Tool development and implementation

In *Chapter 3*, I generated an annotation to support experimental and bioinformatics approaches to studying *P. falciparum* lncRNAs. Several aspects of the annotation expanded and improved upon existing annotations. Firstly, by revisiting annotation, I could incorporate the updated UTR annotations and verify previous annotations and reported lncRNAs using independent sequencing. Secondly, I improved the annotation quality and identified novel lncRNAs using long-read sequencing and manual curation. This annotation is a resource for researchers in the field and is now publicly available to browse or download on PlasmoDB (Amos et al., 2021).

Additionally, I added two aspects that make the annotation more translatable to *in vitro* experiments. LncRNAs were given an evidence-based ranking score generated from collated supportive information. Rankings allow the annotation to be subsetted based on confidence. I also employed a descriptive classification system that classifies lncRNAs by genomic context using eight categories. Compared to simple intergenic and antisense classifiers, these categories can inform applications, such as when selecting an appropriate approach for lncRNA disruption. For example, in *Chapter 5*, I used the annotation to obtain lncRNA sequences, and I used evidence rankings to prioritise targets and feature lists to select interest groups. When considering contingencies for disruption approaches, I used the categorisation system to create the intergenic interest group, which would be targetable by Cas9 knockout approaches.

Although the annotation is an improvement, it is not definitive. Firstly, the annotation was generated with sequencing from asexual blood stages, not the entire parasite life cycle. Secondly, it is based on only two long-read RNA-Seq datasets. As long-read RNA-Seq datasets are generated, they should be used or repurposed to curate the lncRNA annotation. However, it is common for studies to report ncRNA reads as predicted lncRNAs without generating or contributing to the annotation. For example, the recently published linked PacBio and ONT RNA-Seq *P. falciparum* dataset from Shaw et al. would be valuable for lncRNA annotation (Shaw et al., 2022). Manual curation offers a higher-quality approach to annotation, although it is laborious and time-consuming. Future work should examine the applicability of existing computational tools that were developed for lncRNA annotation and analysis in other organisms to *P. falciparum*.

In *Chapter 4*, I explored potential CRISPR-based tools for lncRNA perturbation. I endeavoured to apply Cas13 to *P. falciparum*, a tool that could facilitate RNA knockdown into the parasite. I was unsuccessful in obtaining parasites that expressed Cas13; however, *P. falciparum* is known to have low genetic tractability, and several studies have reported toxicity when utilising Cas13. This toxicity is caused by non-specific RNAse activity causing collateral damage. Considering the immense impact that Cas13 could provide, its application to the parasite should be further investigated. Investigating the application of other Cas13 variants in *P. falciparum*, may be required to identify an enzyme that is tolerated.

The Cas9 and dCpf1 enzymes presented alternative approaches that harnessed existing tools. CRISPR-Cas9 was used to generate a lncRNA knockout, but the system greatly restricted the selection of targets. In *Chapter 5*, CRISPR-dCpf1 was used to facilitate gene interference of lncRNAs, which had a lesser impact on target selection. The approach successfully repressed the expression of lncRNAs in all, but one recrudesced lncRNA-disrupted line. The level of interference varied considerably between biological replicates when using CRISPR-dCpf1, suggesting that this tool should be optimised to achieve more consistent levels of interference and that cloning may be required to isolate clones with high levels of interference.

To support the fitness phenotyping of lncRNA-disrupted parasites, I generated a strain-matched fluorescent line in *Chapter 4*, which I used to quantify the parasite fitness in *Chapter 5*. I developed a vector that rapidly generates parasites expressing mNeonGreen from the endogenous *pfpare* locus. I demonstrated that the line was comparable to existing fluorescent lines. A limitation of the generated lines is that they have a minor fitness defect, which means they can be outcompeted by the wild-type strain, and lines with little to no fitness defects. However, when comparing different lines in parallel in competition assays, the generated reporter lines are effective in quantifying parasite fitness.

6.2 Biological insights

Through the process of developing and implementing tools, I gained insights into the nature of *P. falciparum* lncRNAs, which confirmed and expanded on previous studies. LncRNAs are pervasive, with thousands existing throughout the genome in diverse genomic contexts. Their sequences vary in size and GC content but are usually substantially more AT-rich than coding sequences. LncRNAs rarely contain exons or known RNA motifs. They are predicted to form secondary structures but are not likely to encode peptides. LncRNAs can exhibit highly

stage-specific expression in *P. falciparum* blood stages. Few functional mechanisms have been elucidated for *P. falciparum* lncRNAs, however, the elucidated mechanisms include negatively regulating antisense loci through their transcription, interacting with histones and forming DNA-RNA hybrids at specific loci (Sierra-Miranda et al., 2012; Filarsky et al., 2018; Heinberg et al., 2022). LncRNAs have been associated with roles in sexual differentiation, *var* gene switching and telomere maintenance. Amongst the several lncRNAs that I characterised, I was able to suggest potential roles in drug resistance and gametocytogenesis, although these lncRNAs and their mechanisms require further investigation.

A limitation to these insights is that they are restricted to only some parasite stages, as annotations were generated from RNA sequencing of asexual blood-stage parasites. To obtain a complete understanding of the role of lncRNAs in the parasite, lncRNAs must be investigated across the whole parasite life cycle. There may be gametocyte-, liver-, and mosquito-stage specific lncRNAs with essential biological functions, which are not captured in asexual blood-stage RNA sequencing.

6.3 Outstanding questions

To what extent are lncRNAs regulating the *P. falciparum* transcriptome?

In malaria research, the theories and understanding around transcriptional regulation have differed widely in the past two decades. From a "just-in-time model" supported by stage-specific expression to a post-transcriptional regulation model supported by the dearth of canonical transcription factors to the current understanding that transcriptional regulation occurs in the parasite. It remains unclear, however, what contribution lncRNAs have in transcriptional regulation. Beyond the well-studied examples, do the other thousands of lncRNAs function as transcriptional regulators?

Since the early discoveries of lncRNAs in *Plasmodium*, their role in the parasite has been questioned. Due to the pervasive nature of transcription in the parasite and the high occurrence of bidirectional promoters, scientists argue that they could be a byproduct of transcription. LncRNAs with established functionality (*gdv1*-aslncRNA, *md1*-lncRNA and *var*-aslncRNAs) and the lncRNA-disrupted lines with measurable phenotypes in this thesis suggest otherwise. However, the presence of functional lncRNAs does not preclude the existence of lncRNAs that are transcriptional noise, as the population could be functionally heterogeneous.

How do the features of *P. falciparum* lncRNAs contribute to their function?

It was challenging to draw conclusions from features identified in the characterisation of annotated lncRNAs. The relationship between features such as sequence, genome context, structure and expression with lncRNA function and the conservation of that function remains unclear in *P. falciparum*. Additional features not explored in this study, such as localisation and interactions, also have undetermined associations with function. Relationships between lncRNA features and function have been studied to some degree in other organisms, and these associations are often co-opted to understand lncRNA features in *P. falciparum*. However, these relationships are not universally applicable to all organisms and must be investigated specifically in *P. falciparum*.

The study of these associations in *P. falciparum* will rely on further functional characterisation studies. An analysis that could be undertaken more readily is to explore the level of conservation of lncRNAs and their features between species and strains. If lncRNA loci are well-conserved and maintained in the genome compared to truly intergenic sequences that are not transcribed or act as genomic features, that would provide further evidence of biological significance. Exploring the conservation of lncRNAs in field isolates would also be valuable. Deletions at the *gdv1* locus (including in lncRNAs) were captured in the sequencing of West African *P. falciparum* isolates, demonstrating that local selection and differentiation occur at this locus (Duffy et al., 2018). Conceivably, variation and selection could be occurring at other lncRNA loci and may aid in identifying lncRNAs of biological significance.

What functional mechanisms do lncRNAs exhibit in *P. falciparum*?

LncRNAs are biologically complex because they can interact with many different types of molecules and subsequently regulate gene expression through various mechanisms. The high level of diversity in lncRNA mechanisms identified in mammalian systems may or may not be observed in *P. falciparum* lncRNAs. This will become more apparent as more specific lncRNAs are functionally characterised in the parasite. This thesis supplies the research field with an annotation and perturbation approach that will support these studies; however, complementary tools will be required to facilitate large-scale screens and elucidate the precise mechanisms of specific lncRNAs. Multidisciplinary approaches that integrate perturbation screening with expression profiling and other approaches such as interactions profiling (ChIRP-Seq, LIGR-Seq and ChIRP-MS), localisation visualisation (RNA-FISH) and ribosomal profiling will enable meaningful high-throughput functional characterisation.

The systematic functional characterisation of *P. falciparum* lncRNAs is necessary, but it will require extensive resources and time, like the characterisation of genes, which is still ongoing as 35% of genes lack putative function (Oberstaller et al., 2021).

To what extent do *P. falciparum* lncRNAs contribute to pathogenesis and disease?

Beyond functional characterisation, it will be crucial to understand how *Plasmodium* lncRNAs contribute to pathogenesis in malaria. This understanding could inspire and enable the development of novel therapeutic and vaccine candidates that target lncRNAs or their pathways. Furthermore, lncRNAs can be targeted in several ways, including with siRNAs (in organisms with RNAi), CRISPR-Cas9, antisense oligonucleotides and small molecule compounds. LncRNA-derived interventions are being developed for several cancers and have begun clinical trials. For example, Andes-1537, an antisense oligonucleotide against antisense non-coding mitochondrial RNA, is currently in a clinical trial for multiple solid tumours (Chen et al., 2021). With current interventions under threat, targeting regulatory molecules like lncRNAs may provide innovative approaches to combating the ongoing malaria epidemic.

References

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., Verdine, V., Cox, D. B. T., Kellner, M. J., Regev, A., Lander, E. S., Voytas, D. F., Ting, A. Y., and Zhang, F. (2017). Rna targeting with crispr–cas13. *Nature*, 550(7675):280–284.

Addo-Gyan, D., Matsushita, H., Sora, E., Nishi, T., Yuda, M., Shinzawa, N., and Iwanaga, S. (2022). Chromosome splitting of plasmodium berghei using the crispr/cas9 system. *PLoS One*, 17(2):e0260176.

Adjalley, S. H., Chabbert, C. D., Klaus, B., Pelechano, V., and Steinmetz, L. M. (2016). Landscape and dynamics of transcription initiation in the malaria parasite plasmodium falciparum. *Cell Rep*, 14(10):2463–75.

Adjalley, S. H., Johnston, G. L., Li, T., Eastman, R. T., Ekland, E. H., Eappen, A. G., Richman, A., Sim, B. K., Lee, M. C., Hoffman, S. L., and Fidock, D. A. (2011). Quantitative assessment of plasmodium falciparum sexual development reveals potent transmission-blocking activity by methylene blue. *Proc Natl Acad Sci U S A*, 108(47):E1214–23.

Adjalley, S. H. and Lee, M. C. (2022). *CRISPR/Cas9 editing of the Plasmodium falciparum genome*, page In press. Springer.

Al-Khafif, G. D., El-Banna, R., Khattab, N., Gad Rashed, T., and Dahesh, S. (2018). The immunodetection of non-falciparum malaria in ancient egyptian bones (giza necropolis). *Biomed Res Int*, 2018:9058108.

Alvarez, D. R., Ospina, A., Barwell, T., Zheng, B., Dey, A., Li, C., Basu, S., Shi, X., Kadri, S., and Chakrabarti, K. (2021). The rna structurome in the asexual blood stages of malaria pathogen plasmodium falciparum. *RNA Biol*, 18(12):2480–2497.

Amandio, A. R., Necsulea, A., Joye, E., Mascrez, B., and Duboule, D. (2016). Hotair is dispensible for mouse development. *PLoS Genetics*, 12(12):e1006232.

Amaral, M. S., Maciel, L. F., Silveira, G. O., Olberg, G. G. O., Leite, J. V. P., Imamura, L. K., Pereira, A. S. A., Miyasato, P. A., Nakano, E., and Verjovski-Almeida, S. (2020). Long non-coding rna levels can be modulated by 5-azacytidine in schistosoma mansoni. *Sci Rep*, 10(1):21565.

Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E., and Dzikowski, R. (2015). Antisense long noncoding rnas regulate var gene activation in the malaria parasite plasmodium falciparum. *Proc Natl Acad Sci U S A*, 112(9):E982–91.

Amos, B., Aurrecoechea, C., Barba, M., Barreto, A., Basenko, E., Bażant, W., Belnap, R., Blevins, A. S., Böhme, U., Brestelli, J., Brunk, B. P., Caddick, M., Callan, D., Campbell, L., Christensen, M., Christophides, G., Crouch, K., Davis, K., DeBarry, J., Doherty, R., Duan, Y., Dunn, M., Falke, D., Fisher, S., Flicek, P., Fox, B., Gajria, B., Giraldo-Calderón, G. I., Harb, O. S., Harper, E., Hertz-Fowler, C., Hickman, M., Howington, C., Hu, S., Humphrey, J., Iodice, J., Jones, A., Judkins, J., Kelly, S. A., Kissinger, J. C., Kwon, D. K., Lamoureux, K., Lawson, D., Li, W., Lies, K., Lodha, D., Long, J., MacCallum, R. M., Maslen, G., McDowell, M. A., Nabrzyski, J., Roos, D. S., Rund, S. S. C., Schulman, S., Shanmugasundram, A., Sitnik, V., Spruill, D., Starns, D., Stoeckert, Christian J, J., Tomko, S. S., Wang, H., Warrenfeltz, S., Wieck, R., Wilkinson, P. A., Xu, L., and Zheng, J. (2021). Veupathdb: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res*, 50(D1):D898–D911.

Amulic, B., Salanti, A., Lavstsen, T., Nielsen, M. A., and Deitsch, K. W. (2009). An upstream open reading frame controls translation of var2csa, a gene implicated in placental malaria. *PLoS Pathog*, 5(1):e1000256.

Andrews, S. (2010). Fastqc: A quality control tool for high throughput sequence data.

Armstrong, C. M. and Goldberg, D. E. (2007). An fkbp destabilization domain modulates protein levels in plasmodium falciparum. *Nat Methods*, 4(12):1007–9.

Augagneur, Y., Jaubert, L., Schiavoni, M., Pachikara, N., Garg, A., Usmani-Brown, S., Wesolowski, D., Zeller, S., Ghosal, A., Cornillot, E., Said, H. M., Kumar, P., Altman, S., and Ben Mamoun, C. (2013). Identification and functional analysis of the primary pantothenate transporter, pfpat, of the human malaria parasite plasmodium falciparum. *J Biol Chem*, 288(28):20558–67.

Augagneur, Y., Wesolowski, D., Tae, H. S., Altman, S., and Ben Mamoun, C. (2012). Gene selective mrna cleavage inhibits the development of plasmodium falciparum. *Proc Natl Acad Sci U S A*, 109(16):6235–40.

Ay, F., Bunnik, E. M., Varoquaux, N., Bol, S. M., Prudhomme, J., Vert, J. P., Noble, W. S., and Le Roch, K. G. (2014). Three-dimensional modeling of the p. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res*, 24(6):974–88.

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

Balu, B., Shoue, D. A., Fraser, M. J., 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(45):16391–16396.

Bancells, C. and Deitsch, K. W. (2013). A molecular switch in the efficiency of translation reinitiation controls expression of var2csa, a gene implicated in pregnancy-associated malaria. *Mol Microbiol*, 90(3):472–88.

Baragaña, B., Hallyburton, I., Lee, M. C., Norcross, N. R., Grimaldi, R., Otto, T. D., Proto, W. R., Blagborough, A. M., Meister, S., Wirjanata, G., Ruecker, A., Upton, L. M., Abraham, T. S., Almeida, M. J., Pradhan, A., Porzelle, A., Luksch, T., Martínez, M. S., Luksch, T., Bolscher, J. M., Woodland, A., Norval, S., Zuccotto, F., Thomas, J., Simeons, F., Stojanovski, L., Osuna-Cabello, M., Brock, P. M., Churcher, T. S., Sala, K. A., Zakutansky, S. E., Jiménez-Díaz, M. B., Sanz, L. M., Riley, J., Basak, R., Campbell, M., Avery, V. M., Sauerwein, R. W., Dechering, K. J., Noviyanti, R., Campo, B., Frearson, J. A., Angulo-Barturen, I., Ferrer-Bazaga, S., Gamo, F. J., Wyatt, P. G., Leroy, D., Siegl, P., Delves, M. J., Kyle, D. E., Wittlin, S., Marfurt, J., Price, R. N., Sinden, R. E., Winzeler, E. A., Charman, S. A., Bebrevska, L., Gray, D. W., Campbell, S., Fairlamb, A. H., Willis, P. A., Rayner, J. C., Fidock, D. A., Read, K. D., and Gilbert, I. H. (2015). A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature*, 522(7556):315–20.

Barcons-Simon, A., Cordon-Obras, C., Guizetti, J., Bryant, J. M., and Scherf, A. (2020). Crispr interference of a clonally variant gc-rich noncoding rna family leads to general repression of var genes in plasmodium falciparum. *mBio*, 11(1).

Barker, R. H., J., Metelev, V., Rapaport, E., and Zamecnik, P. (1996). Inhibition of plasmodium falciparum malaria using antisense oligodeoxynucleotides. *Proc Natl Acad Sci U S A*, 93(1):514–8.

Barker, R. H., Metelev, V., Coakley, A., and Zamecnik, P. (1998). Plasmodium falciparum: Effect of chemical structure on efficacy and specificity of antisense oligonucleotides against malaria in vitro. *Exp Parasitol*, 88(1):51–59.

Batugedara, G. and Le Roch, K. G. (2019). Unraveling the 3d genome of human malaria parasites. *Semin Cell Dev Biol*, 90:144–153.

Batugedara, G., Lu, X. M., Abel, S., Chahine, Z., Hristov, B., Williams, D., Hollin, T., Wang, T., Cort, A., Lenz, T., Thompson, T., Prudhomme, J., Tripathi, A. K., Xu, G., Cudini, J., Dogga, S., Lawniczak, M., Stafford Noble, W., Sinnis, P., and Le Roch, K. G. (2022). Deciphering the non-coding code of pathogenicity and sexual differentiation in the human malaria parasite. *bioRxiv*, page 511630.

Baum, J., Papenfuss, A. T., Mair, G. R., Janse, C. J., Vlachou, D., Waters, A. P., Cowman, A. F., Crabb, B. S., and de Koning-Ward, T. F. (2009). Molecular genetics and comparative genomics reveal rnai is not functional in malaria parasites. *Nucleic Acids Res*, 37(11):3788–98.

Baumgarten, S., Bryant, J. M., Sinha, A., Reyser, T., Preiser, P. R., Dedon, P. C., and Scherf, A. (2019). Transcriptome-wide dynamics of extensive m(6)a mrna methylation during plasmodium falciparum blood-stage development. *Nat Microbiol*, 4(12):2246–2259.

Benchling (2020). Benchling [biology software].

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J., and Gluenz, E. (2017). A crispr cas9 high-throughput genome editing toolkit for kinetoplastids. *R Soc Open Sci*, 4(5):170095.

Benham, C. J. (1979). Torsional stress and local denaturation in supercoiled dna. *Proc Natl Acad Sci U S A*, 76(8):3870–4.

Bennink, S., Kiesow, M. J., and Pradel, G. (2016). The development of malaria parasites in the mosquito midgut. *Cell Microbiol*, 18(7):905–18.

Bennink, S., von Bohl, A., Ngwa, C. J., Henschel, L., Kuehn, A., Pilch, N., Weißbach, T., Rosinski, A. N., Scheuermayer, M., Repnik, U., Przyborski, J. M., Minns, A. M., Orchard, L. M., Griffiths, G., Lindner, S. E., Llinás, M., and Pradel, G. (2018). A seven-helix protein constitutes stress granules crucial for regulating translation during human-to-mosquito transmission of plasmodium falciparum. *PLoS Pathog*, 14(8):e1007249.

Berretta, J. and Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep*, 10(9):973–82.

Bester, A. C., Lee, J. D., Chavez, A., Lee, Y. R., Nachmani, D., Vora, S., Victor, J., Sauvageau, M., Monteleone, E., Rinn, J. L., Provero, P., Church, G. M., Clohessy, J. G., and Pandolfi, P. P. (2018). An integrated genome-wide crispra approach to functionalize lncrnas in drug resistance. *Cell*, 173(3):649–664.e20.

Birnbaum, J., Flemming, S., Reichard, N., Soares, A. B., Mesén-Ramírez, P., Jonscher, E., Bergmann, B., and Spielmann, T. (2017). A genetic system to study plasmodium falciparum protein function. *Nat Methods*, 14(4):450–456.

Bogart, S. (2022). Sankeymatic [software].

Boopathi, P. A., Subudhi, A. K., Garg, S., Middha, S., Acharya, J., Pakalapati, D., Saxena, V., Aiyaz, M., Chand, B., Mugasimangalam, R. C., Kochar, S. K., Sirohi, P., Kochar, D. K., and Das, A. (2013). Revealing natural antisense transcripts from plasmodium vivax isolates: evidence of genome regulation in complicated malaria. *Infect Genet Evol*, 20:428–43.

Bot, J. F., van der Oost, J., and Geijsen, N. (2022). The double life of crispr–cas13. *Current Opinion in Biotechnology*, 78:102789.

Bousema, T. and Drakeley, C. (2011). Epidemiology and infectivity of plasmodium falciparum and plasmodium vivax gametocytes in relation to malaria control and elimination. *Clin Microbiol Rev*, 24(2):377–410.

Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of plasmodium falciparum. *PLoS Biol*, 1(1):E5.

Braks, J. A., Mair, G. R., Franke-Fayard, B., Janse, C. J., and Waters, A. P. (2008). A conserved u-rich rna region implicated in regulation of translation in plasmodium female gametocytes. *Nucleic Acids Res*, 36(4):1176–86.

Broadbent, K. M., Broadbent, J. C., Ribacke, U., Wirth, D., Rinn, J. L., and Sabeti, P. C. (2015). Strand-specific rna sequencing in plasmodium falciparum malaria identifies developmentally regulated long non-coding rna and circular rna. *BMC Genomics*, 16(1):454.

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

Bryant, J., C, R., C, S.-B., S, B., J, G., A, S., and L, S. (2017). Crispr/cas9 genome editing reveals that the intron is not essential for var2csa gene activation or silencing in plasmodium falciparum. *mBio*, 8(4):e00729–17.

Bryant, J. M., Baumgarten, S., Dingli, F., Loew, D., Sinha, A., Claës, A., Preiser, P. R., Dedon, P. C., and Scherf, A. (2020). Exploring the virulence gene interactome with crispr/dcas9 in the human malaria parasite. *Mol Syst Biol*, 16(8):e9569.

Bryant, J. M., Baumgarten, S., Glover, L., Hutchinson, S., and Rachidi, N. (2019). Crispr in parasitology: Not exactly cut and dried! *Trends Parasitol*, 35(6):409–422.

Bugnon, L. A., Edera, A. A., Prochetto, S., Gerard, M., Raad, J., Fenoy, E., Rubiolo, M., Chorostecki, U., Gabaldón, T., Ariel, F., Di Persia, L. E., Milone, D. H., and Stegmayer, G. (2022). Secondary structure prediction of long noncoding rna: review and experimental comparison of existing approaches. *Briefings in Bioinformatics*, 23(4).

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

Bunnik, E. M., Polishko, A., Prudhomme, J., Ponts, N., Gill, S. S., Lonardi, S., and Le Roch, K. G. (2014). Dna-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite plasmodium falciparum. *BMC Genomics*, 15(1):347.

Bushell, E., Gomes, A. R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modrzynska, K., Schwach, F., Martin, R. E., Mather, M. W., McFadden, G. I., Parts, L., Rutledge, G. G., Vaidya, A. B., Wengelnik, K., Rayner, J. C., and Billker, O. (2017). Functional profiling of a plasmodium genome reveals an abundance of essential genes. *Cell*, 170(2):260–272.e8.

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

Böhme, U., Otto, T. D., Sanders, M., Newbold, C. I., and Berriman, M. (2019). Progression of the canonical reference malaria parasite genome from 2002-2019. *Wellcome Open Res*, 4:58.

Cai, P., Otten, A. B. C., Cheng, B., Ishii, M. A., Zhang, W., Huang, B., Qu, K., and Sun, B. K. (2020). A genome-wide long noncoding rna crispri screen identifies prancr as a novel regulator of epidermal homeostasis. *Genome Res*, 30(1):22–34.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). Blast+: architecture and applications. *BMC Bioinformatics*, 10:421.

Campbell, B. C., Nabel, E. M., Murdock, M. H., Lao-Peregrin, C., Tsoulfas, P., Blackmore, M. G., Lee, F. S., Liston, C., Morishita, H., and Petsko, G. A. (2020). mgreenlantern: a bright monomeric fluorescent protein with rapid expression and cell filling properties for neuronal imaging. *Proc Natl Acad Sci U S A*, 117(48):30710–30721.

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

Campelo Morillo, R. A., Tong, X., Xie, W., Abel, S., Orchard, L. M., Daher, W., Patel, D. J., Llinás, M., Le Roch, K. G., and Kafsack, B. F. C. (2022). The transcriptional regulator hdp1 controls expansion of the inner membrane complex during early sexual differentiation of malaria parasites. *Nat Microbiol*, 7(2):289–299.

Caro, F., Ahyong, V., Betegon, M., and DeRisi, J. L. (2014). Genome-wide regulatory dynamics of translation in the plasmodium falciparum asexual blood stages. *Elife*, 3.

Caro, F., Miller, M. G., and DeRisi, J. L. (2012). Plate-based transfection and culturing technique for genetic manipulation of plasmodium falciparum. *Malar J*, 11:22.

Carrasquilla, M., Adjalley, S., Sanderson, T., Marin-Menendez, A., Coyle, R., Montandon, R., Rayner, J. C., Pance, A., and Lee, M. C. S. (2020). Defining multiplicity of vector uptake in transfected plasmodium parasites. *Sci Rep*, 10(1):10894.

Carter, R. and Mendis, K. N. (2002). Evolutionary and historical aspects of the burden of malaria. *Clin Microbiol Rev*, 15(4):564–94.

Carter, R. and Miller, L. H. (1979). Evidence for environmental modulation of gametocytogenesis in plasmodium falciparum in continuous culture. *Bull World Health Organ*, 57 Suppl 1(Suppl):37–52.

Carvalho, T. G., Thiberge, S., Sakamoto, H., and Ménard, R. (2004). Conditional mutagenesis using site-specific recombination in plasmodium berghei. *Proc Natl Acad Sci U S A*, 101(41):14931–6.

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2012). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, 28(4):464–9.

Caudron-Herger, M., Rusin, S. F., Adamo, M. E., Seiler, J., Schmid, V. K., Barreau, E., Kettenbach, A. N., and Diederichs, S. (2019). R-deep: Proteome-wide and quantitative identification of rna-dependent proteins by density gradient ultracentrifugation. *Mol Cell*, 75(1):184–199.e10.

Cerutti, N., Marin, A., Massa, E. R., and Savoia, D. (1999). Immunological investigation of malaria and new perspectives in paleopathological studies. *Boll Soc Ital Biol Sper*, 75(3-4):17–20.

Chakrabarti, K., Pearson, M., Grate, L., Sterne-Weiler, T., Deans, J., Donohue, J. P., and Ares, M., J. (2007). Structural rnas of known and unknown function identified in malaria parasites by comparative genomics and rna analysis. *RNA*, 13(11):1923–39.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805.

Chappell, L., Ross, P., Orchard, L., Russell, T. J., Otto, T. D., Berriman, M., Rayner, J. C., and Llinás, M. (2020). Refining the transcriptome of the human malaria parasite plasmodium falciparum using amplification-free rna-seq. *BMC Genomics*, 21(1):395.

Chen, C.-K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the x chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science*, 354(6311):468–472.

Chen, G., Liu, S.-c., Fan, X.-y., Jin, Y.-l., Li, X., and Du, Y.-t. (2022). Plasmodium manipulates the expression of host long non-coding rna during red blood cell intracellular infection. *Parasites & Vectors*, 15(1):182.

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., and Doudna, J. A. (2018). Crispr-cas12a target binding unleashes indiscriminate single-stranded dnase activity. *Science*, 360(6387):436–439.

Chen, Y., Li, Z., Chen, X., and Zhang, S. (2021). Long non-coding rnas: From disease code to drug role. *Acta Pharmaceutica Sinica B*, 11(2):340–354.

Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., Yost, K. E., Kim, J., He, J., and Nevins, S. A. (2018). Promoter of lncrna gene pvt1 is a tumor-suppressor dna boundary element. *Cell*, 173(6):1398–1412. e22.

Chu, C., Quinn, J., and Chang, H. Y. (2012). Chromatin isolation by rna purification (chirp). *J Vis Exp*, .(61):3912.

Chu, C., Zhang, Q., da Rocha, S., Flynn, R., Bharadwaj, M., Calabrese, J., Magnuson, T., Heard, E., and Chang, H. (2015). Systematic discovery of xist rna binding proteins. *Cell*, 161(2):404–416.

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., and Lawrence, J. B. (2009). An architectural role for a nuclear noncoding rna: Neat1 rna is essential for the structure of paraspeckles. *Mol Cell*, 33(6):717–726.

Collins, C. R., Das, S., Wong, E. H., Andenmatten, N., Stallmach, R., Hackett, F., Herman, J. P., Müller, S., Meissner, M., and Blackman, M. J. (2013). Robust inducible cre recombinase activity in the human malaria parasite plasmodium falciparum enables efficient gene deletion within a single asexual erythrocytic growth cycle. *Mol Microbiol*, 88(4):687–701.

Collins, C. R., Hackett, F., Howell, S. A., Snijders, A. P., Russell, M. R., Collinson, L. M., and Blackman, M. J. (2020). The malaria parasite sheddase sub2 governs host red blood cell membrane sealing at invasion. *Elife*, 9.

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). Upsetr: an r package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18):2938–2940.

Coulibaly, A., Diop, M. F., Kone, A., Dara, A., Ouattara, A., Mulder, N., Miotto, O., Diakite, M., Djimde, A., and Amambua-Ngwa, A. (2022). Genome-wide snp analysis of plasmodium falciparum shows differentiation at drug-resistance-associated loci among malaria transmission settings in southern mali. *Front Genet*, 13:943445.

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(8):1548–54.

Cowell, A. N., Istvan, E. S., Lukens, A. K., Gomez-Lorenzo, M. G., Vanaerschot, M., Sakata-Kato, T., Flannery, E. L., Magistrado, P., Owen, E., Abraham, M., LaMonte, G., Painter, H. J., Williams, R. M., Franco, V., Linares, M., Arriaga, I., Bopp, S., Corey, V. C., Gnädig, N. F., Coburn-Flynn, O., Reimer, C., Gupta, P., Murithi, J. M., Moura, P. A., Fuchs, O., Sasaki, E., Kim, S. W., Teng, C. H., Wang, L. T., Akidil, A., Adjalley, S., Willis, P. A., Siegel, D., Tanaseichuk, O., Zhong, Y., Zhou, Y., Llinás, M., Ottilie, S., Gamo, F. J., Lee, M. C. S., Goldberg, D. E., Fidock, D. A., Wirth, D. F., and Winzeler, E. A. (2018). Mapping the malaria parasite druggable genome by using in vitro evolution and chemogenomics. *Science*, 359(6372):191–199.

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., and Zhang, F. (2017). Rna editing with crispr-cas13. *Science*, 358(6366):1019–1027.

Cox, F. E. G. (2010). History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors*, 3(1):5.

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(5):e0178163.

Cudini, J. (2021). *Characterisation of Plasmodium parasite sexual commitment and development*. Thesis.

Cui, L., Fan, Q., and Li, J. (2002). The malaria parasite plasmodium falciparum encodes members of the puf rna-binding protein family with conserved rna binding activity. *Nucleic Acids Res*, 30(21):4607–17.

Cui, L., Lindner, S., and Miao, J. (2015). Translational regulation during stage transitions in malaria parasites. *Ann N Y Acad Sci*, 1342(1):1–9.

Cui, L., Miao, J., Furuya, T., Li, X., Su, X. Z., and Cui, L. (2007). Pfgcn5-mediated histone h3 acetylation plays a key role in gene expression in plasmodium falciparum. *Eukaryot Cell*, 6(7):1219–27.

Cunningham, C. H., Hennelly, C. M., Lin, J. T., Ubalee, R., Boyce, R. M., Mulogo, E. M., Hathaway, N., Thwai, K. L., Phanzu, F., Kalonji, A., Mwandagalirwa, K., Tshefu, A., Juliano, J. J., and Parr, J. B. (2021). A novel crispr-based malaria diagnostic capable of plasmodium detection, species differentiation, and drug-resistance genotyping. *EBioMedicine*, 68:103415.

Cárdenas, P., Esherick, L. Y., Chambonnier, G., Dey, S., Turlo, C. V., Nasamu, A. S., and Niles, J. C. (2022). Genetargeter: automated in silico design for genome editing in the malaria parasite, plasmodium falciparum. *CRISPR J*, 5(1):155–164.

Dandewad, V., Vindu, A., Joseph, J., and Seshadri, V. (2019). Import of human mirna-risc complex into plasmodium falciparum and regulation of the parasite gene expression. *J Biosci*, 44(2):50.

Diffendall, G. M., Barcons-Simon, A., Baumgarten, S., Dingli, F., Loew, D., and Scherf, A. (2023). Discovery of ruf6 ncrna-interacting proteins involved in p. falciparum immune evasion. *Life Sci Alliance*, 6(1).

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21.

Donde, M. J., Rochussen, A. M., Kapoor, S., and Taylor, A. I. (2022). Targeting non-coding rna family members with artificial endonuclease xnazymes. *Commun Biol*, 5(1):1010.

Duffy, C. W., Amambua-Ngwa, A., Ahouidi, A. D., Diakite, M., Awandare, G. A., Ba, H., Tarr, S. J., Murray, L., Stewart, L. B., D'Alessandro, U., Otto, T. D., Kwiatkowski, D. P., and Conway, D. J. (2018). Multi-population genomic analysis of malaria parasites indicates local selection and differentiation at the gdv1 locus regulating sexual development. *Sci Rep*, 8(1):15763.

Engelmann, S., Silvie, O., and Matuschewski, K. (2009). Disruption of plasmodium sporozoite transmission by depletion of sporozoite invasion-associated protein 1. *Eukaryot Cell*, 8(4):640–8.

Engreitz, J., Lander, E. S., and Guttman, M. (2015). Rna antisense purification (rap) for mapping rna interactions with chromatin. *Methods Mol Biol*, 1262:183–197.

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., and Lander, E. S. (2016). Local regulation of gene expression by lncrna promoters, transcription and splicing. *Nature*, 539(7629):452–455.

Epp, C., Li, F., Howitt, C. A., Chookajorn, T., and Deitsch, K. W. (2009). Chromatin associated sense and antisense noncoding rnas are transcribed from the var gene family of virulence genes of the malaria parasite plasmodium falciparum. *RNA*, 15(1):116–27.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048.

Fan, Q., An, L., and Cui, L. (2004). Plasmodium falciparum histone acetyltransferase, a yeast gcn5 homologue involved in chromatin remodeling. *Eukaryot Cell*, 3(2):264–76.

Fan, Y., Shen, S., Wei, G., Tang, J., Zhao, Y., Wang, F., He, X., Guo, G., Shang, X., Yu, X., Ma, Z., He, X., Liu, M., Zhu, Q., Le, Z., Wei, G., Cao, J., Jiang, C., Zhang, Q., and Miller, L. H. (2020). Rrp6 regulates heterochromatic gene silencing via ncrna ruf6 decay in malaria parasites. *mBio*, 11(3):e01110–20.

Fastman, Y., Assaraf, S., Rose, M., Milrot, E., Basore, K., Arasu, B. S., Desai, S. A., Elbaum, M., and Dzikowski, R. (2018). An upstream open reading frame (uorf) signals for cellular localization of the virulence factor implicated in pregnancy associated malaria. *Nucleic Acids Res*, 46(10):4919–4932.

Fennell, C., Babbitt, S., Russo, I., Wilkes, J., Ranford-Cartwright, L., Goldberg, D. E., and Doerig, C. (2009). Pfeik1, a eukaryotic initiation factor 2alpha kinase of the human malaria parasite plasmodium falciparum, regulates stress-response to amino-acid starvation. *Malar J*, 8:99.

Fidock, D. and Wellems, T. (1997). Transformation with human dihydrofolate reductase renders malaria parasites insensitive to wr99210 but does not affect the intrinsic activity of proguanil. *Proc Natl Acad Sci U S A*, 94:10931–6.

Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J. C., and Scherf, A. (2002). A central role for plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J*, 21(4):815–24.

Filarsky, M., Fraschka, S. A., Niederwieser, I., Brancucci, N. M. B., Carrington, E., Carrió, E., Moes, S., Jenoe, P., Bártfai, R., and Voss, T. S. (2018). Gdv1 induces sexual commitment of malaria parasites by antagonizing hp1-dependent gene silencing. *Science*, 359(6381):1259–1263.

Fivelman, Q. L., McRobert, L., Sharp, S., Taylor, C. J., Saeed, M., Swales, C. A., Sutherland, C. J., and Baker, D. A. (2007). Improved synchronous production of plasmodium falciparum gametocytes in vitro. *Molecular and Biochemical Parasitology*, 154(1):119–123.

Flores, M. V., Atkins, D., Wade, D., O'Sullivan, W. J., and Stewart, T. S. (1997). Inhibition of plasmodium falciparum proliferation in vitro by ribozymes. *J Biol Chem*, 272(27):16940–5.

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

Fraschka, S. A., Filarsky, M., Hoo, R., Niederwieser, I., Yam, X. Y., Brancucci, N. M. B., Mohring, F., Mushunje, A. T., Huang, X., Christensen, P. R., Nosten, F., Bozdech, Z., Russell, B., Moon, R. W., Marti, M., Preiser, P. R., Bártfai, R., and Voss, T. S. (2018). Comparative heterochromatin profiling reveals conserved and unique epigenome signatures linked to adaptation and development of malaria parasites. *Cell Host Microbe*, 23(3):407–420.e8.

Frischknecht, F., Martin, B., Thiery, I., Bourgouin, C., and Menard, R. (2006). Using green fluorescent malaria parasites to screen for permissive vector mosquitoes. *Malar J*, 5(1):23.

Gabryszewski, S. J., Dhingra, S. K., Combrinck, J. M., Lewis, I. A., Callaghan, P. S., Hassett, M. R., Siriwardana, A., Henrich, P. P., Lee, A. H., Gnädig, N. F., Musset, L., Llinás, M., Egan, T. J., Roepe, P. D., and Fidock, D. A. (2016). Evolution of fitness cost-neutral mutant pfcrt conferring p. falciparum 4-aminoquinoline drug resistance is accompanied by altered parasite metabolism and digestive vacuole physiology. *PLoS Pathog*, 12(11):e1005976.

Galen, S. C., Borner, J., Martinsen, E. S., Schaer, J., Austin, C. C., West, C. J., and Perkins, S. L. (2018). The polyphyly of plasmodium: comprehensive phylogenetic analyses of the malaria parasites (order haemosporida) reveal widespread taxonomic conflict. *R Soc Open Sci*, 5(5):171780.

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.

Gao, N., Li, Y., Li, J., Gao, Z., Yang, Z., Li, Y., Liu, H., and Fan, T. (2020). Long non-coding rnas: The regulatory mechanisms, research strategies, and future directions in cancers. *Front Oncol*, 10:598817.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., and Bowman, S. (2002). Genome sequence of the human malaria parasite plasmodium falciparum. *Nature*, 419(6906):498–511.

Garg, A., Wesolowski, D., Alonso, D., Deitsch, K. W., Ben Mamoun, C., and Altman, S. (2015). Targeting protein translation, rna splicing, and degradation by morpholino-based conjugates in plasmodium falciparum. *Proc Natl Acad Sci U S A*, 112(38):11935–11940.

Gerber, P. P., Donde, M. J., Matheson, N. J., and Taylor, A. I. (2022). Xnazymes targeting the sars-cov-2 genome inhibit viral infection. *Nature Communications*, 13(1):6716.

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. *Nature Biotechnology*, 32:819.

Glushakova, S., Beck, J. R., Garten, M., Busse, B. L., Nasamu, A. S., Tenkova-Heuser, T., Heuser, J., Goldberg, D. E., and Zimmerberg, J. (2018). Rounding precedes rupture and breakdown of vacuolar membranes minutes before malaria parasite egress from erythrocytes. *Cell Microbiol*, 20(10):e12868.

Gomes, A., Bushell, E., Schwach, F., Girling, G., Anar, B., Quail, M., Herd, C., Pfander, C., Modrzynska, K., Rayner, J., and Billker, O. (2015). A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite. *Cell Host Microbe*, 17(3):404–413.

Gomes, A. R., Marin-Menendez, A., Adjalley, S. H., Bardy, C., Cassan, C., Lee, M. C. S., and Talman, A. M. (2022). A transcriptional switch controls sex determination in plasmodium falciparum. *Nature*, 612:528–533.

Goonewardene, R., Daily, J., Kaslow, D., Sullivan, T. J., Duffy, P., Carter, R., Mendis, K., and Wirth, D. (1993). Transfection of the malaria parasite and expression of firefly luciferase. *Proc Natl Acad Sci U S A*, 90(11):5234–6.

Goudarzi, M., Berg, K., Pieper, L. M., and Schier, A. F. (2019). Individual long non-coding rnas have no overt functions in zebrafish embryogenesis, viability and fertility. *Elife*, 8:e40815.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at embl–ebi. *Nucleic Acids Res*, 38(2):W695–W699.

Graf, M., Bonetti, D., Lockhart, A., Serhal, K., Kellner, V., Maicher, A., Jolivet, P., Teixeira, M. T., and Luke, B. (2017). Telomere length determines terra and r-loop regulation through the cell cycle. *Cell*, 170(1):72–85.e14.

Grossi, E., Raimondi, I., Goñi, E., González, J., Marchese, F. P., Chapaprieta, V., Martín-Subero, J. I., Guo, S., and Huarte, M. (2020). A lncrna-swi/snf complex crosstalk controls transcriptional activation at specific promoter regions. *Nat Commun*, 11(1):936.

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008). The vienna rna websuite. *Nucleic Acids Res*, 36(Web Server issue):W70–W74.

Guerreiro, A., Deligianni, E., Santos, J. M., Silva, P. A. G. C., Louis, C., Pain, A., Janse, C. J., Franke-Fayard, B., Carret, C. K., Siden-Kiamos, I., and Mair, G. R. (2014). Genome-wide rip-chip analysis of translational repressor-bound mrnas in the plasmodium gametocyte. *Genome Biol*, 15(11):493.

Guinet, F., Dvorak, J. A., Fujioka, H., Keister, D. B., Muratova, O., Kaslow, D. C., Aikawa, M., Vaidya, A. B., and Wellems, T. E. (1996). A developmental defect in plasmodium falciparum male gametogenesis. *J Cell Biol*, 135(1):269–78.

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

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

Hacisuleyman, E., Goff, L. A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D. G., Sauvageau, M., and Kelley, D. R. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding rna firre. *Nature Structural & Molecular Biology*, 21(2):198–206.

Haldar, K., Bhattacharjee, S., and Safeukui, I. (2018). Drug resistance in plasmodium. *Nature Reviews Microbiology*, 16(3):156–170.

Hall, N., Karras, M., Raine, J. D., Carlton, J. M., Kooij, T. W., Berriman, M., Florens, L., Janssen, C. S., Pain, A., Christophides, G. K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M. A., Ormond, D., Doggett, J., Trueman, H. E., Mendoza, J., Bidwell, S. L., Rajandream, M. A., Carucci, D. J., Yates, J. R., r., Kafatos, F. C., Janse, C. J., Barrell, B., Turner, C. M., Waters, A. P., and Sinden, R. E. (2005). A comprehensive survey of the plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science*, 307(5706):82–6.

Hamid, N., Xuhua, X., A., H. D., and GoldingB. (2014). The evolution of genomic gc content undergoes a rapid reversal within the genus plasmodium. *Genome*, 57(9):507–511.

Hammam, E., Ananda, G., Sinha, A., Scheidig-Benatar, C., Bohec, M., Preiser, P. R., Dedon, P. C., Scherf, A., and Vembar, S. S. (2019). Discovery of a new predominant cytosine dna modification that is linked to gene expression in malaria parasites. *Nucleic Acids Res*, 48(1):184–199.

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., and Kjems, J. (2013). Natural rna circles function as efficient microrna sponges. *Nature*, 495:384–388.

Haswell, J. R., Mattioli, K., Gerhardinger, C., Maass, P. G., Foster, D. J., Peinado, P., Wang, X., Medina, P. P., Rinn, J. L., and Slack, F. J. (2021). Genome-wide crispr interference screen identifies long non-coding rna loci required for differentiation and pluripotency. *PLoS One*, 16(11):e0252848.

Hawking, F., Worms, M. J., and Gammage, K. (1968). 24- and 48-hour cycles of malaria parasites in the blood; their purpose, production and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 62(6):731–760.

Heinberg, A., Amit-Avraham, I., Mitesser, V., Simantov, K., Goyal, M., Nevo, Y., Kandelis-Shalev, S., Thompson, E., and Dzikowski, R. (2022). A nuclear redox sensor modulates gene activation and var switching in plasmodium falciparum. *Proc Natl Acad Sci U S A*, 119(33):e2201247119.

Hempelmann, E. and Krafts, K. (2013). Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria. *Malar J*, 12:232.

Herrera-Úbeda, C., Marín-Barba, M., Navas-Pérez, E., Gravemeyer, J., Albuixech-Crespo, B., Wheeler, G. N., and Garcia-Fernàndez, J. (2019). Microsyntenic clusters reveal conservation of lncrnas in chordates despite absence of sequence conservation. *Biology (Basel)*, 8(3).

Hickey, G. (2021). Cactus.

Hoeijmakers, W., Miao, J., Schmidt, S., Toenhake, C. G., Shrestha, S., Venhuizen, J., Henderson, R., Birnbaum, J., Ghidelli-Disse, S., Drewes, G., Cui, L., Stunnenberg, H. G., Spielmann, T., and Bártfai, R. (2019). Epigenetic reader complexes of the human malaria parasite, plasmodium falciparum. *Nucleic Acids Res*, 47(22):11574–11588.

Holmes, M. J., Augusto, L. D. S., Zhang, M., Wek, R. C., and Sullivan, W. J., J. (2017). Translational control in the latency of apicomplexan parasites. *Trends Parasitol*, 33(12):947–960.

Homann, O. R. and Johnson, A. D. (2010). Mochiview: versatile software for genome browsing and dna motif analysis. *BMC Biol*, 8:49.

Horrocks, P. and Kilbey, B. J. (1996). Physical and functional mapping of the transcriptional start sites of plasmodium falciparum proliferating cell nuclear antigen. *Mol Biochem Parasitol*, 82(2):207–215.

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

Hoshizaki, J., Adjalley, S. H., Thathy, V., Judge, K., Berriman, M., Reid, A. J., and Lee, M. C. S. (2022a). A manually curated annotation characterises genomic features of p. falciparum lncrnas. *BMC Genomics*, 23(1):780.

Hoshizaki, J., Jagoe, H., and Lee, M. C. S. (2022b). Efficient generation of mneongreen plasmodium falciparum reporter lines enables quantitative fitness analysis. *Front Cell Infect Microbiol*, 12:981432.

Hunt, P., Afonso, A., Creasey, A., Culleton, R., Sidhu, A. B., Logan, J., Valderramos, S. G., McNae, I., Cheesman, S., do Rosario, V., Carter, R., Fidock, D. A., and Cravo, P. (2007). Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. *Mol Microbiol*, 65(1):27–40.

Huynh, N., Depner, N., Larson, R., and King-Jones, K. (2020). A versatile toolkit for crispr-cas13-based rna manipulation in drosophila. *Genome Biol*, 21(1):279.

Ippolito, M. M., Moser, K. A., Kabuya, J.-B. B., Cunningham, C., and Juliano, J. J. (2021). Antimalarial drug resistance and implications for the who global technical strategy. *Current Epidemiology Reports*, 8(2):46–62.

Ishizaki, T., Hernandez, S., Paoletta, M. S., Sanderson, T., and Bushell, E. S. C. (2022). Crispr/cas9 and genetic screens in malaria parasites: small genomes, big impact. *Biochem Soc Trans*, 50(3):1069–1079.

Istvan, E. S., Mallari, J. P., Corey, V. C., Dharia, N. V., Marshall, G. R., Winzeler, E. A., and Goldberg, D. E. (2017). Esterase mutation is a mechanism of resistance to antimalarial compounds. *Nature Communications*, 8(1):14240.

Iwanaga, S., Kaneko, I., Kato, T., and Yuda, M. (2012). Identification of an ap2-family protein that is critical for malaria liver stage development. *PLoS One*, 7(11):e47557.

Jain, A. K., Xi, Y., McCarthy, R., Allton, K., Akdemir, K. C., Patel, L. R., Aronow, B., Lin, C., Li, W., and Yang, L. (2016). Lncpress1 is a p53-regulated lncrna that safeguards pluripotency by disrupting sirt6-mediated de-acetylation of histone h3k56. *Mol Cell*, 64(5):967–981.

Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., Yang, W., Turner, L., Lavstsen, T., and Theander, T. G. (2013). Pfsetvs methylation of histone h3k36 represses virulence genes in plasmodium falciparum. *Nature*, 499(7457):223–227.

Jing, Q., Cao, L., Zhang, L., Cheng, X., Gilbert, N., Dai, X., Sun, M., Liang, S., and Jiang, L. (2018). Plasmodium falciparum var gene is activated by its antisense long noncoding rna. *Front Microbiol*, 9:3117.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). Ncbi blast: a better web interface. *Nucleic Acids Res*, 36(2):W5–W9.

Jones, M. L., Das, S., Belda, H., Collins, C. R., Blackman, M. J., and Treeck, M. (2016). A versatile strategy for rapid conditional genome engineering using loxp sites in a small synthetic intron in plasmodium falciparum. *Sci Rep*, 6:21800.

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

Kafsack, B. F., Rovira-Graells, N., Clark, T. G., Bancells, C., Crowley, V. M., Campino, S. G., Williams, A. E., Drought, L. G., Kwiatkowski, D. P., Baker, D. A., Cortés, A., and Llinás, M. (2014). A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, 507(7491):248–252.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., and Petrov, A. I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding rna families. *Nucleic Acids Res*, 46(D1):D335–D342.

Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., and Gao, G. (2017). Cpc2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res*, 45(W1):W12–W16.

Kannan, S., Altae-Tran, H., Jin, X., Madigan, V. J., Oshiro, R., Makarova, K. S., Koonin, E. V., and Zhang, F. (2022). Compact rna editors with small cas13 proteins. *Nature Biotechnology*, 40(2):194–197.

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermüller, J., Hofacker, I. L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., and Gingeras, T. R. (2007). Rna maps reveal new rna classes and a possible function for pervasive transcription. *Science*, 316(5830):1484–1488.

Karner, H., Webb, C.-H., Carmona, S., Liu, Y., Lin, B., Erhard, M., Chan, D., Baldi, P., Spitale, R. C., and Sun, S. (2020). Functional conservation of lncrna jpx despite sequence and structural divergence. *Journal of Molecular Biology*, 432(2):283–300.

Kayiba, N. K., Yobi, D. M., Devleesschauwer, B., Mvumbi, D. M., Kabututu, P. Z., Likwela, J. L., Kalindula, L. A., DeMol, P., Hayette, M.-P., Mvumbi, G. L., Lusamba, P. D., Beutels, P., Rosas-Aguirre, A., and Speybroeck, N. (2021). Care-seeking behaviour and socio-economic burden associated with uncomplicated malaria in the democratic republic of congo. *Malar J*, 20(1):260.

Kensche, P. R., Hoeijmakers, W. A. M., Toenhake, C. G., Bras, M., Chappell, L., Berriman, M., and Bártfai, R. (2016). The nucleosome landscape of plasmodium falciparum reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Res*, 44(5):2110–2124.

Kim, A., Popovici, J., Vantaux, A., Samreth, R., Bin, S., Kim, S., Roesch, C., Liang, L., Davies, H., and Felgner, P. (2017). Characterization of p. vivax blood stage transcriptomes from field isolates reveals similarities among infections and complex gene isoforms. *Sci Rep*, 7:7761.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nature Biotechnology*, 37(8):907–915.

Kirkman, L. A., Lawrence, E. A., and Deitsch, K. W. (2014). Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. *Nucleic Acids Res*, 42(1):370–9.

Knuepfer, E., Napiorkowska, M., van Ooij, C., and Holder, A. A. (2017). Generating conditional gene knockouts in plasmodium – a toolkit to produce stable dicre recombinase-expressing parasite lines using crispr/cas9. *Sci Rep*, 7(1):3881.

Kolevzon, N., Nasereddin, A., Naik, S., Yavin, E., and Dzikowski, R. (2014). Use of peptide nucleic acids to manipulate gene expression in the malaria parasite plasmodium falciparum. *PLoS One*, 9(1):e86802.

Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., and Hsu, P. D. (2018). Transcriptome engineering with rna-targeting type vi-d crispr effectors. *Cell*, 173(3):665–676.e14.

Konopka, J. K., Task, D., Afify, A., Raji, J., Deibel, K., Maguire, S., Lawrence, R., and Potter, C. J. (2021). Olfaction in anopheles mosquitoes. *Chem Senses*, 46:bjab021.

Kramer, K. J., Kan, S. C., and Siddiqui, W. A. (1982). Concentration of plasmodium falciparum-infected erythrocytes by density gradient centrifugation in percoll. *J Parasitol*, 68(2):336–337.

Kuang, D., Qiao, J., Li, Z., Wang, W., Xia, H., Jiang, L., Dai, J., Fang, Q., and Dai, X. (2017). Tagging to endogenous genes of plasmodium falciparum using crispr/cas9. *Parasites & Vectors*, 10(1):595.

Kyes, S. A., Christodoulou, Z., Raza, A., Horrocks, P., Pinches, R., Rowe, J. A., and Newbold, C. I. (2003). A well-conserved plasmodium falciparum var gene shows an unusual stage-specific transcript pattern. *Molecular Microbiology*, 48(5):1339–1348.

Lagarde, J., Uszczynska-Ratajczak, B., Santoyo-Lopez, J., Gonzalez, J. M., Tapanari, E., Mudge, J. M., Steward, C. A., Wilming, L., Tanzer, A., Howald, C., Chrast, J., Vela-Boza, A., Rueda, A., Lopez-Domingo, F. J., Dopazo, J., Reymond, A., Guigó, R., and Harrow, J. (2016). Extension of human lncrna transcripts by race coupled with long-read high-throughput sequencing (race-seq). *Nature Communications*, 7(1):12339.

Lambros, C. and Vanderberg, J. P. (1979). Synchronization of plasmodium falciparum erythrocytic stages in culture. *J Parasitol*, 65(3):418–20.

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). Crispr interference (crispri) for sequence-specific control of gene expression. *Nat Protoc*, 8(11):2180–96.

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

Le Roch, K. G., Johnson, J. R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S. F., Williamson, K. C., Holder, A. A., Carucci, D. J., Yates, J. R., r., and Winzeler, E. A. (2004). Global analysis of transcript and protein levels across the plasmodium falciparum life cycle. *Genome Res*, 14(11):2308–18.

Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., and Winzeler, E. A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, 301(5639):1503–8.

Lee, M. C. S., Lindner, S. E., Lopez-Rubio, J.-J., and Llinás, M. (2019). Cutting back malaria: Crispr/cas9 genome editing of plasmodium. *Briefings in Functional Genomics*, 18(5):281–289.

Lee, R. A., Puig, H., Nguyen, P. Q., Angenent-Mari, N. M., Donghia, N. M., McGee, J. P., Dvorin, J. D., Klapperich, C. M., Pollock, N. R., and Collins, J. J. (2020). Ultrasensitive crispr-based diagnostic for field-applicable detection of plasmodium species in symptomatic and asymptomatic malaria. *Proc Natl Acad Sci U S A*, 117(41):25722–25731.

Lee, V. V., Judd, L. M., Jex, A. R., Holt, K. E., Tonkin, C. J., Ralph, S. A., Sessions, P. F. d., and Lindner, S. (2021). Direct nanopore sequencing of mrna reveals landscape of transcript isoforms in apicomplexan parasites. *mSystems*, 6(2):e01081–20.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). Upset: Visualization of intersecting sets. *IEEE transactions on visualization and computer graphics*, 20(12):1983–1992.

Li, F., Sonbuchner, L., Kyes, S. A., Epp, C., and Deitsch, K. W. (2008). Nuclear non-coding rnas are transcribed from the centromeres of plasmodium falciparum and are associated with centromeric chromatin. *J Biol Chem*, 283(9):5692–8.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G. P. D. P. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079.

Li, J., Jin, K., Li, M., Mathy, N. W., Gong, A.-Y., Deng, S., Martins, G. A., Sun, M., Strauss-Soukup, J. K., and Chen, X.-M. (2021a). A host cell long noncoding rna nr_033736 regulates type i interferon-mediated gene transcription and modulates intestinal epithelial anti-cryptosporidium defense. *PLoS Pathog*, 17(1):e1009241.

Li, X., Zhou, B., Chen, L., Gou, L.-T., Li, H., and Fu, X.-D. (2017). Grid-seq reveals the global rna–chromatin interactome. *Nature Biotechnology*, 35(10):940–950.

Li, Y., Baptista, R. P., and Kissinger, J. C. (2020). Noncoding rnas in apicomplexan parasites: an update. *Trends Parasitol*, 36(10):835–849.

Li, Y., Baptista, R. P., Sateriale, A., Striepen, B., and Kissinger, J. C. (2021b). Analysis of long non-coding rna in cryptosporidium parvum reveals significant stage-specific qntisense transcription. *Frontiers in Cellular and Infection Microbiology*, 10:608298.

Liang, X., Boonhok, R., Siddiqui, F. A., Xiao, B., Li, X., Qin, J., Min, H., Jiang, L., Cui, L., and Miao, J. (2022). A leak-free inducible crispri/a system for gene functional studies in plasmodium falciparum. *Microbiol Spectr*, 10(3):e0278221.

Liao, Q., Shen, J., Liu, J., Sun, X., Zhao, G., Chang, Y., Xu, L., Li, X., Zhao, Y., Zheng, H., Zhao, Y., and Wu, Z. (2014). Genome-wide identification and functional annotation of plasmodium falciparum long noncoding rnas from rna-seq data. *Parasitol Res*, 113(4):1269–81.

Liao, Y., Smyth, G. K., and Shi, W. (2013). featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930.

Lindner, S. E., Swearingen, K. E., Harupa, A., Vaughan, A. M., Sinnis, P., Moritz, R. L., and Kappe, S. H. (2013). Total and putative surface proteomics of malaria parasite salivary gland sporozoites. *Mol Cell Proteomics*, 12(5):1127–43.

Liu, C., Bai, B., Skogerbø, G., Cai, L., Deng, W., Zhang, Y., Bu, D., Zhao, Y., and Chen, R. (2005). Noncode: an integrated knowledge database of non-coding rnas. *Nucleic Acids Res*, 33(Database issue):D112–5.

Liu, S. J., Horlbeck, M. A., Cho, S. W., Birk, H. S., Malatesta, M., He, D., Attenello, F. J., Villalta, J. E., Cho, M. Y., Chen, Y., Mandegar, M. A., Olvera, M. P., Gilbert, L. A., Conklin, B. R., Chang, H. Y., Weissman, J. S., and Lim, D. A. (2017). Crispri-based genome-scale identification of functional long noncoding rna loci in human cells. *Science*, 355(6320).

Liu, Y., Cao, Z., Wang, Y., Guo, Y., Xu, P., Yuan, P., Liu, Z., He, Y., and Wei, W. (2018). Genome-wide screening for functional long noncoding rnas in human cells by cas9 targeting of splice sites. *Nature Biotechnology*, 36(12):1203–1210.

Llinás, M., Bozdech, Z., Wong, E. D., Adai, A. T., and DeRisi, J. L. (2006). Comparative whole genome transcriptome analysis of three plasmodium falciparum strains. *Nucleic Acids Res*, 34(4):1166–1173.

Lodde, V., Floris, M., Muroni, M. R., Cucca, F., and Idda, M. L. (2022). Non-coding rnas in malaria infection. *Wiley Interdiscip Rev RNA*, 13(3):e1697.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol*, 15(12):550.

Lu, F., Jiang, H., Ding, J., Mu, J., Valenzuela, J. G., Ribeiro, J. M. C., and Su, X.-z. (2007). cdna sequences reveal considerable gene prediction inaccuracy in the plasmodium falciparum genome. *BMC Genomics*, 8(1):255.

Lu, J., Tong, Y., Pan, J., Yang, Y., Liu, Q., Tan, X., Zhao, S., Qin, L., and Chen, X. (2016). A redesigned crispr/cas9 system for marker-free genome editing in plasmodium falciparum. *Parasites & Vectors*, 9(1):198.

Luo, G. Z., Wang, F., Weng, X., Chen, K., Hao, Z., Yu, M., Deng, X., Liu, J., and He, C. (2016). Characterization of eukaryotic dna n(6)-methyladenine by a highly sensitive restriction enzyme-assisted sequencing. *Nat Commun*, 7:11301.

López-Barragán, M. J., Lemieux, J., Quiñones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., and Su, X.-z. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of plasmodium falciparum. *BMC Genomics*, 12(1):587.

López-Jiménez, E., Rojas, A. M., and Andrés-León, E. (2018). *RNA sequencing and Prediction Tools for Circular RNAs Analysis*, pages 17–33. Springer Singapore, Singapore.

López-Urrutia, E., Bustamante Montes, L. P., Ladrón de Guevara Cervantes, D., Pérez-Plasencia, C., and Campos-Parra, A. D. (2019). Crosstalk between long non-coding rnas, micro-rnas and mrnas: deciphering molecular mechanisms of master regulators in cancer. *Frontiers in Oncology*, 9.

Machado, M., Steinke, S., and Ganter, M. (2021). Plasmodium reproduction, cell size, and transcription: how to cope with increasing dna content? *Frontiers in Cellular and Infection Microbiology*, 11.

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(5787):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(2):e1000767.

Marin-Mogollon, C., Salman, A. M., Koolen, K. M. J., Bolscher, J. M., van Pul, F. J. A., Miyazaki, S., Imai, T., Othman, A. S., Ramesar, J., van Gemert, G. J., Kroeze, H., Chevalley-Maurel, S., Franke-Fayard, B., Sauerwein, R. W., Hill, A. V. S., Dechering, K. J., Janse, C. J., and Khan, S. M. (2019). A p. falciparum nf54 reporter line expressing mcherry-luciferase in gametocytes, sporozoites, and liver-stages. *Front Cell Infect Microbiol*, 9:96.

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

Markus, B., Bell, G., Lorenzi, H., and Lourido, S. (2019). Optimizing systems for cas9 expression in toxoplasma gondii. *mSphere*, 4(3):00386–19.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):3.

Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L.-L., Chen, R., Dean, C., Dinger, M. E., Fitzgerald, K. A., Gingeras, T. R., Guttman, M., Hirose, T., Huarte, M., Johnson, R., Kanduri, C., Kapranov, P., Lawrence, J. B., Lee, J. T., Mendell, J. T., Mercer, T. R., Moore, K. J., Nakagawa, S., Rinn, J. L., Spector, D. L., Ulitsky, I., Wan, Y., Wilusz, J. E., and Wu, M. (2023). Long non-coding rnas: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology*.

McCarty, N. S., Graham, A. E., Studená, L., and Ledesma-Amaro, R. (2020). Multiplexed crispr technologies for gene editing and transcriptional regulation. *Nature Communications*, 11(1):1281.

McGlincy, N. J., Meacham, Z. A., Reynaud, K. K., Muller, R., Baum, R., and Ingolia, N. T. (2021). A genome-scale crispr interference guide library enables comprehensive phenotypic profiling in yeast. *BMC Genomics*, 22(1):205.

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K., and Guttman, M. (2015). The xist lncrna interacts directly with sharp to silence transcription through hdac3. *Nature*, 521(7551):232–6.

McLean, K. J. and Niles, J. C. (2022). A transcription factor helps plasmodium falciparum gametocytogenesis take shape. *Trends in Parasitology*, 38(9):722–723.

McManus, K. F., Taravella, A. M., Henn, B. M., Bustamante, C. D., Sikora, M., and Cornejo, O. E. (2017). Population genetic analysis of the darc locus (duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genetics*, 13(3):e1006560.

Meissner, M., Breinich, M. S., Gilson, P. R., and Crabb, B. S. (2007). Molecular genetic tools in toxoplasma and plasmodium: achievements and future needs. *Curr Opin Microbiol*, 10(4):349–356.

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(8):2980–2985.

Menard, K. L., Haskins, B. E., Colombo, A. P., and Denkers, E. Y. (2018). Toxoplasma gondii manipulates expression of host long noncoding rna during intracellular infection. *Sci Rep*, 8(1):15017.

Mercer, T. R., Clark, M. B., Crawford, J., Brunck, M. E., Gerhardt, D. J., Taft, R. J., Nielsen, L. K., Dinger, M. E., and Mattick, J. S. (2014). Targeted sequencing for gene discovery and quantification using rna captureseq. *Nature Protocols*, 9(5):989–1009.

Milner, D. A., J. (2018). Malaria pathogenesis. *Cold Spring Harb Perspect Med*, 8(1):a025569.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. (2020). Pfam: The protein families database. *Nucleic Acids Res*, 49(D1):D412–D419.

Miyazaki, S., Yang, A. S. P., Geurten, F. J. A., Marin-Mogollon, C., Miyazaki, Y., Imai, T., Kolli, S. K., Ramesar, J., Chevalley-Maurel, S., Salman, A. M., van Gemert, G. A., van Waardenburg, Y. M., Franke-Fayard, B., Hill, A. V. S., Sauerwein, R. W., Janse, C. J., and Khan, S. M. (2020). Generation of novel plasmodium falciparum nf135 and nf54 lines expressing fluorescent reporter proteins under the control of strong and constitutive promoters. *Front Cell Infect Microbiol*, 10:270.

Mogollon, C. M., van Pul, F. J. A., Imai, T., Ramesar, J., Chevalley-Maurel, S., de Roo, G. M., Veld, S. A. J., Kroeze, H., Franke-Fayard, B. M. D., Janse, C. J., and Khan, S. M. (2016). Rapid generation of marker free p. falciparum fluorescent reporter lines using modified crispr/cas9 constructs and selection protocol. *PLoS One*, 11(12):e0168362–e0168362.

Mohring, F., Hart, M. N., Rawlinson, T. A., Henrici, R., Charleston, J. A., Diez Benavente, E., Patel, A., Hall, J., Almond, N., Campino, S., Clark, T. G., Sutherland, C. J., Baker, D. A., Draper, S. J., and Moon, R. W. (2019). Rapid and iterative genome editing in the malaria parasite plasmodium knowlesi provides new tools for p. vivax research. *Elife*, 8:e45829.

Molina-Cruz, A., Zilversmit, M. M., Neafsey, D. E., Hartl, D. L., and Barillas-Mury, C. (2016). Mosquito vectors and the globalization of plasmodium falciparum malaria. *Annual Review of Genetics*, 50(1):447–465.

Moon, R. W., Hall, J., Rangkuti, F., Ho, Y. S., Almond, N., Mitchell, G. H., Pain, A., Holder, A. A., and Blackman, M. J. (2013). Adaptation of the genetically tractable malaria pathogen plasmodium knowlesi to continuous culture in human erythrocytes. *Proc Natl Acad Sci U S A*, 110(2):531–536.

Moraes Barros, R. R., Straimer, J., Sa, J. M., Salzman, R. E., Melendez-Muniz, V. A., Mu, J., Fidock, D. A., and Wellems, T. E. (2015). Editing the plasmodium vivax genome, using zinc-finger nucleases. *J Infect Dis*, 211(1):125–129.

Mourier, T., Carret, C., Kyes, S., Christodoulou, Z., Gardner, P. P., Jeffares, D. C., Pinches, R., Barrell, B., Berriman, M., Griffiths-Jones, S., Ivens, A., Newbold, C., and Pain, A. (2008). Genome-wide discovery and verification of novel structured rnas in plasmodium falciparum. *Genome Res*, 18(2):281–92.

Méndez-Mancilla, A., Wessels, H. H., Legut, M., Kadina, A., Mabuchi, M., Walker, J., Robb, G. B., Holden, K., and Sanjana, N. E. (2022). Chemically modified guide rnas enhance crispr-cas13 knockdown in human cells. *Cell Chem Biol*, 29(2):321–327.e4.

Müller, K., Matuschewski, K., and Silvie, O. (2011). The puf-family rna-binding protein puf2 controls sporozoite conversion to liver stages in the malaria parasite. *PLoS One*, 6(5):e19860.

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., and Prasanth, K. V. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA*, 18(8):1487–1499.

Nasamu, A. S., Falla, A., Pasaje, C. F. A., Wall, B. A., Wagner, J. C., Ganesan, S. M., Goldfless, S. J., and Niles, J. C. (2021). An integrated platform for genome engineering and gene expression perturbation in plasmodium falciparum. *Sci Rep*, 11(1):342.

Nessel, T., Beck, J. M., Rayatpisheh, S., Jami-Alahmadi, Y., Wohlschlegel, J. A., Goldberg, D. E., and Beck, J. R. (2020). Exp1 is required for organisation of exp2 in the intraerythrocytic malaria parasite vacuole. *Cell Microbiol*, 22(5):e13168.

Ng, C. L. and Fidock, D. A. (2019). Plasmodium falciparum in vitro drug resistance selections and gene editing. *Methods Mol Biol*, 2013:123–140.

Nishi, T., Shinzawa, N., Yuda, M., and Iwanaga, S. (2021). Highly efficient crispr/cas9 system in plasmodium falciparum using cas9-expressing parasites and a linear donor template. *Sci Rep*, 11(1):18501.

Nissan, H., Ukawuba, I., and Thomson, M. (2021). Climate-proofing a malaria eradication strategy. *Malar J*, 20(1):190.

Nkrumah, L. J., Muhle, R. A., Moura, P. A., Ghosh, P., Hatfull, G. F., Jacobs, W. R., J., and Fidock, D. A. (2006). Efficient site-specific integration in plasmodium falciparum chromosomes mediated by mycobacteriophage bxb1 integrase. *Nat Methods*, 3(8):615–21.

Noonpakdee, W., Pothikasikorn, J., Nimitsantiwong, W., and Wilairat, P. (2003). Inhibition of plasmodium falciparum proliferation in vitro by antisense oligodeoxynucleotides against malarial topoisomerase ii. *Biochem Biophys Res Commun*, 302(4):659–64.

Oberstaller, J., Otto, T. D., Rayner, J. C., and Adams, J. H. (2021). Essential genes of the parasitic apicomplexa. *Trends parasitol*, 37(4):304–316.

Oduola, A. M., Milhous, W. K., Weatherly, N. F., Bowdre, J. H., and Desjardins, R. E. (1988). Plasmodium falciparum: induction of resistance to mefloquine in cloned strains by continuous drug exposure in vitro. *Exp Parasitol*, 67(2):354–360.

Ogwan'g, R. A., Mwangi, J. K., Githure, J., Were, J. B., Roberts, C. R., and Martin, S. K. (1993). Factors affecting exflagellation of in vitro-cultivated plasmodium falciparum gametocytes. *Am J Trop Med Hyg*, 49(1):25–29.

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

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

Pazos Obregón, F., Soto, P., Lavín, J. L., Cortázar, A. R., Barrio, R., Aransay, A. M., and Cantera, R. (2018). Cluster locator, online analysis and visualization of gene clustering. *Bioinformatics*, 34(19):3377–3379.

Pennisi, E. (2012). Genomics. encode project writes eulogy for junk dna. *Science*, 337(6099):1159, 1161.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time rt-pcr. *Nucleic Acids Res*, 29(9):e45.

Pfander, C., Anar, B., Schwach, F., Otto, T. D., Brochet, M., Volkmann, K., Quail, M. A., Pain, A., Rosen, B., Skarnes, W., Rayner, J. C., and Billker, O. (2011). A scalable pipeline for highly effective genetic modification of a malaria parasite. *Nat Methods*, 8(12):1078–82.

Plewka, P., Thompson, A., Szymanski, M., Nuc, P., Knop, K., Rasinska, A., Bialkowska, A., Szweykowska-Kulinska, Z., Karlowski, W. M., and Jarmolowski, A. (2018). A stable trna-like molecule is generated from the long noncoding rna gut15 in arabidopsis. *RNA Biol*, 15(6):726–738.

Ponnudurai, T., Leeuwenberg, A. D., and Meuwissen, J. H. (1981). Chloroquine sensitivity of isolates of plasmodium falciparum adapted to in vitro culture. *Trop Geogr Med*, 33(1):50–4.

Ponting, C. P. and Haerty, W. (2022). Genome-wide analysis of human long noncoding rnas: a provocative review. *Annual Review of Genomics and Human Genetics*, 23(1):153–172.

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

Ponts, N., Harris, E. Y., Lonardi, S., and Le Roch, K. G. (2011). Nucleosome occupancy at transcription start sites in the human malaria parasite: a hard-wired evolution of virulence? *Infect Genet Evol*, 11(4):716–24.

Portugaliza, H. P., Llorà-Batlle, O., Rosanas-Urgell, A., and Cortés, A. (2019). Reporter lines based on the gexp02 promoter enable early quantification of sexual conversion rates in the malaria parasite plasmodium falciparum. *Sci Rep*, 9(1):14595.

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., and Grummt, I. (2015). Lncrna khps1 regulates expression of the proto-oncogene sphk1 via triplex-mediated changes in chromatin structure. *Mol Cell*, 60(4):626–636.

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(8):e73783.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing crispr as an rna-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–83.

Quansah, E., Chen, Y., Yang, S., Wang, J., Sun, D., Zhao, Y., Chen, M., Yu, L., and Zhang, C. (2023). Crispr-cas13 in malaria parasite: Diagnosis and prospective gene function identification. *Frontiers in Microbiology*, 14:1076947.

Quinlan, A. R. and Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

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

Radfar, A., Méndez, D., Moneriz, C., Linares, M., Marín-García, P., Puyet, A., Diez, A., and Bautista, J. M. (2009). Synchronous culture of plasmodium falciparum at high parasitemia levels. *Nature Protocols*, 4(12):1899–1915.

Rajaram, K., Liu, H. B., and Prigge, S. T. (2020). Redesigned tetr-aptamer system to control gene expression in plasmodium falciparum. *mSphere*, 5(4):00457–20.

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

Rapaport, E., Misiura, K., Agrawal, S., and Zamecnik, P. (1992). Antimalarial activities of oligodeoxynucleotide phosphorothioates in chloroquine-resistant plasmodium falciparum. *Proc Natl Acad Sci U S A*, 89(18):8577–80.

Razavi Vakhshourpour, S., Nateghpour, M., Shahrokhi, N., Motevalli Haghi, A., Mohebali, M., and Hanifian, H. (2022). Potential of rh5 antisense on plasmodium falciparum proliferation abatement. *Iran J Parasitol*, 17(4):525–534.

Read, M. and Hyde, J. E. (1993). Simple in vitro cultivation of the malaria parasite plasmodium falciparum (erythrocytic stages) suitable for large-scale preparations. *Methods Mol Biol*, 21:43–55.

Ribeiro, J. M., Garriga, M., Potchen, N., Crater, A. K., Gupta, A., Ito, D., and Desai, S. A. (2018). Guide rna selection for crispr-cas9 transfections in plasmodium falciparum. *Int J Parasitol*, 48(11):825–832.

Rinn, J. L. and Chang, H. Y. (2020). Long noncoding rnas: molecular modalities to organismal functions. *Annual Review of Biochemistry*, 89(1):283–308.

Ritchie, M. R. H. (2019). Malaria. *Our World in Data*.

Rocamora, F., Zhu, L., Liong, K. Y., Dondorp, A., Miotto, O., Mok, S., and Bozdech, Z. (2018). Oxidative stress and protein damage responses mediate artemisinin resistance in malaria parasites. *PLoS Pathog*, 14(3):e1006930.

Ross, L. S., Dhingra, S. K., Mok, S., Yeo, T., Wicht, K. J., Kümpornsin, K., Takala-Harrison, S., Witkowski, B., Fairhurst, R. M., Ariey, F., Menard, D., and Fidock, D. A. (2018). Emerging southeast asian pfcrt mutations confer plasmodium falciparum resistance to the first-line antimalarial piperaquine. *Nat Commun*, 9(1):3314.

Rovira-Graells, N., Crowley, V. M., Bancells, C., Mira-Martínez, S., Ribas de Pouplana, L., and Cortés, A. (2015). Deciphering the principles that govern mutually exclusive expression of plasmodium falciparum clag3 genes. *Nucleic Acids Res*, 43(17):8243–57.

RTS, S. C. T. P. (2015). Efficacy and safety of rts,s/as01 malaria vaccine with or without a booster dose in infants and children in africa: final results of a phase 3, individually randomised, controlled trial. *Lancet*, 386(9988):31–45.

Rudlaff, R. M., Kraemer, S., Marshman, J., and Dvorin, J. D. (2020). Three-dimensional ultrastructure of plasmodium falciparum throughout cytokinesis. *PLoS Pathog*, 16(6):e1008587.

Ruiz, J. L., Tena, J. J., Bancells, C., Cortés, A., Gómez-Skarmeta, J. L., and Gómez-Díaz, E. (2018). Characterization of the accessible genome in the human malaria parasite plasmodium falciparum. *Nucleic Acids Res*, 46(18):9414–9431.

Ruiz-Orera, J. and Albà, M. M. (2019). Translation of small open reading frames: roles in regulation and evolutionary innovation. *Trends Genet*, 35(3):186–198.

Russell, A. J. C., Sanderson, T., Bushell, E., Talman, A. M., Anar, B., Girling, G., Hunziker, M., Kent, R. S., Martin, J. S., Metcalf, T., Montandon, R., Pandey, V., Pardo, M., Roberts, A. B., Sayers, C., Schwach, F., Choudhary, J. S., Rayner, J. C., Voet, T., Modrzynska, K. K., Waters, A. P., Lawniczak, M. K. N., and Billker, O. (2023). Regulators of male and female sexual development are critical for the transmission of a malaria parasite. *Cell Host Microbe*, 31(2):305–319.e10.

Rutledge, G. G., Böhme, U., Sanders, M., Reid, A. J., Cotton, J. A., Maiga-Ascofare, O., Djimdé, A. A., Apinjoh, T. O., Amenga-Etego, L., Manske, M., Barnwell, J. W., Renaud, F., Ollomo, B., Prugnolle, F., Anstey, N. M., Auburn, S., Price, R. N., McCarthy, J. S., Kwiatkowski, D. P., Newbold, C. I., Berriman, M., and Otto, T. D. (2017). Plasmodium malariae and p. ovale genomes provide insights into malaria parasite evolution. *Nature*, 542(7639):101–104.

Sakamoto, H., Thiberge, S., Akerman, S., Janse, C. J., Carvalho, T. G., and Ménard, R. (2005). Towards systematic identification of plasmodium essential genes by transposon shuttle mutagenesis. *Nucleic Acids Res*, 33(20):e174–e174.

Salcedo-Amaya, A. M., van Driel, M. A., Alako, B. T., Trelle, M. B., van den Elzen, A. M., Cohen, A. M., Janssen-Megens, E. M., van de Vegte-Bolmer, M., Selzer, R. R., Iniguez, A. L., Green, R. D., Sauerwein, R. W., Jensen, O. N., and Stunnenberg, H. G. (2009). Dynamic histone h3 epigenome marking during the intraerythrocytic cycle of plasmodium falciparum. *Proc Natl Acad Sci U S A*, 106(24):9655–60.

Saraf, A., Cervantes, S., Bunnik, E. M., Ponts, N., Sardiu, M. E., Chung, D. W., Prudhomme, J., Varberg, J. M., Wen, Z., Washburn, M. P., Florens, L., and Le Roch, K. G. (2016). Dynamic and combinatorial landscape of histone modifications during the intraerythrocytic developmental cycle of the malaria parasite. *J Proteome Res*, 15(8):2787–801.

Sarma, K., Levasseur, P., Aristarkhov, A., and Lee, J. T. (2010). Locked nucleic acids (lnas) reveal sequence requirements and kinetics of xist rna localization to the x chromosome. *Proc Natl Acad Sci U S A*, 107(51):22196–201.

Scherf, A., Wahlgren, M., Moll, K., and Kaneko, A. (2013). Methods in malaria research. *EVIMalaR*, 6th edition.

Schudoma, C. (2018). mdg.py.

Serrano-Durán, R., López-Farfán, D., and Gómez-Díaz, E. (2022). Epigenetic and epitranscriptomic gene regulation in plasmodium falciparum and how we can use it against malaria. *Genes*, 13(10):1734.

Shaner, N. C., Lambert, G. G., Chammas, A., Ni, Y., Cranfill, P. J., Baird, M. A., Sell, B. R., Allen, J. R., Day, R. N., Israelsson, M., Davidson, M. W., and Wang, J. (2013). A bright monomeric green fluorescent protein derived from branchiostoma lanceolatum. *Nat Methods*, 10(5):407–409.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B. J. (2016). Global mapping of human rna-rna interactions. *Mol Cell*, 62(4):618–626.

Shaw, P. J., Kaewprommal, P., Wongsombat, C., Ngampiw, C., Taechalertpaisarn, T., Kamchonwongpaisan, S., Tongsima, S., and Piriyapongsa, J. (2022). Transcriptomic complexity of the human malaria parasite plasmodium falciparum revealed by long-read sequencing. *PLoS One*, 17(11):e0276956.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). Seqkit: a cross-platform and ultrafast toolkit for fasta/q file manipulation. *PLoS One*, 11(10):e0163962.

Sheokand, P. K., Narwal, M., Thakur, V., and Mohmmed, A. (2021). Glms mediated knock-down of a phospholipase expedite alternate pathway to generate phosphocholine required for phosphatidylcholine synthesis in plasmodium falciparum. *Biochemical Journal*, 478(18):3429–3444.

Shinzawa, N., Nishi, T., Hiyoshi, F., Motooka, D., Yuda, M., and Iwanaga, S. (2020). Improvement of crispr/cas9 system by transfecting cas9-expressing plasmodium berghei with linear donor template. *Commun Biol*, 3(1):426.

Siegel, T. N., Hon, C.-C., Zhang, Q., Lopez-Rubio, J.-J., Scheidig-Benatar, C., Martins, R. M., Sismeiro, O., Coppée, J.-Y., and Scherf, A. (2014). Strand-specific rna-seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in plasmodium falciparum. *BMC Genomics*, 15(1):150.

Sierra-Miranda, M., Delgadillo, D. M., Mancio-Silva, L., Vargas, M., Villegas-Sepulveda, N., Martinez-Calvillo, S., Scherf, A., and Hernandez-Rivas, R. (2012). Two long non-coding rnas generated from subtelomeric regions accumulate in a novel perinuclear compartment in plasmodium falciparum. *Mol Biochem Parasitol*, 185(1):36–47.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, 7:539.

Silvie, O., Briquet, S., Müller, K., Manzoni, G., and Matuschewski, K. (2014). Post-transcriptional silencing of uis4 in plasmodium berghei sporozoites is important for host switch. *Molecular Microbiology*, 91(6):1200–1213.

Simantov, K., Goyal, M., and Dzikowski, R. (2022). Emerging biology of noncoding rnas in malaria parasites. *PLoS Pathog*, 18(7):e1010600.

Simon, M. D., Wang, C. I., Kharchenko, P. V., West, J. A., Chapman, B. A., Alekseyenko, A. A., Borowsky, M. L., Kuroda, M. I., and Kingston, R. E. (2011). The genomic binding sites of a noncoding rna. *Proc Natl Acad Sci U S A*, 108(51):20497–502.

Singer, M., Marshall, J., Heiss, K., Mair, G. R., Grimm, D., Mueller, A.-K., and Frischknecht, F. (2015). Zinc finger nuclease-based double-strand breaks attenuate malaria parasites and reveal rare microhomology-mediated end joining. *Genome Biol*, 16(1):249.

Sinha, A., Hughes, K. R., Modrzynska, K. K., Otto, T. D., Pfander, C., Dickens, N. J., Religa, A. A., Bushell, E., Graham, A. L., Cameron, R., Kafsack, B. F. C., Williams, A. E., Llinas, M., Berriman, M., Billker, O., and Waters, A. P. (2014). A cascade of dna-binding proteins for sexual commitment and development in plasmodium. *Nature*, 507(7491):253–257.

Siregar, J. E., Kurisu, G., Kobayashi, T., Matsuzaki, M., Sakamoto, K., Mi-ichi, F., Watanabe, Y., Hirai, M., Matsuoka, H., Syafruddin, D., Marzuki, S., and Kita, K. (2015). Direct evidence for the atovaquone action on the plasmodium cytochrome bc1 complex. *Parasitol Int*, 64(3):295–300.

Siwo, G. H., Tan, A., Button-Simons, K. A., Samarakoon, U., Checkley, L. A., Pinapati, R. S., and Ferdig, M. T. (2015). Predicting functional and regulatory divergence of a drug resistance transporter gene in the human malaria parasite. *BMC Genomics*, 16(1):115.

Skinner-Adams, T. S., Hawthorne, P. L., Trenholme, K. R., and Gardiner, D. L. (2003). Gateway™ vectors for plasmodium falciparum transfection. *Trends Parasitol*, 19(1):17–18.

Smargon, A. A., Cox, D. B. T., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., Abudayyeh, O. A., Essletzbichler, P., Shmakov, S., Makarova, K. S., Koonin, E. V., and Zhang, F. (2017). Cas13b is a type vi-b crispr-associated rna-guided rnase differentially regulated by accessory proteins csx27 and csx28. *Mol Cell*, 65(4):618–630.e7.

Smith, T. G., Walliker, D., and Ranford-Cartwright, L. C. (2002). Sexual differentiation and sex determination in the apicomplexa. *Trends Parasitol*, 18(7):315–323.

SnapGene (2022). Snapgene software.

Sorber, K., Dimon, M. T., and DeRisi, J. L. (2011). Rna-seq analysis of splicing in plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic Acids Res*, 39(9):3820–35.

Spillman, N. J., Beck, J. R., Ganesan, S. M., Niles, J. C., and Goldberg, D. E. (2017). The chaperonin tric forms an oligomeric complex in the malaria parasite cytosol. *Cellular Microbiology*, 19(6):e12719.

Sridhar, B., Rivas-Astroza, M., Nguyen, T. C., Chen, W., Yan, Z., Cao, X., Hebert, L., and Zhong, S. (2017). Systematic mapping of rna-chromatin interactions in vivo. *Curr Biol*, 27(4):602–609.

Stanway, R. R., Bushell, E., Chiappino-Pepe, A., Roques, M., Sanderson, T., Franke-Fayard, B., Caldelari, R., Golomingi, M., Nyonda, M., Pandey, V., Schwach, F., Chevalley, S., Ramesar, J., Metcalf, T., Herd, C., Burda, P.-C., Rayner, J. C., Soldati-Favre, D., Janse, C. J., Hatzimanikatis, V., Billker, O., and Heussler, V. T. (2019). Genome-scale identification of essential metabolic processes for targeting the plasmodium liver stage. *Cell*, 179(5):1112–1128.e26.

Stephens, M. (2016). False discovery rates: a new deal. *Biostatistics*, 18(2):275–294.

Stojic, L., Lun, A. T., Mangei, J., Mascalchi, P., Quarantotti, V., Barr, A. R., Bakal, C., Marioni, J. C., Gergely, F., and Odom, D. T. (2018). Specificity of rnai, lna and crispri as loss-of-function methods in transcriptional analysis. *Nucleic Acids Res*, 46(12):5950–5966.

Stokes, B. H., Dhingra, S. K., Rubiano, K., Mok, S., Straimer, J., Gnädig, N. F., Deni, I., Schindler, K. A., Bath, J. R., Ward, K. E., Striepen, J., Yeo, T., Ross, L. S., Legrand, E., Ariey, F., Cunningham, C. H., Souleymane, I. M., Gansané, A., Nzoumbou-Boko, R., Ndayikunda, C., Kabanywanyi, A. M., Uwimana, A., Smith, S. J., Kolley, O., Ndounga, M., Warsame, M., Leang, R., Nosten, F., Anderson, T. J., Rosenthal, P. J., Ménard, D., and Fidock, D. A. (2021). Plasmodium falciparum k13 mutations in africa and asia impact artemisinin resistance and parasite fitness. *Elife*, 10:e66277.

Straimer, J., Lee, M. C., Lee, A. H., Zeitler, B., Williams, A. E., Pearl, J. R., Zhang, L., Rebar, E. J., Gregory, P. D., Llinás, M., Urnov, F. D., and Fidock, D. A. (2012). Site-specific genome editing in plasmodium falciparum using engineered zinc-finger nucleases. *Nat Methods*, 9(10):993–998.

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

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 6(7):e21800.

Talman, A. M., Blagborough, A. M., and Sinden, R. E. (2010). A plasmodium falciparum strain expressing gfp throughout the parasite's life-cycle. *PLoS One*, 5(2):e9156.

Tavares, R. C. A., Pyle, A. M., and Somarowthu, S. (2019). Phylogenetic analysis with improved parameters reveals conservation in lncrna structures. *Journal of Molecular Biology*, 431(8):1592–1603.

Tehlan, A., Bhowmick, K., Kumar, A., Subbarao, N., and Dhar, S. K. (2022). The tetrameric structure of plasmodium falciparum phosphoglycerate mutase is critical for optimal enzymatic activity. *Journal of Biological Chemistry*, 298(3):101713.

Theron, M., Cross, N., Cawkill, P., Bustamante, L. Y., and Rayner, J. C. (2018). An in vitro erythrocyte preference assay reveals that plasmodium falciparum parasites prefer type o over type a erythrocytes. *Sci Rep*, 8(1):8133.

Thommen, B. T., Passecker, A., Buser, T., Hitz, E., Voss, T. S., and Brancucci, N. M. B. (2022). Revisiting the effect of pharmaceuticals on transmission stage formation in the malaria parasite plasmodium falciparum. *Frontiers in Cellular and Infection Microbiology*, 12:802341.

Toenhake, C. G. and Bártfai, R. (2019). What functional genomics has taught us about transcriptional regulation in malaria parasites. *Briefings in Functional Genomics*, 18(5):290–301.

Toenhake, C. G., Fraschka, S. A., Vijayabaskar, M. S., Westhead, D. R., van Heeringen, S. J., and Bartfai, R. (2018). Chromatin accessibility-based characterization of the gene regulatory network underlying plasmodium falciparum blood-stage development. *Cell Host Microbe*, 23(4):557–569.e9.

Trager, W. and Jensen, J. B. (1976). Human malaria parasites in continuous culture. *Science*, 193(4254):673–675.

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., and Bubulya, P. A. (2010). The nuclear-retained noncoding rna malat1 regulates alternative splicing by modulating sr splicing factor phosphorylation. *Mol Cell*, 39(6):925–938.

Trotta, E. (2014). On the normalization of the minimum free energy of rnas by sequence length. *PLoS One*, 9(11):e113380.

Turner, A. W., Wong, D., Khan, M. D., Dreisbach, C. N., Palmore, M., and Miller, C. L. (2019). Multi-omics approaches to study long non-coding rna function in atherosclerosis. *Frontiers in Cardiovascular Medicine*, 6.

Tusting, L. S., Willey, B., Lucas, H., Thompson, J., Kafy, H. T., Smith, R., and Lindsay, S. W. (2013). Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis. *Lancet*, 382(9896):963–72.

Upadhyay, R., Bawankar, P., Malhotra, D., and Patankar, S. (2005). A screen for conserved sequences with biased base composition identifies noncoding rnas in the a–t rich genome of plasmodium falciparum. *Molecular and Biochemical Parasitology*, 144(2):149–158.

Usui, M., Prajapati, S. K., Ayanful-Torgby, R., Acquah, F. K., Cudjoe, E., Kakaney, C., Amponsah, J. A., Obboh, E. K., Reddy, D. K., Barbeau, M. C., Simons, L. M., Czesny, B., Raiciulescu, S., Olsen, C., Abuaku, B. K., Amoah, L. E., and Williamson, K. C. (2019). Plasmodium falciparum sexual differentiation in malaria patients is associated with host factors and gdv1-dependent genes. *Nature Communications*, 10(1):2140.

Valgardsdottir, R., Chiodi, I., Giordano, M., Rossi, A., Bazzini, S., Ghigna, C., Riva, S., and Biamonti, G. (2008). Transcription of satellite iii non-coding rnas is a general stress response in human cells. *Nucleic Acids Res*, 36(2):423–434.

van den Hoff, M. J., Moorman, A. F., and Lamers, W. H. (1992). Electroporation in 'intracellular' buffer increases cell survival. *Nucleic Acids Res*, 20(11):2902.

VanWye, J. D. and Haldar, K. (1997). Expression of green fluorescent protein in plasmodium falciparum. *Mol Biochem Parasitol*, 87(2):225–229.

Voorberg-van der Wel, A. M., Zeeman, A.-M., Nieuwenhuis, I. G., van der Werff, N. M., Klooster, E. J., Klop, O., Vermaat, L. C., Kumar Gupta, D., Dembele, L., Diagana, T. T., and Kocken, C. H. M. (2020). A dual fluorescent plasmodium cynomolgi reporter line reveals in vitro malaria hypnozoite reactivation. *Commun Biol*, 3:7.

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(9):915–918.

Walker, M. P. and Lindner, S. E. (2019). Ribozyme-mediated, multiplex crispr gene editing and crispr interference (crispri) in rodent-infectious plasmodium yoelii. *J Biol Chem*, 294(24):9555–9566.

Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T., Corcoran, L. M., Burkot, T. R., and Carter, R. (1987). Genetic analysis of the human malaria parasite plasmodium falciparum. *Science*, 236(4809):1661–1666.

Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., Jiang, Z., Xu, J., Liu, Q., and Cao, X. (2014). The stat3-binding long noncoding rna lnc-dc controls human dendritic cell differentiation. *Science*, 344(6181):310–313.

Wang, Y., Gong, A. Y., Ma, S., Chen, X., Strauss-Soukup, J. K., and Chen, X. M. (2017). Delivery of parasite cdg7_flc_0990 rna transcript into intestinal epithelial cells during cryptosporidium parvum infection suppresses host cell gene transcription through epigenetic mechanisms. *Cell Microbiol*, 19(11):12760.

Wanidworanun, C., Nagel, R. L., and Shear, H. L. (1999). Antisense oligonucleotides targeting malarial aldolase inhibit the asexual erythrocytic stages of plasmodium falciparum. *Molecular and Biochemical Parasitology*, 102(1):91–101.

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

Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z., and Sanjana, N. E. (2020). Massively parallel cas13 screens reveal principles for guide rna design. *Nature Biotechnology*, 38(6):722–727.

Wessler, S. R. (2005). Homing into the origin of the ap2 dna binding domain. *Trends Plant Sci*, 10(2):54–56.

WHO (2022). World malaria report. Report Licence: CC BY-NC-SA 3.0 IGO., World Health Organisation.

Williams, T. N. (2006). Human red blood cell polymorphisms and malaria. *Current Opinion in Microbiology*, 9(4):388–394.

Wilson, D. W., Crabb, B. S., and Beeson, J. G. (2010). Development of fluorescent plasmodium falciparum for in vitro growth inhibition assays. *Malar J*, 9(1):152.

Winkle, M., El-Daly, S. M., Fabbri, M., and Calin, G. A. (2021). Noncoding rna therapeutics - challenges and potential solutions. *Nature Reviews Drug Discovery*, 20(8):629–651.

Wong, L. S., Wei, L., Wang, G., Law, C. T., Tsang, F. H., Chin, W. C., Ng, I. O., and Wong, C. M. (2022). In vivo genome-wide crispr activation screening identifies functionally important long noncoding rnas in hepatocellular carcinoma. *Cell Mol Gastroenterol Hepatol*, 14(5):1053–1076.

Wu, H., Yin, Q.-F., Luo, Z., Yao, R.-W., Zheng, C.-C., Zhang, J., Xiang, J.-F., Yang, L., and Chen, L.-L. (2016). Unusual processing generates spa lncrnas that sequester multiple rna binding proteins. *Mol Cell*, 64(3):534–548.

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(4):973–977.

Xiao, B., Yin, S., Hu, Y., Sun, M., Wei, J., Huang, Z., Wen, Y., Dai, X., Chen, H., Mu, J., Cui, L., and Jiang, L. (2019). Epigenetic editing by crispr/dcas9 in plasmodium falciparum. *Proc Natl Acad Sci U S A*, 116(1):255–260.

Xing, J., Liu, H., Jiang, W., and Wang, L. (2021). Lncrna-encoded peptide: functions and predicting methods. *Frontiers in Oncology*, 10:622294.

Xu, C., Zhou, Y., Xiao, Q., He, B., Geng, G., Wang, Z., Cao, B., Dong, X., Bai, W., Wang, Y., Wang, X., Zhou, D., Yuan, T., Huo, X., Lai, J., and Yang, H. (2021). Programmable rna editing with compact crispr–cas13 systems from uncultivated microbes. *Nature Methods*, 18(5):499–506.

Xu, D., Cai, Y., Tang, L., Han, X., Gao, F., Cao, H., Qi, F., and Kapranov, P. (2020). A crispr/cas13-based approach demonstrates biological relevance of vlinc class of long non-coding rnas in anticancer drug response. *Sci Rep*, 10(1):1794.

Yang, M., Shang, X., Zhou, Y., Wang, C., Wei, G., Tang, J., Zhang, M., Liu, Y., Cao, J., and Zhang, Q. (2021). Full-length transcriptome analysis of plasmodium falciparum by single-molecule long-read sequencing. *Front Cell Infect Microbiol*, 11:631545.

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7sk small nuclear rna inhibits the cdk9/cyclin t1 kinase to control transcription. *Nature*, 414(6861):317–322.

Yao, R.-W., Wang, Y., and Chen, L.-L. (2019). Cellular functions of long noncoding rnas. *Nature Cell Biology*, 21(5):542–551.

Yeoh, L. M., Lee, V. V., McFadden, G. I., and Ralph, S. A. (2019). Alternative splicing in apicomplexan parasites. *mBio*, 10(1).

Yin, Q. F., Yang, L., Zhang, Y., Xiang, J. F., Wu, Y. W., Carmichael, G. G., and Chen, L. L. (2012). Long noncoding rnas with snorna ends. *Mol Cell*, 48(2):219–30.

Yin, S., Fan, Y., He, X., Wei, G., Wen, Y., Zhao, Y., Shi, M., Wei, J., Chen, H., Han, J., Jiang, L., and Zhang, Q. (2020). The cryptic unstable transcripts are associated with developmentally regulated gene expression in blood-stage plasmodium falciparum. *RNA Biol*, 17(6):828–842.

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J. L., De, S., Huarte, M., Zhan, M., Becker, K. G., and Gorospe, M. (2012). Lincrna-p21 suppresses target mrna translation. *Mol Cell*, 47(4):648–655.

Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T., and Kaneko, I. (2010). Transcription factor ap2-sp and its target genes in malarial sporozoites. *Mol Microbiol*, 75(4):854–63.

Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G. R., Janse, C. J., Waters, A. P., Kato, T., and Kaneko, I. (2009). Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol*, 71(6):1402–1414.

Zetsche, B., Gootenberg, J., Abudayyeh, O., Slaymaker, I., Makarova, K., Essletzbichler, P., Volz, S., Joung, J., van der Oost, J., Regev, A., Koonin, E., and Zhang, F. (2015). Cpf1 is a single rna-guided endonuclease of a class 2 crispr-cas system. *Cell*, 163(3):759–771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., and Gootenberg, J. S. (2017). Multiplex gene editing by crispr–cpf1 using a single crrna array. *Nature Biotechnology*, 35(1):31–34.

Zhang, M., Fennell, C., Ranford-Cartwright, L., Sakthivel, R., Gueirard, P., Meister, S., Caspi, A., Doerig, C., Nussenzweig, R. S., Tuteja, R., Sullivan, W. J., J., Roos, D. S., Fontoura, B. M., Ménard, R., Winzeler, E. A., and Nussenzweig, V. (2010). The plasmodium eukaryotic initiation factor-2alpha kinase ik2 controls the latency of sporozoites in the mosquito salivary glands. *J Exp Med*, 207(7):1465–1474.

Zhang, M., Mishra, S., Sakthivel, R., Rojas, M., Ranjan, R., Sullivan, W. J., J., Fontoura, B. M., Ménard, R., Dever, T. E., and Nussenzweig, V. (2012). Pk4, a eukaryotic initiation factor 2 alpha (eif2alpha) kinase, is essential for the development of the erythrocytic cycle of plasmodium. *Proc Natl Acad Sci U S A*, 109(10):3956–61.

Zhang, M., Wang, C., Otto, T. D., Oberstaller, J., Liao, X., Adapa, S. R., Udenze, K., Bronner, I. F., Casandra, D., Mayho, M., Brown, J., Li, S., Swanson, J., Rayner, J. C., Jiang, R. H. Y., and Adams, J. H. (2018). Uncovering the essential genes of the human malaria parasite plasmodium falciparum by saturation mutagenesis. *Science*, 360(6388):7847.

Zhang, Q., Siegel, T. N., Martins, R. M., Wang, F., Cao, J., Gao, Q., Cheng, X., Jiang, L., Hon, C. C., Scheidig-Benatar, C., Sakamoto, H., Turner, L., Jensen, A. T., Claes, A., Guizetti, J., Malmquist, N. A., and Scherf, A. (2014). Exonuclease-mediated degradation of nascent rna silences genes linked to severe malaria. *Nature*, 513(7518):431–435.

Zhang, W. W., Lypaczewski, P., and Matlashewski, G. (2017). Optimized crispr-cas9 genome editing for leishmania and its use to target a multigene family, induce chromosomal translocation, and study dna break repair mechanisms. *mSphere*, 2(1):00340–16.

Zhang, Z., Chen, J., Zhu, Z., Zhu, Z., Liao, X., Wu, J., Cheng, J., Zhang, X., Mei, H., and Yang, G. (2021). Crispr-cas13-mediated knockdown of lncrna-gacat3 inhibited cell proliferation and motility, and induced apoptosis by increasing p21, bax, and e-cadherin expression in bladder cancer. *Frontiers in Molecular Biosciences*, 7:627774.

Zhao, Y., Wang, F., Wang, C., Zhang, X., Jiang, C., Ding, F., Shen, L., and Zhang, Q. (2020). Optimization of crispr/cas system for improving genome editing efficiency in plasmodium falciparum. *Front Microbiol*, 11:625862.

Zhu, S., Li, W., Liu, J., Chen, C. H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., Yuan, P., Brown, M., Liu, X. S., and Wei, W. (2016). Genome-scale deletion screening of human long non-coding rnas using a paired-guide rna crispr-cas9 library. *Nat Biotechnol*, 34(12):1279–1286.

Appendix A

Additional information for *Chapter 3*

A.1 Supplemental data

The extended annotation table, which includes the annotation and features for each lncRNA can be accessed (Hoshizaki et al., 2022a) under Additional File 3.

Table A.1 Alignment of 5′ and 3′ homologous regions of the *pfpare* locus in *P. falciparum* strain sequences

Table A.1 Long-read RNA sequencing information

Dataset	N50 read length (bp)	Longest read (bp)	Total sequence (Mb)
Pf nanopore 1	1306	10245	375.7
Pf nanopore 2*	2008	12804	277.7

*(Lee et al., 2021).

A.2 Data accessibility

The data underlying this article are available at ArrayExpress (long-read RNA sequencing, E-MTAB-11766) and European Nucleotide Archive (short-read RNA sequencing, ERP104547). The annotation is available on PlasmoDB as a dataset that can be viewed on the genome browser or downloaded as a gff file.

A.3 Full author list and contributions

Johanna Hoshizaki (JH), Sophie Adjalley (SA), Vandana Thathy (VT), Kim Judge (KJ), Matthew Berriman (MB), Adam Reid (AR) and Marcus Lee (ML). JH, AR and ML conceived and designed the experiments. VT prepared the *P. falciparum* short-read samples for sequencing. SA prepared the *P. falciparum* samples for long-read sequencing, and KJ coordinated the sequencing completed by Sanger Scientific Operations. AR mapped the sequencing reads, JH performed the manual curation and generated the annotation. AR, MB and ML supervised the work. JH wrote the manuscript.

Dr Ross Waller and Dr Scott Chisholm completed the proteomics analysis of the putative coding lncRNAs.

A.4 Acknowledgements

I thank Dr Chris Newbold for supporting the generation of short-read sequencing. I thank Emma Betteridge, Alexander Dove and Sanger Scientific Operations for their assistance in generating the long-read RNA sequencing. I thank Dr Ulrike Böehme and Dr Lia Chappell for their guidance and expertise in manual curation and lncRNA annotation, respectively. I thank Dr Mengquan Yang and Dr Qingfeng Zhang for providing RNA transcript data. I also thank Dr Lia Chappell for sharing updated UTR annotations and short-read RNA sequencing for *P. falciparum* and *P. knowlesi*.

Appendix B

Additional information for *Chapter 4*

B.1 Supplemental data

Figure B.1 Alignment of 5′ and 3′ homologous regions of the *pfpare* locus in *P. falciparum* strain sequences
Figure B.2-B.6 Maps of the Cas9, PspCas13b, Conditional PspCas13b with and without U6psp and RfxCas13d plasmids
Table B.1 Guide RNAs used in *Chapter 4*
Table B.2 Primers used in *Chapter 4*

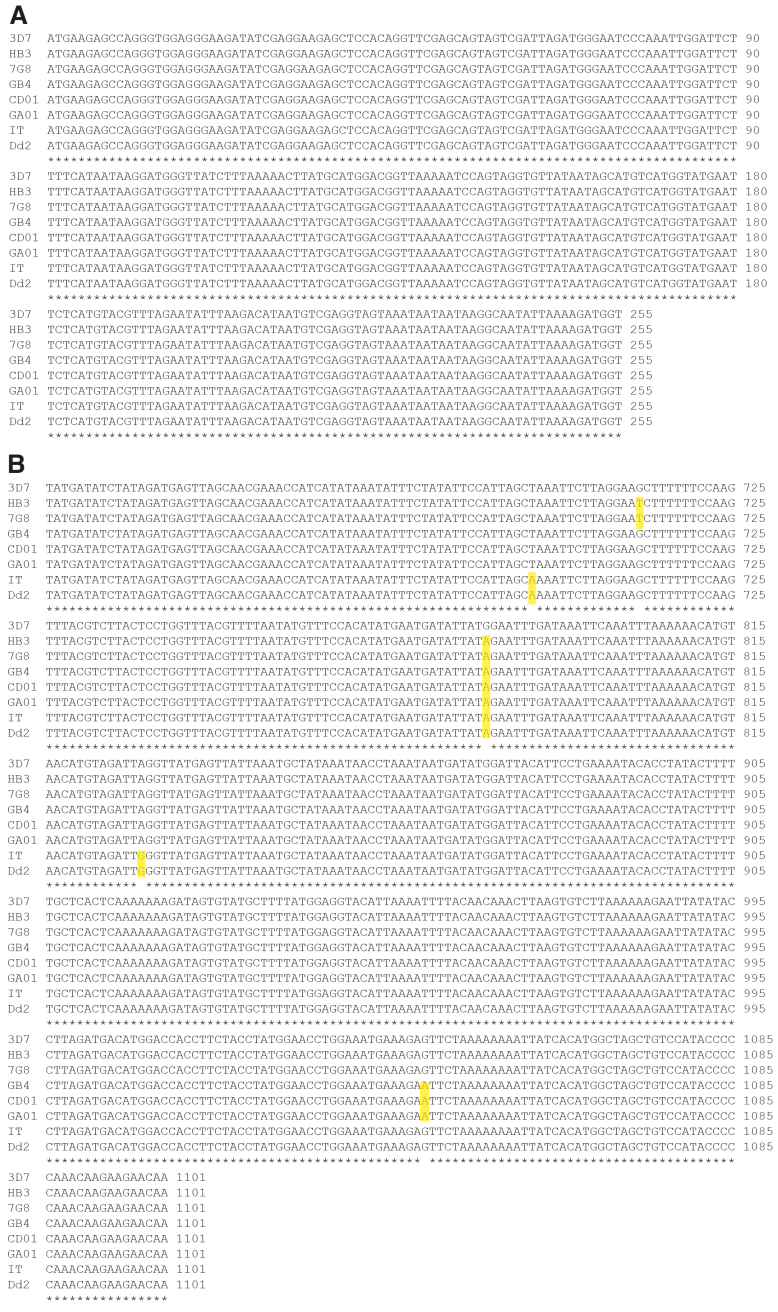


Fig. B.1 Alignment of 5′ and 3′ homologous regions of the *pfpare* locus in *P. falciparum* strain sequences. The DNA sequences of the 5′ (**A**) and 3′ (**B**) *pfpare* homologous regions used in pDC2-coCas9-*pare*-BSD-mNeonGreen were obtained for *P. falciparum* 3D7, HB3, 7G8, GB4, CD01, GA01, IT and Dd2 strains from PlasmoDB (Amos et al., 2021). Clustal Omega was used to create a multiple sequence alignment (Goujon et al., 2010; Sievers et al., 2011). Single nucleotide polymorphisms are highlighted with yellow boxes.

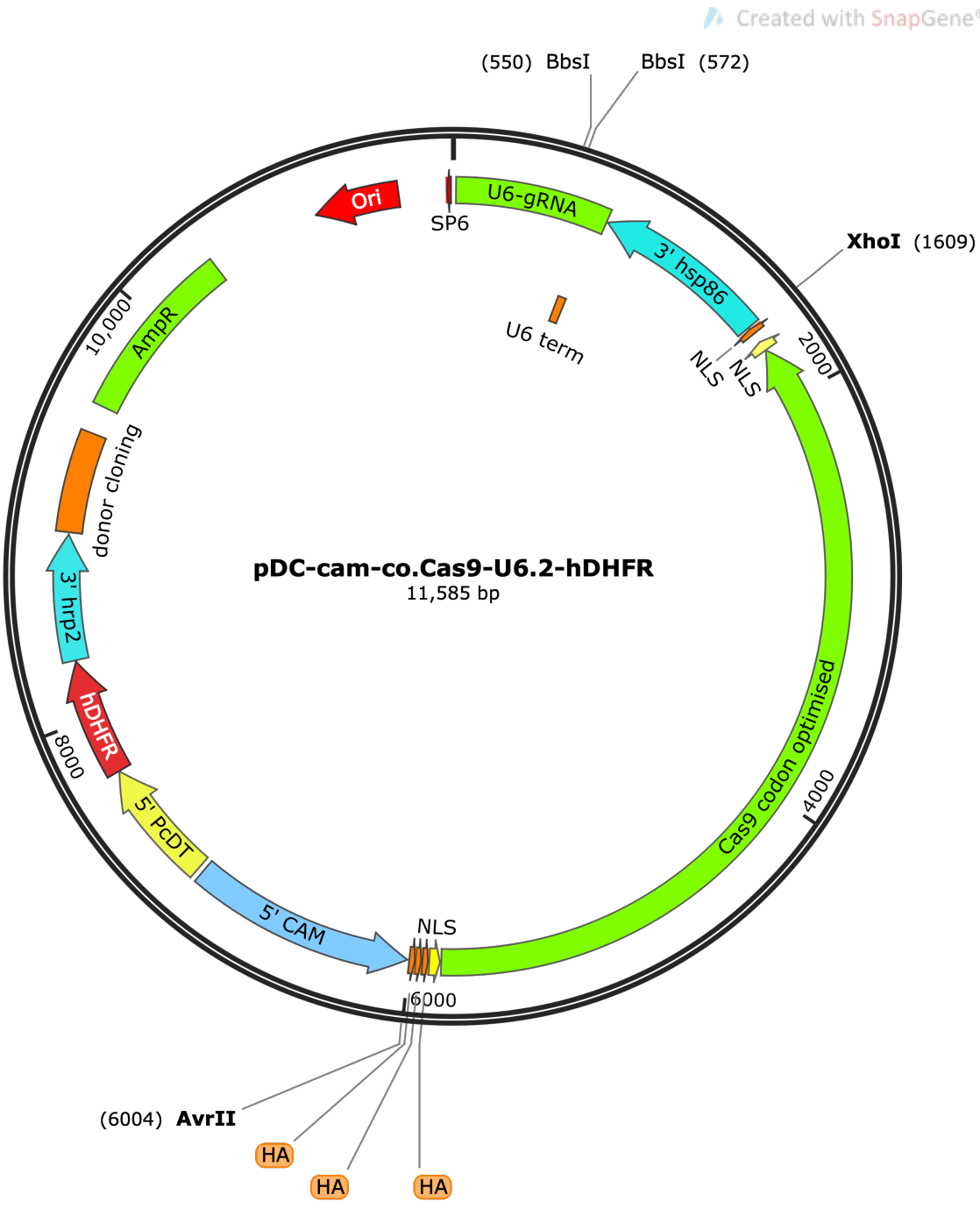


Fig. B.2 Map of CRISPR-Cas9 plasmid. Created in SnapGene (SnapGene, 2022).

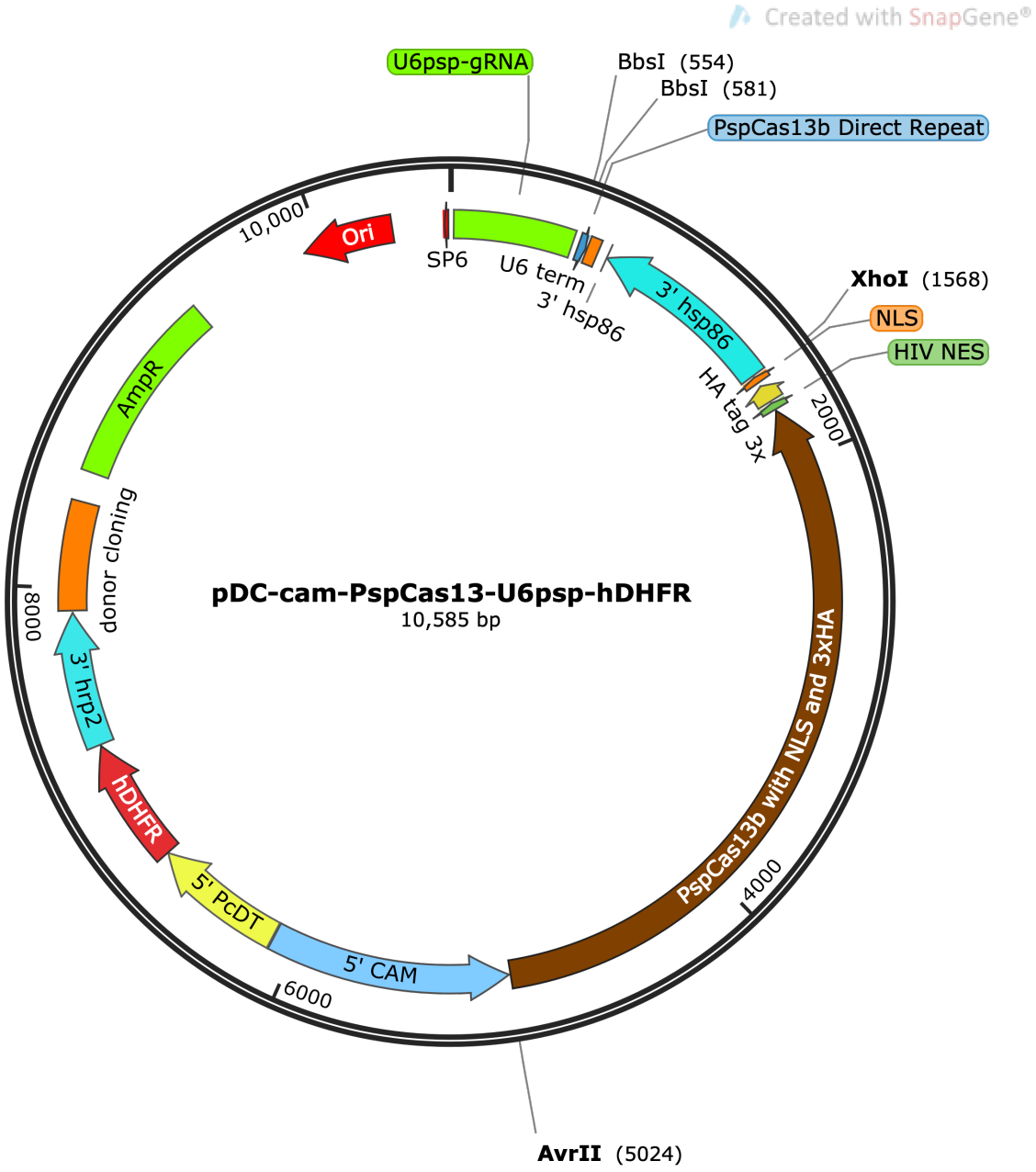


Fig. B.3 Map of CRISPR-PspCas13b plasmid. Created in SnapGene (SnapGene, 2022).

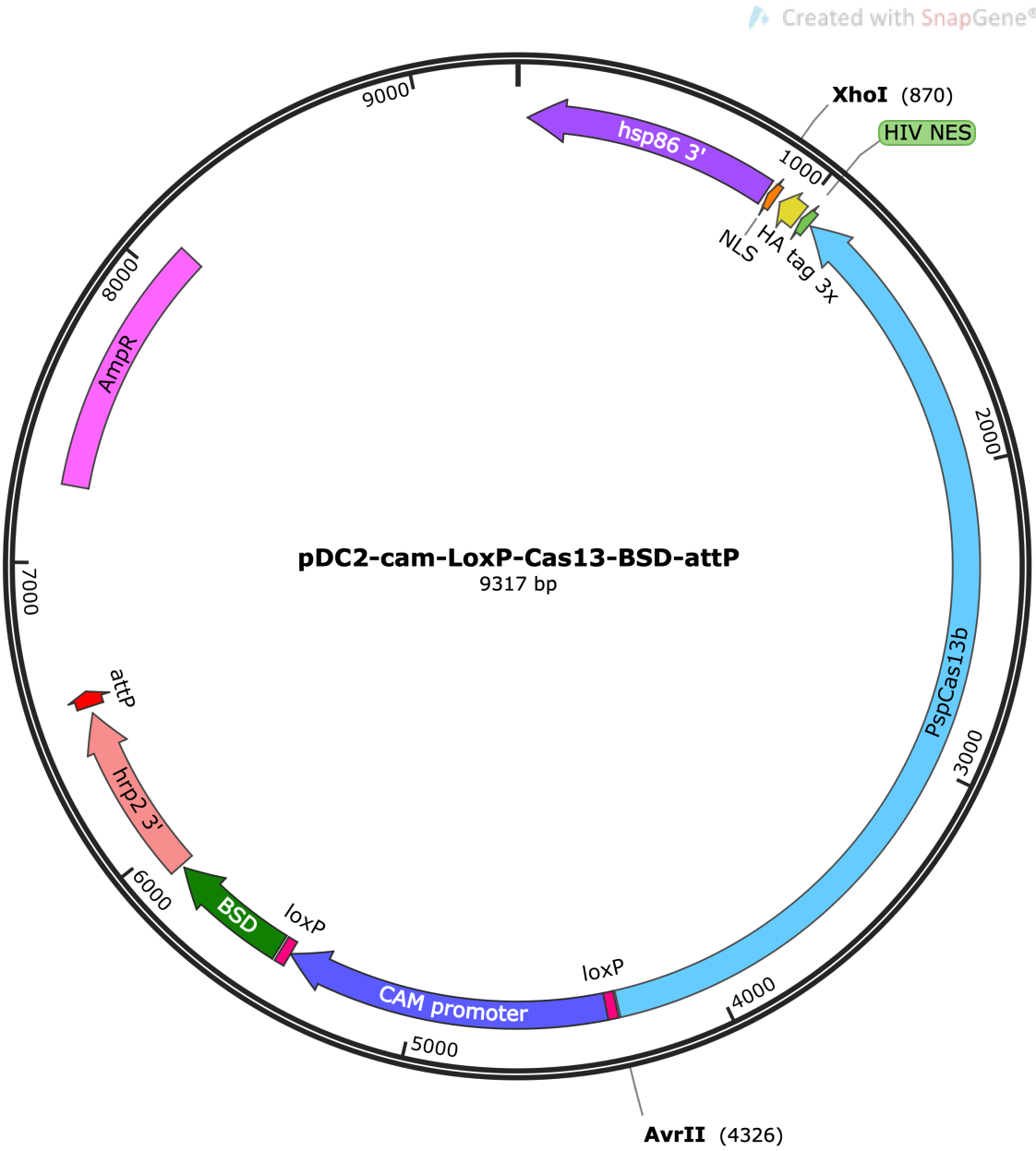


Fig. B.4 Map of conditional CRISPR-PspCas13b plasmid. Created in SnapGene (SnapGene, 2022).

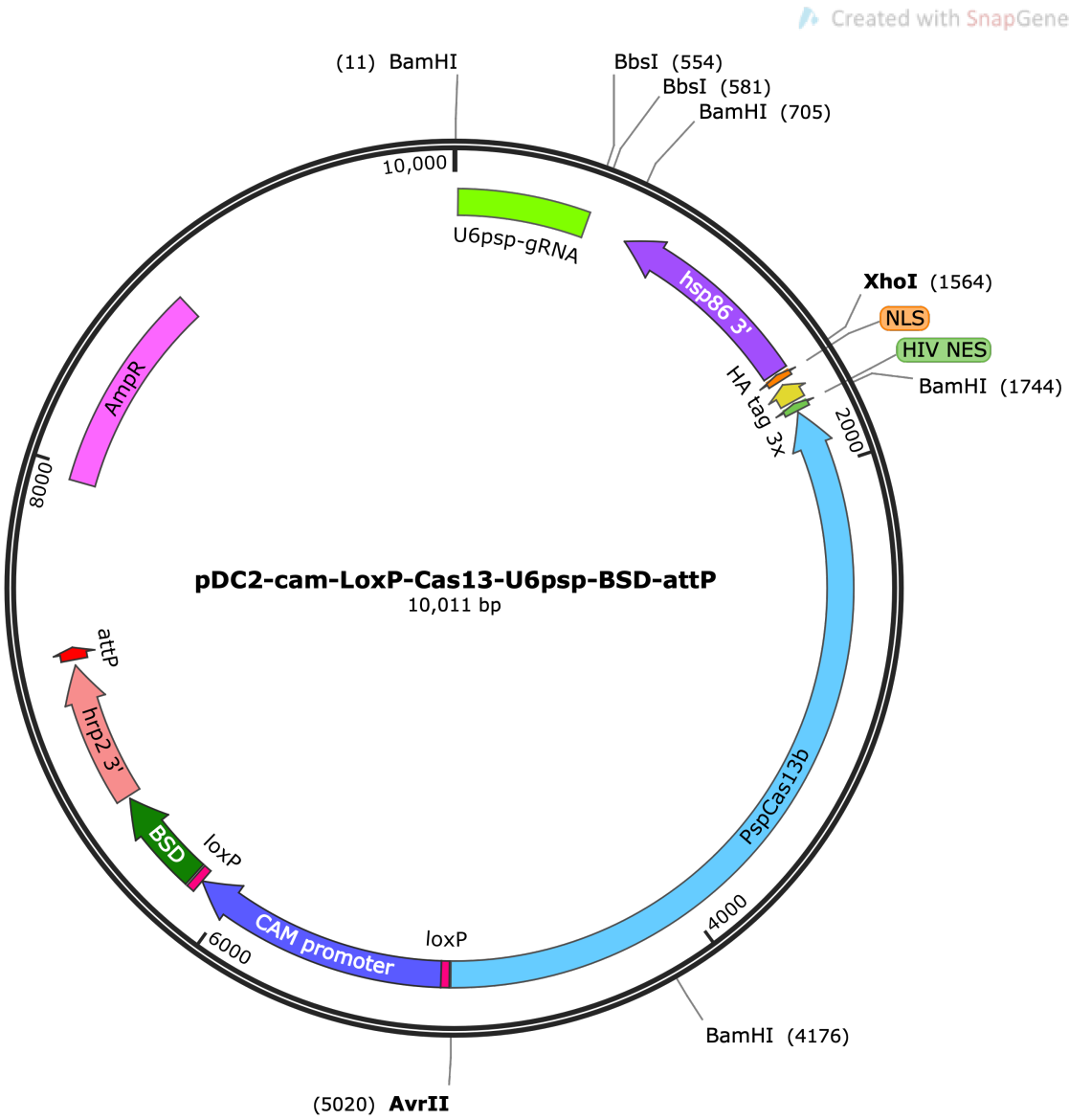


Fig. B.5 Map of conditional CRISPR-PspCas13b with U6 cassette plasmid. Created in SnapGene (SnapGene, 2022).

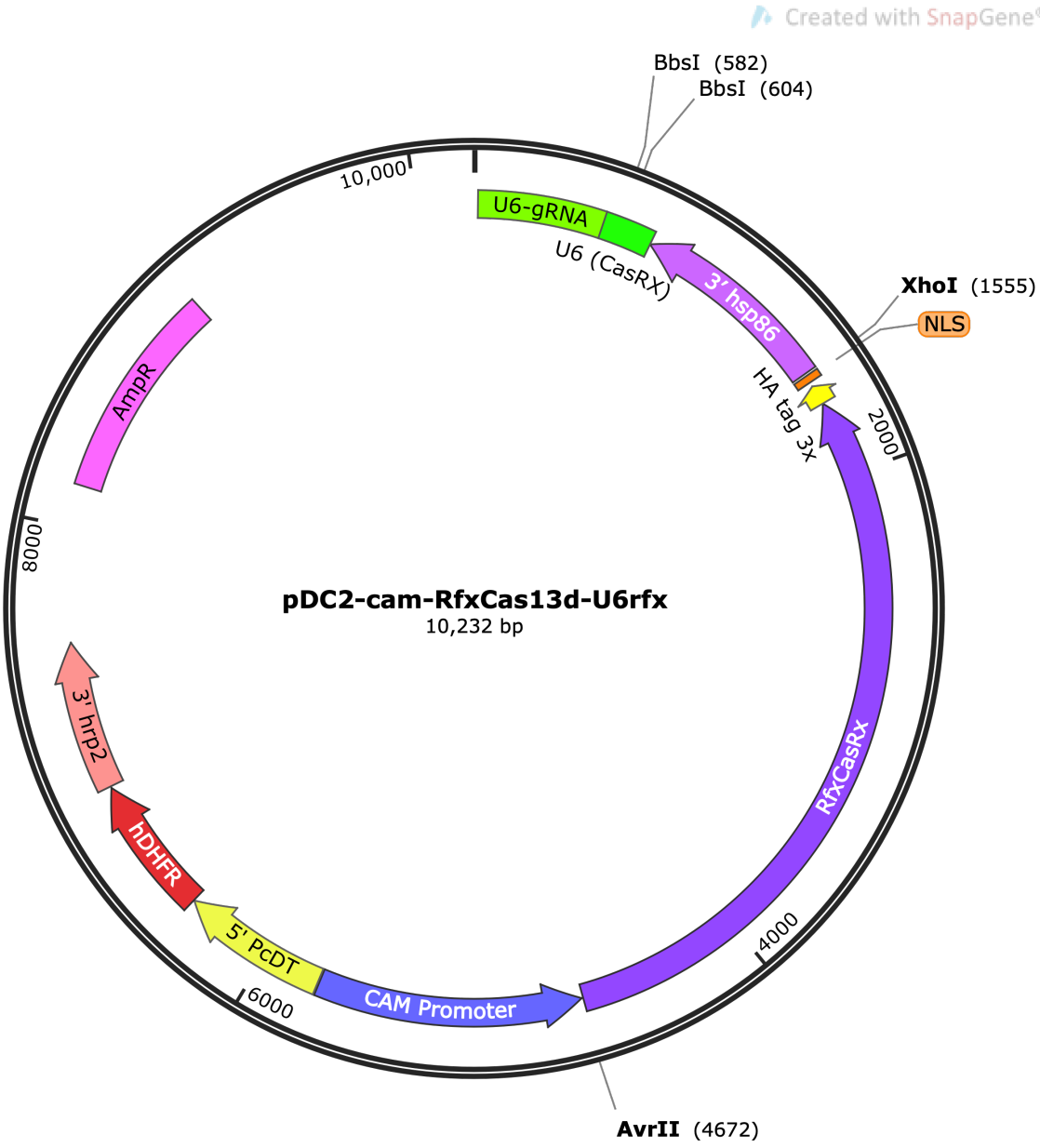


Fig. B.6 Map of CRISPR-RfxCas13d plasmid. Created in SnapGene (SnapGene, 2022).

Table B.1 Guide RNAs for *Chapter 4*

Enzyme	Target	Number	Sequence (5′-3′)
Cas13	*gfp*	1	CCCCGGTGAACAGCTCCTCGCCCTTGCTCA
		2	TGTGGCCGTTTACGTCGCCGTCCAGCTCGA
		3	TTGCCGTAGGTGGCATCGCCCTCGCCCTCG
		4	GTAGCGGCTGAAGCACTGCACGCCGTAGGT
		5	GTGCGCTCCTGGACGTAGCCTTCGGGCATG
		6	GGTTCACCAGGGTGTCGCCCTCGAACTTCA
		7	TGTTCTGCTGGTAGTGGTCGGCGAGCTGCA
		8	TCCCGGCGGCGGTCACGAACTCCAGCAGGA
	pfpare	1	TATTCCCATTTTCTTTCGTCACATTTCAAA
		2	TAAGGATGGGTTATCTTTAAAAACTTATGC
		3	CATAATGTCGAGGTAGTAAATAATAATAAG
		4	TTGAGGAGTTTAATAAAAATGGATATTCAT
		5	TTAAGAACCCATATAAGACAATTTGATGAT
		6	AAATCAAAAGATAATAATAATAAATTAAAT
	non-targeting (NT)	1	GTAATGCCTGGCTTGTCGACGCATAGTCTG
Cas9	PF3D7lncRNA_0019	1	AAAGAAAAAAATATGTGATA
		2	GTGGTAAATATATATTGTAT
		3	TTTATTTCCTCCTTTTTCTT
	PF3D7lncRNA_0337	1	GGTACTTACAGCGATTGAGA
		2	AGATGCATCGATGTGGTGAA
		3	TATACTGATTATGATAAACC
		4	ATAGGTAATCAAATTGGTGT
	PF3D7lncRNA_0567	1	CATAAGAATATATGTGAAGT
		2	TTATGCAATGCTTATCAATA
		3	ATGCTTATCAATATGGTTGT
		4	TATGGTTGTGGGTTTGTAGA
		5	ATGGAATTTCTTCTTTTGTT
	pfpare	1	GGACAGTCAGAAGGATGGAA

Different overhangs were added to the forward and reverse gRNA sequences depending on the plasmid: 5′-CACC-forward and 5′-CAAC-reverse for PspCas13b, 5′-AAAC and 5′-AAAA for RfxCas13d, and 5′-TATT and 5′-AAAC for Cas9.

Table B.2 Primers for *Chapter 4*

Name	Target	Sequence
p35	Cas9 U6 (f)	AAGCACCGACTCGGTGCCAC
p36	Other U6 cassettes (f)	GTTGTGTGGAATTGTGAGCGG
p37	CAM promoter (f)	TTTTAAACTAGAAAAGGAATAAC
p38	3'Hsp86 (r)	TATTGGGGTGATGATAAAATG
p261	5'PcDT (f)	CATTTGCTAGTAACAATTGTGTAGTGC
p282	3'hrp2 (f)	AACATATGTTAAATATTTATTTCTC
p283	pDC2 backbone (f)	AGGGTTATTGTCTCATGAGCGG
p675	5'PcDT (r)	CACGAGAAAAAGGGGATGCC
p1786	PspCas13b (f)	GTAGCGCCGGATCACTTTCT
p1787	PspCas13b (f)	GTACAGAAATGGATGAGGCT
p1788	PspCas13b (r)	GCTGATCTGCCTGTTCCTGG
p1797	PspCas13b for Gibson (f)	GTACATAAATATATTATATAACTCGAGACTCGAGTTACCCCCTTC
p1798	PspCas13b for Gibson (r)	GTATGCTATACGAAGTTATCCTAGGATGAACATCCCCGCTCTG
p1799	SDM of bbsI in BSD (f)	CATCCCCATCTCTGAGGACTACAGCGTCGCC
p1800	SDM of bbsI in BSD (r)	GGCGACGCTGTAGTCCTCAGAGATGGGGATG
p1998	5'HR PF3D7lncRNA_0337 (f)	CGAAAAGTGCCACCTGACGTCACATATATATGTGTATGACCTAAAAG
p1999	5'HR PF3D7lncRNA_0337 (r)	GAGGTACCGAGCTCGAATTCAGATACGAAGAAGGATAATAAC
p2000	3'HR PF3D7lncRNA_0337 (f)	GCGAATTAGCTAAGCATGCGGGCCCAGCATATCAAATATCATATACC
p2001	3'HR PF3D7lncRNA_0337 (r)	ATATGGGAATTTCCTTATAGGGCCCTCTATATTCATCATGTTTTG
p2012	5'HR PF3D7lncRNA_0567 (f)	CGAAAAGTGCCACCTGACGTCAGTATGCTATAAAAAAAATGA
p2013	5'HR PF3D7lncRNA_0567 (r)	GAGGTACCGAGCTCGAATTCTGAAAGCTATTTTTTCTACA
p2014	3'HR PF3D7lncRNA_0567 (f)	GCGAATTAGCTAAGCATGCGGGCCCTCATAAAACTATTACAAATGATTTATAAG
p2015	3'HR PF3D7lncRNA_0567 (r)	ATATGGGAATTTCCTTATAGGGCCCCATAATTTTTTCTACAATTTTTAATAAG
p2022	5'HR PF3D7lncRNA_0019 (f)	CGAAAAGTGCCACCTGACGTCTGTAAAGATTCATTTATTTCC
p2023	5'HR PF3D7lncRNA_0019 (r)	GAGGTACCGAGCTCGAATTCCTATATAATAAATGTTATACTCAATAATC
p2024	3'HR PF3D7lncRNA_0019 (f)	GCGAATTAGCTAAGCATGCGGGCCCAGGTAAATTTCAACATCTGA
p2025	3'HR PF3D7lncRNA_0019 (r)	ATATGGGAATTTCCTTATAGGGCCCACTGTCTACGAATAACCTAAAG
p2066	RfxCas13d for Gibson (f)	ACCTAATAGAAATATATCACCTAGGATGGC
p2067	RfxCas13d for Gibson (r)	CTCTTGGGTCCAGAACTTCCA
p2068	RfxCas13d (f)	GTATACACTTATGATGAGTTTAAA
p2069	RfxCas13d (f)	ATTTCGTAATATACAGATACTATATAG
p2070	RfxCas13d (f)	CGAATTAAAAGCATTAGCAG
p2071	RfxCas13d (f)	TTAGTAGTGTAACAAAATTATGTG
p2072	U6rfx for Gibson (f)	CATATTAAGTATATAATATTGAAACACCGAA
p2073	U6rfx for Gibson (r)	ATCAAATAGCATGCCTGCAG
p2074	5'HR PF3D7lncRNA_0337 (ex,f)	TCACATTGAATTGTAAATTATTA
p2075	3'HR PF3D7lncRNA_0337 (ex,r)	AAATATTTTAAGCCTGACTTTTTTAATCATTCTTC
p2076	5'HR PF3D7lncRNA_0567 (ex,f)	ATAATGTAGTAGAAAATTACATATATT
p2077	3'HR PF3D7lncRNA_0567 (ex,r)	GAAAAGTAATAGAAAATTTTCTTATACTTTTGGATGC
p2078	5'HR PF3D7lncRNA_0019 (ex,f)	TAAATGTAATAATATAATATACATATTACATA
p2079	3'HR PF3D7lncRNA_0019 (ex,r)	TCCTTCTCAAAAATAAATAAACTGATATAAAAATGTATA
p2115	BSD cassette (f)	ATAGTTGCCATTGGTAACGA
p2116	NeonGreen (f)	GGAACTTAACTTAAAATCAA
p2117	NeonGreen (f)	TGCCGATTGGTGCAGGTCAA
p2118	3'HR *pfpare* (r)	AGTAAGACGTAAACTTGGAA
p2170	LncRNA3 knockout (f)	TTATATTTGTAGAAAAAATAGC
p2171	LncRNA3 knockout (r)	ATCATTTGTAATAGTTTTATGA
p2172	Upstream lncRNA3 (ex,f)	ATAATGTAGTAGAAAATTACATATATTTT
p2173	Downstream lncRNA3 (ex,r)	GTAATAGAAAATTTTCTTATACTTTTGGATGC
p2174	5'UTR *pfpare* (ex,f)	TGCACTTGTTTTACATTTTTATATT
p2175	3'UTR *pfpare* (ex,r)	TGTAACATCACTAATTAATTTATTAA
p2176	NeonGreen (f)	AGAACCGAAAATATGAAGCT

f: forward; r:reverse; ex: external; HR: homology region.

B.2 Full author list and contributions

B.2.1 CRISPR-Cas13 development and Cas9-mediated lncRNA knock-out work

Authors: Johanna Hoshizaki (JH), Sophie Adjalley (SA) and Marcus Lee (ML). JH, SA and ML conceived and designed the experiments. SA generated the PspCas13b plasmid. JH generated the conditional-CRISPR-PspCas13b and RfxCas13d constructs and performed the experiments and analysis. SA and ML supervised the work. Thank you to Mukul Rawat for completing western blots.

B.2.2 mNeonGreen fluorescent lines

Authors: Johanna Hoshizaki (JH), Hannah Jagoe (HJ) and Marcus Lee (ML). JH and ML conceived and designed the experiments. HJ generated the *pfpare*-targeting CRISPR-Cas9 construct (pDC2-coCas9-*pare*-2A-BSD). JH generated the construct for mNeonGreen integration. JH performed the experiments and analysis. ML supervised the work. JH wrote the manuscript.

Appendix C

Additional information for *Chapter 5*

C.1 Supplemental data

Figure C.1 Map of CRISPR-dLbCpf1Sir2a plasmid
Figure C.2 PCA plot of the RNA-Seq timecourse experiment
Figure C.3 Genes antisense to lncRNAs or sharing their bidirectional promoter are not differentially expressed in lncRNA-disrupted mutants
Table C.1 Discarded lncRNA targets due to the lack of specific gRNAs
Table C.2 Guide RNAs used in *Chapter 5*
Table C.3 qPCR primers used in *Chapter 5*
Table C.4 Extended list of differentially expressed genes in lncRNA-disrupted mutants
Table C.5 GO term analysis of differentially expressed genes in lncRNA-disrupted mutants

Fig. C.1 Map of CRISPR-dLbCpf1Sir2a plasmid. Created in SnapGene (SnapGene, 2022).

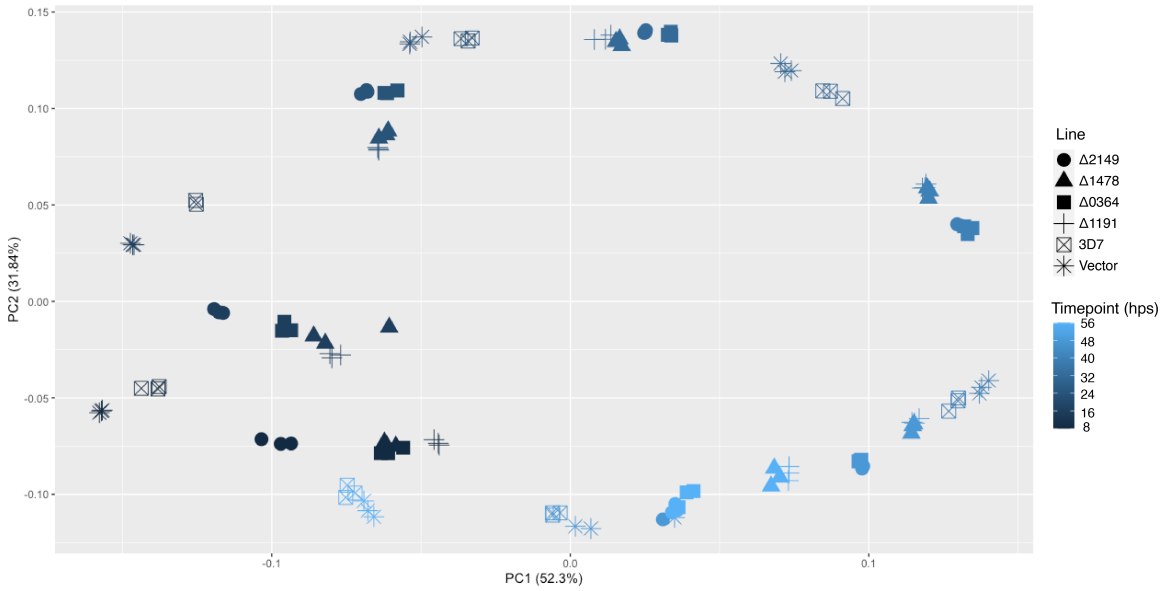


Fig. C.2 PCA plot of the RNA-Seq timecourse experiment. Parasite lines are denoted by shape and the timepoint post-synchronisation is denoted by colour.

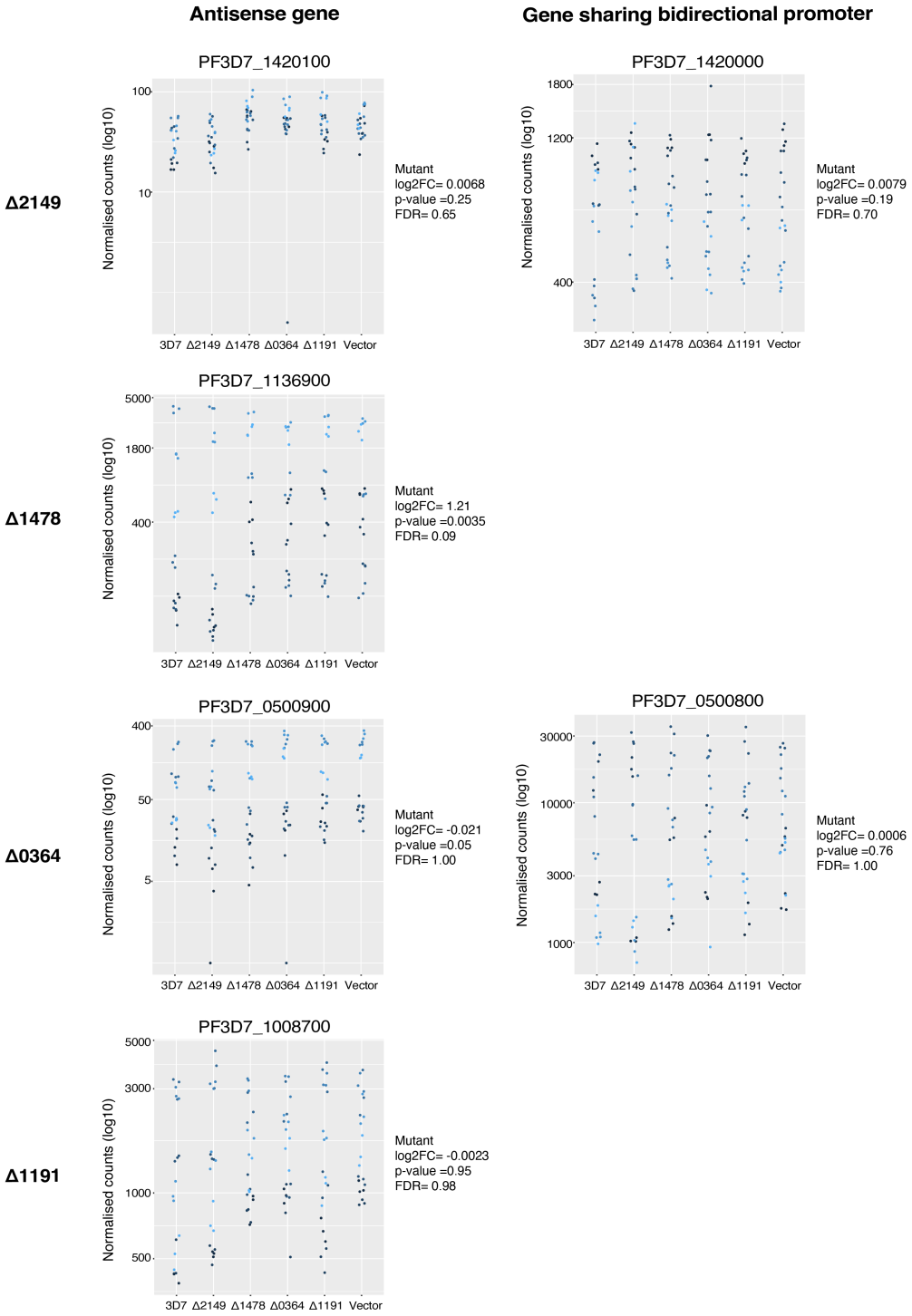


Fig. C.3 Genes antisense to lncRNAs or sharing their bidirectional promoter are not differentially expressed in lncRNA-disrupted mutants. Each row shows the antisense gene and gene sharing a bidirectional promoter for the targeted lncRNA in a lncRNA-disrupted mutant. Normalised expression counts are shown for all queried lines. The log2FC, p-values and FDR values are shown for the mutant line of interest versus 3D7, calculated during differential expression analysis.

Table C.1 Discarded lncRNA targets due to the lack of specific gRNAs

Interest group	LncRNA	Phenotype
Association with drug resistance	PF3D7lncRNA_0624	BI-2536, Suloctidil & MMV007181
	PF3D7lncRNA_0778	(multiple mutations) BI-2536
Asexual blood-stage specific	PF3D7lncRNA_0261	Late schizonts
Gametocyte-stage specific	PF3D7lncRNA_1872	Males
Long-length	PF3D7lncRNA_1523	6417 bp
Intergenic	PF3D7lncRNA_0064	Intergenic
	PF3D7lncRNA_0270	Intergenic
	PF3D7lncRNA_0624	Intergenic
	PF3D7lncRNA_0632	Intergenic
	PF3D7lncRNA_1143	Intergenic

Table C.2 Guide RNAs used in *Chapter 5*

Target	Number	Sequence (5'-3')
PF3D7lncRNA_0270	1	tacacaatatattttctagtatt
	2	ttcatatataattacttgcgtaa
	3	taattaaaatgggactataaaaa
	4	tataaataaattacttcttgaca
PF3D7lncRNA_0322	1	tagatgtgttgaattatttattt
	2	tcccttgttcatattctatattt
	3	ttctatcttatattctttctttt
	4	atttcatatattgtataccatgt
PF3D7lncRNA_0652	1	tattctatatataatattgtact
	2	ttgtatgtaagaggagaataata
	3	atgaaaaaataatatatattagg
	4	tgtagtacacaaaaattcatatt
PF3D7lncRNA_0706	1	cagaataaataaaagatgaaaga
	2	gttataatgttttcataaaaaaa
	3	ttgcctctttatgatattaaata
	4	ctacattttgctgaaaggcattt
PF3D7lncRNA_1059	1	cccccaatatatagtaaatctaa
	2	tgtatatatctttaatatgttat
	3	tatgtatttaattttatttaatt
	4	tgaataatattttgaagaaatag
PF3D7lncRNA_1633	1	tatatatattcttccacttgttg
	2	tgtagtaatactcgattaaagat
	3	taaatgtttactattatttacat
	4	tgtaacaatattattacgcgtaa
PF3D7lncRNA_1901	1	tgttatatatataaagagatgc
	2	gcatatcatcaatattaaaaata
	3	tatatggaaatttcttttctctt
	4	atcttaattatgtacatataaaa
PF3D7lncRNA_2081	1	taacattatttattttaatgtac
	2	taaattggtaaaaaatatttcta
	3	atatacccgtgaatataaatata
	4	atatatcaagtaatatattcttg
PF3D7lncRNA_0329	1	acgtgtttattataatatatatt
	2	tataaaaacaatatagcgatttt
	3	tacatttaatatgtatagaaaat
	4	aaaattaccgcaatttatatata
PF3D7lncRNA_0355	1	tgaacatattataattaataaaa
	2	ataatctaaacatatgtaaatac
	3	aacccaaagtctggtatgacaaa
	4	tgtgtattataaatgtggattaa
PF3D7lncRNA_0594	1	ataacattcataaatacaaacct
	2	agatatgtcataatgaataacat
	3	tgcatattgaagttatataaagt
	4	aaaaataagttcattaaatgtgt
PF3D7lncRNA_0710	1	gaaaaaaaaaaccttacaatat
	2	caatattatagaaatatattttc
	3	atatttaacataaaactttgta
	4	caagataaaattattatacaaaa
PF3D7lncRNA_0750	1	aaaaagtcaagaataaacatatt
	2	aaatgaacaagcttgattggagc
	3	ttatatttgattaatacatatgc
	4	cgtgggtacatacacctcatata
PF3D7lncRNA_0942	1	aaaaagcgtttattcatttatat
	2	gaaggtgtattggatatttatta
	3	tcttaaaattacataagaatttt
	4	aacctgaaattaaagcccatata
PF3D7lncRNA_1168	1	ttcaagttattttatatattata
	2	aatattatctaataaatgtttat
	3	tatcacatatatgtatatattat
	4	acatttaaactgttgatatttta
PF3D7lncRNA_2112	1	atatccgtgtatattgtatgata
	2	actatatgtttcactataattgt
	3	aaaatataaaaatcataaaaata
	4	ttacatcatatataaaattcaaa

Target	Number	Sequence (5'-3')
PF3D7lncRNA_2247	1	tttgaaaatttgaacgaatgaat
	2	tgtgtatacaataaaagaaatat
	3	cctttcaaatgtaaatgaaaaaa
	4	tacttgaattaaatttaaaataa
PF3D7lncRNA_2285	1	atgcatacatttaattaaaaatt
	2	ttaaagaggatgtaaataaatta
	3	agattaaagaaataattataact
	4	tgtcttgtataaataatgaagat
PF3D7lncRNA_0123	1	atgtcgaaatatctttctgttca
	2	ttattgttaccgtgtatgattct
	3	tatattaataataatcaaagaca
	4	tatcgcaataactctaaaataga
PF3D7lncRNA_0324	1	tatatagaatctaatttatacaa
	2	tataacatatgaaattttaaagt
	3	aatattgcatgtcaaaacattaa
	4	tatatatgtatttaaaatttaat
PF3D7lncRNA_0441	1	aaaggaaagaaaaaaatcttaat
	2	taattcattctagagttatattt
	3	aaacaaaacatatatgaacatat
	4	ctcatatattttatatttgttat
PF3D7lncRNA_0536	1	aagatcatataaaaatatatatt
	2	ttaatatcttgacctttaaata
	3	ttattatgcacttgtttgctttt
	4	aatcttgttttcaaaaaaaagaa
PF3D7lncRNA_0765	1	taagcctagagcataaattcccc
	2	tttaatctaacagtttcatgatt
	3	aacaaaaatacaacagtataaaa
	4	aggcaattttaatatatatataa
PF3D7lncRNA_1081	1	tttgtgtggacacatatatataa
	2	tgtttgttatcacgtccaaaatg
	3	taacctcaaaattgttttcccat
	4	ttatgtccttatcaatattattc
PF3D7lncRNA_1295	1	acaattaatttgtattatatatg
	2	tatatataatatatgtggtacgc
	3	gtataattaatctttatgtaaaa
PF3D7lncRNA_1422	1	ttttgaggtaatgaataatttca
	2	aaaaaatatagtatggtcttatt
	3	tctatatttacacaggtatggat
	4	atatggacatatgataaaattat
PF3D7lncRNA_1429	1	ggagttcatgatatcaaagagtt
	2	tgtcctcatgtgtatgatgaaag
	3	aaataaataaatgactacatttt
PF3D7lncRNA_1639	1	attgaatataaaattttcttaca
	2	taaaaaaaccttaggaacaatta
	3	attataattttctgacatcttgg
PF3D7lncRNA_2266	1	tgttctcaacttgtatatttaaa
	2	taaattcattattagtatttatt
	3	ctaaaatctgttaaaaattataa
	4	agtggatgtagaagtaaataata
PF3D7lncRNA_0054	1	taatacacatgcaatgatataat
	2	atgaccatataattaatataagt
	3	aatgccagttgtcttatatatat
	4	cgataatattataaaagataaaa
PF3D7lncRNA_0349	1	gaaaatataatatccaagtttcc
	2	tcttttccttatcctattcattt
	3	ttgaaataacaaataaattgttt
	4	tttcctattttaatatttcatat
PF3D7lncRNA_0486	1	tcatatccaaataaaaatattag
	2	aatattctaataaaaaagtgaat
	3	ttattaaatataaatgaaataat
	4	tgattacattataaaaatatata
PF3D7lncRNA_0752	1	tctatattcctaatataatataa
	2	tactttggttattccttatgtt
	3	ctatataataaatattattatat
	4	ataaactgtgtggatatatttat

Target	Number	Sequence (5'-3')
PF3D7lncRNA_0940	1	tgaataaattataaattcatata
	2	tcttttattttatataataatat
	3	ttttatataattattaactcgta
PF3D7lncRNA_1090	1	tctatagattaataatattcgtt
	2	ttttaaaaaattaaatataaaca
	3	aataacaggtgtattttattata
	4	atacgaaaatatatagtaacata
PF3D7lncRNA_1150	1	ttattcaatgtgtatacatataa
	2	cttaaatataataattatatata
	3	ttatcatatttattattaaaata
	4	aaagaatttaatgccttttaaaa
PF3D7lncRNA_1216	1	tttaaatgtacttatcaactttt
	2	tttataaatatatacacgtggta
	3	tttcaagaattattattagaagt
	4	tatatgtttaattattatatatt
PF3D7lncRNA_1556	1	ttttatcttattcattttctttt
	2	cctttccttatcaggtaaattac
	3	atttgcgtagatatatatatttt
	4	taataatgaatactgacttcatt
PF3D7lncRNA_2172	1	ttatatattagtgactaatatat
	2	atattccttatatatatgtaaac
	3	tttaaaggaagaaaagcgattaa
	4	tgagatatgtgagaataaggtac
PF3D7lncRNA_1093	1	atatgtataatatttcaaaataa
	2	gatactttataataccatactaa
	3	taataatatattatttcatcata
	4	tattcttttcttttaataaacat
PF3D7lncRNA_1824	1	tccacctaggcatataaatcaat
	2	taatattcaaatatatacattat
	3	catatgggaaaaaattaaaatat
	4	tattttatttattcttcaccctt
PF3D7lncRNA_1936	1	tcattttaattaattaatctctt
	2	taaaaccgagttaaaagtgatta
	3	tatgcaaattattcatacaataa
	4	taaaacaatttattttattattc
PF3D7lncRNA_1960	1	ctgaatcacctattaattctact
	2	attatatcttcatcctctgtttt
	3	tataaattatcaacaatagatgt
	4	atttcattttgtcatcactatta
PF3D7lncRNA_2277	1	tataaataattttgattaactct
	2	gttcctattattatatattattc
	3	gaaggatagctatgaaataataa
	4	aagaaaccgttatgaagaaaagg
Non-Targeting	1	ggccgtgaaaatgaagacttact
	2	acatgatgcgaatgaaagagagt
	3	aattacagaaattgtcagtgcct
	4	ttgctaaaggcaaaggtaaagca
MFR3	1	ttgtatatcaaactatagtatga
	2	attacatataataacatttcggt
	3	aaatataggaatgaaacaattaa
	4	aatggttgtgtatcaaaaatgaa

An 5′-AGAT overhang was added to the forward sequences, and a 5′-AAAA overhang was added to the reverse sequence.

Table C.3 qPCR Primers used in *Chapter 5*

Target	Primers (Forward and Reverse)	Efficiency (%)
18S rRNA*	GCTGACTACGTCCCTGCCC ACAATTCATCATCATATCTTTCAATCGGTA	103
Cyclophilin	AAACGGGAGATCCTTCAGGT AAGGACATGGGACAGTGGTT	131
mfr3	GCTCAAGGGCTCTTTGTATCT TGAAGAAGGAAGGGAAATCATGT	97
LncRNA_0736	AGGAAGGTATAATACATGTTCCGTT GGGACACATGACTACGACAAA	101
LncRNA_1690	GATTCTTTGAGTAAACGCAGTGAG CTAACCTTAGCTGGCTCTGAAG	90
LncRNA_2149	AACTGTCCATCTTTATCATCCGT ATGACGAGACTCAATGTACCATAC	91
LncRNA_0055	TGTATGCATGGGTGTGTGTAA GAAAGGGAGAAATAACTTATCGTTGTAG	104
LncRNA_0364	GCCCATCATTCCAAATCCAAA GCGATAACAGTTCAGGATGAAAG	101
LncRNA_0456	AAGTTTGTGTCCATTGAAGATTCC TGGGTGTTTATGTTTACATGTGC	106
LncRNA_0787	GGAGATGCCACAGGTTTATG CACATGTGCATACATACATACATAC	103
LncRNA_1259	TCGCGTGTCCTTACCATTTATC CCTAGCCATGCGCCTTAAA	102
LncRNA_2353	GTTGTAGGCAGCGTTATT ACATGCCGTACTATAACAAA	117
LncRNA_614	AAACATTACAAGCTAAAAGTTTTCCT ACAATTTATTACAGCTCAGGACAA	93
LncRNA_1478	TGTCCTCATGTGTATGATGAAAGA ACTCTTTGATATCATGAACTCCAA	95
LncRNA_1191	CATTACAAGCTAAAAGTTTTCCTT ACAATTTATTACAGCTCAGGACA	104
dCpf1_1	CCTTCACCGGCTTCTTTGATA CGGGTCAGATTCTCGTTGATAC	108
dCpf1_2	TGCTGGACAAGAAGGAGAAG GTGCACCACCTGAGAGATATAG	99

Primer efficiencies were between 90-110. *Primer obtained from Usui et al. (2019).

Table C.4 Extended list of differentially expressed genes in lncRNA-disrupted mutants

2149		**1478**		**346**		**1191**	
FDR<0.01	0.01<FDR<0.05	FDR<0.01	0.01<FDR<0.05	FDR<0.01	0.01<FDR<0.05	FDR<0.01	0.01<FDR<0.05
PF3D7_0202400	PF3D7_0109000	PF3D7_0109000	PF3D7_0102500	PF3D7_0532400	PF3D7_0202400	PF3D7_0222100	PF3D7_0110000
PF3D7_0209500	PF3D7_0113400	PF3D7_0114000	PF3D7_0110000	PF3D7_0532600	PF3D7_0207600	PF3D7_0221900	PF3D7_0114100
PF3D7_0413400	PF3D7_0204000	PF3D7_0202400	PF3D7_0113800	PF3D7_0533100	PF3D7_0207700	PF3D7_0311700	PF3D7_0202400
PF3D7_0421500	PF3D7_0206900	PF3D7_0209000	PF3D7_0204000	PF3D7_0601500	PF3D7_0412600	PF3D7_0413400	PF3D7_0209500
PF3D7_0423800	PF3D7_0210600	PF3D7_0209500	PF3D7_0206900	PF3D7_0601700	PF3D7_0532500	PF3D7_0413500	PF3D7_0412400
PF3D7_0424100	PF3D7_0212600	PF3D7_0214600	PF3D7_0210600	PF3D7_0617500	PF3D7_0533000	PF3D7_0414500	PF3D7_0412600
PF3D7_0532400	PF3D7_0214600	PF3D7_0311700	PF3D7_0212600	PF3D7_0601200	PF3D7_1035200	PF3D7_0425600	PF3D7_0421500
PF3D7_0532800	PF3D7_0221500	PF3D7_0314700	PF3D7_0213500	PF3D7_0601600	PF3D7_1133300	PF3D7_0532400	PF3D7_0532600
PF3D7_0533000	PF3D7_0308300	PF3D7_0413400	PF3D7_0316000	PF3D7_0617400		PF3D7_0532800	PF3D7_1029100
PF3D7_0533100	PF3D7_0316000	PF3D7_0422900	PF3D7_0323400	PF3D7_0809100		PF3D7_0533000	PF3D7_1031300
PF3D7_0601200	PF3D7_0323400	PF3D7_0532400	PF3D7_0402300	PF3D7_0936900		PF3D7_0533100	PF3D7_1040200
PF3D7_0601400	PF3D7_0404000	PF3D7_0533100	PF3D7_0413500	PF3D7_1030100		PF3D7_0711900	PF3D7_1200500
PF3D7_0822900	PF3D7_0412400	PF3D7_0600100	PF3D7_0414500	PF3D7_1228600		PF3D7_1118900	PF3D7_1231200
PF3D7_0919200	PF3D7_0414500	PF3D7_0601200	PF3D7_0414800	PF3D7_1240700		PF3D7_1200600	PF3D7_1304700
PF3D7_1007300	PF3D7_0414800	PF3D7_0617500	PF3D7_0414900	PF3D7_1301200		PF3D7_1216500	PF3D7_1319800
PF3D7_1035300	PF3D7_0414900	PF3D7_0700600	PF3D7_0415800	PF3D7_1457500		PF3D7_1240700	PF3D7_1372500
PF3D7_1133400	PF3D7_0415800	PF3D7_0700700	PF3D7_0424100	PF3D7_1028500		PF3D7_1240900	PF3D7_1438800
PF3D7_1200600	PF3D7_0422900	PF3D7_0700800	PF3D7_0515600	PF3D7_1200600		PF3D7_1301200	PF3D7_1471800
PF3D7_1231200	PF3D7_0423600	PF3D7_0713100	PF3D7_0525800	PF3D7_1240400		PF3D7_1302100	PF3D7_1476500
PF3D7_1243700	PF3D7_0424200	PF3D7_0822900	PF3D7_0532600	PF3D7_1240900		PF3D7_1335200	
PF3D7_1335200	PF3D7_0424300	PF3D7_0828800	PF3D7_0532800	PF3D7_1372500		PF3D7_1346700	
PF3D7_1351700	PF3D7_0519000	PF3D7_0911100	PF3D7_0533000			PF3D7_1404600	
	PF3D7_0525800	PF3D7_0919200	PF3D7_0600900				
	PF3D7_0532600	PF3D7_0919300	PF3D7_0613900				
	PF3D7_0613900	PF3D7_1007300	PF3D7_0615100				
	PF3D7_0618000	PF3D7_1035300	PF3D7_0720400				
	PF3D7_0713100	PF3D7_1128600	PF3D7_0724900				
	PF3D7_0725400	PF3D7_1133400	PF3D7_0725400				
	PF3D7_0805200	PF3D7_1149500	PF3D7_0805200				
	PF3D7_0805300	PF3D7_1216500	PF3D7_0805300				
	PF3D7_0808200	PF3D7_1231200	PF3D7_0808200				
	PF3D7_0814700	PF3D7_1302100	PF3D7_0816800				
	PF3D7_0818100	PF3D7_1335200	PF3D7_0818100				
	PF3D7_0828800	PF3D7_1346700	PF3D7_0831300				
	PF3D7_0911100	PF3D7_1351700	PF3D7_0904200				
	PF3D7_0919300	PF3D7_1476500	PF3D7_0913900				
	PF3D7_0936900		PF3D7_0914100				
	PF3D7_1003600		PF3D7_0928100				
	PF3D7_1014100		PF3D7_0936900				
	PF3D7_1020100		PF3D7_1003600				
	PF3D7_1028900		PF3D7_1020100				
	PF3D7_1029100		PF3D7_1026400				
	PF3D7_1035400		PF3D7_1026600				
	PF3D7_1035500		PF3D7_1028900				
	PF3D7_1036500		PF3D7_1031000				
	PF3D7_1104900		PF3D7_1035400				
	PF3D7_1125700		PF3D7_1036500				
	PF3D7_1222700		PF3D7_1038400				
	PF3D7_1229800		PF3D7_1104900				
	PF3D7_1231300		PF3D7_1125700				
	PF3D7_1232500		PF3D7_1146800				
	PF3D7_1247700		PF3D7_1200600				
	PF3D7_1251200		PF3D7_1205100				
	PF3D7_1301600		PF3D7_1206500				
	PF3D7_1310700		PF3D7_1226100				
	PF3D7_1312500		PF3D7_1229800				
	PF3D7_1342500		PF3D7_1231300				
	PF3D7_1361800		PF3D7_1232500				
	PF3D7_1372500		PF3D7_1243700				
	PF3D7_1401900		PF3D7_1247700				
	PF3D7_1404600		PF3D7_1301600				
	PF3D7_1423300		PF3D7_1310700				
	PF3D7_1455300		PF3D7_1312500				
	PF3D7_1468400		PF3D7_1319800				
			PF3D7_1327300				
			PF3D7_1342500				
			PF3D7_1369300				
			PF3D7_1400100				
			PF3D7_1401900				
			PF3D7_1403000				
			PF3D7_1404600				
			PF3D7_1408800				
			PF3D7_1411000				
			PF3D7_1416700				
			PF3D7_1417700				
			PF3D7_1426500				
			PF3D7_1430800				
			PF3D7_1431100				
			PF3D7_1438800				
			PF3D7_1454900				
			PF3D7_1455300				
			PF3D7_1466500				
			PF3D7_1468400				
			PF3D7_1473400				

Table C.5 GO term analysis of differentially expressed genes in lncRNA-disrupted mutants

Δ2149	Δ1478	Δ0346	Δ1191
movement in host environment	movement in host environment	obsolete pathogenesis	biological process involved in symbiotic interaction
biological process	entry into host	cell-cell adhesion	biological process involved in interspecies interaction between organisms
biological process involved in symbiotic interaction	biological process	cell adhesion	translocation of molecules into host
biological process involved in interspecies interaction between organisms	cytoskeleton-dependent intracellular transport	production of molecular mediator of immune response	translocation of peptides or proteins into host
biological process involved in interaction with host	cell-cell adhesion	immunoglobulin production	biological process involved in interaction with host
cell motility	cell adhesion	cytoadherence to microvasculature, mediated by symbiont protein	biological process
actin filament organization	obsolete pathogenesis	adhesion of symbiont to host	cAMP biosynthetic process
cell-cell adhesion	cell motility	biological process	obsolete pathogenesis
cell adhesion	reproductive process	regulation of immune response	
actin filament-based process	reproduction	biological adhesion	
actin cytoskeleton organization	biological process involved in symbiotic interaction	regulation of immune system process	
actin filament-based transport	biological process involved in interspecies interaction between organisms	immune system process	
vesicle transport along actin filament	actin filament-based transport	biological process involved in symbiotic interaction	
vesicle cytoskeletal trafficking	vesicle transport along actin filament	biological process involved in interspecies interaction between organisms	
establishment of vesicle localization	biological process involved in interaction with host	biological process involved in interaction with host	
vesicle localization	vesicle cytoskeletal trafficking	exit from host cell	
actin filament-based movement	chromosome organization involved in meiot	exit from host	
supramolecular fiber organization			
cytoskeleton-dependent intracellular transport			
protein insertion into membrane			

C.2 Full author list and contributions

Authors: Johanna Hoshizaki (JH), Sophie Adjalley (SA), Adam Reid (AR) and Marcus Lee (ML). JH, AR, SA and ML conceived and designed the experiments. JH performed the experiments and analysis. ML supervised the work.

Juliana Cudini and Mara Lawniczak analysed single-cell RNA-Seq to identify stage-enriched lncRNAs that were used in target selection.

C.3 Acknowledgements

I thank Emma Carpenter and Megan Pierce for their assistance in setting up the drug assays.

Appendix D

Publications and Preprints

First-author works

Hoshizaki, J., Adjalley, S., Thathy, V., Judge, K., Berriman, M., Reid, AJ., & Lee, MCS. A manually curated annotation characterises genomic features of *P. falciparum* lncRNAs. BMC Genomics. 23, 780 (2022).

Hoshizaki, J., Jagoe, H., & Lee, MCS. Efficient generation of mNeonGreen *Plasmodium falciparum* reporter lines enables quantitative fitness analysis. Frontiers in Cellular and Infection Microbiology. 12, 981432 (2022).

Hoshizaki, J., & Lee, MCS. Scientists on a RAMPAGE to find apicomplexan transcription start sites. Nature Reviews Microbiology 19, 483 (2021).

Other works

Smith, C., Henrici, R., Karpiyevich, M., Ansbro, MR., Hoshizaki, J., van der Heden van Noort, G., Ascher, D., Sutherland, C., Lee, MCS., & Artavanis-Tsakonas, K. Drug resistance-associated mutations in *Plasmodium* UBP-1 disrupt ubiquitin hydrolysis. bioRxiv 2022.09.15.508122

Kümpornsin, K., Kochakarn, T., Yeo, T., Luth, MR., Pearson, RD., Hoshizaki, J., Schindler, KA., Mok, S., Park, H., Uhlemann, A-C., Cubel, SM., Franco, V., Gomez-Lorenzo, MG., Gamo, FJ., Winzeler, EA., Fidock, D., Chookajorn, T., & Lee, MCS. Generation of a mutator parasite to drive resistome discovery in *Plasmodium falciparum*. bioRxiv 2022.08.23.504974