Advancing Functional Genomics in *P. falciparum* Through Systems Biology Approaches

Sean T. Windle

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2025

Reading Committee:
John Aitchison, Chair
Alexis Kaushansky
Christoph Grundner

Program Authorized to Offer Degree:
Pathobiology

University of Washington

Abstract

Advancing Functional Genomics in *P. falciparum* Through Systems Biology Approaches

Sean T. Windle

Chair of the Supervisory Committee:

John Aitchison

Department of Pediatrics

Plasmodium falciparum, the parasite responsible for the deadliest form of malaria, has a complex lifecycle involving various host cell environments in both human and mosquito hosts. The parasite must tightly regulate the expression of each of its ~5500 genes at each stage in order to adapt to its current environment while continuing development. Roughly half of these genes are essential to the asexual blood stage, making it challenging to study gene function and regulation in subsequent stages. Thus, I established a multi-stage transcriptomic dataset that encompasses the complete lifecycle of the parasite. This allows for the comparison of genes across each of the lifecycle stages, helping to generate new hypotheses. To better validate stage-specific phenotypes, I adapted a dimerizable Cre recombinase system using loxP-flanked stop elements for leak-free controllable over-expression and Cas9-directed knockouts. Using a high-fidelity Cas9 under this inducible system, we were able to conditionally perform genome editing by homology-directed repair, editing at specific timepoints and lifecycle stages. I further improved the versatility of this system by establishing a scalable cloning approach for pairing synthesized gRNA-donor sequences. Expanding this system to recently developed *Plasmodium* CRISPRa and CRISPRi systems, stable and scalable CRISPR screens can be generated for over- and under-expression of numerous genes. I

generated two conditional knockouts of uncharacterized genes, identifying blood and gametocyte phenotypes. I then explored the ability to screen for transcriptional regulators of parasite development and artemisinin resistance using small-scale CRISPRa and CRISPRi screens. I envision these novel approaches will aid in further characterizing the parasite genome across its lifecycle, and ultimately help in developing new antimalarials and vaccines.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am incredibly grateful to all of the people who supported me throughout my PhD journey. I want to thank my advisor, John Aitchison, for allowing me to pursue a completely independent project, on a topic completely outside the focus of the lab. John provided both the support necessary for me to succeed, as well as the freedom to explore my ideas.

I am also thankful to all of my fellow students and colleagues in the Pathobiology community, especially Pathobiology cohort, Nicole, Anna, Natasha, and Tanvi. I do not know if I could have completed my PhD without their encouragement and moral support.

Of course I must thank my dearest cat, Jazzy, for being the most loyal little monster. I enjoy nothing more than to come home to her incessant meowing.

Finally, I would like to thank my parents for inspiring me to pursue an academic career. Their endless support, thoughtful feedback, and extremely gentle criticisms have helped me to get to where I am today. Thank you for being the very best cheerleaders.

Chapter 1 – Introduction

1.1 Global Burden of Malaria

Malaria is a life-threatening vector-borne disease that has had an enormous global health burden in humans throughout history [1]. The disease, primarily caused by the apicomplexan parasite, *Plasmodium falciparum*, is responsible for over 600,000 deaths and 200 million cases globally per year [2]. The disease is endemic to over 80 countries, particularly Sub-Saharan Africa where cases and mortality are greatest [3]. This has continued to greatly affect countries in low and middle-income countries, where antimalarial controls, treatment, and vaccines are least accessible. Since the beginning of the COVID-19 pandemic in 2020, progress towards reducing the burden of malaria has halted, with many efforts and interventions being disrupted by the pandemic [4]. This has led cases and mortality to increase from 2019, pushing the goal of eradication even further away.

1.2 The *P. falciparum* Lifecycle

The *P. falciparum* parasite has a complex lifecycle split between an *Anopheles* mosquito vector and a mammalian host [Figure 1.1]. Starting with an infectious mosqusito bite, sporozoites are injected into the dermis of a mammalian host (i.e., humans) before moving to the bloodstream [5]. From there, the parasites are able to access the liver, where they invade hepatocytes and initiate the liver stage of infection [6]. The parasites begin to develop within the hepatocytes over the course of seven days, replicating to form over tens of thousands of merozoites per infected hepatocyte, before finally rupturing the hepatocyte and being released into the bloodstream [7]. Merozoites that are released from the liver go on to infect erythrocytes, initiating the blood stage of infection. Parasites begin developing into ring-shaped trophozoites (ring stage), mature trophozoites as parasites grow in size and become more metabolically active (trophozoite stage), and

then begin rapid nuclear division to form many individual merozoites (schizont stage) [8]. At the end of the cycle (~48 hours), the merozoites rupture their host red blood cell and go on to invade a new red blood cell. Parasites can continuously infect, replicate, and lyse red blood cells, ultimately reaching a high parasitemia over 5% and leading to anemia in

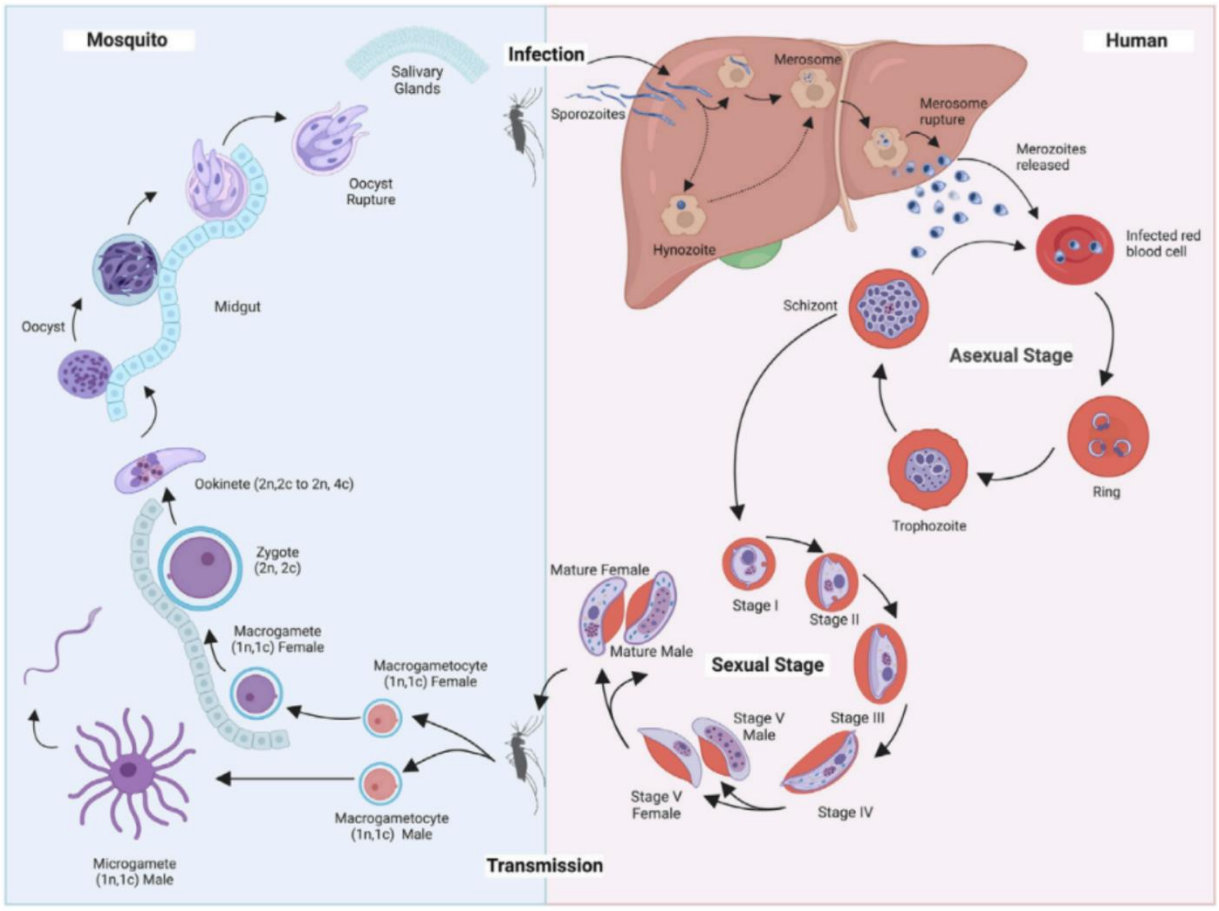

Figure 1.1 The *Plasmodium falciparum* lifecycle
The parasite lifecycle of P. falciparum consists of four general stages: the Liver Stage, the Asexual Blood Stage, the Sexual Gametocyte Stage, and the Mosquito Stage. Adapted from Chahine & Roche, 2022.

its host [9].

During the blood stage, a subset of parasites break from their asexual cycle, and begin differentiation into a sexual gametocyte form (gametocyte stage) [10]. These immature gametocytes are sequestered to the bone marrow and spleen, where they can also infect and develop in erythroblasts [11, 12]. These gametocytes continue to develop over 9-12 days into either a distinct 'male' form or a 'female' form, before releasing back

into the bloodstream. These mature forms are then ingested by the mosquito upon blood feeding on the human host, initiating the sexual phase of the lifecycle [13]. While in the mosquito midgut, environmental changes such as pH and temperature trigger gametogenesis. Male gametocytes begin dividing into eight microgametes carrying flagella, while female gametocytes create a single macrogamete. The macrogamete is fertilized by a microgamete forming a diploid zygote, beginning meiosis and differentiation into motile ookinetes. Ookinetes cross the midgut epithelia to establish as oocysts that will begin rapid replication into hacploid sporozoites. The oocysts rupture and thousands of sporozoites migrate to the mosquito salivary gland, allowing for transmission upon subsequent mosquito bite, thus completing the parasite lifecycle [5].

1.3 Antimalarial Controls

Bed nets and antimalarial drugs have been the primary means of controlling malaria in endemic regions [14]. Insecticide-treated bed nets (ITNs) have had a tremendous impact in reducing mortality and the spread of malaria, as well as helping control vector populations [15]. ITNs have been shown to decrease malaria mortality by up to 20% and cases of malaria by up to 50%, providing one of the most cost-effective global health interventions [16]. However, despite widespread ITN use in endemic regions, millions of malaria cases persist. Malaria treatment involves a combination of antimalarials. Chloroquine, artemisinin, and various partner drugs have had a tremendous impact in treating malaria and decreasing the burden of disease [17]. While chloroquine has been used globally since the early 1900's, it has been phased out of many regimens due to widespread chloroquine resistance throughout both Southeast Asia and Africa. Most current treatments rely on a combination of an artemisinin-based drug paired with one or more partner drugs [18]. These regimens are known as artemisinin-based combination therapies (ACTs). Artemisinin was discovered in the 1960s, and offered effective, fast-acting, and safe treatment for malaria [19]. While artemisinin derivatives are highly

effective, they have short half-lives and are eliminated from the body much faster than other classes of antimalarials [20]. To ensure complete clearance of malaria parasites, ideal partner drugs have longer half-lives and a separate mechanism of action from artemisinins.

However, as ACTs became widely used to treat *P. falciparum,* the parasite has continued to evolve and develop resistance to these antimalarials [21]. In Southeast Asia, drug resistance to artemisinin and respective partner drugs has become widespread. It is thought that resistance is largely conferred through mutations to *Kelch13, mdr1, plasmepsin 2-3,* and *crt,* and a combination of other genes [22–26]. The exact mechanisms for resistance are still poorly understood, particularly where there is resistance to both partner drugs in ACTs [27]. Furthermore, there are large genomic differences between Southeast Asian parasites and African parasites, with many parasite mutations found in resistant Southeast Asian parasites having little to no effect in parasites from Africa. A better understanding of the mechanisms of resistance, as well as the compensatory mutations that are necessary for resistance to arise, is essential in combating widespread antimalarial resistance. A combination of modern bioinformatics and genetic approaches will be key to helping elucidate these questions.

1.4 Vaccines

Only recently have vaccines, such as RTS-S and R21, begun to be distributed in Sub-Saharan Africa [28]. These vaccines are composed of recombinant circumsporozoite protein (CSP), an abundant parasite protein found on the surface of sporozoites, fused to a surface antigen from HepB virus [29]. CSP is necessary for binding and invasion of hepatocytes by the sporozoite and has high immunogenicity in mammalian hosts, making for a strong vaccine target antigen [30]. Although RTS-S and R21 show promise in reducing cases and mortality, the efficacy of these vaccines has been mixed, particularly over the years following vaccination [31–34]. An alternative approach to malaria vaccine

development involves attenuating whole parasites by irradiation or knocking out genes essential for liver stage development [35]. These attenuated vaccines have had mixed success in clinical trials of their own but may be the key to creating a highly efficacious, long-lasting malaria vaccine [36, 37]. The design of genetically attenuated vaccines is particularly challenging, however, as liver stage-specific genes must be knocked out without disrupting function up until the liver stage [38]. Identification of these genes has been primarily through bioinformatic analyses of RNA or protein expression to identify those abundantly expressed solely in the liver stage [39]. This approach has identified several strong candidates for attenuation. Another vaccine approach is to immunize people to proteins expressed in gametocytes to block transmission upon uptake by the mosquito [40–42]. These vaccines are known as transmission-blocking vaccines. However, as above, the main limitation is identifying genes predominantly expressed in gametocytes and are specifically involved in transmission to the mosquito. Vaccine development is yet another area where novel approaches to genetic modification and genetic screening could rapidly advance the field.

1.5 Characterization of Proteins in *P. falciparum* and Current Limitations

Both vaccine and drug development are highly dependent on identifying essential parasite proteins. In order to study the gametocyte stage, mosquito stage, or liver stage in the laboratory setting, one must start from asexual blood stage cultures and proceed through the lifecycle until the stage of interest. Disruption of genes essential to the blood stage precludes advancement to any subsequent stages, and thus prohibiting functional characterization in those stages. Consequently, almost all characterization of gene expression and essentiality has focused on the asexual blood stage [43]. Another major challenge comes from the AT-rich genome of *P. falciparum*, with 82% of DNA being A's and T's, complicating molecular genetics [44].

Many labs choose to study mouse malaria strains, such as *P. yoelii* and *P. berghei* as models for *P. falciparum*. However, there is relatively low genetic conservation between these species, and protein functions are not necessarily conserved [45]. Another key limitation, particularly in *P. falciparum*, is that genetic manipulation depends on transfection of plasmid, and transfection efficiencies are low, often requiring 20+ days to recover to visible parasitemias following drug selection [46, 47]. For a plasmid to reach the parasite nucleus and be maintained, the plasmid must overcome multiple membranes and barriers from the host RBC and the parasite. These inefficiencies make genome-wide screens prohibitively expensive and challenging. Additionally, *Plasmodium* lacks the pathways necessary for many other forms of gene editing and gene perturbations, such as the Dicer and RISC components needed for RNAi [48, 49].

1.6 Genetic Tools

Genetic manipulation in *Plasmodium* has been critical to a better understanding of parasite biology. Only in the mid-1990's did the field establish plasmid transfections as a means of editing the parasite genome [50]. Prior to this, malaria genetic crosses were a key tool in matching phenotypes to genotypes and understanding gene function [51]. A number of recombinant progeny can be produced by crossing two parasite lines with a divergent phenotype of interest (i.e., drug susceptibility). Then by measuring the phenotype of each progeny, these phenotypes can be mapped to genotypes and narrowed down to a specific locus. The first genetic cross was performed using the parasite lines HB3 and 3D7, helping to identify the gene responsible for pyrimethamine resistance, *pfhdhfr-ts* [52–54]. Another genetic cross was performed between the parasite lines HB3 and Dd2, linking chloroquine resistance to a vacuole membrane protein that was then aptly named chloroquine resistance transporter (*pfcrt*) [55, 56]. A number of other genes and polymorphisms involved in drug resistance, RBC invasion, and mosquito infection have been identified by genetic crosses, such as the genes *pfmdr1,*

pfRh5, and *pfs47* respectively [51, 57–59]. Genetic crosses in malaria were extremely tedious to perform, as they required completing the entire parasite lifecycle with the use of non-human primates. This was an expensive, time-consuming, and technically challenging process, and so only a small number of genetic crosses have been performed to date. Recently however, a humanized mouse model using transplanted human hepatocytes and red blood cells has simplified the process tremendously [60, 61]. There have now been more *P. falciparum* genetic crosses performed in the mouse model than in non-human primates [62–65]. A primary limitation for genetic crosses is the inability to measure beyond the genes and polymorphisms of the two parental lines. Additionally, phenotypes are often the result of more than one specific mutation, and genetic crosses require a very large number of distinct recombinant progeny to identify secondary or tertiary loci [65]. However, genetic crosses still fill an important gap in identifying new genotype-phenotype relationships and have often helped to identify phenotypes beyond their original intent.

While genetic crosses provide a genome-wide screening tool, specific targeted mutations have been introduced through homologous recombination. With the ability to transfect plasmids, one could create a specific mutation in a gene of interest [e.g., knockouts (KOs), single nucleotide polymorphisms (SNPs), etc.] and using a selectable marker, select for parasites that integrate the new sequence and the drug cassette from homologous regions [66]. This was typically performed after building strong evidence that the mutation would have the predicted effect, as from a genetic cross or genome-wide association studies (GWAS) data. This method was eventually replaced with the development of CRISPR/Cas9 technology [67].

Until recently, there had been no genome-wide screens to identify essential genes in *Plasmodium*. This is in large part due to the challenges in designing such a system and the limitations inherent in *Plasmodium* biology. However, in 2017, a genome-wide knockout screen targeting 2,578 genes in *P. berghei* was performed, demonstrating that

~45% of genes were essential for blood stage growth, while an additional 18% resulted in slowed growth [68]. In comparison to other organisms, 45% is a high percentage of essential genes, particularly since the measured essentiality only represents one stage of the parasite lifecycle. To further understand essentiality across the lifecycle, the ~1300 blood stage viable clones generated in the prior screen were fed to mosquitos to identify gametocyte and mosquito essential genes, and parasites were allowed to reinfect mice by mosquito bite to identify liver stage essential genes [69]. The majority of the genes tested were in fact dispensable throughout the entire lifecycle, while a small subset resulted in slow-growth or were essential to mosquito or liver stage development. However, these were inherently all genes that were dispensable to blood stage growth.

Subsequently, a PiggyBac transposase screen was performed in *P. falciparum* using over 9,000 individual transfections to establish >38,000 unique transposon insertions **s**throughout the genome [Figure 1.2] [70]. Assuming random integration, this would be sufficient for saturation of the genome. However, only parasites that recover over the 10+ days following selection can be isolated and the integration sites identified by sequencing. Insertions in coding sequences should produce functional gene knockouts, which if essential, would not survive and thus not be sequenced. By measuring which insertions survive, the essential genes can be deduced from their absence. Roughly ~60% of genes were found to be essential to blood stage growth. This was the first genome-wide screen for essentiality in *P. falciparum*, offering numerous insights and further validating the findings of the *P. berghei* screen. However, there is a bias in transposon screens for preferential insertion in non-coding regions, increasing the likelihood of false positives [71, 72].

Much work has been done using the existing genetic approaches. However, the introduction of CRISPR/Cas9 to the malaria field has rapidly improved the field's ability to better characterize the parasite's genome.

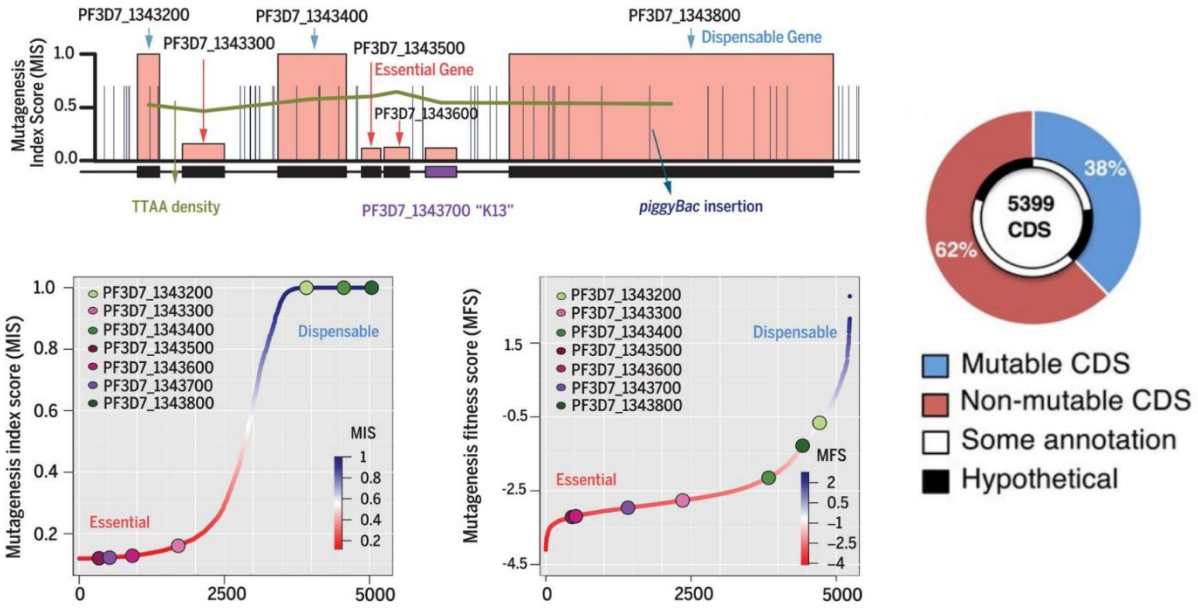

Figure 1.2 PiggyBac transposon screen in *P. falciparum*
The piggyBac transposon screen contained >38,000 identified transposon insertion sites. The frequency of insertions for each gene are used to measure essentiality, with insertion-free genes most likely to essential. 62% of genes were determined to be essential, with 38% dispensable. Adapted from Zhang et al, 2018.

1.7 CRISPR/Cas9 System for Genetic Editing

CRISPR/Cas9 is a technology based upon a bacterial acquired immune system that mediates resistance to bacteriophage infection [73, 74]. The key elements of this system include a DNA double strand endonuclease, a CRISPR-associated protein (typically Cas9), that is guided to specific DNA sequences by a complexed guideRNA (gRNA). The gRNA consists of two components: a CRISPR RNA (crRNA) encoding a specific 20-base pair sequence (protospacer), which anneals or is fused with the trans-activating CRISPR RNA (tracrRNA) [Figure 1.3]. The tracrRNA sequence helps bind to the Cas9 protein, forming a ribonucleoprotein, while the variable protospacer dictates the specific site to cut. For simpler design, the crRNA and crRNA can be expressed as a single gRNA, or sgRNA, using a PolIII promoter such as the U6 promoter [75]. Although a number of bacterial species carry Cas9 proteins, the most widely used for gene editing purposes is the Streptococcus pyogenes Cas9 (SpCas9). The only requirement for the DNA cleavage site is the presence

of a protospacer-adjacent motif (PAM) following the protospacer sequence in the target genome. In the case of SpCas9, the PAM sequence is “NGG”, where N is any nucleotide and G is a guanine. The absence of a correct PAM precludes Cas9 cutting at the specific gRNA site.

Following the initial description of the bacterial CRISPR/Cas9 system, it was immediately recognized that this system could be harnessed to provide site-specific endonuclease activity at any desired genomic location by simply generating a gRNA (or sgRNA) encoding that sequence [76, 77]. In most organisms, the resulting double strand cleavage triggers the activation of cellular non-homologous end-joining pathways (NHEJ), frequently generating insertions or deletions (indels) [78]. Alternatively, if a donor DNA template encoding a desired sequence flanked by regions that are homologous to either side of the cleavage site is included, the cell may utilize homology-directed repair (HDR) pathways to introduce the desired sequence into the site. While most organisms preferentially repair double-strand breaks through the highly efficient NHEJ pathway, *Plasmodium* lacks this repair pathway and thus does not generate indels following the initial cut [44, 49]. Thus, a repair template must be included to allow for HDR of the double strand break. Although the absence of NHEJ hinders our ability to generate knockouts by indel-mediated frameshift mutations, HDR editing is often highly efficient (up to ~100%) in *Plasmodium* [79]. This method of genetic modification has become standard for generating insertions, deletions, and mutations in the *Plasmodium* genome.

Although Cas9 is frequently used for generating mutations by double-strand DNA cleavage, the protein can also be modified to simply direct effectors to a specific genomic site [80]. Point mutations D10A and H840A in SpCas9 are sufficient for completely inhibiting endonuclease activity, creating an endonuclease-dead Cas9 (dCas9). dCas9 can then be fused with various effector domains or proteins, allowing for a variety of epigenetic modifications [81–83]. Fusion of dCas9 to transcriptional activator domains would allow for over-expression (CRISPRa), while fusion to transcriptional repressor

domains would allow for inhibition of expression (CRISPRi). Even the binding of dCas9 alone can be sufficient for knockdown of gene expression through steric hindrance and interference with the RNA polymerase [84]. A number of dCas9 fusions have been established for mammalian systems, plant systems, as well as for *P. falciparum* [85–88]. Unlike Cas9-mediated gene editing, CRISPRa and CRISPRi do not depend on any DNA repair, and thus only need to be provided with a gRNA targeting the gene of interest. This provides an accessible, scalable system for gene perturbation.

Other prominent gene editing tools have been built based on the Cas9 system [Figure 1.3]. Base-editing is one approach that relies on a modified Cas9 that generates single-strand DNA breaks ("nickase") fused to a cytidine deaminase [89, 90]. The gRNA can be targeted to specific sequences to direct the cytidine deaminase to create cytidine-to-uridine mutations within the gRNA-targeted sequence. The corresponding nick triggers repair of the uridine through mismatch repair pathways, resulting in a C-to-T mutation. This approach offers the ability to mutate specific bases and amino acids using only the specified gRNA with no donor repair template. In addition to certain amino-acid substitutions, C-to-T mutations are also able to create stop codons for gene disruption [91]. Four codon changes can result in a stop codon if edited by cytidine deaminase: CAA, CAG, CGA, and TGG (via reverse strand). This would provide an alternative means of generating premature stop codons using gRNAs, allowing for scalable knockout screens.

Similarly, prime-editing is another approach in which a nicking Cas9 is fused to a reverse transcriptase, and paired with an sgRNA with extended sequence corresponding to one's desired mutations [92, 93]. The extended sgRNA acts as a priming sequence for the reverse transcriptase, causing the DNA to be repaired to this newly specified sequence. Prime-editing allows for specific insertions and deletions (<50 bp in size) to be made, while only needing an extended gRNA in combination with the nicking Cas9 to be

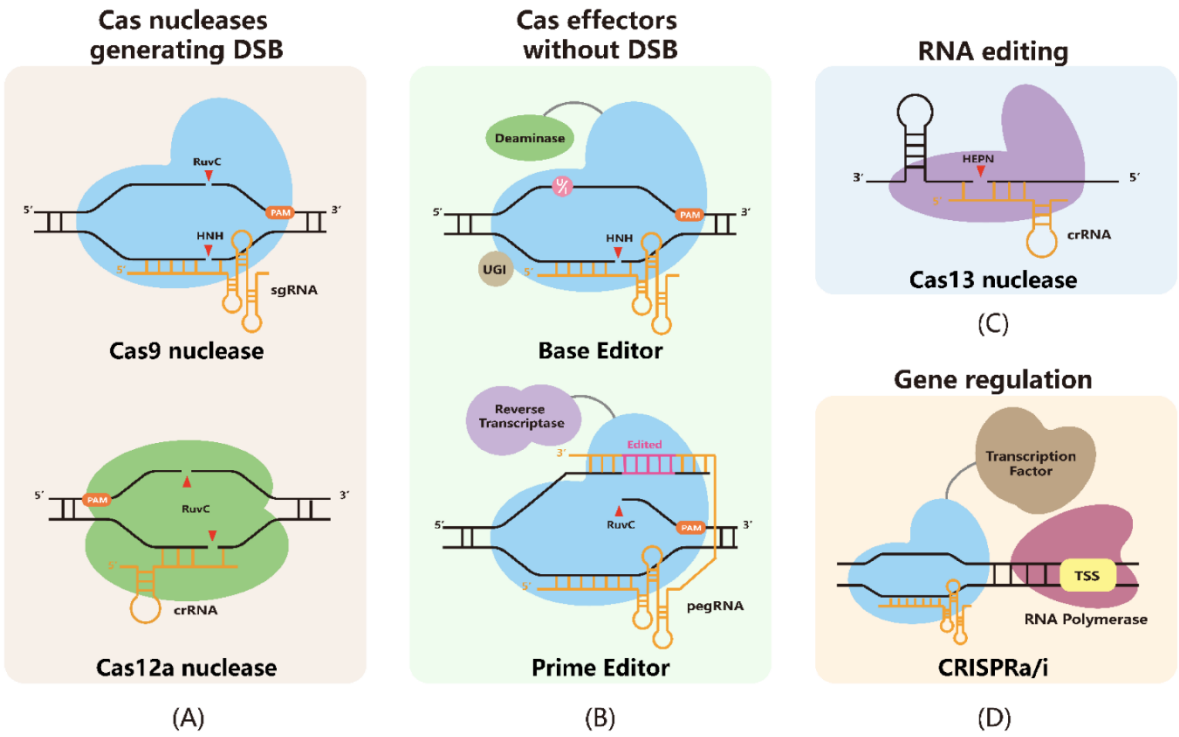

Figure 1.3 CRISPR-based technologies
Many CRISPR-based tools have been developed for various purposes. Cas nucleases, such as Cas9, allow for targeted double-strand breaks, prompting repair by the cell. Base editors and prime editors use modified Cas9-nickases and fusions to mutate proximal bases. Cas13 nuclease allows for RNA editing, binding and cleaving RNA instead of DNA. CRISPRa/i (activision / inhibition) are able to modulate expression through effectors and histone acetylation without permanent DNA editing.

provided. This system offers more versatility in the types of mutations that can be introduced, while still not depending on double stranded breaks, an additional donor repair template, or homology directed repair.

Although the number of gene editing tools entering the field of molecular biology is rapidly growing, there has yet to be widespread adoption of many of these tools in malaria research. However, one group developed a CRISPRa- and CRISPRi-based system by fusion of dCas9 to a histone acetyltransferase domain (GCN5) or a histone deacetylase domain (Sir2a) respectively [88]. These fusions allowed for specific over- and under-expression of malaria genes via gRNA, providing a means of scalably perturbing multiple genes in parallel. CRISPRa and CRISPRi screens have become a core tool for identifying new gene functions, both through gain-of-function screens and loss-of-function screens [94].

1.8 Conditional Gene Expression Systems

Even with the implementation of CRISPR/Cas9 tools, there are still a number of challenges in understanding gene function across the lifecycle. In particular, genes that are essential for blood stage survival and development cannot be studied in subsequent stages of development, as any knockout would be lethal to the parasite, not allowing the parasite to reach the stage of interest. Essential genes constitute roughly 60% of the parasite genome [70]. To better understand the role of these blood stage-essential genes, conditional genetic systems are necessary.

Several conditional systems have been developed for molecular biology. The Cre/loxP system has been used for decades, largely for mouse studies and mammalian cell lines [95, 96]. This system derived from the bacteriophage P1, involves Cre, a recombinase, that specifically mediates recombination between pairs of loxP sites, resulting in deletion of sequences flanked by the sites. By expressing Cre under the control of cell type-specific promoters, the deletion of loxP-flanked sequences can be restricted to the cell type of interest. Additionally, this system is one-directional—once the loxP-flanked DNA has been excised, it cannot be reversed. This system can also be employed to turn on gene expression (rather than turning it off) by inserting loxP-flanked stop elements (lox-STOP-lox, or LSL) in the 5'-untranslated regions of genes, which will prevent gene expression until Cre expression excises the loxP-flanked STOP element [97]. This approach has been used for inducible control of the gametocyte-regulating transcription factor, AP2G, allowing for controllable gametocyte induction [98].

Common systems used for conditional expression are the Tet-Off and Tet-On inducible systems [99]. These systems are tunable with the addition of tetracycline or a tetracycline derivative such as doxycycline. The Tet-Off system relies on a tetracycline-dependent transactivator protein which allows for repression at a Tet-responsive element upstream the gene of interest when drug is present [100]. In the case of Tet-On, the system uses a reverse transactivator which only binds in the presence of the drug [101].

These systems have been used for inducible expression of transgenes, as well as inducible over and underexpression [99].

Most of these systems retain some amount of expression and are considered to be "leaky" [102–104]. Incomplete knockdown is not necessarily bad, however. Complete knockout of an essential gene may cause immediate cell death and prohibit further study of the phenotype. Incomplete knockdowns may still allow for identification of phenotypes while preventing some deleterious effects that might obscure the phenotype [105]. However, some applications for conditional expression depend on leak-free expression, such as Cas9 expression, where even a small amount of Cas9 could lead to a permanent genomic change.

Given the numerous ways in which Cas9 can be used to perturb a gene (knockout, overexpression, underexpression, introduction of specific mutations, etc.), there is great value in establishing a controllable system for Cas9, dCas9, and various effectors [106, 107]. In cases where temporal control of knockouts is required, such as in the case of differentiated cells, an inducible system may be necessary. A conditional Cas9 system enables phenotyping across different stages and cell types.

An early design involved splitting Cas9 into two individually expressed proteins that could be chemically dimerized into a functional Cas9 [108]. This relied on fusing the N-terminal Cas9 fragment to an FRB domain and fusing the C-terminal Cas9 fragment to an FKBP domain. These domains rapidly dimerize in the presence of rapamycin, allowing for quick induction of Cas9 activity. However, leakiness due to auto-dimerization poses a problem for preventing mutations and editing prematurely [109].

Another system utilized the autoinhibition mechanism of a BH3 peptide and the BCL-xL protein, each fused to Cas9 [110, 111]. The interaction of the BH3 and BCL-xL domains of the fusion protein block Cas9 activity. However, in the presence of the BCL-xL inhibitor, A115, the BH3:BCL-xL interaction is disrupted, recovering Cas9's enzymatic activity. This allows for controllable Cas9 activity within minutes of A115 addition, as well

as tunable activity by A115 concentration. This system has been expanded to a number of effector functions, including CRISPRa, CRISPRi, base-editing, and prime-editing [112].

Rather than relying on direct control of the Cas9 protein's activity, an alternative approach is to control the expression of Cas9 [113, 114]. The simplest system would be one in which Cas9 is only expressed in the cell type of interest. This would only require a cell type-specific promoter to drive Cas9 expression, at the expense of versatility. If one is only interested in a specific stage or cell type, this approach would be sufficient. Alternatively, chemically-inducible promoters can be used. Tet-inducible Cas9 has been established in mammalian cells and is widely available [115, 116]. The Tet-inducible Cas9 relies on the Tet-operon and activation through doxycycline. TetR has also been used for activating sgRNA RNA expression, allowing for inducible sgRNAs. Similarly, Cre/loxP strategies can be used with an interfering gene and terminator to irreversibly turn on Cas9 [117, 118]. The disadvantage of expression-based systems for Cas9 is the temporal delay in Cas9 activity. Turning on expression is not immediate, and production of protein can be delayed by hours.

There are now numerous genetic tools to explore, each with their own advantages and disadvantages. Applying these tools to malaria will help fill a large gap in the field's current technologies for characterizing the genome.

1.9 Bioinformatics Approaches for Malaria Research

Bioinformatics has been an essential component in identifying drug resistance, gene function, and even in the design of new vaccines [39, 69, 119–124]. In the case of examining drug resistance, genetic crosses involving two parental lines with divergent phenotypes depended on quantitative trait locus analysis to identify specific loci that might confer the phenotype of interest [51, 125]. This classical approach led to the identification of mutations in the chloroquine resistance transporter gene, *pfcrt*, as well as a number of other critically important SNPs [126–130]. Similarly, genome-wide

association studies using patient isolates helped to identify a number of genes associated with drug susceptibility [131–133]. With advancements to genome sequencing, entire genomes can now be sequenced inexpensively and compared, as well as entire transcriptomes [44, 134, 135]. Microarray datasets were pivotal in identifying genes involved in blood stage development and gametocyte development, though RNA-seq has since overtaken microarray as the primary means of transcriptomic analysis [136–138]. Analysis of differential gene expression has become standard practice now, helping identify subtle differences between stages of development, different strains, and knockout lines [139–141]. This has only expanded with advancements in single-cell RNA-sequencing, which provides for transcriptomic profiles for singular cells [142, 143]. A handful of single-cell transcriptomic datasets have been established, producing a "Malaria Cell Atlas" from blood stage up to mosquito stages of development [123, 144–146]. As our ability to generate extensive genomic, transcriptomic, and proteomic datasets increases, so does the need for computational tools to help analyze these complex data.

Despite *Plasmodium* only having ~5500 genes, most of these genes have yet to be characterized. Much of our current knowledge and gene annotations are derived from orthologs in other organisms such as yeast, plants, and mammalian cell lines [147–149]. This relies on the assumption that genes that have shared ancestry retain their function. Gene ontology-based annotation attributes gene identities and functions across species, offering a means of annotating *P. falciparum* genes *in silico.* However, *Plasmodium* has diverged considerably from any well-characterized organism [44]. For instance, *Toxoplasma gondii,* another apicomplexan parasite, is evolutionarily separated from *Plasmodium* by hundreds of millions of years [150–152]. Even within the same genus (*Plasmodium*), only ~80% of genes are shared between *P. falciparum* and *P. berghei [153, 154]*. Furthermore, gene sequence similarity between *P. falciparum* and *P. berghei* orthologs is ~75%. Due to these limitations, orthology and gene ontology are only able to

provide us with strong gene candidates to pursue experimentally, rather than providing conclusive gene functions.

With the rapid adaptation of transcriptomics, there are hundreds of unique transcriptomic datasets now available [149]. One dataset consisted of over 1,000 microarray transcriptomic profiles generated from unique *P. falciparum* clinical isolates from patients across Africa and Southeast Asia [155]. Comparative genomics paired with parasite clearance for each isolate provided valuable insights into specific transcriptional features associated with artemisinin susceptibility, and identified genes involved in the unfolded protein response as a possible mediator for artemisinin resistance.

Further work has been done to identify transcriptomic patterns associated with artemisinin resistance, including the development of a gene regulatory network (GRN) model built on the prior microarray dataset [Figure 1.4] [156–159]. By taking advantage of the heterogeneity that exists between patient isolates, gene co-expression was identified and clustered together. Then, expression of these clusters can be attributed to a selection of genes predicted to be transcriptional regulators (TRs) using regularized regression methods. The produced GRN model attributes weights to genes predicted to be regulated by a given TR. When expression of a TR is changed, its targets change by the given weight. The GRN model was able to accurately predict gene expression in 4000-4500 genes using just the expression data of the TRs. The model accurately predicted expression for both clinical isolates and *in vitro* samples. As the only GRN model in *Plasmodium*, this model provides novel insights into how *Plasmodium* regulates its genome. Additionally, by combining established knowledge of drug susceptibility, parasite development, and gene function with predictions from the GRN model, we can generate novel hypotheses on how genes might influence parasite biology.

1.10 Applications of Bioinformatics and Molecular Tools in P. falciparum

In *Plasmodium*, limitations of genetic manipulation and transfection have hindered the field's progress in designing new therapies and vaccines. Although the existing systems allow for study of essentiality at different stages of the lifecycle, numerous challenges exist in implementing these. For example, in order to introduce loxP sites for Cre-mediated knockout, one would have to clone a gRNA and donor sequence, potentially for both the 5' and 3' loxP insertion. This may involve multiple PCRs, ligations or assemblies, and sequencing before the plasmids are ready for transfection. Transfection then would take roughly 30 days before parasites fully recover from drug selection, and then these parasites would need to be cloned. This process from beginning to end can take many months and may not work. To alleviate many of these problems, conditional Cas9 systems relying solely on a gRNA would provide a means of perturbing genes at scale and with stage-specificity.

In the following chapters, I describe efforts to further elucidate gene function through both computational and molecular approaches. First, utilizing the vast amount of transcriptomic data available, I sought to create a compiled transcriptomic profile for the main stages of development in *P. falciparum* to better predict essentiality across the lifecycle. These analyses complement our findings from our gene regulatory network model, presenting novel hypotheses to be pursued. Second, building on available tools, I constructed a scalable system for conditional Cas9 gene editing. In this system, Cas9 can be induced in the blood stage as well as the gametocyte stage, allowing for gene knockout or the introduction of specific mutations. Lastly, I present the development and application of an accessible system for medium-throughput CRISPRa and CRISPRi screens. These screens can be applied to any number of phenotypes, allowing for expansive identification of gene-phenotype relationships. Ultimately, the goal of this work is to establish new approaches for rapidly and accessibly identifying gene function and essentiality across the lifecycle.

Chapter 2 – Exploring *P. falciparum* Transcriptomics for Characterization Across All Stages of The Lifecycle

2.1 Introduction

Numerous transcriptomic datasets from *Plasmodium* have been established to date, largely focusing on the asexual blood stage [149]. Although single-cell RNAseq has provided insight into the heterogeneity between cells across the lifecycle, the lack of read depth and the inherent sparsity of the data make quantifying lowly expressed genes a challenge [123, 143, 145]. However, bulk RNA-seq is able to capture many more transcripts at low abundance and has now been performed on all stages of the *P. falciparum* lifecycle [149, 160]. These datasets offer a glimpse into the expression profiles across the entire lifecycle, allowing for comparative analyses to better understand gene regulation and gene function. However, these datasets have been generated in different labs, with different protocols, and most importantly, in dramatically different cellular stages, complicating efforts to compare across these timepoints. Given the value that transcriptomics offers, there is a need for better comparative tools that can provide new hypotheses for *Plasmodium* research.

I first sought to compile a complete bulk RNA-seq transcriptomics dataset covering all of the major stages of the parasite lifecycle. This compiled dataset would provide a complete transcriptomic picture of the parasite lifecycle, while highlighting differences between each of the stages and timepoints. Building on these data and the gene regulatory network model our lab developed, I explored gene features that could be predictive of gene function and essentiality across the lifecycle. This work provides a better understanding of the complete parasite lifecycle, as well as key regulators throughout, that can be used to guide future directions.

2.2 RNA-seq Transcriptome Analysis Across the *P. falciparum* Lifecycle

I first compiled as many high quality datasets as possible for each respective stage of the parasite lifecycle [43, 149, 161–178]. Efforts have been made to do this for the asexual and sexual stages, but none have included bulk RNAseq data from the mosquito and liver stages [179]. I began by taking the existing published asexual and sexual datasets, resulting in ~25 individual transcriptomic datasets (Table 2.1). To remove heterogeneity, I filtered out samples from patient isolates and less frequently used parasite lines. The *P. falciparum* liver stage transcriptome has only been fully quantified in one dataset covering sporozoites, day 4, day 5, and day 6 of liver stage [160]. Together, these datasets are representative of all the major stages of the *P. falciparum* parasite cycle.

Table 2.1 RNA-seq transcriptomic datasets for P. falciparum

Stages	Contributor	Brief Citation	Year
BS	Karine LeRoch	Bunnik et al.	2013
BS	Gwladys I. Bertin	Kamaliddin et al	2021
BS	Zbynek Bozdech	Kucharski and Tripathi et al.	2020
BS + MS	Victor G. Corces	Gomez-Diaz	2017
BS	Chris I. Newbold	Otto et al.	2011
BS	Joseph L DeRisi	Caro et al.	2014
BS	Manuel Llinas	Chappell et al. 2020	2020
MS	Stephen L. Hoffman	Eappen et al. 2022	2022
BS + GS	Xin-zhuan Su	Lopez-Barragan et al.	2011
BS	Wieteke Hoeijmakers	Hoeijmakers et al.	2014
BS	Arnab Pain	Subudhi et al. 2020	2020
BS	Chris I. Newbold	Otto et al.	2010
BS	Michael F. Duffy	Josling et al 2015	2015
BS	Henk Stunnenberg	Bartfai et al.	2010
GS	Richard Bartfai	Lasonder et al.	2016
BS	Michael F. Duffy	Tonkin-Hill et al 2018	2018
BS	Aubrey Cunnington	Lee et al.	2018
BS + MS	Artur Scherf	Zanghi et al.	2017
BS	Michael F.Duffy	Tang et al. 2020	2020
MS	Stefan Kappe	Lindner et al.	2019
BS + MS	Stephen L. Hoffman	Hoffmann et al.	2013
BS + GS	Xin-zhuan Su	Lopez-Barragan et al.	2011
BS	John Adams	Zhang et al. 2021	2021
BS	Richard Bartfai	Toenhake et al.	2018
BS	Tim W Gilberger	Wichers et al. 2019	2019
MS + LS	Stefan Kappe	Zhangi et al. 2024	2024
GS	John Adams	Chawla et al. 2023	2023

In order for these datasets to be compared, I explored several methods for normalization. Under ideal circumstances, every RNA-seq sample would contain a spike-in control, be carried out together, and be performed using the same methodology [180, 181]. Unfortunately, none of those conditions are met when comparing the existing publicly available datasets. Furthermore, due to the complex biological differences that exist across the lifecycle, direct comparisons of expression are inherently flawed. For instance, if one normalizes by total reads or counts per million (CPM), in cases where parasites have globally low levels of expression, such as in early rings, or comparing between highly divergent transcriptomes, differential expression analysis may miss real signals while picking up noise. As for simply leaving the data untransformed, this has a tendency to dramatically over- or under-estimate the extremes and would ignore read depths especially comparing new and old datasets. Ultimately, there is no perfect method for differential expression that can be applied to these datasets without violating key assumptions, thus I pursued a balanced approach that uses class-specific quantile normalization, which would retain many of the features distinct to each stage while accounting for read depth [182, 183].

As a means of simplifying comparisons between stages, I chose broad timepoints that encompassed all development within that stage. This led to 3 blood stage timepoints (Ring, Troph, Schizont), 3 sexual stage timepoints (Early Gametocyte, Male Gametocyte, Female Gametocyte), 2 mosquito timepoints (Oocyst, Sporozoite), and 3 liver stage timepoints (Day 4 LS, Day 5 LS, Day 6 LS). Each dataset was then quantile normalized by class (i.e., Stage timepoint) and the mean was taken for each timepoint across the

Product Description	Gene ID	Ring	Troph	Schizont	Gam	Male	Female	Oocyst	Sporo	Liver D4	Liver D5	Liver D6
thrombospondin-related sporozoite	PF3D7_0104000	3.1	10.3	22.1	38.9	113.5	165.8	2140.5	5471.4	10.5	5.0	15.9
conserved Plasmodium membrane	PF3D7_0104100	5.3	14.0	56.4	1.8	11.3	26.6	614.1	3907.7	26.3	17.2	33.0
StAR-related lipid transfer protein	PF3D7_0104200	302.5	110.3	280.1	28.8	19.9	27.5	282.6	300.8	259.9	349.7	264.8
ubiquitin carboxyl-terminal hydrolase	PF3D7_0104300	280.7	78.8	109.0	7.3	65.7	21.9	196.8	52.4	209.5	242.5	189.9
4-hydroxy-3-methylbut-2-enyl	PF3D7_0104400	23.4	71.3	94.8	40.5	165.7	296.9	43.5	32.3	163.8	237.6	350.8
conserved protein, unknown function	PF3D7_0104500	73.1	142.6	90.3	102.1	34.3	51.9	0.0	0.3	452.2	563.3	375.1
conserved Plasmodium protein,	PF3D7_0104600	4.3	8.5	7.1	3.5	4.7	4.6	0.0	1.1	46.2	31.5	32.5
major facilitator superfamily-related	PF3D7_0104700	0.6	0.6	1.4	2.2	5.4	13.0	99.9	1.6	0.0	7.1	7.0
novel putative transporter 1, putative	PF3D7_0104800	18.1	25.0	63.0	29.5	95.3	103.8	144.7	66.8	18.8	20.6	30.8

High

Low

Figure 2.1 Example of compiled transcriptomics across the *P. falciparum* lifecycle
Genes and their normalized transcripts for each stage of development are sortable and viewable as a heatmap.

datasets. The averages for each timepoint were compiled into a unified dataset of mean expression values across the parasite lifecycle [Figure 2.1].

To identify genes differentially expressed in any stage, fold change between each stage was calculated for every gene. The fold change was further sorted to identify genes primarily expressed in one stage, setting a threshold of >3-fold above the next highest stage. Male gametocytes contain the highest number of genes primarily expressed in only that stage, with 313 genes >3 fold higher than the next highest stage.

These stage timepoints can be further consolidated into the four distinct stages: Blood Stage (BS), Gametocyte Stage (GS), Mosquito Stage (MS), and Liver Stage (LS). Rather than use the average, the maximum level of expression from each sub-stage was used. The above analysis was repeated with these four consolidated stages, classifying ~1950 genes as broadly stage-specific [Figure 2.2]. 484 genes were primarily expressed (>3-fold) in MS, 239 in LS, 776 in GS, and 424 in BS.

Gene ID	Product Description	BS ESSENTIAL	BS	GS	MS	LS	FOLD CHANGE	STAGE EXCLUSIVE
PF3D7_0102200	ring-infected erythrocyte surface	0	9283.7	5.3	0.9	1.1	1744.8	BLOOD
PF3D7_1001500	early transcribed membrane protein	1	14183.9	11.2	0.2	2.1	1265.2	BLOOD
PF3D7_0518800	secreted ookinete protein, putative	1	11.0	10991.8	0.3	1.3	998.9	GAMS
PF3D7_1031000	25 kDa ookinete surface antigen	0	54.3	51210.1	42.5	1.8	942.7	GAMS
PF3D7_1454900	conserved Plasmodium protein,	1	29.0	27227.8	15.3	3.4	939.8	GAMS
PF3D7_0207300	serine repeat antigen 8	0	2.6	3.5	5724.2	6.8	839.4	MOZZY
PF3D7_1316300	conserved Plasmodium protein,	1	14.5	18006.4	23.9	2.1	753.5	GAMS
PF3D7_1334900	MSP7-like protein, fragment,	1	134.6	0.0	0.2	0.2	701.1	BLOOD
PF3D7_0825700	conserved Plasmodium protein,	0	3.6	1995.3	0.2	1.2	549.7	GAMS
PF3D7_0825800	conserved Plasmodium protein,	1	7.5	5190.6	0.3	9.9	523.4	GAMS
PF3D7_1335900	thrombospondin-related anonymous	1	5.1	103.3	53700.5	27.0	520.0	MOZZY
PF3D7_1400900	Plasmodium exported protein (PHISTa),	1	0.2	0.1	95.9	0.1	501.0	MOZZY

Figure 2.2 Example of consolidated stage-specific gene expression
The gene expression for each broad stage of the P. falciparum lifecycle, showing blood stage essentiality and fold-change of the most expressed stage expression compared to the next highest stage.

These subsets can be further analyzed by Gene Ontology (GO) enrichment analysis for biological pathways and molecular function [184]. In LS-exclusive genes, pathways involved in biotin metabolism and pyruvate metabolism are most enriched, while the most enriched LS molecular functions are fatty acid synthesis and oxidoreductase activity. These results are in agreement with findings that *Plasmodium* parasites rely on fatty acid metabolism for development in late liver stage infection, as well as help explain the interactions between lipid peroxidation and the parasite [69, 185–187].

Using K-means clustering of the expression from each timepoint from the lifecycle, 7 distinct clusters were generated [Figure 2.3]. Cluster 1 contains genes primarily expressed in sporozoites through the end of liver stage. Cluster 3 contains genes expressed exclusively in human host environments and not in the mosquito, while Cluster 5 consists of MS-specific genes not expressed in the human host. Cluster 6 is largely specific to the late sexual stages and the oocyst, likely involved in sexual biology.

The blood and gametocyte stages were most correlated with each other, with the mosquito stage being most distinct [Figure 2.3B]. The liver stage was moderately correlated with blood and gametocyte. Given the non-mammalian host setting, as well as the unique sexual biology occurring, it was expected that the mosquito stages would be least correlated with the other stages.

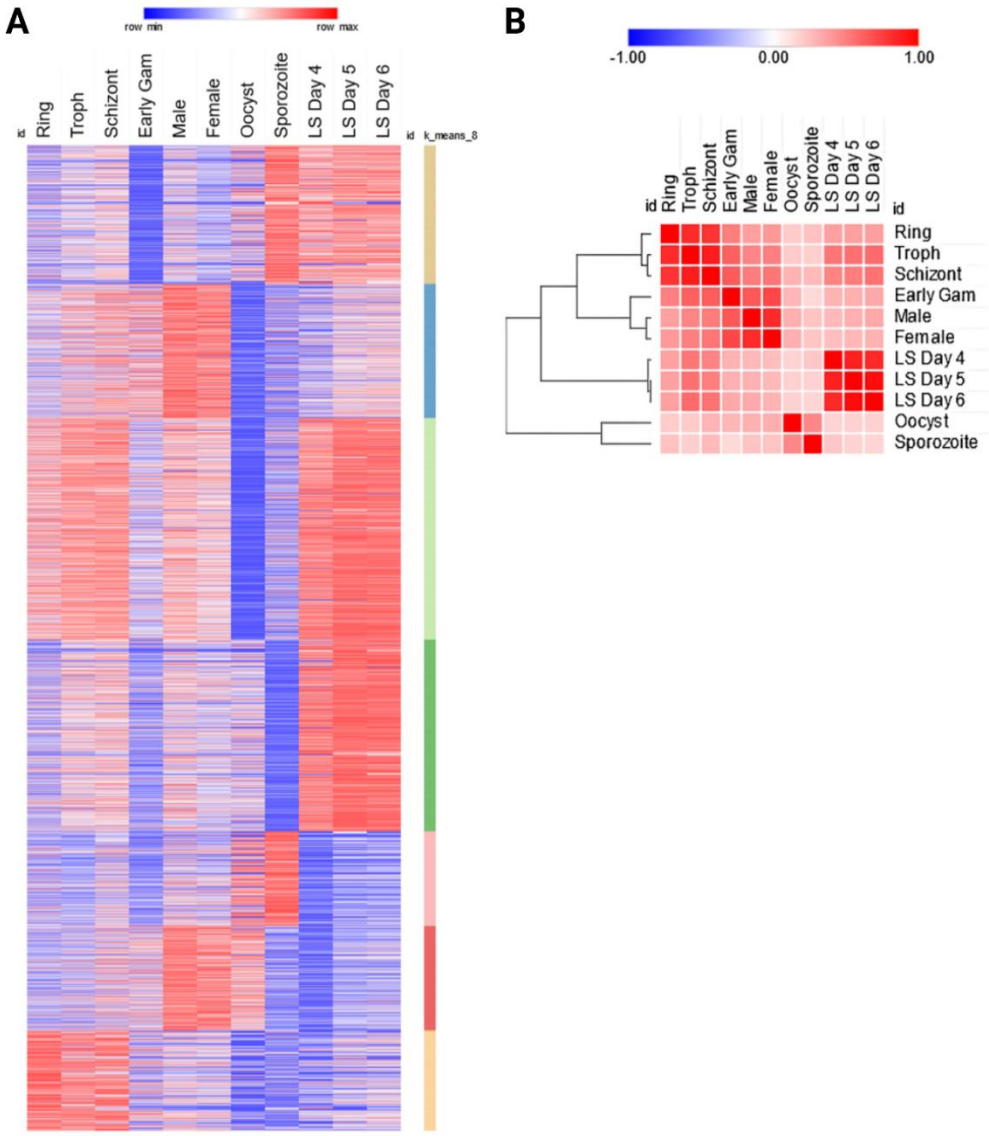

Figure 2.3 Patterns of expression across all stages of the *P. falciparum* lifecycle

A) Clustered heatmap of gene expression for each stage of development.

B) Correlation matrix of gene expression for all stages of development.

2.3 Predicting Gene Essentiality from Bioinformatics Data

Given the number of challenges with experimentally validating essentiality throughout the parasite lifecycle, I sought to identify gene features that could be predictive of essentiality. Gene features such as gene expression and proteomics can be specific to the stage of development, while other features apply to all stages, as in orthology or non-synonymous to synonymous SNP ratios. I first started by comparing the gene expression of essential and dispensable genes in the blood stage, where essentiality has been tested by a PiggyBac screen [Figure 2.4A].

The mean expression level for peak expression in blood stage was 19.51 for essential genes and 12.55 for dispensable genes ($p<0.001$). This pattern was most apparent at the highest levels of expression. In the 250 most expressed genes in BS, 84% are BS essential. Inversely, BS essential genes have lower levels of non-synonymous SNP frequencies from field data [Figure 2.4B]. In LS, BS, and GS, the 250 most abundantly expressed genes are all more likely to be essential in the BS (84%, 74%, 87% respectively), which is not true for the MS (54%) [Figure 2.4C]. On average, only 25% of the top 250 genes were shared, predominantly made up of housekeeping genes like histones. This would suggest that BS gene essentiality is not independent across the lifecycle and is indicative of functions critical to survival within the human host, but not the mosquito host.

To better understand what features are most predictive of essentiality, I trained a random forest classifier using the following features: expression data, proteomics data, non-synonymous SNPs, SNPs with stop codons, total SNPs, and orthology. The model is able to correctly classify BS essentiality with 71% accuracy and 73% precision [Figure 2.4D]. However, the model performs best at the extreme ends, suggesting that these features are predictive of BS essentiality once beyond a threshold. The top feature for prediction was expression, followed by non-synonymous SNPs, agreeing with previous observations. Although each stage contains its own unique biology, it is likely that the

evolutionary pressures that have led to these differences act on each stage of development, allowing these features to be used to generate predictions on essentiality across the lifecycle [188–190].

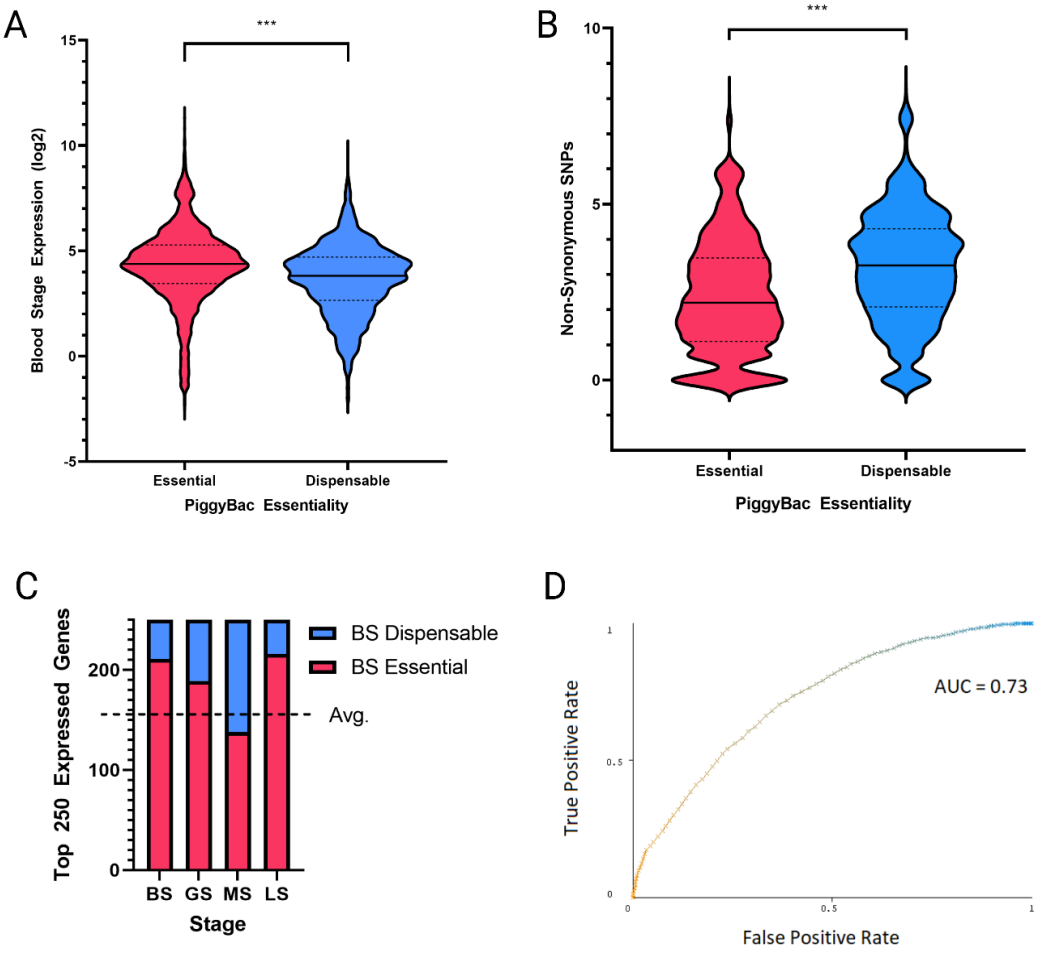

Figure 2.4 Predictors of blood stage gene essentiality

A) Blood stage expression (log2) in essential genes and dispensable genes ($p < 0.001$).

B) Frequencies of non-synonymous SNPs in essential genes or dispensable from field isolates ($p < 0.001$).

C) The frequency of blood stage-essential genes in the top 250 expressed genes for all four main stages of the parasite lifecycle.

D) Results from a random forest classifier to predict essentiality using available gene features, including transcriptomics, proteomics, non-synonymous SNPs, SNPs with stop codons, total SNPs, and orthology.

Development of genetically attenuated *Plasmodium* vaccines requires identification of genes that are necessary for liver stage development. This poses a major challenge, as any knockout generated must be grown into gametocytes, fed to mosquitos, and finally infect a humanized mouse by mosquito bite or dissected sporozoites. Further,

if the gene is essential for any stage other than liver stage, it would be impossible to produce any attenuated sporozoites for vaccination. By using available essentiality data in addition to our compiled transcriptomic dataset, we can select ideal candidates that are least likely to have a phenotype outside of the liver stage. Following the assumption that higher expression is indicative of necessity, I sorted all genes by liver stage expression. These genes can be further filtered by removing all previously putative BS-essential genes by PiggyBac screen. Of the BS-dispensable genes, there are 80 LS genes with expression >3 any other stage. These genes are ideal candidates for genetically attenuated sporozoites, and should be a high priority for future vaccine development.

2.4 Gene Regulatory Network Analysis for Uncovering Regulators of Parasite Biology

Gene regulatory networks are part of many genetic pathways and often are found at the center of developmental regulation [191–193]. Gene regulatory network models seek to use these generated transcriptomic datasets to better understand the underlying regulators of gene expression. These models often are built from large transcriptomic datasets. However, despite the availability of numerous transcriptomic datasets, no GRN models had been established in *Plasmodium*. Using over 1,000 microarray transcriptomes from patient isolates, our lab created a GRN model able to accurately predict genome-wide transcriptomics [Figure 2.5A] [159].

The GRN model is able to accurately predict both blood stage and gametocyte stage expression data, despite being trained on parasites isolated from blood samples [Figure 2.5B]. It is important to note that despite circulating blood samples primarily consisting of asexual stage parasites, gametocytes can still be found at low levels in circulating blood. Ultimately, late stage gametocytes must be present in the blood for it to pass to its mosquito host. Further, the GRN model can accurately predict targets of transcriptomic regulators, allowing for artificial gene “perturbations” (over-expression or under-expression) to measure the effects on its targets [Figure 2.5C]. I sought to test the

model's ability to predict TR targets in gametocytes using the AP2-G gene, a master regulator of gametocyte induction and development [Figure 2.6A], identifying 44 putative targets of AP2-G. Numerous labs have explored the effects of AP2-G perturbation and have generated transcriptomic datasets from over-expression and knockout of AP2-G [194–197]. Among the genes found to be differentially expressed following AP2-G over-expression, 30 are found to be targets of AP2-G by the GRN model [Figure 2.6B]. Despite not having complete overlap, there is significant enrichment ($p<0.001$).

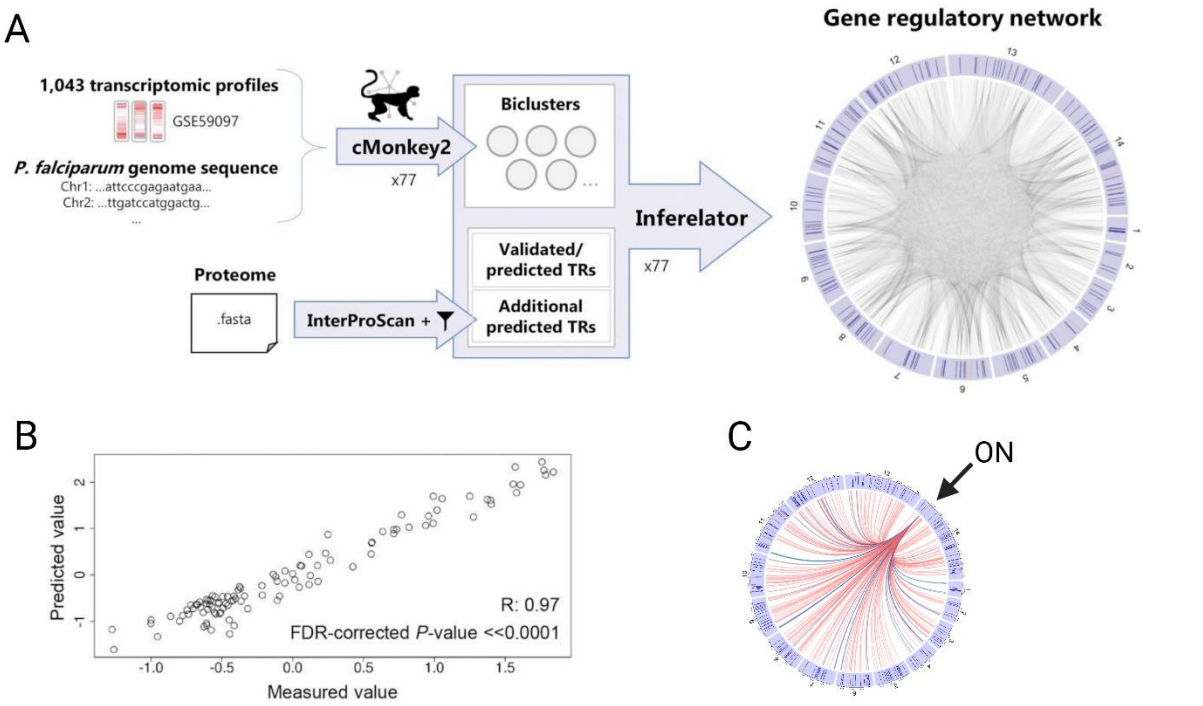

Figure 2.5 Outline of the EGRIN gene regulatory network model for *P. falciparum*
A) 1,043 microarray field isolate datasets were used to create the model. 258 putative transcriptional regulators are assigned predicted weights for each gene target.
B) The model is highly predictive of expression (R = 0.97) and can predict transcriptomes with high accuracy.
C) An example of artificially "perturbing" a TR by over-expression.

The differences can be partially attributed to the artificial induction of AP2-G, as well as the unique environment of *in vitro* parasites. It is important to note that the GRN model is trained on *in vivo* patient isolates which have yet to be lab adapted and surely encounter other environmental signals not found *in vitro*. These findings highlight the power of the GRN model in not just accurately predicting blood stage transcriptomes, but

gametocyte transcriptomes as well. Additionally, the model is able to accurately predict targets of transcriptional regulators for both blood stage and gametocyte stage.

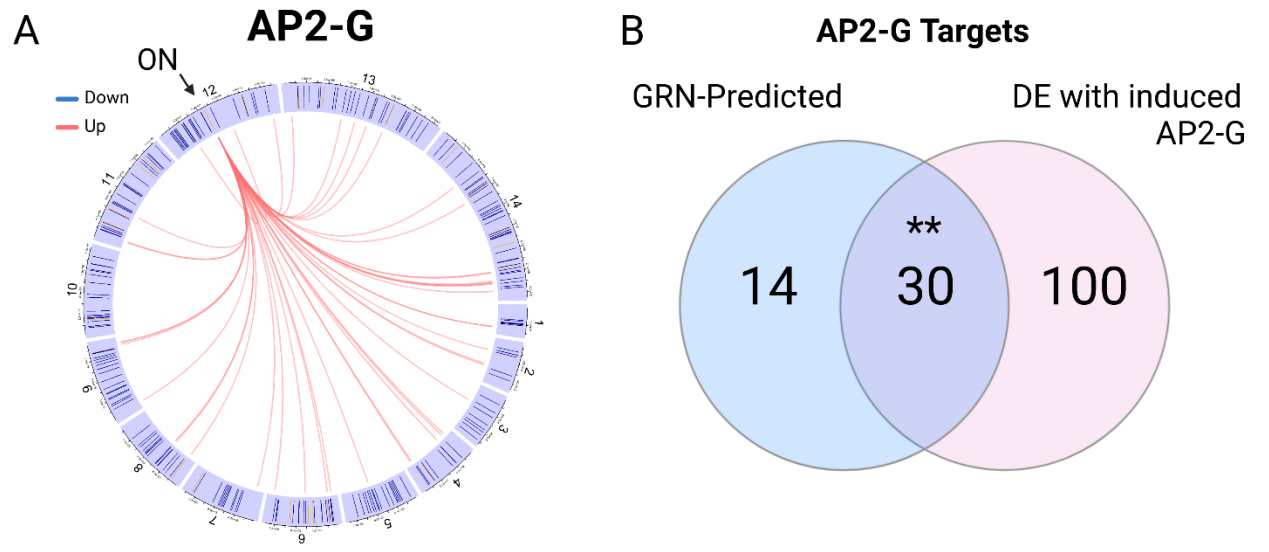

Figure 2.6 Predicted AP2-G Targets from the EGRIN gene regulatory network model
A) AP2-G is predicted by the EGRIN gene regulatory network model to positively regulate 44 genes.
B) Comparison of differentially expressed genes following AP2-G overexpression and genes predicted by the EGRIN GRN model. There is overlap of 30 genes (68% of the model's predictions) between the model's predicted genes and differentially expressed genes

While exploring gene expression patterns of artemisinin resistance, we observed an interesting subset of slow-clearing patient isolates following drug treatment. Slow-clearing parasites were more likely to have a wider, more diverse transcriptome, and there was greater heterogeneity between slow-clearing parasites. Using the GRN model, these diverse transcriptomes can be traced back to predicted transcriptional regulators that may be acting to create heterogeneous transcriptomic profiles. We identified three subsets of transcriptional regulators based on the GRN model's target predictions. One subset consists of TRs that are predicted to regulate hundreds or thousands of genes, which we refer to as 'Hub Regulators'. Another subset consists of TRs that correlate with high amounts of heterogeneity, which we call 'Variability Inducers', and inversely, some TRs which contain very narrow transcriptomes which we call 'Variability Suppressors' [Table 2.2]. However, without experimental validation, it is impossible to say these TRs

are completely causative, or simply correlative, with these patterns. In subsequent chapters, I will discuss efforts to experimentally perturb these putative transcriptional regulators and explore the effects of these perturbations.

Table 2.2 A subset of putative transcriptional regulators for hubs, variability-promoters, and variability-suppressors

Putative Transcriptional Hub Regulators	
Gene ID	**Description**
PF3D7_1239300	Conserved Plasmodium protein, unknown function
PF3D7_0621000	Conserved Plasmodium protein, unknown function
PF3D7_1326700	Conserved Apicomplexan protein, unknown function
PF3D7_0501800	Chromatin assembly factor 1 subunit A
Putative Transcriptional Variability-Promoter	
Gene ID	**Description**
PF3D7_0809900	JmjC domain-containing protein 1
PF3D7_1449300	Transcription factor IIIb subunit, putative
PF3D7_0522200	Transcription initiation factor TFIID subunit 10, putative
PF3D7_1239300	Conserved Apicomplexan protein, unknown function
Putative Transcriptional Variability-Suppressor	
Gene ID	**Description**
PF3D7_1433300	Histone-binding protein RBBP7
PF3D7_0525000	Zinc finger protein, putative
PF3D7_021600	DEAD/DEAH box helicase, putative
PF3D7_0519500	CCR4 domain-containing protein 1, putative

2.5 Discussion

I explored the transcriptomic profiles of all stages of the *P. falciparum* lifecycle. By combining datasets from across the lifecycle, comparisons between each stage can be made for the first time. A small subset of LS-expressed genes are strictly unique to the LS, while the majority are expressed across the lifecycle. There are also a number of genes abundantly expressed in all of the human host environments, implying shared biology between these stages. The most enriched set of LS genes are involved in fatty acid metabolism and acyl-CoA pathways, agreeing with prior findings in *Plasmodium* liver stage biology [185, 187, 198].

Given the importance of essential genes in drug and vaccine development, I investigated patterns of essentiality using established BS essentiality data. Numerous

features offer predictive power for identifying essentiality, including abundant gene expression and SNP frequencies from field isolates. By applying these findings to other parasite stages of development, we may better predict essentiality across the lifecycle.

I have presented a novel multi-stage dataset that can be used for further functional genomics analysis. These results offer new insights into parasite biology and gene function across the lifecycle. As improvements in informatics and transcriptomics are made, these approaches will become more valuable. The EGRIN gene regulatory network model offers a number of meaningful insights into transcriptional regulation, both in the blood stage and gametocyte stage. This model can be paired with known transcriptomic data and biology to generate new hypotheses. Ultimately however, experimental validation of all of these findings will be necessary.

2.6 Materials & Methods

Transcriptomic Dataset Handling

All RNA-seq datasets were acquired from PlasmoDB, NCBI, or directly from the original lab. For analysis, only datasets containing transcriptomes derived from the NF54 lineage, including NF54, 3D7, or any genetically modified parasites from that lineage were used. Reads were aligned using HISAT2 to the reference genome of NF54, generating transcriptomes for 5499 unique gene transcripts [199]. Samples and datasets were normalized by class-specific quantile normalization and merged into a single dataset. All further data transformations were done in Excel or Prism.

Transcriptomics Analysis

Fold change between stages was done in Excel. Gene Ontology enrichment analysis was performed using ShinyGo 0.81 [200]. Heatmaps were generated using Morpheus (https://software.broadinstitute.org/morpheus/). Clustering was performed using Morpheus or Weka [201, 202]. All statistical analyses were performed in GraphPad Prism 8 (GraphPad Software) using an unpaired t-test. Datasets for training predictive models

were acquired from PlasmoDB. Random forest modeling was performed using scikit-learn in Python or using Weka [203].

Chapter 3 – Development and Application of a Conditional Cas9 System for Homology-Directed Gene Editing Across the *P. falciparum* Lifecycle

A version of this chapter has been submitted for publication.

Sean T. Windle, Maxwell L. Neal, Fred D. Mast, Stefan H. I. Kappe, and John D. Aitchison

3.1 Introduction

CRISPR/Cas9 has transformed the entire field of biological research, greatly simplifying both genetic manipulation and the use of genetic screens [51, 67, 79, 204]. Cas9 is a powerful tool for targeted editing of a desired locus, allowing for the creation of indel-mediated knockouts as well as larger edits via homology-directed repair. As an endonuclease, Cas9 works by creating a double-stranded break at a target site specified by a 20-bp guide RNA sequence proximal to a short recognition sequence (PAM). In some cases, double-stranded breaks can occur in other locations in the genome, leading to deleterious off-target effects [205, 206]. A number of efforts have been made to improve the efficiency of various aspects of Cas9 editing, including: improved indel rates, improved HDR rates, lower off-target effects, expanded PAM recognition, and a number of Cas9-effector fusions with their own unique functions [207–210]. Despite the rapid advancements in the field of molecular biology, the malaria field has lagged in adopting these tools. And yet, many of these tools would provide great value to the malaria field due to the many limitations and challenges inherent in current approaches to genetic manipulation in malaria research.

Cas9 editing has become the default approach for editing the *Plasmodium* genome. *Plasmodium's* lack of NHEJ pathways is a mixed blessing for editing in the parasite. Although the creation of simple indel knockouts is unavailable, HDR editing is highly efficient in this organism and allows for large gene insertions, deletions, or replacements using a donor sequence with homology arms [49]. This means, however, that even creating simple gene knockouts requires multiple steps of plasmid cloning. The sgRNA

encoding the unique 20bp gRNA sequence must be inserted into the plasmid, as well as a donor sequence containing both a 5' and 3' homology arm of 50bp or greater in length. For gene knockouts, it is common to use the 5' and 3' UTRs as homology arms, omitting the coding sequence of the gene. However, due to the AT-rich genome, these UTRs are typically 85% AT-rich and are appreciably more challenging to amplify and clone than most DNA. Although Gibson assembly cloning methods can simplify this process by including 20-30bp of homology on the primers and cloning in one step, mutations and inefficiency still pose a major problem [211]. Nevertheless, generating a gene knockout in *Plasmodium* from beginning to end can take many months and is not a trivial undertaking.

Another layer of difficulty is added by the extremely low efficiencies of electroporation and plasmid uptake by *P. falciparum* [212]. Although other *Plasmodium* species have considerably higher rates of plasmid uptake by transfection, *P. falciparum*'s is sufficiently low that an unsuccessful transfection is considered normal. This can lead to confounding results when trying to knockout a gene, as there is no way of differentiating between a knockout that has a deleterious phenotype and a failed transfection. A solution to this problem was the development of conditional expression systems. A number have been developed for *P. falciparum* and *P. berghei*, including the TetR-DOZI aptamer system, the glmS ribozyme system, and the dimerizable Cre system [213–215]. Although each has their advantages and disadvantages, they are all commonly used for exploring knockout phenotypes in the blood stage, as well as other stages of the lifecycle. However, in all cases, generating these mutants requires a large amount of work and can lead to confounding results due to the limitations of the systems. There is still a pressing need to establish an accessible and effective system for studying phenotypes across the lifecycle.

I tested multiple inducible systems for Cas9 with the goal of creating a tool for inducible control of Cas9 activity and gene editing across the parasite lifecycle. The most successful strategy utilizes a dimerizable Cre recombinase and lox-STOP-lox (LSL) cassette

inserted into the Cas9 5'-UTR region. I further modified this system to allow for paired gRNA-donor sequences that could be commercially synthesized. Finally, I tested two candidates predicted by bioinformatics analyses to be essential to gametocyte development. This platform offers an accessible means of generating mutations in blood and gametocyte stages that can be used for gene knockouts followed by phenotyping.

3.2 Exploring Inducible Systems for the Conditional Expression of Cas9

A number of inducible Cas9 systems are now available for use in mammalian systems [106–108, 110, 111]. Presumably, many of these systems could be adapted for use in *Plasmodium* research. Inducible promoters have been around for many years and have been applied to thousands of different applications, including Cas9 systems. Unfortunately, the majority of these inducible promoters are incompatible with *Plasmodium*. However, the TetR-DOZI system, which involves addition of degradation aptamers to the 3' and/or 5' end, is well established, and efforts have been made to minimize the aptamers' tendency to recombine, reducing background expression. My initial approach involved using this system to create a controllable Cas9.

To easily measure expression of Cas9, I created a Cas9-GFP fusion. Building upon an established TetR-DOZI plasmid, I replaced the existing aptamer-controlled coding sequence with Cas9-GFP, and also added a U6 promoter driving expression of a donor sequence for a validated gene knockout [Figure 3.1A]. Following transfection, the GFP signal was examined by fluorescence microscopy. However, even in the absence of anhydrotetracycline, GFP signal was clearly visible and sequencing of the target gene showed complete gene knockout [Figure 3.1B]. Sequencing of the aptamers showed no recombination, so it is likely that the background expression of Cas9 is simply too high for inducible activation.

An alternative approach to inducibly controlling Cas9 is through direct interactions on the actual Cas9 protein. The benefit of this approach is that there is little delay

between chemical induction and protein activity. One such system is the ciCas9, or chemically inducible Cas9 [Figure 3.1C] [110, 111]. This system relies on an interaction between Bcl-xL and a BH3 peptide that are fused to separate ends of Cas9. In the absence of the Bcl-xL inhibitor A115, Bcl-xL and BH3 bind tightly, thereby disrupting Cas9 activity. When A115 is added, the inhibiting interaction is broken and Cas9 becomes catalytically active. An added benefit is that this interaction is tunable and highly responsive to A115 and requires only a small modification to the Cas9 expression vector, i.e. replacement of the SpCas9 coding sequence with ciCas9 sequence in our one plasmid system.

I therefore created a ciCas9 plasmid encoding the same gRNA as in the previous experiment. Following transfection, I tested six concentrations of A115, ranging from 0.1 uM to 10 uM with a 24h exposure. However, all six concentrations had an immediate effect on parasite growth, and at 10 uM resulted in near complete parasite killing [Figure 3.1D]. This was particularly interesting as *Plasmodium* contains no Bcl-xL orthologs or proteins containing similar structures. A recent finding, however, identified Bcl-xL as an essential host protein for *P. falciparum* survival in red blood cells, and a previous finding

identified Bcl-xL inhibitors that prohibit *Plasmodium* liver stage infection [216–218]. Thus, the ciCas9 approach to inducible Cas9 was abandoned.

I explored using the dimerizable Cre recombinase approach as a means of

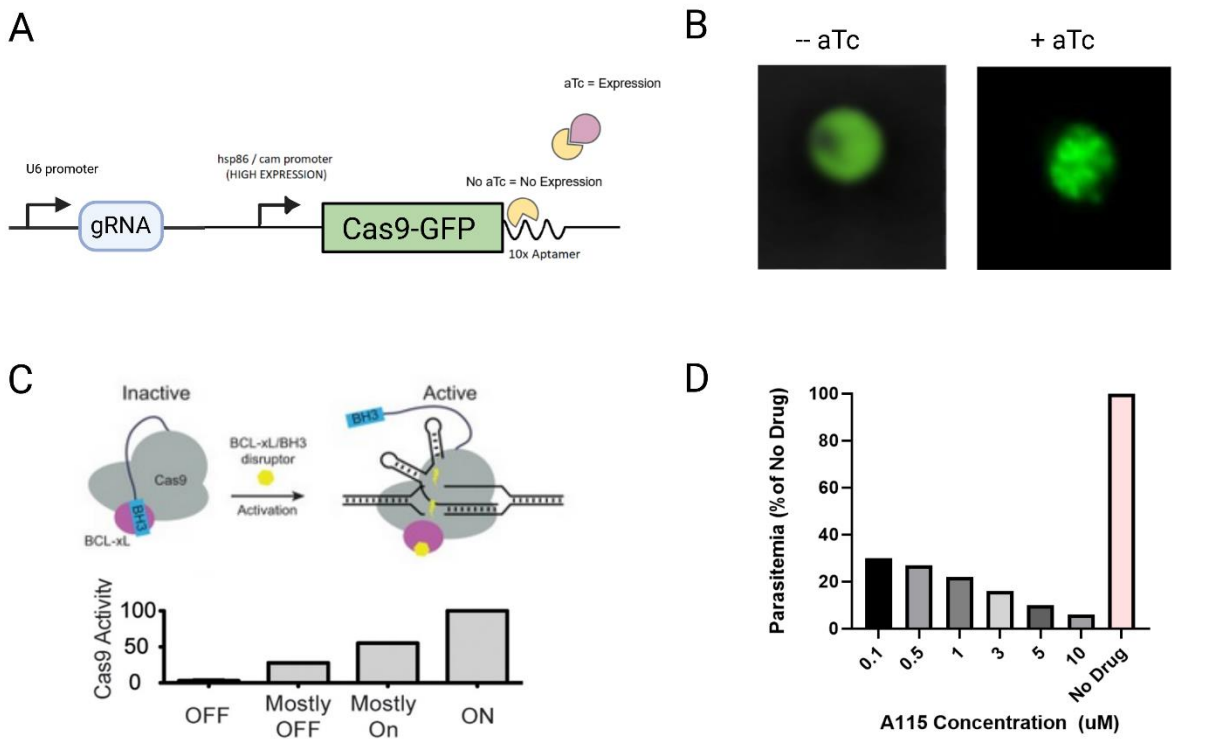

Figure 3.1 TetR-DOZI-inducible Cas9 and BCL-xL chemically inducible Cas9 systems
A) Design containing U6 promoter driving gRNA and hsp86 promoter driving Cas9-GFP with and without aTc. Without aTc, Cas9-GFP mRNA is degraded. With aTc present, TetR-DOZI stops degrading mRNA, allowing Cas9 to become active.
B) GFP visible in both -aTc and +aTc treated cells.
C) Schematic showing BCL-xL/BH3 interaction blocking Cas9 activity. Addition of BCL-xL inhibitor A115 disrupts the BCL-xL interaction, activating Cas9.
D) Parasite growth relative to control in presence of varying concentrations of A115.

controlling Cas9 expression. Building upon the design used for inducible AP2-G activation, I sought to create a DiCre rapamycin-inducible Cas9 system by inserting the loxP-hDHFR-TERM-loxP cassette upstream of Cas9 [196]. Concurrently, another group developed an inducible CRISPRa and CRISPRi system that utilized DiCre to turn on dCas9 [219]. Coincidentally, they had also tested the TetR aptamer system and similarly concluded that background expression was too high.

Rather than redesign a plasmid, I adopted their recently published system which utilized GFP and the hsp86 terminator flanked by loxP sites located upstream of dCas9. In this system, the hsp86 promoter drives expression of GFP until rapamycin-mediated Cre excises the GFP and terminator, allowing expression of dCas9. I first began by replacing dCas9 with catalytically active high fidelity Cas9, SpCas9-HF. To better visualize changes in activity, I replaced GFP with a codon optimized mNeonGreen (mNG). I then moved to the hsp86-gfp-Cas9 cassette to a plasmid containing the U6 promoter while removing Esp3I restriction sites needed for gRNA cloning, creating the DCL3-Cas9 plasmid [Figure 3.2A].

Initial testing of this system in an NF54-DiCre parasite line showed highly effective control of Cas9 induction and excision of the GFP [Figure 3.2B]. Rapamycin was most effective at 50nM for 6h while showing no toxicity. Complete loss of the mNeonGreen signal often required until the next red blood cell cycle, as mNG is highly stable. Within 48 hours from initial treatment, roughly 70% of parasites had lost mNG signal [Figure 3.2C]. By 96 hours, 96% of parasites showed no mNG signal. Despite washing rapamycin off at hour 6, it is possible that residual rapamycin persists, continuing to convert parasites from mNG+ to mNG-. As it is not ideal to leave rapamycin for longer than necessary, this was a welcomed feature. Cas9 expression by qRT-PCR increased 239-fold over 96h following rapamycin treatment, suggesting near leak-free control of Cas9 [Figure 3.2D].

As a proof of concept to test inducible HDR, I added a validated gRNA and donor sequence targeting Circumsporozoite Protein (CSP), a protein dispensable for both blood and gametocyte stages. This donor consists of ~1050bp encoding ~525bp of the CSP 5'UTR and ~525bp of the 3'UTR but omitting the CSP coding sequence. Addition of rapamycin should remove mNG while turning on Cas9, allowing for Cas9 to cut at the CSP gRNA site and forcing repair through the HDR donor sequence [Figure 3.3A]. The donor specifically

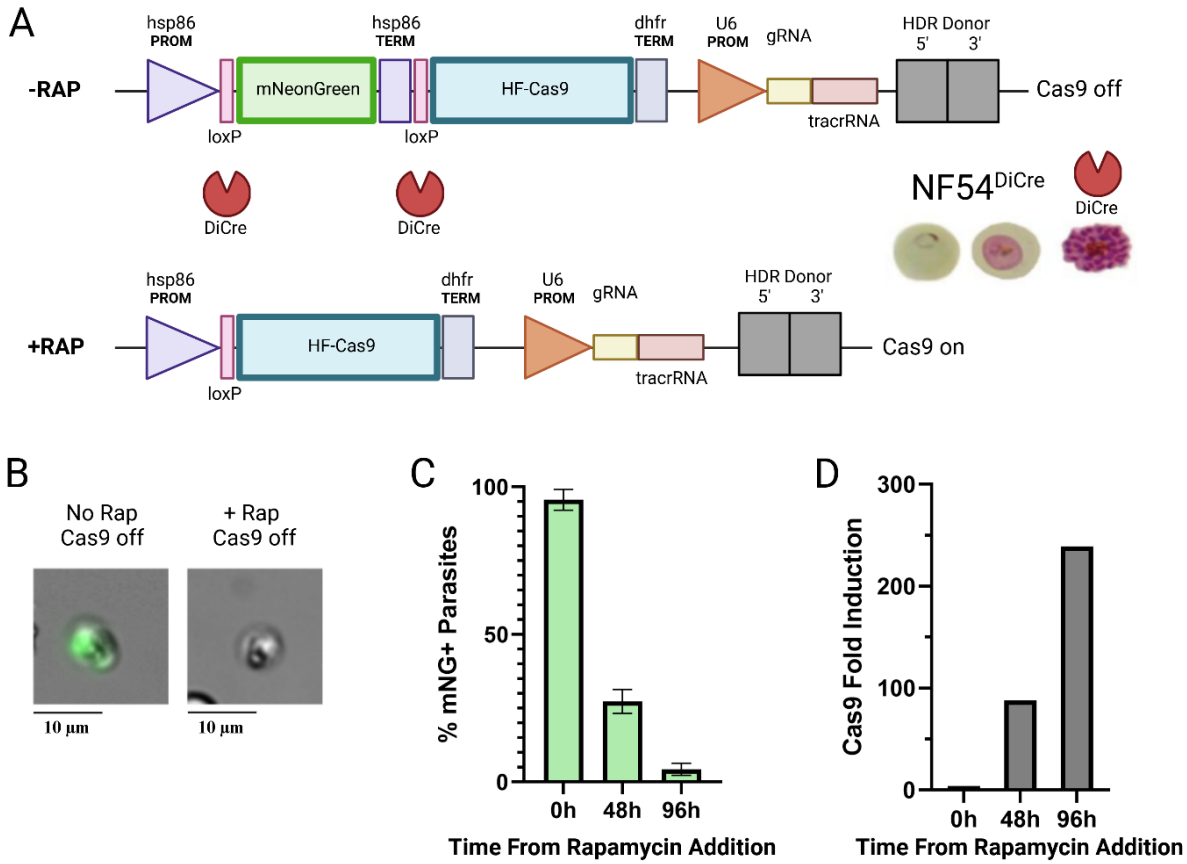

Figure 3.2 A DiCre-inducible system for Cas9 expression

A) Schematic showing the DiCre-inducible Cas9 system. The Hsp86 promoter is followed by a loxP-flanked mNeonGreen and Hsp86 terminator, creating a loxP-STOP-loxP cassette upstream of a high-fidelity Cas9. When used in the NF54^{DiCre} parasite line, addition of rapamycin activates DiCre, excising the STOP element between the two loxP sites and activating expression of Cas9. The plasmid also contains a U6 promoter driving gRNA expression, allowing for targeted cleavage by Cas9, and an HDR donor template that allows for repair of the cleaved DNA.

B) Fluorescent microscopy of representative trophozoites before and after addition of rapamycin, showing that mNG expression has been extinguished. Bar – 10 µm.

C) Percent of mNG-positive parasites at 0h, 48h, and 96h following addition of 50nM rapamycin. Parasites were tightly synchronized to early ring stage before rapamycin addition. Data shown represents the means of three replicates ± SEM.

D) Fold-induction of Cas9 relative to uninduced at 0h, 48h, and 96h following addition of 50nM rapamycin. Expression was measured by SYBR Green RT-qPCR assay for Cas9 mRNA at respective timepoints.

does not contain any sequence from the gRNA, preventing continuous cutting once repaired. If parasites do not repair with the donor, the double-strand break will eventually lead to cell death. However, *Plasmodium* parasites do not appear to have another form of efficient repair, likely leading to much higher HDR rates than in mammalian systems [49].

Synchronized ring-stage parasites were induced for 6h at 50nM Rapamycin before washing. Parasites were sequenced each cycle (48h) to determine the rate of editing over time. At day 0 (uninduced), there was minimal editing from WT sequence to KO. By day 2, ~20% of parasites were edited. By day 4, parasites had converted to ~85% KO sequences [Figure 3.3B]. This temporal delay may be in part due to the inducible system, as well as the activity of parasite repair pathways.

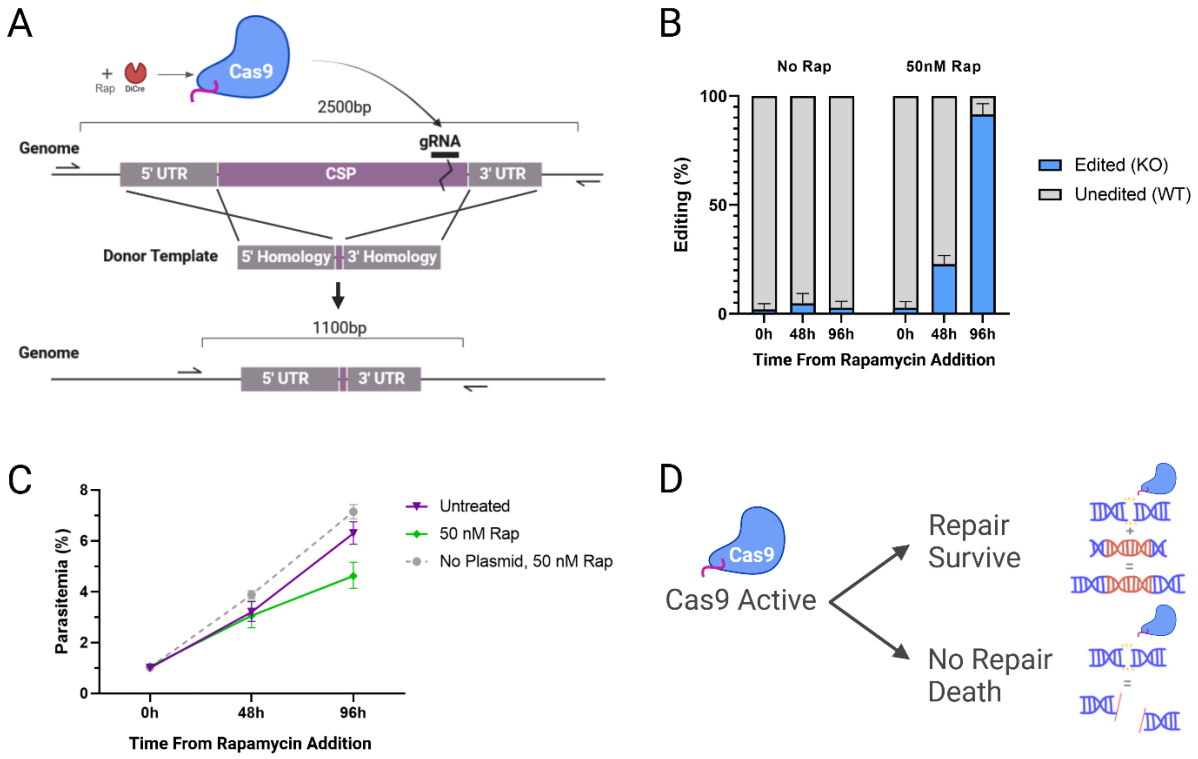

Figure 3.3 DiCre-inducible CSP editing in blood stage parasites

A) Rapamycin activates DiCre, which in turn activates Cas9 expression and gRNA-directed cleavage at the 3' end of the CSP gene. A donor template with homology to the 5' and 3' UTRs of CSP but lacking the coding sequences is used for repair, creating a CSP knockout. Two external primers are used for amplification of the unedited (~2500 bp) or edited (~1100 bp) CSP locus, allowing for quantification of editing rates following induction.

B) Inducible editing rates at 0h, 48h, and 96h after adding 50nM rapamycin or DMSO. Correctly edited alleles contain the intended CSP deletion, while unedited alleles are WT CSP. Data shown represents the means from three experimental replicates ± SEM. (* $p < 0.05$).

C) Growth rates following addition of either 50nM rapamycin or 0nM (untreated) to tightly synchronized ring stage parasites at 1% starting parasitemia. Untreated plasmid-free parasites were included as a control. Data shown represent the means from three or more replicates ± SEM.

D) Diagram showing the binary outcomes following Cas9-induced double strand breaks in the genome. Parasites that repair the break using the donor template survive, while parasites that fail to repair die.

In this system, DiCre must first be activated, allowing for binding and cleavage of loxP sites, removing the mNG-terminator. This in turn shifts expression from mNG to Cas9. Cas9 must then localize to the gRNA site and create a double-strand break. Only then can the parasite repair this break using the donor sequence and HDR.

Although there was no immediately recognizable toxicity, I was interested in whether Cas9 activity had any effect on growth or cell death. At 48h, there was no significant difference in parasitemia or morphology. At 96h, a ~35% decrease in parasitemia was observed, indicating some amount of toxicity or growth defects due to Cas9 [Figure 3.3C]. As the double-strand break repair must occur before replication, parasites may not survive or replicate in cases where Cas9 cleavage occurs before or during replication [Figure 3.3D]. This is particularly relevant, as HDR machinery is most active in stages of active replication. It is possible that parasites must sacrifice growth or replication for repair of the cut DNA.

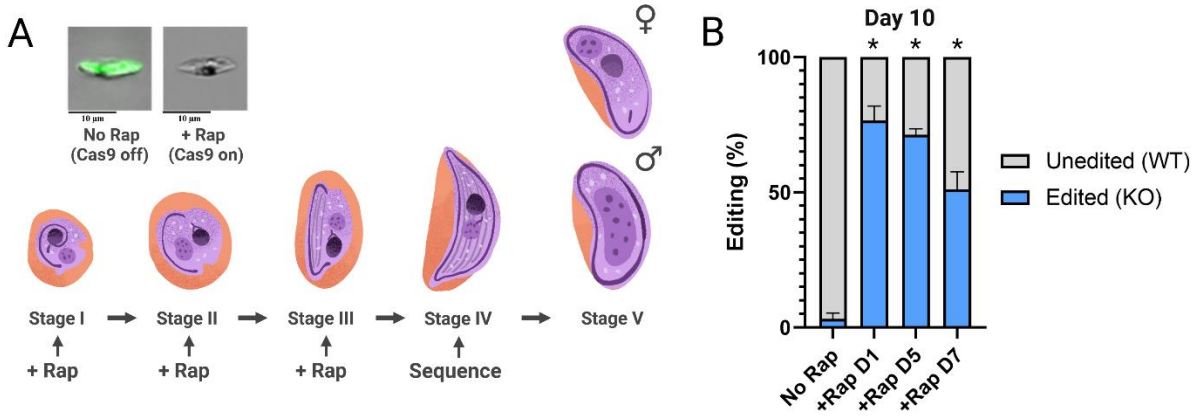

Figure 3.4 DiCre-inducible CSP editing in gametocytes

A) Timeline for inducible Cas9 editing in gametocytes. 50nM rapamycin or DMSO (No Rap) was added on Day 1, Day 5, or Day 7 following gametocyte induction. DNA was isolated from parasites on Day 10 and sequenced.

B) Efficiency of CSP editing to the KO allele in gametocytes at Day 1, Day 5, or Day 7 after addition of rapamycin or DMSO, as measured on Day 10. Data shown represents the means from three or more replicates ± SEM. (* $p < 0.05$)

These high editing efficiencies showed promise for conditional editing in other stages of the parasite. I next wanted to test whether inducible editing could occur in gametocytes. Rapamycin was added to committed gametocytes at 3 timepoints (D1, D5, D7) and measured until Day 10 [Figure 3.4A]. Gametocyte cultures were treated with

GlcNAc for 5 days beginning on day 0 to selectively kill BS parasites. Gametocytes retain some of their HDR capabilities despite not actively dividing. By Day 10, up to 80% of gametocytes were successfully edited to CSP KO [Figure 3.4B]. Although a small amount of background may be present from residual BS parasites, the majority of gametocytes are able to perform HDR and appear morphologically normal.

However, there are some limitations with this system. As it relies on an episomal plasmid, drug selection must be continued to ensure the plasmid is maintained. Additionally, it was observed that long periods (70+ days) in culture led to higher uninduced editing. These edited parasites still retained mNG expression, as flow sorted mNG parasites had the same background editing rate. One possible explanation for this is that the plasmids are being maintained as complex concatemeric forms. It has been shown that plasmids maintained episomally form large concatemers, effectively containing multiple copies [220]. With stronger concentrations of BSD selection, parasites had higher plasmid copy numbers. As this system depends on maintaining the plasmid until induction, it is mandatory to continuously select for parasites carrying the plasmid. This may allow for the creation of larger and more complex concatemers that could unpredictably recombine or erroneously express the Cas9. However, induction of the remaining unedited parasites is still effective, and if the mutation were deleterious, it would be selected against. As CSP is dispensable for the BS, there is little to no selection against these background mutations.

Despite these limitations, these results show that the DiCre-inducible Cas9 is a promising system for stage-specific gene disruption. Although this approach requires no more additional work than normal Cas9 HDR knockouts would, it is still impractical to design Cas9 editing plasmids at large scale.

3.3 Improving the Accessibility and Scalability of the Inducible Cas9 System

To improve the scalability of the system, we sought to overcome the limitations associated with cloning individual gRNAs and inserting two fragments encoding the donor sequence's homology arms, a process that requires multiple rounds of PCRs, ligations, and sequencing for each target site and can take weeks to complete.

Recent CRISPR-based approaches in *S. cerevisiae* and *T. gondii* utilized short (<300bp) gene fragments to perform HDR screens [221]. This strategy allows one to synthesize the 20bp gRNA as well as its respective HDR donor, pairing the synthesis and cloning steps. However, in these studies, very short (30-75bp) homology arms were required in order to fit into the <300bp fragments. Advances in DNA synthesis now allow for the affordable synthesis of 500–1000 bp DNA fragments, overcoming this limitation. To adapt this approach, we modified the DiCre-inducible Cas9 plasmid by removing the 85 bp tracrRNA originally following the gRNA sequence (Figure 3.5A). This modification was necessary for contiguous pairing the gRNA and donor sequences without intervening sequences between the U6 promoter, the gRNA, and the tracrRNA. The tracrRNA can then either be included with the gRNA-donor synthesis, or cloned separately using Type IIS restriction enzymes.

Given the affordability of DNA synthesis, we opted to include the tracrRNA in all synthesized gRNA-donor pairs. An additional benefit of this approach is the potential for generating gRNA-donor libraries targeting tens or hundreds of genes simultaneously, providing a scalable and efficient alternative for high-throughput genome editing. We assessed the efficiency of HDR-mediated gene deletion using the same gRNA and CSP donor, but varying homology arm lengths (~80bp, ~200bp, or ~500bp for each arm) (Figure 3.5B). At 96h, all 3 donor sizes had significant increases in editing efficiency with knockout rates at 15%, 49%, and 80% respectively. Notably, the 80bp homology arms resulted in significantly lower editing efficiency compared to 200bp or 500bp. It is typical that reduced homology arm length results in reduced HDR in other systems, and in this

case, this reduced efficiency may also be compounded by a limited accessibility of the episomal repair template for HDR, particularly when homology arms are short [222–224].

Although donor DNA synthesis is an affordable choice, the highly AT-rich nature of *P. falciparum*'s genome presents a challenge in synthesis, often restricting homology arm lengths to ~100bp. Despite this constraint, commercial synthesis streamlines molecular cloning, and facilitates the development of pooled or arrayed screens. Altogether, these results demonstrate that DiCre-inducible Cas9 is a robust and scalable system that can be used for efficient editing even with small donor templates.

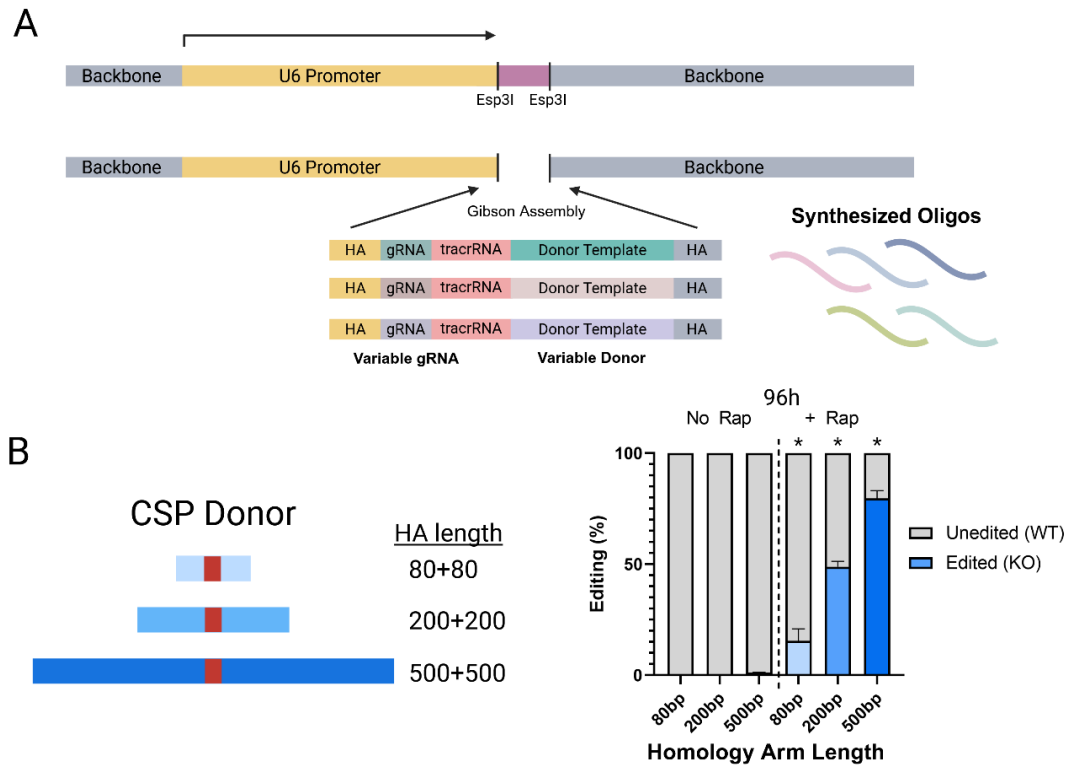

Figure 3.5 Adaptations to the DiCre-inducible Cas9 system for scalable approaches

A) The DiCre-Cas9 plasmid has had the tracrRNA removed, allowing for insertion of a complete gRNA-tracrRNA-donor template oligo by Gibson assembly. Each synthesized oligo contains 25-30bp homology arms (5' and 3') for cloning by Gibson assembly, a specific 20bp gRNA, an 82bp universal tracrRNA, and a variable-length donor template. gRNA-tracrRNA-donor oligos can be synthesized individually, in oligo pools, or in arrays.

B) Testing CSP editing efficiencies using donor homology arms of varying lengths. Three donor templates with homology arm lengths of 80bp, 200bp, or 500bp were used to delete the open reading frame of CSP. Tightly synchronized rings were treated with 50nM rapamycin or DMSO and sequenced 96h later. Data shown represent the means from three replicates ± SEM. (* $p < 0.05$)

3.4 Measuring the Phenotype of Gene Candidates in Blood and Gametocyte Stage using DiCre-inducible Cas9

Using transcriptomics datasets and a gene regulatory network model we developed, we identified two previously uncharacterized genes, PF3D7_1454900 and PF3D7_1316300, that are highly expressed in gametocytes with expression higher in female gametocytes (Figure 3.6A) [179]. Both genes were putatively found to be putatively indispensable in a previous PiggyBac transposon screen [70]. PF3D7_1454900 encodes a 154 amino acid residue protein, while PF3D7_1316300 encodes a 104 amino acid residue protein. Both genes are moderately conserved among *Plasmodium* species, with at least 50% sequence similarity between their *P. berghei* and *P. yoelii* orthologs.

Taking advantage of the accessibility of paired gRNA-donor synthesis, we designed gRNAs paired with the longest HDR donor possible for DNA synthesis (400bp for PF3D7_1454900, and 250bp for PF3D7_1316300) (Figure 3.6B). Rather than whole gene deletion, we engineered donor sequences to generate a small 8bp deletion and frameshift mutation, while simultaneously disrupting the gRNA site. To assess editing efficiency in blood stage parasites, we induced Cas9 with rapamycin in 0-3h synchronized ring stage parasites and measured parasitemia and editing efficiencies at 96h (Figure 3.6C). If the frameshift is not deleterious, even low efficiency editing should eventually result in a largely pure population of edited parasites. By 96h post-rapamycin, parasites appeared morphologically normal for both PF3D7_1454900 and PF3D7_1316300. However, in both cases, the parasitemia was significantly reduced, with decreases of 73% and 65%, respectively (Figure 3.6D). As observed with CSP, parasites continued to grow normally until the following cycle, at which point they either arrested in growth or stalled in development. The proportion of edited parasites in both PF3D7_1454900 and PF3D7_1316300 was 15.7% and 22.3% respectively, suggesting some amount of negative selection may be acting on the mutants (Figure 3.6E). Considering that both conditional

knockouts appeared morphologically unaffected, there may be some impairment in the ability for merozoites to invade, leading to a more sizeable decrease in parasitemia.

To assess gametocyte phenotypes, gametocyte commitment was induced by overgrowth, and a pure gametocyte population was selected for using 50mM GlcNAc. Parasites were treated with rapamycin at Day 5 following removal of GlcNAc and analyzed

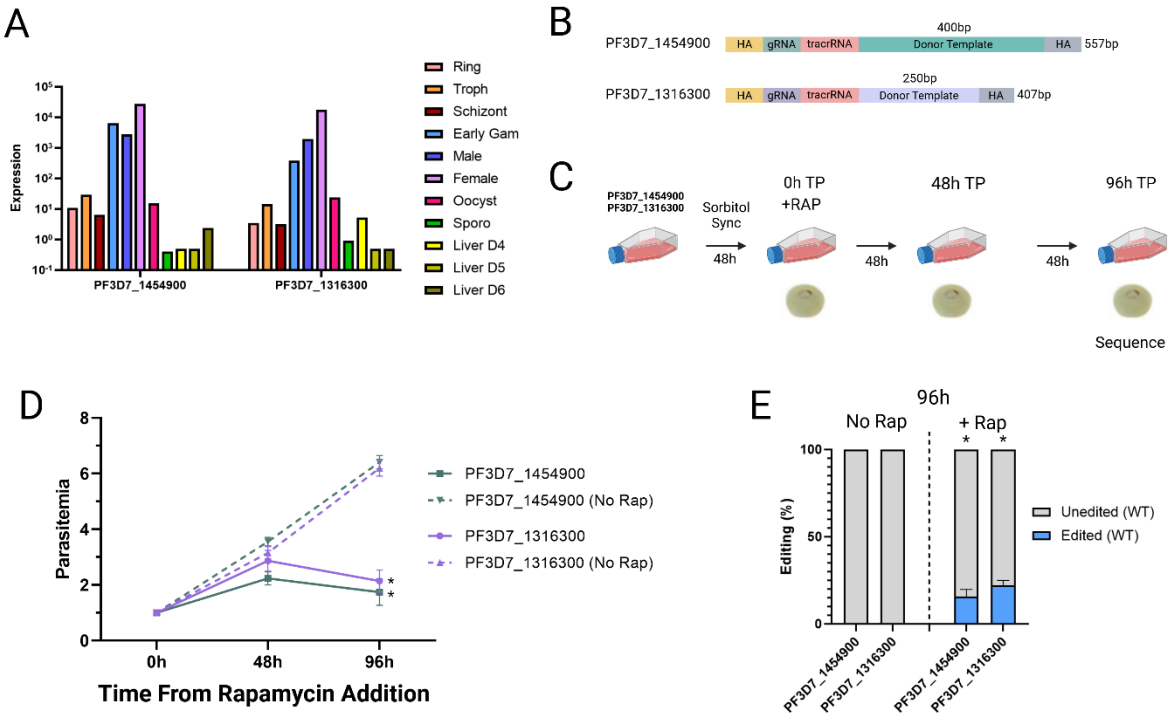

Figure 3.6 Measuring blood stage essentiality of two gene candidates using DiCre-inducible Cas9

A) Normalized gene expression over the parasite lifecycle from compiled transcriptomic datasets for PF3D7_1454900 and PF3D7_1316300. We used 11 broad timepoints, with 3 timepoints in the blood stage, 3 timepoints in the gametocyte stage, 2 timepoints in mosquitos, and 3 timepoints in the liver stage. Both genes were putatively essential by PiggyBac transposon screen (7).

B) Synthesized gRNA-donor pairs for gene candidates PF3D7_1454900 and PF3D7_1316300. The synthesized oligo for PF3D7_1454900 used a 400bp donor template (200bp for 5', 200bp for 3'), with a total length of 557bp. The synthesized oligo for PF3D7_1316300 used a 250bp donor template (141bp for 5', 109bp for 3'), with a total length of 407bp.

C) Timeline for testing blood stage phenotypes. Each transfection was first tightly synchronized to early ring stage parasites before treatment with 50nM rapamycin. Cultures were maintained for 96h (2 cycles), at which time parasitemia was assessed and sequencing was performed to quantify successful editing.

D) Parasitemia following rapamycin-induced targeting of PF3D7_1454900 and PF3D7_1316300 at 96h post treatment. Data shown represent the means from three replicates ± SEM. (* $p < 0.05$)

E) Efficiency of PF3D7_1454900 and PF3D7_1316300 editing to generate the KO alleles in blood stage parasites 96h post rapamycin treatment, or no rapamycin treatment. Data shown represent the means from three replicates ± SEM. (* $p < 0.05$)

at Day 14, prior to peak expression (Figure 3.7A). For both genes, rapamycin treatment significantly reduced the number of viable gametocytes (Figure 3.7B). While some loss may be due to gRNA inefficiencies, a higher proportion of surviving gametocytes would be expected if this were the primary factor. Morphologically, many parasites resembled dying gametocytes in late stages (Stages 4 to 5) of gametocyte development, with the most pronounced effect in female gametocytes for both genes. These findings suggest essentiality of PF3D7_1316300 and PF3D7_1454900 in mid-to-late gametocytes, particularly in female gametocytes.

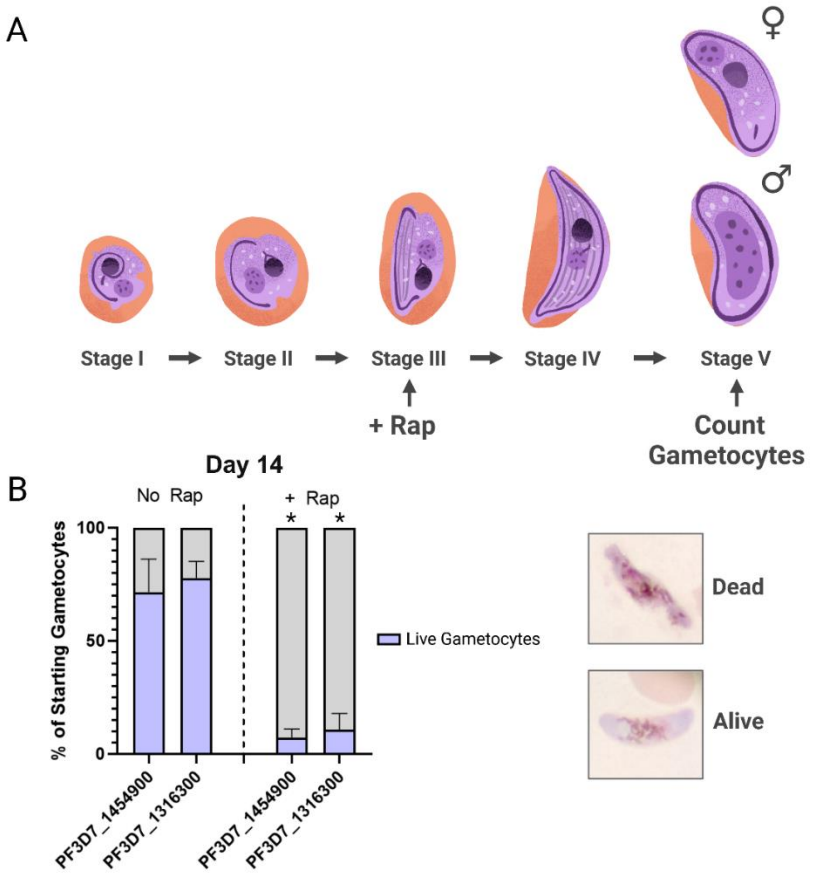

Figure 3.7 Measuring gametocyte-stage phenotypes in parasites with knockouts of two gene candidates using DiCre-inducible Cas9

A) Timeline for rapamycin treatment and sequencing to measure gametocyte phenotypes. Induced gametocytes were treated with 50mM GlcNAc for 5 days. On Day 5, 50nM of rapamycin was added and gametocytes were cultured as normal until Day 14, at which point gametocytes were counted by Giemsa staining.

B) Proportion of starting (Day 5) gametocytes with or without rapamycin induction, counted on Day 14. Live gametocytes retained normal morphology, whereas dead gametocytes began to break apart, condense, or display abnormal morphologies. Data shown represent the means from three replicates ± SEM. (* $p < 0.05$)

3.5 Discussion

The development of conditional systems for genetic manipulation in *P. falciparum* has expanded functional gene characterization across its lifecycle. We sought to improve upon these strategies by developing a conditional Cas9 system that enables introduction of any desired mutation in an accessible and scalable manner. Rather than requiring multiple independent cloning steps, both the gRNA and its respective donor can be commercially synthesized and cloned in a single step. Using the absence of mNeonGreen signal as an indirect marker of Cas9 expression, parasite phenotypes can be measured by fluorescence microscopy or flow cytometry, as well as sorted by FACS. This system enables inducible editing in both blood and gametocyte stage parasites, can be applied using small homology arms for Cas9 donor templates, and can be readily scaled up using a paired gRNA-donor approach. We confirmed that the DiCre-Cas9 system facilitates conditional knockout of genes and provides measurable phenotypes in both blood and gametocyte stages. Furthermore, we identified stage-specific essentiality phenotypes in two previously uncharacterized *P. falciparum* genes, PF3D7_1316300 and PF3D7_1454900.

Both conditional knockouts exhibited modest blood stage phenotypes, but had stronger defects in gametocyte stages. These results are in agreement with transcriptomic data, showing peak expression in mid to late gametocytes, particularly in female gametocytes, where we observed the most severe phenotype. Notably, a recent study identified the *P. yoelii* ortholog of PF3D7_1454900 (PY17X_1322400) as a putative apical polar ring (APR) partner protein using TurboID labeling of APR2 and ARA1 [225]. These proteins localize to the apical polar ring in *P. yoelii* ookinetes*,* suggesting a role in microtubule structure and regulation. PY17X_1322400 knockout parasites displayed decreased ookinete motility and near complete deficiency in oocyst development in the mosquito midgut.

While the genes appeared dispensable in blood and gametocyte stages in *P. yoelii*, this may be due to inherent differences in parasite biology between *P. falciparum* and *P. yoelii*. For instance, gametocytes in *P. falciparum* require an additional 7 to 9 days to mature, and undergo more distinct morphological development than *P. yoelii*. Additionally, *P. falciparum* gametocytes form a complete inner membrane complex with elongated microtubules whereas *P. yoelii* forms incomplete inner membrane complexes during gametocyte development [226, 227]. Based on the available evidence, PF3D7_1454900 may be involved in microtubule formation during brief periods of merozoite invasion and throughout gametocyte development, consistent with the observed phenotypes. However, more thorough study of both PF3D7_1316300 and PF3D7_1454900, is needed to elucidate their precise functions.

This study also evaluated the efficiency of different homology arm lengths for homology-directed repair and editing. CSP editing was least efficient with 80 bp homology arms and most efficient with 500 bp arms, consistent with findings from *P. berghei* and mammalian Cas9 editing systems [223, 224]. Although the donor templates are provided as episomal plasmids, they likely form concatemeric structures with unknown copy numbers [220]. While *P. falciparum* lacks non-homologous end joining pathways and preferentially repairs double-strand breaks via HDR, its efficiency remains lower than that of *S. cerevisiae*, where near-complete editing can be achieved with just 50 bp homology arms [222, 228]. A limitation of this system is the inability to completely distinguish poor repair efficiency from deleterious phenotypes when using donor templates with shorter homology arms, an issue also observed in *S. cerevisiae* [229]. This can be partially addressed by examining additional timepoints and continuing rapamycin treatment, while measuring parasitemia and active growth. Editing efficiency may also be improved by implementing a dual-gRNA strategy, allowing for the complete deletion of the target sequence and subsequent repair by HDR (i.e., 5’ and 3’ ends) [230, 231].

Currently, the DiCre Cas9 system is limited to parasite lines already expressing DiCre, which has only been established in NF54 / 3D7 lines [232–234]. Due to the large size of the DiCre cassette, incorporating it into our plasmid would likely be impractical. However, sequential transfections – first introducing a plasmid carrying the DiCre cassette, followed by the DiCre Cas9 plasmid, with dual selection – could potentially increase its applicability.

There is great value in a tool capable of conditionally editing the parasite genome across the entire lifecycle. However, extending this system to the mosquito and liver stages would require genomic integration to ensure efficient passage through the mosquito vector. One potential approach is to use Bxb1 integrase with attP/attB sites to stably integrate the plasmid, including the gRNA and donor template. However, the only established DiCre-attB parasite line at this time is deficient in gametocyte development, prohibiting study of any stages outside of the blood stage [235, 236].

A recently published approach introduced a barcoded loxP intron into the start of an open reading frame to generate a frameshift mutation, disrupting the open reading frame upon DiCre induction [237]. This system also utilized commercial DNA synthesis , allowing for scalable genetic screens coupled to barcode sequencing. However, this system is inherently limited to frameshift knockouts and still requires generating the initial parasite lines before conducting experiments. Additionally, the introduction of the loxP-intron-loxP may alter function or be deleterious in certain genes. Nevertheless, this method provides a complementary approach that allows for scalable phenotyping of essential genes across the parasite lifecycle.

Our system represents the first the inducible Cas9-HDR genome editing tool for *P. falciparum*. Beyond gene deletions, this system is also amenable to precise genome modifications, including the introduction of SNPs, for functional analysis of resistance mutations. This system could also be adapted for genetic screens, either in a pooled or an arrayed format, to edit multiple genes or sites simultaneously.

3.6 Methods

Plasmid construction

The DiCre-Cas9 plasmid was assembled by cloning using Gibson assembly, combining the BSD cassette and loxP-GFP-loxP-dCas9 cassette with the backbone containing a U6 promoter and ampicillin resistance cassette. The BSD and loxP-GFP-loxP-dCas9 cassettes were amplified from the DiCre-dCas9-effector plasmids, kindly gifted by Jun Miao [219]. The U6-amp backbone was amplified from the Linker3 plasmid, kindly gifted by Ashley Vaughan [238]. Both amplicons contained at least 20-bp of homology to each other. dCas9 was then replaced with a high-fidelity Cas9 using Gibson assembly. The HF-Cas9 was amplified from the VP12 plasmid, kindly gifted by Keith Joung [239]. A codon-optimized mNeonGreen was amplified from a plasmid kindly gifted by Ashley Vaughan. GFP was replaced with the mNeonGreen by Gibson assembly, using at least 20-bp of homology to each other.

For gRNA cloning of single guides, two complementary oligos were designed, with a 5' TATT overhang and a 3' CTGC overhang, allowing for insertion into Esp3I-digested plasmids. Oligos were first annealed by mixing 1uL of 10x NEB T4 ligase buffer and 1uL 100uM of each primer in a 10uL reaction using following conditions: 95C for 5 minutes, gradual decrease to 25C at a rate of 0.1C per second. Annealed oligos were then diluted 1:200 with water. At least 50 ng of Esp3I-digested plasmid was used with 1uL of diluted annealed oligos, mixed with 0.5uL of NEB Quick Ligase, 5uL of 2x Quick Ligation mix, remainder water for a 10uL reaction. The reaction was left at room temperature for 5-30 minutes before transformation. 2uL of ligation mix was used for transformations.

For donor sequence cloning, the 5' HDR homology arms and 3' HDR homology arms were amplified from NF54 genomic DNA, with overhangs for Gibson assembly. Plasmids were either digested by a single-cutter enzyme or amplified with respective overhangs for Gibson assembly. The three components were mixed with 2x Gibson assembly Master

mix or 2x NEBuilder Master mix with a ratio of 1 vector: 5 insert 1: 5 insert 2, using a minimum of 50ng of linear vector. 1-3uL of mix was used for transformations.

For combined gRNA-donor cloning, the paired gRNA and donor were synthesized by commercial DNA synthesis (Twist Biosciences or GenScript), containing at least 20 basepairs of Gibson homology arms, the gRNA, tracrRNA, and respective donor sequence. An Esp3I-digested DCL3 plasmid with the tracrRNA removed was used for Gibson assembly, with a minimum of 50ng of linear vector and 5x of the synthesized oligo. 1-3uL of mix was used for transformations.

Parasite culture

P. falciparum parasites were maintained in complete medium consisting of RPMI-1640, 25mM HEPES, 2.4mM sodium bicarbonate, 1% Albumax II, and 2% gentamicin, as previously described [240]. Parasites were kept at 37C in a 90% N, 5% CO2, 5% O2 environment. Hematocrit was maintained between 4-5% using washed O+ human donor red blood cells. Synchronizations were performed using 5% sorbitol to achieve 0-3h synchronization, as previously described [241]. Parasitemia and morphology were checked by Giemsa stain. For routine culture, an NF54 parasite line that was passed through the complete parasite lifecycle was used to ensure viability at each stage of development. For all DiCre experiments, an NF54-DiCre line capable of mosquito transmission was used [233]. The NF54-DiCre line was kindly gifted by Moritz Treeck.

Gametocytes were induced starting with tightly synchronized ring stage parasites at 1% parasitemia at 5% hematocrit. Cultures were kept at 37C in a 90% N, 5% CO2, 5% O2 environment, using a slide warmer during media changes. Media consisted of RPMI-1640, 25mM HEPES, 2.4mM sodium bicarbonate, and 10% human sera. Media was changed daily for 14 days, adding pre-warmed media and ensuring not to disrupt the settled cells. For pure gametocyte cultures, 50mM of GlcNAc was added to culture for 5 days before being removed.

For induction of Cas9, 50nM of rapamycin was added to cultures before being removed 6h later and replaced with complete media.

Transfection

Transfections were performed by ring-stage electroporation using a Bio-Rad Gene-Pulser II Electroporator set to 0.31kV and 950 µF capacitance, as previously described [50]. 50-100ug of total plasmid DNA was used per transfection, mixed with cytomix (120 mM KCl, 0.15 mM CaCl2, 2mM EGTA, 5 mM MgCl2, 10 mM K2HPO4/KH2PO4 25 mM HEPES, pH 7.6). Parasites were synchronized by sorbitol to attain approximately 5% parasitemia at ring-stage. The following day, parasites were pelleted, washed with cytomix, and mixed 1:1 with the plasmid-cytomix solution. Parasites were transferred to a 0.2cm electroporation cuvette before being electroporated. Electroporated parasites were immediately moved to pre-warmed flasks with 6mL CM and 125uL fresh packed RBCs. The flask was then gassed with 90N/5O/5CO2 mixture before being placed in an incubator.

Media was changed for two days (one complete parasite cycle) before drug was added. For original L3 plasmids, 2.5 nM WR99210 was used and maintained in media for 5 days. For all other plasmid transfections, 2.5 µg/mL Blasticidin was used and maintained in media for 6 days. Parasites were maintained with media changes every other day after until visible by Giemsa smear. For DCL3 transfections, if no mNeonGreen signal could be detected using fluorescence microscopy after parasite recovery, Blasticidin was returned to media until fluorescent parasites were pure.

DNA Extractions

Genomic DNA from parasites was isolated using GeneJET Whole Blood Genomic DNA Purification Kits (Thermo Fisher). Prior to isolating, a minimum of 250uL packed infected red blood cells were spun down and resuspended with 1x PBS + 0.015% saponin, set at room temperature for 5-10 minutes, and then pelleted. The supernatant was discarded and the remaining parasite pellet was used following the manufacturer's

protocol. For the final elution, 50uL of warm (37-50C) water was used in place the provided elution buffer, incubating for 5 minutes at room temperature before spinning at 13,000 x g for 1.5 minutes. The purified DNA was then quantified by Nanodrop.

DNA Sequencing

Plasmids and transfected parasites were amplified using PrimeStar or PrimeStar Max polymerase. To measure genomic editing, 50ng of isolated genomic DNA and primers external to the repaired DNA were used. PCR products were purified by column purification or gel extraction followed by column purification.

All sequencing was done by Plasmidsaurus. For plasmid cloning, whole plasmid sequencing was used. For PCR sequencing, linear amplicon sequencing was used, allowing for quantification of the editing and non-edited DNA from the raw reads. Edits were quantified by by counting the total number of occurrences of the edited and unedited DNA in raw reads.

RT-qPCR

RT-qPCR for quantification of Cas9 expression was performed using the Sigma-Aldrich One-Step SYBR Green Quantitative RT-qPCR Kit. 100pg of genomic DNA was used. Two validated primers internal to Cas9 were used, with final concentrations of 200 nM, following the manufacturer's protocol.

Fluorescent Microscopy

Infected red blood cells were stained in PBS with 10 µM Hoechst 33342 and incubated at 37C for 10 minutes before washing twice and being resuspended in complete media. Cells were then moved to a slide and a coverslip was added. The cells were viewed using a Nikon Eclipse fluorescent microscope at 400-1000x total magnification. For Hoechst detection, the UV filter was used and for GFP or mNeonGreen detection, the blue filter was used. Parasitemias were quantified from the presence or absence of Hoechst and mNeonGreen, depending on the application.

Chapter 4 – Development and Application of a Conditional CRISPRa and CRISPRi System for Scalable Genetic Screens

A version of this chapter is in preparation for publication.

Sean T. Windle, Maxwell L. Neal, Fred D. Mast, Alexis Kaushansky, Stefan H. I. Kappe, and John D. Aitchison

4.1 Introduction

Although CRISPR/Cas9 in *Plasmodium* has become a routine tool for gene characterization, no large-scale CRISPR screens have been conducted to date. With the parasite containing no NHEJ pathways, traditional CRISPR indel screens are inaccessible in *Plasmodium* [49, 73]. Although the DiCre-inducible Cas9 alleviates many of these limitations in *Plasmodium*, there is still no way of perturbing gene expression in a scalable system. Further, despite the wide use of CRISPRa and CRISPRi strategies in mammalian fields, few groups have made use of them in *Plasmodium*.

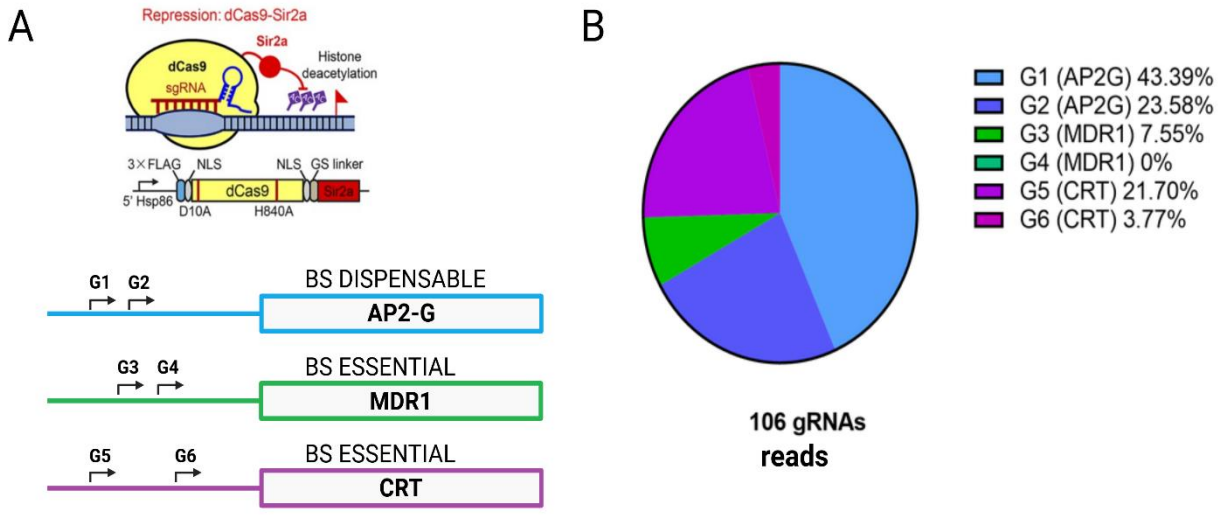

Figure 4.1 CRISPRi pilot experiments using characterized genes

A) *Plasmodium* CRISPRi relies on a dCas9-Sir2a fusion for targeted knockdowns. Two gRNAs were designed for each of the three genes, AP2-G, MDR1, and CRT. AP2-G is known to be dispensable in blood stage, while MDR1 and CRT are essential.

B) The recovered parasites were sequenced. AP2 gRNAs were enriched, while MDR1 and CRT gRNAs were depleted.

The majority of CRISPR screens are genome-wide, often using the same established libraries [242–244]. Many of these available libraries have been validated and are iterations of the early genome-wide libraries following the rise of CRISPR/Cas9 technologies. In the case of human CRISPR screens, there are tens of thousands of unique gRNAs, with many gRNAs targeting different regions of the same gene. This increases confidence that a given gene perturbation is having an effect on the phenotype of interest, at the cost of large library size [75, 242]. CRISPR screens have been applied to numerous biological questions, ranging from cancer biology to agriculture [87, 107, 207]. They have also been further adapted for single-cell experiments, as in Perturb-seq and Crop-seq [245, 246].

The quality of a validated gRNA from an established library is much higher than if one were to choose a random gRNA, as these gRNAs have already gone through initial testing [242, 244]. Despite the versatility of CRISPR gRNAs, the vast majority of gRNAs are ineffective [73, 74, 80, 175]. This is in part due to secondary structure in the RNA molecules and the unique features of a given gRNA site in the genome [73, 247, 248]. For instance, open chromatin may increase efficiency, while methylated DNA could decrease efficiency [249]. While there have been numerous machine learning approaches to predicting gRNA efficiency, they tend not to be generalizable across species or even cell types [250–252]. Thus, it is always advisable to include multiple gRNAs per gene since it is never guaranteed that any individual gRNA will lead to efficient Cas9 function at the desired locus.

The choice of gRNA is further complicated by the type of CRISPR screen being performed. For indel KO screens, ideal targets sit between ~20-50% through the coding sequence and outside of any introns [253]. In the case of CRISPRa and CRISPRi screens, gRNAs are most effective when directed to regions upstream of the start codon or near the transcription start site [254–256]. However, an effective CRISPRi gRNA may not be an

effective CRISPRa gRNA, and vice versa. Ultimately, it is not uncommon for any given gRNA to be unsuccessful.

To explore the feasibility of CRISPRa and CRISPRi screens in *Plasmodium,* I first tested a small pool of gRNAs with CRISPRi. We identified a number of putative transcriptional regulators that may be involved in artemisinin resistance and parasite development. These targets are ideal candidates for overexpression and underexpression screens. Using DiCre-inducible CRISPRa and CRISPRi systems, I generated a stable small-scale screen that could be used to identify mediators of drug resistance and growth phenotypes.

4.2 Establishing Scalable CRISPRa and CRISPRi Systems in *P. falciparum*

CRISPRi and CRISPRa allow for targeted underexpression and overexpression. The *Plasmodium*-adapted systems rely on histone deacetylase (Sir2a) and histone acetyltransferase (GCN5) domains, respectively, to modulate expression of genes targeted by Cas9 binding [88, 257, 258]. However, the likelihood (or propensity) of these perturbations to be selected for has not been quantified. Using well-characterized genes, I tested CRISPRi's capabilities for knockdown through pooled plasmids [24, 26, 57, 70, 126, 130, 133, 162]. I targeted two essential genes MDR1 and CRT, as well as AP2-G (a non-essential gene), with two gRNAs each targeting near the predicted transcription start site, and then measured the frequency of each gRNA following recovery from transfection [Figure 4.1A]. If the knockdown on the genes is deleterious, these gRNAs should be selected against. I hypothesized that gRNAs for MDR1 and CRT, as essential genes, would be underrepresented while AP2-G would be overrepresented. Five out of the six gRNAs were detectable by sequencing [Figure 4.1B]. gRNAs targeting AP2-G were the most abundant, with MDR1 gRNAs only making up 7.6% of the total reads and CRT gRNAs having 25% of the reads. Although CRT's relative abundance was higher than expected, it

is not surprising to see a gRNA have no effect. As a preliminary test for CRISPRi screening, this provided preliminary evidence that essential genes could be selected against.

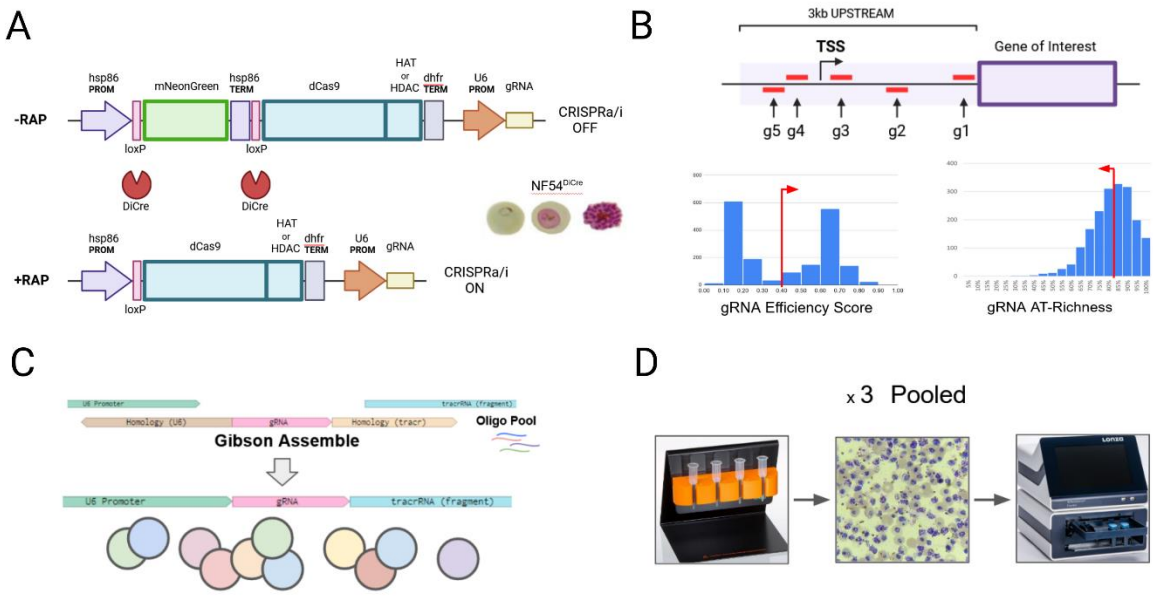

Figure 4.2 DiCre inducible CRISPRa and CRISPRi screen design

A) DiCre-mediated induction of CRISPRa or CRISPRi using a loxP-flanked mNeonGreen and terminator, which block expression of the dCas9-fusion until addition of rapamycin.

B) gRNA selection criteria. gRNAs are chosen anywhere between -3kb (upstream) and +100 (downstream) of the start codon. At least one gRNA is chosen near the transcription start site. gRNAs have a minimum of 0.4 efficiency score, and a maximum AT-richness of 80%.

C) Gibson assembly strategy for pooled cloning of gRNA oligo pools.

D) Outline of transfections. Synchronized schizonts are purified by magnetic column until completely mature. Mature schizonts are then electroporated by Lonza 4D.

A shortcoming of these original experiments was that the system employed two plasmids, one encoding the gRNA and the other encoding the dCas9-fusion protein. In the case of pooled transfections, it is important that all parasites receive the dCas9 plasmid while only receiving one gRNA per parasite. This is impossible to achieve without performing consecutive transfections (i.e., first dCas9 plasmid, followed by gRNA plasmids). To simplify the system, I moved both the dCas9-fusions from CRISPRa and CRISPRi to my established DiCre-inducible Cas9 plasmid, replacing SpCas9-HF. I refer to these plasmids as DCL3-CRISPRa and DCL3-CRISPRi. As in the DiCre-inducible Cas9 system,

here the dCas9-fusions can be induced by the addition of rapamycin. This system could now be applied to perform scalable CRISPR screens.

4.3 Design of Inducible CRISPRa and CRISPRi Screens for Perturbing Transcriptional Regulators

Transcriptional regulators (TRs), as identified by the GRN model, are ideal candidates for CRISPRa and CRISPRi screens [94, 259]. Many of these identified regulators tightly control development throughout the parasite lifecycle [197, 260, 261]. Some of these regulators appear to be involved in control of genes important for drug resistance [63, 156, 157, 262]. We hypothesized that a number of these TRs are involved in large developmental checkpoints controlling many genes (hub regulators), while others may be providing a means for parasites to survive drug exposure. One subset of identified TRs correlates with increased transcriptional variability, potentially offering a means to survive drug through heterogeneity. This “bet-hedging” strategy can be found in many biological systems, including in *Plasmodium* [263–268]. For example, gametocyte induction can also potentially be seen as a bet-hedging strategy, where a small subset of parasites commit to differentiation in response to biological triggers.

We selected a set of 44 total putative TRs, including all 27 known AP2-family transcription factors, 14 TRs of unknown function, and 3 well-characterized transcriptional regulators. By applying inducible CRISPRa and CRISPRi screens to these TRs, we can examine the effects of perturbing transcriptional regulators both for basic parasite development and on the response to antimalarial treatments [Figure 4.2A]. Although the optimal gRNA candidate for CRISPRa and CRISPRi is often near the transcription start site, this is not always the case. To ensure I selected at least one successful gRNA, five gRNAs were selected per gene, spread between -3000bp upstream to +100bp downstream of the start codon [Figure 4.2B]. gRNAs were only chosen if they exceeded a total efficiency score (combined score from CRISPRater and the Doench

model) of 0.40 or higher [256, 269]. Similarly, only gRNAs with AT-composition of 80% or lower were selected. Lastly, 5 non-targeting (nowhere in the genome) scrambled control gRNAs were selected. In total, there were 225 unique gRNAs. These gRNAs were then commercially synthesized and cloned into my DCL3-CRISPRa and DCL3-CRISPRi plasmids by assembly cloning, establishing a library of 225 unique plasmids [Figure 4.2C].

I next explored the optimal transfection approach for transfecting a pooled plasmid library. An ideal transfection would result in every gRNA being represented equally and individually in the parasite pool. However, the limitations of *P. falciparum* transfection complicate this. One group had identified schizont transfection as the optimal method for preventing multiple plasmids from being taken up, while ring-stage transfection led to multiplicity [270]. Another group successfully transfected hundreds of pooled plasmids using similar schizont transfection protocols [271]. I tested the efficiency of schizont transfection using both the Lonza 4D nucleofector and the BioRad electroporator. In both cases, I tightly synchronized parasites to 0-3h rings by sorbitol, before isolating mature schizonts by magnetic column. Schizonts were allowed to grow until complete maturity before transfection. The Lonza 4D resulted in faster recovery by ~2 days, signifying a roughly 3-4x increase in uptake efficiency. Thus, I moved forward with the Lonza 4D nucleofector approach for pooled transfections.

Using prior transfections as an estimate of efficiency, I calculated the average number of unique plasmid uptakes by the time needed to recover from a transfection. Given a known rate of growth (3.7x every 48 hours for NF54-DiCre), the number can be calculated from the number of starting parasites, the growth rate, and the number of cycles needed to recover from transfection. Based on this calculation, ~170 unique plasmids are taken up per transfection. Thus, in order to have representation of every gRNA, at least 3 transfections would need to be performed.

To establish the parasite library pool, I performed 3 separate transfections of 100ug of total plasmid DNA into mature purified schizonts using the Lonza 4D for both

the DCL3-CRISPRa and DCL3-CRISPRi plasmid pools [Figure 4.2D]. Transfections took on average 15 days for parasites to be detected by Giemsa smear. In order to have complete coverage of gRNAs, I pooled each of the three transfections to create one parasite pool for both CRISPRa and CRISPRi. I then sequenced gRNAs from the two populations to measure the frequency of each gRNA. The CRISPRa pool had ~85% representation of the total gRNA pool, while CRISPRi had ~88%. As each gene was represented on average 4.25 gRNAs, this coverage was deemed sufficient for proceeding with screens [Figure 4.3A].

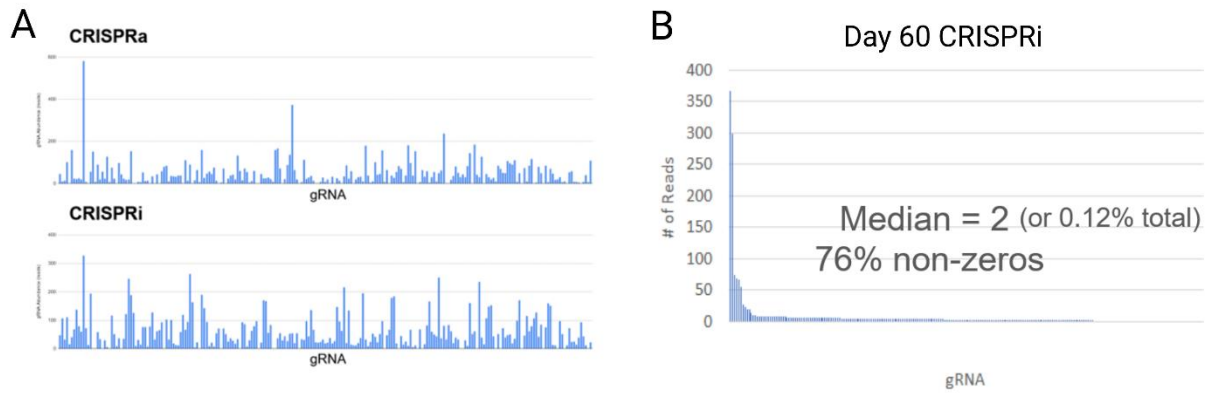

Figure 4.3 Coverage of gRNAs from episomal CRISPRa and CRISPRi plasmid pools
A) Distribution of gRNAs for CRISPRa and CRISPRi pools following 3 pooled transfections and drug selection.
B) Distribution of gRNAs for CRISPRi after 30 additional days in culture.

However, after 60 days of *in vitro* growth, the plasmid population had shifted significantly, with many gRNAs dropping out [Figure 4.3B]. With only ~50% of gRNAs detectable and the majority of reads attributable to the top 5% of gRNAs, this pool seemed insufficient for genetic screening. One possible explanation for this loss of guide RNAs is that plasmids are not inherently stable as episomes. Transfected parasites can also contain very large concatemeric plasmid structures. As these plasmids must be maintained by continued drug selection, it is possible there is some selective pressure for maintaining the most stable, most resistant concatemeric plasmid structures. The concatemers may also increase the amount of heterogeneity within the population, as some may have higher dCas9 and gRNA expression than others.

To establish a stable plasmid pool, I explored approaches that could help maintain the plasmids as stable single copies. I first tested a centromere-containing plasmid, which

includes a 1.5kb fragment of the *P. falciparum* Chr5 centromere. This approach has been used previously to create parasites stably maintaining large foreign DNA episomally. I added this centromere sequence to my DCL3-CRISPRa and DCL3-CRISPRi plasmids. However, the centromere sequence is very AT-rich (95%) and not highly stable for growth in *E. coli*. Ligation and Gibson assembly approaches for cloning the gRNA pool into the centromere plasmid led to largely recombined plasmids, with at best ~15% of the reads containing the full plasmid and cloned gRNA. Ultimately, this was deemed too low for pooled *P. falciparum* transfections.

I next pursued integration approaches that have been established in *P. falciparum*. Bxb1 integrase relies on the presence of attB and attP sequences, with one being on a

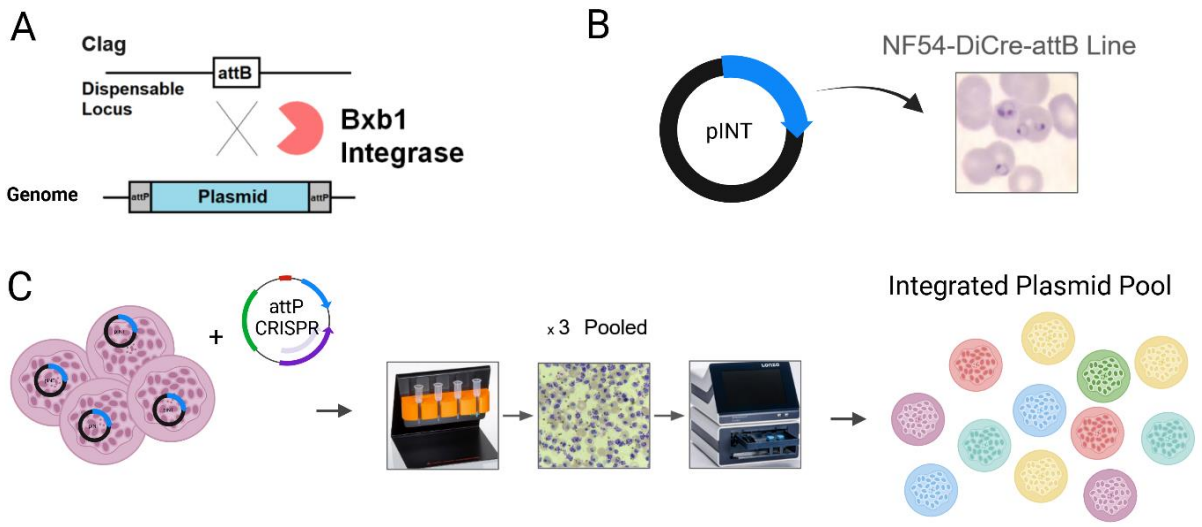

Figure 4.4 Integration approach for stably integrating CRISPRa and CRISPRi plasmids
A) Bxb1 integrase strategy. Plasmids carrying attP can be stably integrated into the attB landing site in a dispensable locus via Bxb1 integrase.
B) Pre-loading of the pINT Bxb1 expression plasmid into an NF54-DiCre-attB line, with DiCre located in the Pfs47 locus and attB located in the Clag6 locus.
C) pINT pre-loaded NF54-DiCre-attB parasites were transfected with CRISPR plasmids carrying attP, allowing for rapid integration into the genome. This was done three times and pooled to establish CRISPRa and CRISPRi

genome landing site and the other found on the plasmid [Figure 4.4A] [272]. By adding a Bxb1-expressing plasmid (pINT) to transfections, attP plasmids can be stably integrated into the attB landing site, forming attL and attR sites that can no longer recombine. This strategy has the added benefit of increasing parasite recovery from transfections, likely

due to the stability of the drug cassette. Fortunately, a NF54-DiCre-attB parasite line was recently developed, which includes both the DiCre cassette stably integrated into the Pfs47 locus, and attB stably integrated into the Clag6 locus [236]. This line would allow us to integrate the plasmids, while still being able to utilize DiCre for the inducible element.

To ensure rapid integration of the plasmid library, I first established an NF54-DiCre-attB line that expressed Bxb1 episomally by transfecting the pINT plasmid by itself. I then added attP sites to both the DCL3-CRISPRa and DCL3-CRISPRi plasmids and recreated gRNA pools in these plasmids. Now, as soon as any plasmid carrying attP reaches the nucleus from transfection, it will be immediately integrated into the attB landing site. This strategy solves multiple problems. First, this ensures a single-copy integration event, preventing multiple gRNAs from potentially being expressed in a single cell and reducing heterogeneity. Second, this maintains the plasmid sequence stably, without the risk of loss after initial drug selection, eliminating the need for continued selection. Lastly, this approach increases the efficiency of transfections for plasmid uptake into the parasite.

To test this approach with the CRISPR pools, I transfected the DCL3-attP-CRISPR pools into NF54-DiCre-attB parasites pre-loaded with pINT in triplicate. Parasites were even more rapidly recovered, with parasites being visible by Giemsa smear by up to day 12. The three transfections were then pooled and sequenced. With episomal CRISPRi plasmids, only 83% of gRNAs were represented by Day 45, with many gRNAs only being identified in single-digit reads [Figure 4.5A]. With the integrated CRISPRi library, the entire gRNA library was fully represented [Figure 4.5B]. This distribution was maintained, even after 20 days of growth, indicating improved stability. A small number of gRNAs were still greatly over-represented. It is possible that these are simply episomal plasmids that are being maintained as multiple copies, in addition to the stably integrated plasmid. Targeted sequencing of the integration site revealed that all parasites in the pool contained an integrated plasmid. These parasite pools can now be induced and applied to genetic screens for any phenotypes of interest.

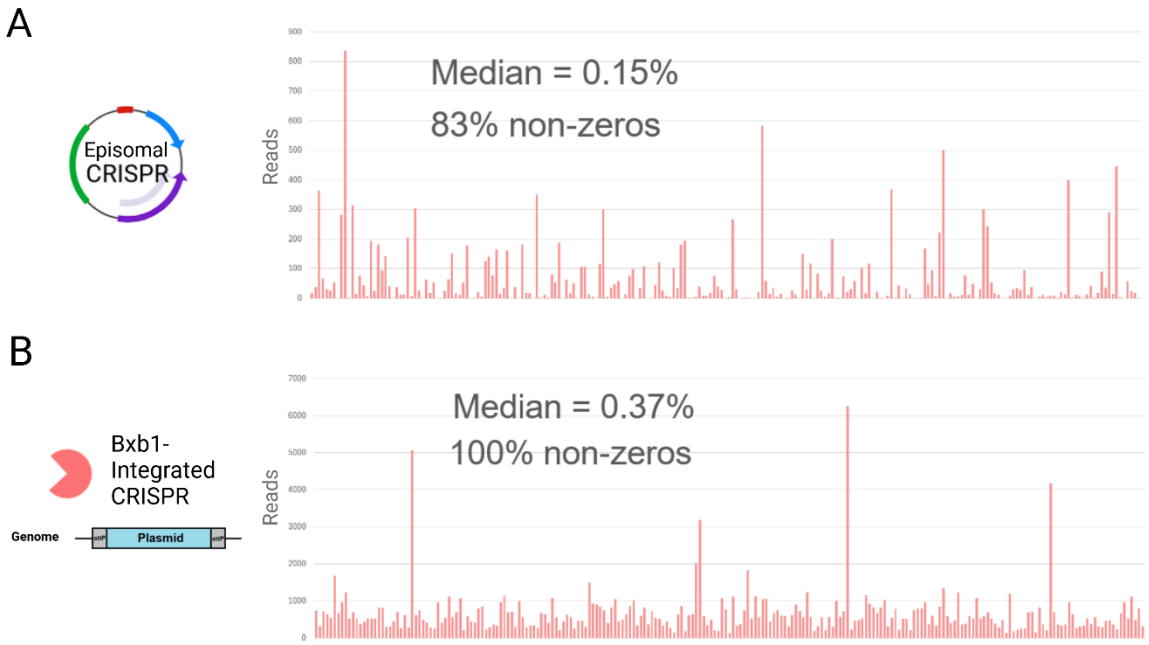

Figure 4.5. gRNA Coverage of CRISPRi Screens
A) Day 45 gRNA distribution from transfected episomal CRISPRi plasmids.
B) Day 45 gRNA distribution from transfected and Bxb1 integrated CRISPRi plasmids.

4.4 Identifying Essential Genes by CRISPR Screening

Now that an integrated CRISPRi pool has been established, I sought to identify genes that are essential to BS growth. By knocking down expression of an essential gene, parasite growth may be impacted, resulting in dropout of gRNAs targeting those genes in the sequenced pools. Although essentiality of BS genes has been determined by the PiggyBac screen, this screen has been shown to occasionally have false positives and false negatives. However, our screen is inducible and can directly measure the effects of the perturbation, unlike the PiggyBac screen which depends on saturation of the genome and dropouts to identify essentiality. Using the integrated CRISPRi pool, I induced dCas9-Sir2a expression by adding rapamycin to highly synchronized parasites before washing 48h later. I cultured these induced parasites normally for 6 additional days (3 cycles) and

sequenced ring-stage parasites. The sequencing results were then compared to the uninduced pool to identify gRNAs that are enriched or depleted [Figure 4.6A].

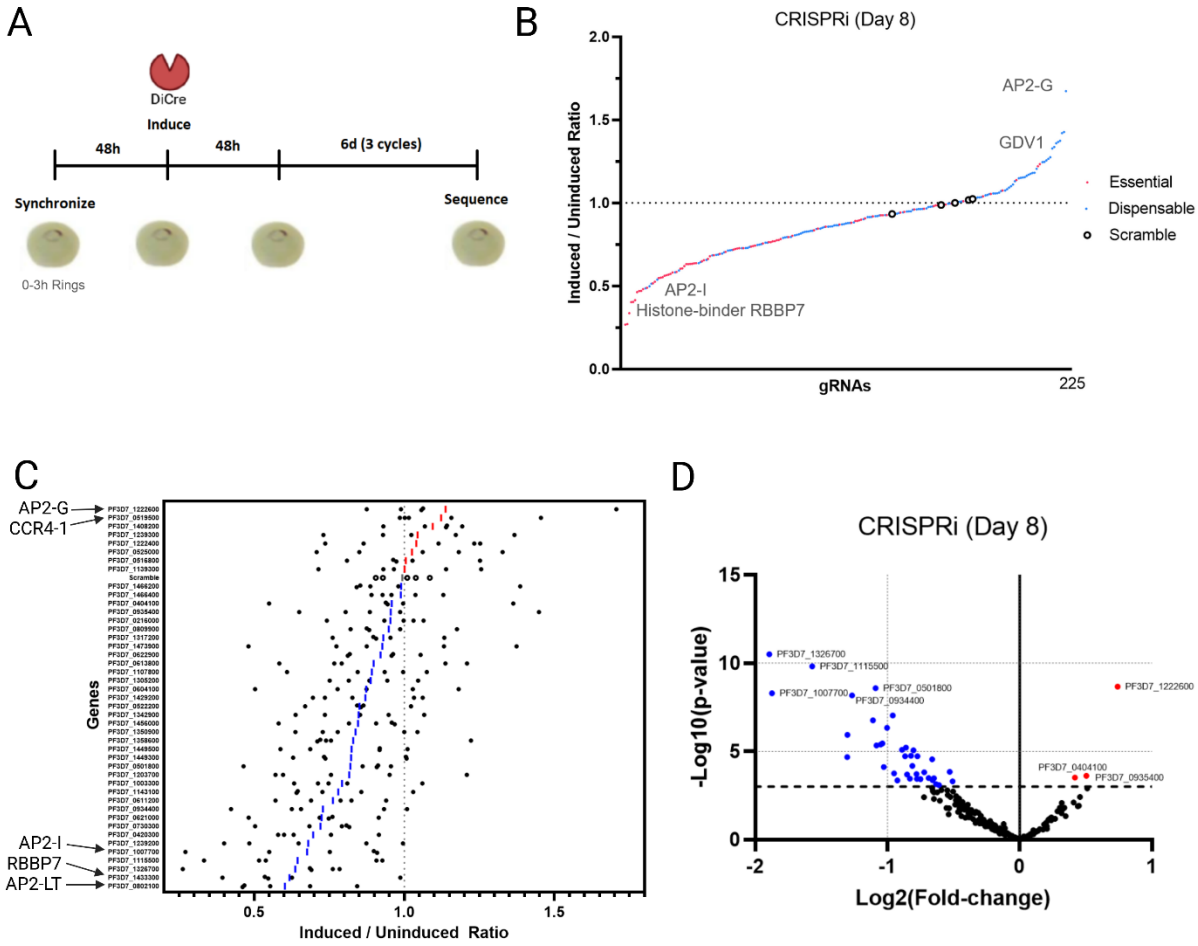

Figure 4.6 Essentiality screen of transcriptional regulators using an induced CRISPRi pool

A) Timeline for induction of CRISPRi. Parasites are tightly synchronized to 0-3h rings by sorbitol before adding 50nM rapamycin for 6h. Parasites are sequenced on day 8.

B) Distribution of gRNAs normalized to the mean gRNA scramble control reads, then as a ratio of induced and uninduced. piggyBac-essential genes are marked in red and dispensable genes in blue.

C) The gRNA distribution sorted by average relative abundance, with blue indicating depletion and red enrichment. Each point represents a single gRNA.

D) A volcano plot of Day 8 CRISPRi gRNAs with log-fold-change on the x-axis and log-p-value on the y-axis. Significantly depleted or enriched gRNAs are indicated by blue (depleted) or red (enriched), $p < 0.05$.

The scrambled gRNA controls had no effect on growth, with an average of 0.99-fold relative abundance before normalization [Figure 4.6B]. However, for each of the targeted genes, there was considerable variation in sequence abundance between each

of the five gRNAs, with some gRNAs having no change and others having large decreases or increases in abundance. This may simply be the result of how the gRNAs were selected, e.g., gRNAs that were closer to the transcription start site might have been more likely to have an effect, while more distant gRNAs may have had little or no effect. Many genes overlap or sit directly adjacent to other genes in the *P. falciparum* genome.

Table 4.1 CRISPRi essentiality and neighboring gene proximity

Genes	Strand	Ratio	Ess.	5' Neighbor	5' Dist.	Ess.	3' Neighbor	3' Dist.	Ess.
PF3D7_0216000	Forward	0.954	Disp.	PF3D7_0215900	930	Disp.	PF3D7_0216100	71	Ess.
PF3D7_0404100	Reverse	0.959	Disp.	PF3D7_0404200	609	Ess.	PF3D7_0404000	415	Ess.
PF3D7_0420300	Forward	0.697	Ess.	PF3D7_0420200	3776	Ess.	PF3D7_0420400	570	Ess.
PF3D7_0501800	Forward	0.821	Ess.	PF3D7_0501700	1534	Disp.	PF3D7_0501900	-449	Disp.
PF3D7_0516800	Reverse	1.005	Disp.	PF3D7_0516900	4105	Ess.	PF3D7_0516700	514	Disp.
PF3D7_0519500	Reverse	1.123	Disp.	PF3D7_0519600	1319	Ess.	PF3D7_0519400	1940	Ess.
PF3D7_0522200	Forward	0.851	Disp.	PF3D7_0522100	41	Ess.	PF3D7_0522300	-251	Ess.
PF3D7_0525000	Forward	1.026	Disp.	PF3D7_0524900	3811	Disp.	PF3D7_0525100	1498	Ess.
PF3D7_0604100	Reverse	0.873	Disp.	PF3D7_0604200	1622	Ess.	PF3D7_0604000	413	Ess.
PF3D7_0611200	Forward	0.763	Disp.	PF3D7_0611100	690	Ess.	PF3D7_0611300	528	Ess.
PF3D7_0613800	Forward	0.898	Ess.	PF3D7_0613700	3184	Ess.	PF3D7_0613900	19	Disp.
PF3D7_0621000	Forward	0.725	Ess.	PF3D7_0620700	1729	Ess.	PF3D7_0621100	28	Ess.
PF3D7_0622900	Reverse	0.921	Disp.	PF3D7_0623000	3339	Ess.	PF3D7_0622800	-589	Ess.
PF3D7_0730300	Forward	0.721	Disp.	PF3D7_0730200	7027	Ess.	PF3D7_0730400	2257	Disp.
PF3D7_0802100	Reverse	0.603	Ess.	PF3D7_0802200	3411	Disp.	PF3D7_0802000	43	Disp.
PF3D7_0809900	Reverse	0.949	Disp.	PF3D7_0810000	4523	Ess.	PF3D7_0809800	788	Disp.
PF3D7_0934400	Forward	0.73	Ess.	PF3D7_0934300	597	Ess.	PF3D7_0934500	-144	Ess.
PF3D7_0935400	Forward	0.955	Disp.	PF3D7_0935300	3403	Ess.	PF3D7_0935500	18758	Ess.
PF3D7_1003300	Reverse	0.793	Ess.	PF3D7_1003400	414	Disp.	PF3D7_1003200	-28	Ess.
PF3D7_1007700	Reverse	0.677	Ess.	PF3D7_1007800	4139	Disp.	PF3D7_1007600	-725	Ess.
PF3D7_1107800	Reverse	0.892	Ess.	PF3D7_1107900	7103	Disp.	PF3D7_1107700	-299	Ess.
PF3D7_1115500	Forward	0.645	Ess.	PF3D7_1115400	666	Disp.	PF3D7_1115600	-518	Disp.
PF3D7_1139300	Reverse	1.001	Disp.	PF3D7_1139500	10663	Disp.	PF3D7_1139200	738	Ess.
PF3D7_1143100	Reverse	0.78	Ess.	PF3D7_1143200	3666	Disp.	PF3D7_1143000	33	Ess.
PF3D7_1203700	Reverse	0.817	Ess.	PF3D7_1203900	4944	Ess.	PF3D7_1203600	-449	Ess.
PF3D7_1222400	Forward	1.04	Disp.	PF3D7_1222300	2490	Ess.	PF3D7_1222500	183	Ess.
PF3D7_1222600	Forward	1.138	Disp.	PF3D7_1222500	3467	Ess.	PF3D7_1222700	3297	Ess.
PF3D7_1239200	Forward	0.681	Ess.	PF3D7_1239100	5060	Ess.	PF3D7_1239300	1984	Disp.
PF3D7_1239300	Forward	1.045	Disp.	PF3D7_1239200	1984	Ess.	PF3D7_1239400	329	Disp.
PF3D7_1305200	Forward	0.888	Disp.	PF3D7_1305100	672	Ess.	PF3D7_1305300	-131	Disp.
PF3D7_1317200	Forward	0.931	Disp.	PF3D7_1317100	6211	Ess.	PF3D7_1317300	-320	Disp.
PF3D7_1326700	Forward	0.637	Ess.	PF3D7_1326600	2513	Disp.	PF3D7_1326800	71	Ess.
PF3D7_1342900	Reverse	0.849	Disp.	PF3D7_1343000	1295	Ess.	PF3D7_1342800	53	Ess.
PF3D7_1350900	Reverse	0.838	Ess.	PF3D7_1351000	2056	Disp.	PF3D7_1350800	123	Disp.
PF3D7_1358600	Forward	0.83	Disp.	PF3D7_1358500	971	Ess.	PF3D7_1358700	396	Ess.
PF3D7_1408200	Reverse	1.095	Disp.	PF3D7_1408300	6528	Disp.	PF3D7_1408100	1478	Disp.
PF3D7_1429200	Reverse	0.871	Disp.	PF3D7_1429300	1268	Disp.	PF3D7_1429100	-373	Ess.
PF3D7_1433300	Forward	0.618	Ess.	PF3D7_1433200	1764	Disp.	PF3D7_1433400	-272	Ess.
PF3D7_1449300	Forward	0.823	Ess.	PF3D7_1449200	1499	Ess.	PF3D7_1449400	-207	Ess.
PF3D7_1449500	Forward	0.825	Ess.	PF3D7_1449400	3850	Ess.	PF3D7_1449600	841	Ess.
PF3D7_1456000	Forward	0.846	Disp.	PF3D7_1455900	3409	Ess.	PF3D7_1456100	-83	Disp.
PF3D7_1466200	Reverse	0.99	Disp.	PF3D7_1466300	2560	Ess.	PF3D7_1466100	616	Ess.
PF3D7_1466400	Reverse	0.989	Disp.	PF3D7_1466500	4878	Ess.	PF3D7_1466300	1007	Ess.
PF3D7_1473900	Forward	0.928	Disp.	PF3D7_1473800	1591	Disp.	PF3D7_1474000	596	Disp.
Averages		**0.865**			**2984**			**790**	

I examined the distance between our target genes and their upstream and downstream neighbors [Table 4.1]. The average distance between our target gene and its upstream neighbor is 2984 bp, whereas the average distance from our target gene and its downstream neighbor is 790 bp. 15 of our target genes had a directly overlapping transcript with a neighboring gene. Notably, all but 5 target genes had at least one essential neighbor, with 19 genes having two essential neighbors.

The most abundant gRNA was AP2-G, whereas the most depleted gRNA was AP2-LT [Figure 4.6C]. Roughly one fifth of the gRNAs were significantly (positively or negatively) changed from the uninduced population as calculated by negative binomial distribution [Figure 4.6D]. Only three gRNAs were significantly enriched.

By including essentiality data from the PiggyBac screen, I compared gRNA fold-changes between BS essential genes and BS dispensable genes. The average ratio for dispensable genes was 1.025-fold decreased from the uninduced, i.e., nearly completely neutral. A small number of gRNAs targeting dispensable genes were significantly depleted, while others were significantly enriched. In contrast, the relative abundance for essential genes decreased by ~21% following induction. The ten most depleted gRNAs were all essential genes, including AP2-I and five other AP2 transcription factors previously identified as essential. The most depleted gRNA was AP2-LT, a transcription factor that has been shown to interact with PfGCN5 and has been linked to late schizont stage genes necessary for merozoite invasion [273, 274].

The PiggyBac screen mutagenesis index scores of the most depleted genes is very low (bottom 20%), indicating high likelihood of essentiality. These results from our CRISPRi screen strongly agree with the PiggyBac screen. However, there were some genes which contradicted their reported essentiality. For instance, PF3D7_0522200, or transcription initiation factor subunit 10, was found to be dispensable by PiggyBac screen, but had significantly depleted gRNAs ($p<0.001$). These findings complement the established PiggyBac screen, offering a controlled measure of essentiality.

4.5 Identifying Mediators of Artemisinin Resistance and Survival

Having established the CRISPRa and CRISPRi TR screens, I was interested in how perturbation of these regulators might affect survival to antimalarials. By over- and under-expressing these specific TRs, we are also potentially affecting the targets of these TRs, many of which may also be contributing to drug resistance. As these TRs were initially selected in part due to expression patterns indicative of artemisinin resistance, pairing these screens with drug selection may provide insights into genetic factors involved in resistance.

As before, I induced the dCas9 fusions by rapamycin in 0-3h synchronized rings. After one cycle (48h), I added 700nM of the anti-malarial drug DHA for 6h [Figure 4.7A]. This concentration has previously been used to differentiate between resistant and sensitive parasites, with 95%+ of sensitive parasites dying in response to drug [Figure 4.7B]. Following CRISPR induction, this should better divide gene perturbations that contribute to resistance or hinder resistance when measuring the surviving population. By day 10 (8 days after DHA), parasites had recovered from DHA and were sequenced at majority ring stage to eliminate schizont biases.

When DHA was added to the induced CRISPRi parasite pool, the majority of gRNAs were depleted relative to the non-targeting controls [Figure 4.7C]. However, the variability of gRNA abundance in the controls was much higher, limiting our analysis. A small number of genes were enriched compared to the controls following DHA treatment. The most depleted gRNA targeted JmjC-1, a histone demethylase [275]. Applying the DHA selection to the CRISPRa library, many more gRNAs were enriched, with roughly half of gRNAs more abundant than the scrambled gRNA controls [Figure 4.7D]. Once again, JmjC-1 was a top hit and was enriched in the DHA-treated cells, providing additional evidence that JmjC-1 mediates survival to DHA. One of the other most enriched genes was AP2-HS, a transcription factor upregulated under stress conditions and in febrile temperatures

[276, 277]. Other labs, as well as our own, have also connected temperature stress to DHA survival [157, 176, 278, 279].

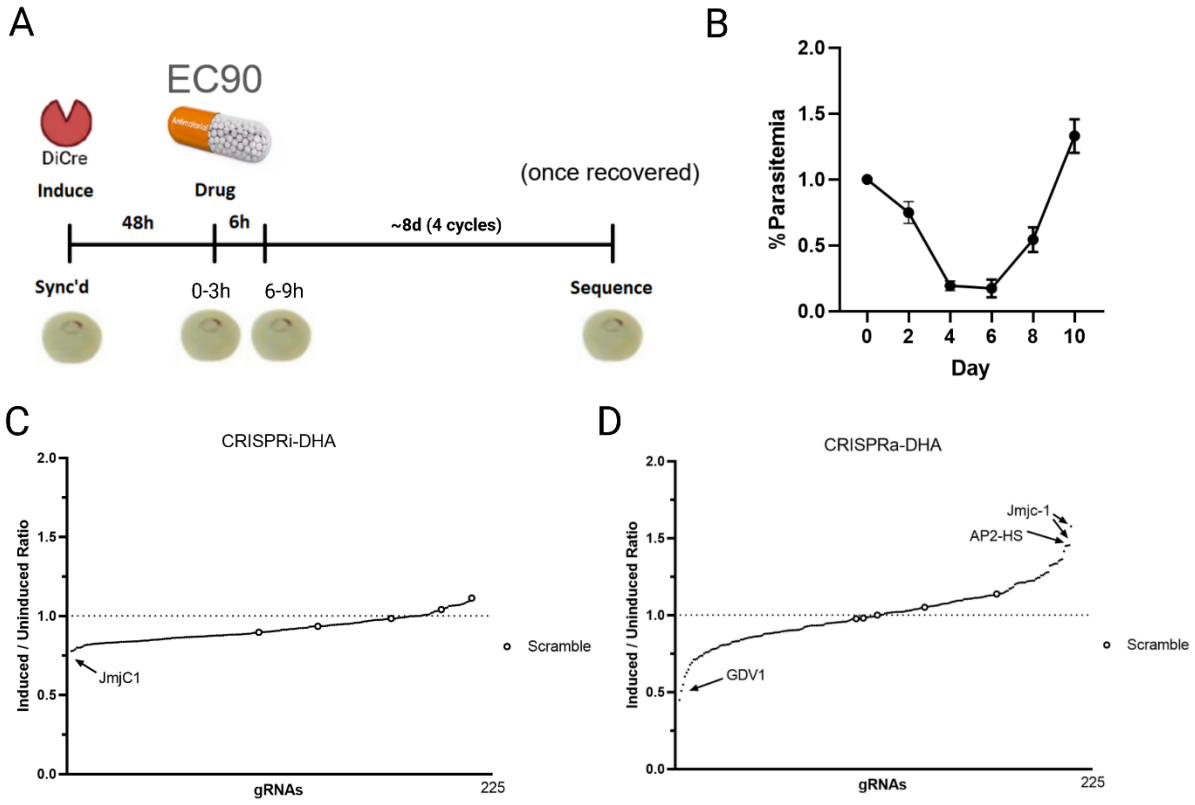

Figure 4.7 DHA selection screen of transcriptional regulators using induced CRISPRi and CRISPRa pools
A) Timeline for CRISPRi screen with 700nM DHA for 6h. Parasites are cultured until recovery to starting parasitemia, then sequenced.
B) Parasitemia of untransfected NF54 parasites following exposure to DHA.
C) Distribution of gRNAs normalized to the mean gRNA scramble control reads, then as a ratio of induced and uninduced following recovery from CRISPRi-DHA.
D) Distribution of gRNAs normalized to the mean gRNA scramble control reads, then as a ratio of induced and uninduced following recovery from CRISPRa-DHA.

JmjC-1 has been found to be involved in development and pathogenesis [275, 280]. Interestingly, JmjC-1 was found to be a variability-inducer and hub regulator of expression in our GRN model, where the gene is predicted to regulate over 1,000 genes. Of note, JmjC-1 is one of the strongest positive regulators of Pfmdr1, PfCRT, and SOD2, all of which have been linked to DHA resistance [176, 281–283]. Additionally, JmjC-1 has been found to be over-expressed in DHA-resistant parasites [156, 158]. Further, in Cambodian

isolates, JmjC-1 mutations appear to have undergone recent population expansion or selection [284].

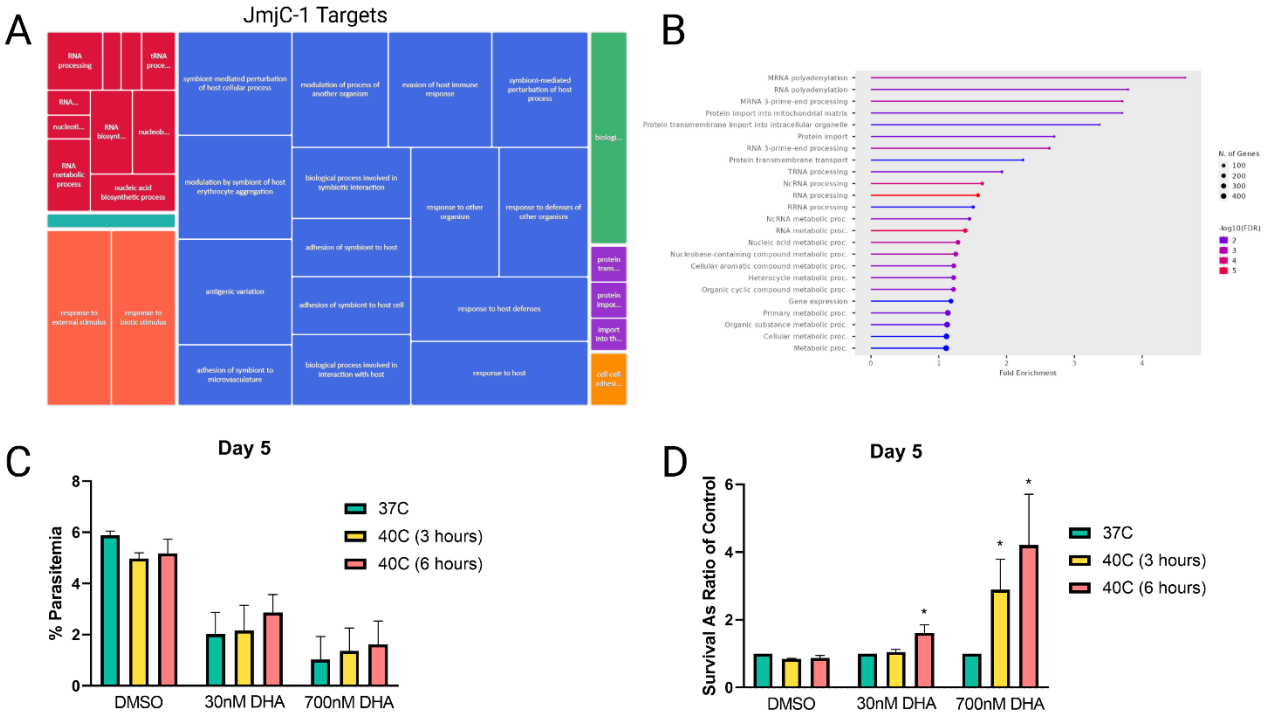

Figure 4.8 JmjC-1 targets and parasite survival to DHA following heat stress
A) Gene ontology enrichment analysis of JmjC-1's predicted targets by the EGRIN gene regulatory network model. Each rectangle represents a biological process, with color indicating similarity. Size is dependent on the p-value of enrichment.
B) Biological processes sorted by fold-enrichment, with color denoting p-value.
C) Day 5 parasitemia of NF54 parasites incubated at 37C, 40C for 3 hours, or 40C for 6 hours before treatment with DMSO, 30nM DHA, or 700nM DHA. Data shown represent the means from three replicates ± SEM.
D) Day 5 parasitemia as a ratio of the 37C control. Data shown represent the means from three replicates ± SEM. (* $p < 0.05$).

Using gene ontology (GO) enrichment analysis, I analyzed JmjC-1's predicted gene targets from the EGRIN model, identifying biological processes related to host processes, cell adhesion, protein import, response to external stimuli, and RNA processes. [Figure 4.8A]. The most enriched biological processes were related to RNA processing and protein transport [Figure 4.8B]. These findings are in agreement with known functions of JmjC-1, as well as the results of the EGRIN model identifying numerous protein channels as targets [262, 275]. Response to external stimuli and stress was also enriched by GO analysis,

suggesting JmjC-1 may aid in rapidly adapting to abrupt changes, such as heat treatment or antimalarials, similar to AP2-HS, another top hit from the CRISPRa-DHA screen.

To explore the relationship between elevated temperature and drug sensitivity, I examined parasite survival to DHA with and without heat treatment. I incubated tightly synchronized ring-stage NF54 parasites at 37C for 6 hours, 40C for 3 hours and 37C for 3 hours, or 40C for 6 hours. Then I treated each culture with DMSO, 30nM DHA, or 700nM DHA for 6 hours after. Parasites decreased in parasitemia in all conditions following DHA treatment at 30nM and 700nM DHA [Figure 4.8C]. However, parasites incubated at 40C first had significantly higher Day 5 parasitemia following 700nM DHA, with 6 hour treatment having a 4.2-fold increase in survival compared to 37C-incubated parasites [Figure 4.8D]. These results suggest febrile temperatures may induce a stress response, allowing for parasites to better survive drug pressure. Additionally, as heat-treated parasites have been shown to have increased transcriptomic heterogeneity, JmjC-1 may be acting as a variability inducer as predicted by the EGRIN model [277].

4.6 Discussion

These CRISPR TR screens allowed for both essentiality and drug phenotyping in the blood stage. This is also the first application of CRISPR screens in the malaria parasite, serving as a framework for future, more expansive screens. Additional characterization of top hits from both CRISPRa and CRISPRi screens could help elucidate the mechanisms behind the shown phenotypes. The relative ease of cloning pooled gRNA simplifies the work needed to prepare plasmid libraries that target tens, hundreds, or thousands of genes. This work also presents a novel strategy for efficient integration of hundreds of pooled plasmids, which can potentially be used for other applications. This strategy would likely be necessary for any efforts in creating a genome-wide CRISPR screen, as it overcomes many of the limitations inherent in genetic manipulation of *P. falciparum*.

The top hit from both CRISPRa and CRISPRi DHA screens was JmjC-1, a histone demethylase. The second hit from the CRISPRa screen was AP2-HS, a transcription factor involved in regulating heat shock response. Previous findings from other groups have identified relationships between heat shock and drug survival, offering a unifying theme [157, 176, 276]. Given how malaria parasites are very often exposed to febrile temperatures as part of the symptoms of malaria, the parasite may have evolved a mechanism of survival in these conditions. One possibility is that parasites diversify as a result of heat stress, allowing for a small subset to survive otherwise lethal treatment to drug. This bet-hedging strategy has been proposed for numerous biological functions in *Plasmodium,* including stage development [265, 266]. Notably, most *in vitro* measures of drug susceptibility (i.e. EC50 assays) are done at 37C, and not 40-42C, the temperatures found in malaria patients during fever. It may be worth revising current *in vitro* assays to include additional incubation temperatures, as this may identify previously undetected mechanisms of drug survival.

This screen could still be further explored to identify TRs important for processing specific metabolites or amino acids, or any other phenotypes of interest. Recent findings have also identified dormancy and recrudescence phenotypes in response to artemisinin. Applying these screens to measure those phenotypes may lead to further insights into the specific mechanism driving drug resistance.

It is important to note that these screens are primarily designed around identifying hits rather than providing objective measurements on every gene. Within any given gene, there were gRNAs that had significant phenotypes and others that had no effect at all. This is to be expected, however, as none of these gRNAs have been validated. Early mammalian CRISPR screens suffered from a similar problem, where the majority of gRNAs had little to no effect. As we are not able to directly measure the rates of these perturbations in a pool, we cannot differentiate between a gRNA that is effective but has no phenotype, or a non-effective gRNA. Furthermore, as this system works by perturbing

genes by histone acetylation at the genomic site, there is a high possibility of polar effects, where nearby genes are also affected. This is particularly notable when genes are in close proximity to each other or possess overlapping antisense sequences, with many of these neighbors being essential. Given the close proximity of neighboring genes, some of the measured effects of perturbation may be incorrectly attributed to one gene, when it is in fact due to a neighboring gene.

It is also important to note this system could only be used in a specific parasite line, NF54-DiCre-attB, which contains the required DiCre for induction and attB site for plasmid integration. Unfortunately, this line produces gametocytes inefficiently and incompletely, likely due to excessive continuous culture and idiosyncrasies of this specific clone. Additionally, this is a laboratory-grown line that has undoubtedly adapted to its *in vitro* conditions. The phenotypes in this line may therefore not be representative of parasites found in endemic regions.

As these represent the first CRISPR screens in malaria, numerous limitations are to be expected. The scope was small, targeting only 44 different genes, and the number of high quality gRNAs available for any of these genes was limited. This could be solved by adapting Cas9 variants, such as NG-Cas9 or SpRY-Cas9, for malaria research, work that is currently ongoing in many labs. Future screens could alternatively be done by creating arrayed screens, where each well contains multiple gRNAs for one specific gene. Many additional improvements are likely necessary for malaria genetic screens to catch up to mammalian and yeast fields. Ultimately, this work serves as a pilot for future CRISPR screens in the malaria field.

4.7 Materials and Methods

Plasmid cloning

To create the DCL3-CRISPRa and DCL3-CRISPRi plasmids, the DCL3-Cas9 plasmid was amplified to omit the HF-SpCas9, while the dCas9-GCN5 and dCas9-Sir2a were amplified to insert by Gibson assembly cloning. These plasmids were further modified to include a Bxb1 attP site, amplified from an existing plasmid. In order to create CRISPRa and CRISPRi libraries, gRNAs were first synthesized by IDT with 20bp homology arms to insert between the U6 promoter and tracrRNA found on the DCL3 plasmid. The plasmid was digested by Esp3I and oligos cloned in by Gibson assembly. Pools were established either by plating and scraping transformed colonies, or by direct inoculation of liquid culture following transformation. All plasmids for transfection were isolated using Machery-Nagel Endotoxin-Free maxipreps and then ethanol precipitated. DNA was resuspended to 5ug/uL in warm TE buffer.

DNA Sequencing

Plasmids and transfected parasites were amplified using PrimeStar Max polymerase, following the following parameters:

98C for 1 minute

98C for 10 seconds*

55-60C for 10 seconds*

64C for 30-60 seconds*

64C for 3 minutes

* Cycled 20-35 times

For amplifying gRNA libraries, the minimum number of cycles needed to attain visible PCR product by gel electrophoresis was first tested and then used for amplification of screens. For each parasite sample to be sequenced for pooled screens, PCRs were done using barcoded primers, allowing for up to 20 samples to be barcoded by PCR, then

pooled for sequencing. PCR products were purified by column purification or gel extraction followed by column purification.

All sequencing was performed by Plasmidsaurus. For sequencing to verify successful cloning, whole plasmid sequencing was used. For sequencing modified genomic DNA by Bxb1, amplicon sequencing was used. For sequencing gRNA screens and the library, Plasmidsaurus custom sequencing was used.

Parasite culture

P. falciparum parasites were maintained in complete medium consisting of RPMI-1640, 25mM HEPES, 2.4mM sodium bicarbonate, 1% Albumax II, and 2% gentamicin. Parasites were kept at 37C in a 90% N, 5% CO2, 5% O2 environment. Hematocrit was maintained between 4-5% using washed O+ human donor red blood cells. Synchronizations were performed using 5% sorbitol to achieve 0-3h synchronization. Parasitemia and morphology were checked by 10% Giemsa stain. For routine culture, an NF54 parasite line that was passed through the complete parasite lifecycle was used to ensure viability at each stage of development. For initial DiCre experiments, an NF54-DiCre line was kindly gifted by Moritz Treeck. For DiCre-attB experiments, an NF54-DiCre-attB line was kindly gifted by Josh Beck.

CRISPR screens were done by first synchronizing transfected parasites at 1-3% parasitemia to 0-3h rings using consecutive 5% sorbitol treatment. Following synchronization, parasites were induced by adding DMSO or 50nM of rapamycin for 6 hours. Parasites were then cultured as normal until the desired timepoint. For DHA screens, induced parasites were treated one full cycle (48hr) following rapamycin.

For heat treatment of parasites, cultures were synchronized to 0-3h rings by consecutive 5% sorbitol treatments, and then moved to a 40C water bath to ensure even heat distribution. Following heat treatment, parasites were either treated with DMSO, 30nM DHA, or 700nM DHA for 6 hours. Parasitemia was monitored by Giemsa smear.

Transfection

Single-plasmid and small pool transfections were performed by ring-stage electroporation using a BioRad Gene-Pulser II Electoporator set to 0.31kV and 950 µF capacitance. 50ug of total plasmid DNA was used per transfection, mixed with cytomix (120 mM KCl, 0.15 mM CaCl2, 2mM EGTA, 5 mM MgCl2, 10 mM K2HPO4/KH2PO4 25 mM HEPES, pH 7.6). Parasites were synchronized by sorbitol to attain ~5% parasitemia at ring-stage. The following day, parasites were pelleted, washed with cytomix, and mixed 1:1 with the plasmid-cytomix solution. Parasites were transferred to a 0.2cm electoporation cuvette before being electroporated. Electroporated parasites were immediately moved to pre-warmed flasks with 6mL CM and 250uL fresh RBCs.

Pooled plasmid transfections were performed by mature schizont electroporation using a Lonza 4D Nucleofector and Protocol FP158. 100ug total pooled plasmid DNA was used per transfection, resuspended into 100 µl of P3 Primary Cell 4D-Nucleofector X Kit. Tightly synchronized mature schizonts were isolated by first performing sorbitol synchronizations over two parasite cycles, then performing a magnetic column purification at parasite hour 45-47 to attain pure schizonts. Parasites were cultured in 5mL CM and carefully monitored until fully mature schizonts were abundant. Parasites were pelleted and 10uL of pure mature schizonts were mixed with the 100uL plasmid-P3 solution before being electroporated. Electroporated parasites were immediately moved to pre-warmed flasks with 6mL CM and 250uL fresh RBCs.

Following transfection, flasks were gassed and incubated overnight. Media was changed for two days (one complete parasite cycle) before drug was added. For all DCL3 transfections, 2.5 ug/mL Blasticidin was used. For pINT transfections, 250 ug/mL G418 was used. For CRISPRi pilot pools, 2.5 ug/mL Blasticidin and 2.5 nM WR99210 was used. For plasmids not Bxb1 integrated, drug was maintained indefinitely to ensure plasmid maintenance. For Bxb1-integrated plasmids, drug was only maintained for 6 days. Cultures were maintained at 4-5% hematocrit. Cultures were cut by 80% every other week

to replenish with fresh blood. Media was changed every other day until parasites were visible by Giemsa smear.

DNA Extractions

Genomic DNA from parasites was isolated using GeneJET Whole Blood Genomic DNA Purification Kits (Thermo Fisher. Prior to isolating, a minimum of 250uL packed infected red blood cells were spun down and resuspended with 1x PBS + 0.015% saponin, set at room temperature for 5-10 minutes, and then pelleted. The supernatant was discarded and the remaining parasite pellet was used following the manufacturer's protocol. For the final elution, 50uL of warm (37-50C) water was used in place the provided elution buffer, incubating for 5 minutes at room temperature before spinning at 13,000 x g for 1.5 minutes. The purified DNA was then quantified by Nanodrop.

Fluorescent Microscopy

Infected red blood cells were stained with 1% DAPI and incubated at 37C for 10 minutes before washing twice and being resuspended in complete media. Cells were then moved to a slide and a coverslip was added. The cells were viewed using a Nikon Eclipse fluorescent microscope at 400-1000x total magnification. For DAPI detection, the UV filter was used and for GFP or mNeonGreen detection, the blue filter was used. Parasitemias were quantified from the presence or absence of DAPI and mNeonGreen, depending on application.

Chapter 5 – Conclusion

5.1 Future Directions

Advancements in molecular tools and computational approaches have opened up new possibilities in all fields of biology. In the last decade, there have been thousands of CRISPR-based screens covering mammalian systems, plants, bacteria, yeast, and now *Plasmodium.* These approaches are distinctly important for helping expand our limited knowledge in malaria research. The malaria field has traditionally lagged behind research on mammals and yeast, but that need not always be the case. By applying the latest technologies to malaria research, we can greatly accelerate our mission to eradicate malaria, using newly found knowledge to design new antimalarials and vaccines.

Computational biology and bioinformatics have become increasingly important for malaria research, coinciding with the accessibility of generating large genomic and transcriptomic datasets. There are now tens of thousands of unique genomic datasets coming from field isolates, with many more originating from *in vitro* culture. The sheer volume of data is challenging to work with but provides that much more knowledge. Particularly, identifying patterns of evolution will be key to preempting the rise of antimalarial resistance. Many of the current SNPs thought to be conferring drug resistance originated in the Southeast Asian *P. falciparum* background, and do not naturally occur in African parasite populations. However, this may not always be the case. By carefully analyzing what changes are occurring at the genome level, the field can better combat drug resistance.

Similarly, the increased use of transcriptomics has contributed greatly to malaria research. In particular, the adaptation of single-cell transcriptomics to malaria has provided numerous insights into gene regulation at single-cell resolution. This has opened

the door to examining heterogeneity within a population of parasites. There are numerous applications to this new technology. Applying single-cell transcriptomics to field samples may help deconvolve the complexity of gametocyte development, drug resistance, and multi-strain infections by measuring both heterogeneity and transcriptomics. In these cases, parasites may be intentionally diversifying, as a means of ensuring survival of the population. Single-cell transcriptomics would be able to help zero in on how much heterogeneity there is, and how it might be contributing to differentiation, or drug resistance.

Pairing single-cell transcriptomics with CRISPR screens, such as the CRISPRa and CRISPRi screens, would allow for "perturb-seq" experiments. This would involve perturbing a subset of genes (such as transcription factors) and measuring the transcriptomic output following perturbation. However, numerous challenges still exist in order to begin effectively performing perturb-seq experiments. Currently, even the best single-cell datasets in *P. falciparum* have limited reads per cell, and measure a limited number of genes expressed per cell. As a result, achieving the statistical power necessary to identify gene-gene interactions from the perturbations is very difficult. One would likely need thousands of cells per perturbation and may not ever be able to attain the resolution required to identify most of the gene-gene interactions. Once these challenges have been overcome, perturb-seq experiments could provide countless insights into what genes are being regulated and how.

The CRISPR screens presented here offer a pooled, accessible means of over- and under-expressing tens, hundreds, or potentially thousands of different genes in parallel. However, in a pooled format, changes to morphology cannot be easily determined as a result of the CRISPR perturbation. One solution would be to apply a morphological-based screen to identify morphology phenotypes associated with the gene perturbation. For instance, applying the visual cell sorting approach to identify differentiating gametocytes, could help to identify gametocyte-inducing transcriptional regulators.

An alternative solution would be to establish an arrayed CRISPR screen for *Plasmodium.* Utilizing 96-well electroporation systems, 96 different genes (one per well) could be perturbed via CRISPRa, CRISPRi, or HDR KO. In yeast, a genome-wide arrayed gene deletion screen has been established. The establishment of these seminal screens led to numerous phenotype screens, including biological function screens, environmental stress screens, and drug screens. Although the piggyBac transposon screen has provided some of these features, the isolated parasites are inherently dispensable for the blood stage and cannot be studied further in subsequent stages. Hence, the ideal arrayed screen for *Plasmodium* would offer a means of inducibly controlling the perturbation. In such a system, one could measure growth or morphology phenotypes in blood stage, gametocyte stage, or even potentially mosquito and liver stages. Additionally, isolating the knockout is trivial once an interesting phenotype has been identified. An arrayed screen of this sort would provide great value to the field by helping prevent wasteful efforts in creating individual knockouts and allowing screens to be performed accessibly.

There is a great need in adapting some of the latest molecular tools, such as prime-editing or helicase-assisted continuous editing. These tools still rely on Cas9 for targeting, but do not depend on HDR to implement the mutations. These advancements may trivialize many of the current difficulties in targeted editing. HDR for instance, requires providing a gRNA and donor. In many cases, it is simply not feasible to design effective donor templates for DNA synthesis, and thus requires multiple cloning steps. There would be great value in molecular tools that could knockout or mutate specific genes using only a gRNA. This would allow for stronger screens, and potentially help identify resistance mutants that would otherwise be inaccessible.

Similarly, improvements to the CRISPRa and CRISPRi systems in *Plasmodium* are still needed. The current iterations (dCas9-GCN5 and dCas9-Sir2a respectively) more closely resemble some of the early generation CRISPRa and CRISPRi tools in mammalian systems, which depend on single domains for acetylation or deacetylation. By adapting

some of the approaches used for modern mammalian CRISPRa and CRISPRi, such as the synergistic activation mediator (SAM) approach, stronger over- and under-expression effects can be done. The SAM approach utilizes a modified sgRNA stemloop to attract transcriptional activators to the stemloop of the gRNA, and thus the target location, for gene activation. Along with the various kinds of Cas9 effectors and functions, expanded PAM recognition would provide many more viable gRNAs to select from. This is a major limitation with the current system, as SpCas9 can only target NGG PAMs. This is particularly problematic as *Plasmodium* has 82% AT-rich DNA and even higher in the UTRs, where CRISPRa and CRISPRi must be targeted towards. Consequently, it is currently not uncommon to find less than 5 viable gRNAs in the 3kb region upstream of a gene. The current systems could be adapted to use NG-Cas9 (targeting an NG PAM) and SpRY (nearly PAMless targeting), or any of the newer variants. This could potentially provide 5-20x more gRNAs to select, helping ensure that there are effective gRNAs available.

Applying the conditional systems to other platforms, such as piggyBac, may provide another avenue for establishing genome-wide knockout screens. In the case of an inducible transposon screen, one could put the piggyBac transposase under conditional expression, preventing insertions from occurring until the stage of interest, and using the parasite's replication rather than successful plasmid uptakes as the means of increasing transposon counts. Here, one could simply transfect the inducible piggyBac, grow to high parasitemia, and induce in the stage of interest to immediately generate a completely saturated transposon screen. This would surely help in alleviating some of the current limitations of the piggyBac screens and reduce the number of false positives and false negatives. One could also better measure growth phenotypes, as sequencing can be done every cycle to identify fast, slow, or normal growing mutations.

5.2 Concluding Remarks

The implementation of computational biology, transcriptomics, and CRISPR/Cas9 systems have helped transform the broader field of biology. These tools provide a means of characterizing genomes as well as identifying the basic biological processes underlying any phenotype. The work discussed in my dissertation involved applying some of these approaches to *P. falciparum*, where they have been largely underutilized. I compiled and analyzed numerous datasets on gene expression to establish a unified dataset that encompasses all of the stages of the parasite lifecycle. Adapting approaches from external fields, I developed inducible Cas9, CRISPRa, and CRISPRi toolsets that could be applied to identify inaccessible phenotypes in *P. falciparum*. I envision the work described here will provide a framework for future genetic screens in *Plasmodium*, as well as aiding in the design of new vaccines and antimalarials, ultimately bringing us closer to eradicating malaria.

References

1. Talapko, J., Škrlec, I., Alebić, T., Jukić, M., and Včev, A. (2019). Malaria: the past and the present. Microorganisms *7*.
2. Venkatesan, P. (2024). The 2023 WHO World malaria report. Lancet Microbe *5*, e214.
3. Poespoprodjo, J. R., Douglas, N. M., Ansong, D., Kho, S., and Anstey, N. M. (2023). Malaria. Lancet *402*, 2328–2345.
4. Weiss, D. J., Bertozzi-Villa, A., Rumisha, S. F., Amratia, P., Arambepola, R., Battle, K. E., Cameron, E., Chestnutt, E., Gibson, H. S., Harris, J., et al. (2021). Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: a geospatial modelling analysis. Lancet Infect. Dis. *21*, 59–69.
5. Sinnis, P., and Coppi, A. (2007). A long and winding road: the Plasmodium sporozoite's journey in the mammalian host. Parasitol. Int. *56*, 171–178.
6. Goswami, D., Minkah, N. K., and Kappe, S. H. I. (2022). Malaria parasite liver stages. J. Hepatol. *76*, 735–737.
7. Scheiner, M., Burda, P.-C., and Ingmundson, A. (2024). Moving on: How malaria parasites exit the liver. Mol. Microbiol. *121*, 328–340.
8. Cowman, A. F., Healer, J., Marapana, D., and Marsh, K. (2016). Malaria: biology and disease. Cell *167*, 610–624.
9. White, N. J. (2022). Severe malaria. Malar. J. *21*, 284.
10. Ngotho, P., Soares, A. B., Hentzschel, F., Achcar, F., Bertuccini, L., and Marti, M. (2019). Revisiting gametocyte biology in malaria parasites. FEMS Microbiol. Rev. *43*, 401–414.
11. Venugopal, K., Hentzschel, F., Valkiūnas, G., and Marti, M. (2020). Plasmodium asexual growth and sexual development in the haematopoietic niche of the host. Nat. Rev. Micro. *18*, 177–189.
12. Neveu, G., Richard, C., Dupuy, F., Behera, P., Volpe, F., Subramani, P. A., Marcel-Zerrougui, B., Vallin, P., Andrieu, M., Minz, A. M., et al. (2020). Plasmodium falciparum sexual parasites develop in human erythroblasts and affect erythropoiesis. Blood *136*, 1381–1393.
13. Meibalan, E., and Marti, M. (2017). Biology of malaria transmission. Cold Spring Harb. Perspect. Med. *7*.
14. Kar, S., and Kar, S. (2010). Control of malaria. Nat. Rev. Drug Discov. *9*, 511–512.

15. Böhmert, A. L., Logan, R. A. E., Portwood, N. M., Hartke, J., and Ingham, V. A. (2024). A descriptive review of next-generation insecticide-treated bed nets for malaria control. Front. Malar. *2*.
16. Pryce, J., Richardson, M., and Lengeler, C. (2018). Insecticide-treated nets for preventing malaria. Cochrane Database Syst. Rev. *11*, CD000363.
17. Tse, E. G., Korsik, M., and Todd, M. H. (2019). The past, present and future of anti-malarial medicines. Malar. J. *18*, 93.
18. Nosten, F., and White, N. J. (2007). Artemisinin-based combination treatment of falciparum malaria. Am. J. Trop. Med. Hyg. *77*, 181–192.
19. Wang, J., Xu, C., Wong, Y. K., Li, Y., Liao, F., Jiang, T., and Tu, Y. (2019). Artemisinin, the Magic Drug Discovered from Traditional Chinese Medicine. Engineering *5*, 32–39.
20. Benakis, A., Paris, M., Loutan, L., Plessas, C. T., and Plessas, S. T. (1997). Pharmacokinetics of artemisinin and artesunate after oral administration in healthy volunteers. Am. J. Trop. Med. Hyg. *56*, 17–23.
21. White, N. J., and Chotivanich, K. (2024). Artemisinin-resistant malaria. Clin. Microbiol. Rev. *37*, e0010924.
22. Wang, S., Huang, F., Yan, H., Yin, J., and Xia, Z. (2023). A review of malaria molecular markers for drug resistance in Plasmodium falciparum and Plasmodium vivax in China. Front. Cell Infect. Microbiol. *13*, 1167220.
23. Ménard, D., Khim, N., Beghain, J., Adegnika, A. A., Shafiul-Alam, M., Amodu, O., Rahim-Awab, G., Barnadas, C., Berry, A., Boum, Y., et al. (2016). A Worldwide Map of Plasmodium falciparum K13-Propeller Polymorphisms. N. Engl. J. Med. *374*, 2453–2464.
24. Duraisingh, M. T., and Cowman, A. F. (2005). Contribution of the pfmdr1 gene to antimalarial drug-resistance. Acta Trop. *94*, 181–190.
25. Bopp, S., Magistrado, P., Wong, W., Schaffner, S. F., Mukherjee, A., Lim, P., Dhorda, M., Amaratunga, C., Woodrow, C. J., Ashley, E. A., et al. (2018). Plasmepsin II-III copy number accounts for bimodal piperaquine resistance among Cambodian Plasmodium falciparum. Nat. Commun. *9*, 1769.
26. Fidock, D. A., Nomura, T., Talley, A. K., Cooper, R. A., Dzekunov, S. M., Ferdig, M. T., Ursos, L. M., Sidhu, A. B., Naudé, B., Deitsch, K. W., et al. (2000). Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. Mol. Cell *6*, 861–871.

27. Wicht, K. J., Mok, S., and Fidock, D. A. (2020). Molecular Mechanisms of Drug Resistance in Plasmodium falciparum Malaria. Annu. Rev. Microbiol. *74*, 431–454.
28. Hammershaimb, E. A., and Berry, A. A. (2024). Pre-erythrocytic malaria vaccines: RTS,S, R21, and beyond. Expert Rev. Vaccines *23*, 49–52.
29. Collins, K. A., Snaith, R., Cottingham, M. G., Gilbert, S. C., and Hill, A. V. S. (2017). Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. Sci. Rep. *7*, 46621.
30. Singh, A. P., Buscaglia, C. A., Wang, Q., Levay, A., Nussenzweig, D. R., Walker, J. R., Winzeler, E. A., Fujii, H., Fontoura, B. M. A., and Nussenzweig, V. (2007). Plasmodium circumsporozoite protein promotes the development of the liver stages of the parasite. Cell *131*, 492–504.
31. Olotu, A., Fegan, G., Wambua, J., Nyangweso, G., Leach, A., Lievens, M., Kaslow, D. C., Njuguna, P., Marsh, K., and Bejon, P. (2016). Seven-Year Efficacy of RTS,S/AS01 Malaria Vaccine among Young African Children. N. Engl. J. Med. *374*, 2519–2529.
32. Tinto, H., Otieno, W., Gesase, S., Sorgho, H., Otieno, L., Liheluka, E., Valéa, I., Sing'oei, V., Malabeja, A., Valia, D., et al. (2019). Long-term incidence of severe malaria following RTS,S/AS01 vaccination in children and infants in Africa: an open-label 3-year extension study of a phase 3 randomised controlled trial. Lancet Infect. Dis. *19*, 821–832.
33. Datoo, M. S., Dicko, A., Tinto, H., Ouédraogo, J.-B., Hamaluba, M., Olotu, A., Beaumont, E., Ramos Lopez, F., Natama, H. M., Weston, S., et al. (2024). Safety and efficacy of malaria vaccine candidate R21/Matrix-M in African children: a multicentre, double-blind, randomised, phase 3 trial. Lancet.
34. Birkett, A., Miller, R. S., and Soisson, L. A. (2022). The Importance of Exercising Caution When Comparing Results from Malaria Vaccines Administered on the EPI Schedule and on a Seasonal Schedule. Am. J. Trop. Med. Hyg. *107*, 1356.
35. Goswami, D., Minkah, N. K., and Kappe, S. H. I. (2019). Designer parasites: genetically engineered plasmodium as vaccines to prevent malaria infection. J. Immunol. *202*, 20–28.
36. Nunes-Cabaço, H., Moita, D., and Prudêncio, M. (2022). Five decades of clinical assessment of whole-sporozoite malaria vaccines. Front. Immunol. *13*, 977472.
37. Abuelazm, M. T., Elzeftawy, M. A., Kamal, M. A., Badr, H., Gamal, M., Aboulgheit, M., Abdelazeem, B., Abd-Elsalam, S., and Abouzid, M. (2024). Protective efficacy and safety of radiation-attenuated and chemo-attenuated Plasmodium Falciparum

sporozoite vaccines against controlled and natural malaria infection: a systematic review and meta-analysis of randomized controlled trials. Infection.
38. Vaughan, A. M., Wang, R., and Kappe, S. H. I. (2010). Genetically engineered, attenuated whole-cell vaccine approaches for malaria. Hum Vaccin *6*, 107–113.
39. Kumar, H., Frischknecht, F., Mair, G. R., and Gomes, J. (2015). In silico identification of genetically attenuated vaccine candidate genes for Plasmodium liver stage. Infect. Genet. Evol. *36*, 72–81.
40. Canepa, G. E., Molina-Cruz, A., Yenkoidiok-Douti, L., Calvo, E., Williams, A. E., Burkhardt, M., Peng, F., Narum, D., Boulanger, M. J., Valenzuela, J. G., et al. (2018). Antibody targeting of a specific region of Pfs47 blocks Plasmodium falciparum malaria transmission. npj Vaccines *3*, 26.
41. Duffy, P. E. (2021). Transmission-Blocking Vaccines: Harnessing Herd Immunity for Malaria Elimination. Expert Rev. Vaccines *20*, 185–198.
42. Salinas, N. D., Ma, R., Dickey, T. H., McAleese, H., Ouahes, T., Long, C. A., Miura, K., Lambert, L. E., and Tolia, N. H. (2023). A potent and durable malaria transmission-blocking vaccine designed from a single-component 60-copy Pfs230D1 nanoparticle. npj Vaccines *8*, 124.
43. 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. [version 2; peer review: 3 approved]. Wellcome Open Res. *4*, 58.
44. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498–511.
45. Frech, C., and Chen, N. (2011). Genome comparison of human and non-human malaria parasites reveals species subset-specific genes potentially linked to human disease. PLoS Comput. Biol. *7*, e1002320.
46. Gardiner, D. L., Skinner-Adams, T. S., Spielmann, T., and Trenholme, K. R. (2003). Malaria transfection and transfection vectors. Trends Parasitol. *19*, 381–383.
47. Wang, S., Zeng, W., Zhao, W., Xiang, Z., Zhao, H., Yang, Q., Li, X., Duan, M., Li, X., Wang, X., et al. (2022). Comparison of in vitro transformation efficiency methods for Plasmodium falciparum. Mol. Biochem. Parasitol. *247*, 111432.
48. Mueller, A.-K., Hammerschmidt-Kamper, C., and Kaiser, A. (2014). RNAi in Plasmodium. CPD *20*, 278–283.

49. Lee, A. H., Symington, L. S., and Fidock, D. A. (2014). DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum. Microbiol. Mol. Biol. Rev. *78*, 469–486.
50. Wu, Y., Sifri, C. D., Lei, H. H., Su, X. Z., and Wellems, T. E. (1995). Transfection of Plasmodium falciparum within human red blood cells. Proc. Natl. Acad. Sci. USA *92*, 973–977.
51. Su, X.-Z., Lane, K. D., Xia, L., Sá, J. M., and Wellems, T. E. (2019). Plasmodium Genomics and Genetics: New Insights into Malaria Pathogenesis, Drug Resistance, Epidemiology, and Evolution. Clin. Microbiol. Rev. *32*.
52. 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*, 1661–1666.
53. Cowman, A. F., Morry, M. J., Biggs, B. A., Cross, G. A., and Foote, S. J. (1988). Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. Proc. Natl. Acad. Sci. USA *85*, 9109–9113.
54. Zolg, J. W., Plitt, J. R., Chen, G. X., and Palmer, S. (1989). Point mutations in the dihydrofolate reductase-thymidylate synthase gene as the molecular basis for pyrimethamine resistance in Plasmodium falciparum. Mol. Biochem. Parasitol. *36*, 253–262.
55. Vaidya, A. B., Morrisey, J., Plowe, C. V., Kaslow, D. C., and Wellems, T. E. (1993). Unidirectional dominance of cytoplasmic inheritance in two genetic crosses of Plasmodium falciparum. Mol. Cell. Biol. *13*, 7349–7357.
56. Vaidya, A. B., Muratova, O., Guinet, F., Keister, D., Wellems, T. E., and Kaslow, D. C. (1995). A genetic locus on Plasmodium falciparum chromosome 12 linked to a defect in mosquito-infectivity and male gametogenesis. Mol. Biochem. Parasitol. *69*, 65–71.
57. Duraisingh, M. T., Roper, C., Walliker, D., and Warhurst, D. C. (2000). Increased sensitivity to the antimalarials mefloquine and artemisinin is conferred by mutations in the pfmdr1 gene of Plasmodium falciparum. Mol. Microbiol. *36*, 955–961.
58. Hayton, K., Gaur, D., Liu, A., Takahashi, J., Henschen, B., Singh, S., Lambert, L., Furuya, T., Bouttenot, R., Doll, M., et al. (2008). Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of Plasmodium falciparum invasion. Cell Host Microbe *4*, 40–51.

59. Molina-Cruz, A., Garver, L. S., Alabaster, A., Bangiolo, L., Haile, A., Winikor, J., Ortega, C., van Schaijk, B. C. L., Sauerwein, R. W., Taylor-Salmon, E., et al. (2013). The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. Science *340*, 984–987.
60. Vaughan, A. M., Pinapati, R. S., Cheeseman, I. H., Camargo, N., Fishbaugher, M., Checkley, L. A., Nair, S., Hutyra, C. A., Nosten, F. H., Anderson, T. J. C., et al. (2015). Plasmodium falciparum genetic crosses in a humanized mouse model. Nat. Methods *12*, 631–633.
61. Vendrely, K. M., Kumar, S., Li, X., and Vaughan, A. M. (2020). Humanized mice and the rebirth of malaria genetic crosses. Trends Parasitol. *36*, 850–863.
62. Kumar, S., Li, X., McDew-White, M., Reyes, A., Delgado, E., Sayeed, A., Haile, M. T., Abatiyow, B. A., Kennedy, S. Y., Camargo, N., et al. (2022). A Malaria Parasite Cross Reveals Genetic Determinants of Plasmodium falciparum Growth in Different Culture Media. Front. Cell Infect. Microbiol. *12*, 878496.
63. Kanai, M., Mok, S., Yeo, T., Shears, M. J., Ross, L. S., Jeon, J. H., Narwal, S., Haile, M. T., Tripathi, A. K., Mlambo, G., et al. (2024). Identification of the drug/metabolite transporter 1 as a marker of quinine resistance in a NF54×Cam3.II P. falciparum genetic cross. BioRxiv.
64. Kane, J., Li, X., Kumar, S., Button-Simons, K. A., Vendrely Brenneman, K. M., Dahlhoff, H., Sievert, M. A. C., Checkley, L. A., Shoue, D. A., Singh, P. P., et al. (2024). A Plasmodium falciparum genetic cross reveals the contributions of pfcrt and plasmepsin II/III to piperaquine drug resistance. MBio *15*, e0080524.
65. Button-Simons, K. A., Kumar, S., Carmago, N., Haile, M. T., Jett, C., Checkley, L. A., Kennedy, S. Y., Pinapati, R. S., Shoue, D. A., McDew-White, M., et al. (2021). The power and promise of genetic mapping from Plasmodium falciparum crosses utilizing human liver-chimeric mice. Commun. Biol. *4*, 734.
66. Duraisingh, M. T., Triglia, T., and Cowman, A. F. (2002). Negative selection of Plasmodium falciparum reveals targeted gene deletion by double crossover recombination. Int. J. Parasitol. *32*, 81–89.
67. Ghorbal, M., Gorman, M., Macpherson, C. R., Martins, R. M., Scherf, A., and Lopez-Rubio, J.-J. (2014). Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. Nat. Biotechnol. *32*, 819–821.
68. Bushell, E., Gomes, A. R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modrzynska, K., Schwach, F., Martin, R. E., et al. (2017). Functional profiling of a

plasmodium genome reveals an abundance of essential genes. Cell *170*, 260–272.e8.

69. Stanway, R. R., Bushell, E., Chiappino-Pepe, A., Roques, M., Sanderson, T., Franke-Fayard, B., Caldelari, R., Golomingi, M., Nyonda, M., Pandey, V., et al. (2019). Genome-Scale Identification of Essential Metabolic Processes for Targeting the Plasmodium Liver Stage. Cell *179*, 1112–1128.e26.
70. Zhang, M., Wang, C., Otto, T. D., Oberstaller, J., Liao, X., Adapa, S. R., Udenze, K., Bronner, I. F., Casandra, D., Mayho, M., et al. (2018). Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis. Science *360*.
71. Li, M. A., Pettitt, S. J., Eckert, S., Ning, Z., Rice, S., Cadiñanos, J., Yusa, K., Conte, N., and Bradley, A. (2013). The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. Mol. Cell. Biol. *33*, 1317–1330.
72. Yoshida, J., Akagi, K., Misawa, R., Kokubu, C., Takeda, J., and Horie, K. (2017). Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. Sci. Rep. *7*, 43613.
73. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.
74. Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.
75. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.
76. Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., Cheng, A. W., Trevino, A. E., Konermann, S., Chen, S., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. *32*, 670–676.
77. Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelson, T., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.
78. Xue, C., and Greene, E. C. (2021). DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. Trends Genet. *37*, 639–656.

79. Wagner, J. C., Platt, R. J., Goldfless, S. J., Zhang, F., and Niles, J. C. (2014). Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. Nat. Methods *11*, 915–918.
80. Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.
81. Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. Cell *159*, 647–661.
82. Levy, S., Somasundaram, L., Raj, I. X., Ic-Mex, D., Phal, A., Schmidt, S., Ng, W. I., Mar, D., Decarreau, J., Moss, N., et al. (2022). dCas9 fusion to computer-designed PRC2 inhibitor reveals functional TATA box in distal promoter region. Cell Rep. *38*, 110457.
83. Wang, C., Qu, Y., Cheng, J. K. W., Hughes, N. W., Zhang, Q., Wang, M., and Cong, L. (2022). dCas9-based gene editing for cleavage-free genomic knock-in of long sequences. Nat. Cell Biol. *24*, 268–278.
84. 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*, 1173–1183.
85. Alerasool, N., Segal, D., Lee, H., and Taipale, M. (2020). An efficient KRAB domain for CRISPRi applications in human cells. Nat. Methods *17*, 1093–1096.
86. Zhang, Y., Yin, C., Zhang, T., Li, F., Yang, W., Kaminski, R., Fagan, P. R., Putatunda, R., Young, W.-B., Khalili, K., et al. (2015). CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. Sci. Rep. *5*, 16277.
87. Moradpour, M., and Abdulah, S. N. A. (2020). CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. Plant Biotechnol. J. *18*, 32–44.
88. Xiao, B., Yin, S., Hu, Y., Sun, M., Wei, J., Huang, Z., Wen, Y., Dai, X., Chen, H., Mu, J., et al. (2019). Epigenetic editing by CRISPR/dCas9 in Plasmodium falciparum. Proc. Natl. Acad. Sci. USA *116*, 255–260.
89. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

90. Rees, H. A., and Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788.
91. Billon, P., Bryant, E. E., Joseph, S. A., Nambiar, T. S., Hayward, S. B., Rothstein, R., and Ciccia, A. (2017). CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. Mol. Cell *67*, 1068–1079.e4.
92. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149–157.
93. Anzalone, A. V., Koblan, L. W., and Liu, D. R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. *38*, 824–844.
94. Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., Lu, T., Maroc, L., Norman, T. M., Song, B., et al. (2022). High-content CRISPR screening. Nat. Rev. Methods Primers *2*.
95. Smith, A. J., De Sousa, M. A., Kwabi-Addo, B., Heppell-Parton, A., Impey, H., and Rabbitts, P. (1995). A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. Nat. Genet. *9*, 376–385.
96. Kim, H., Kim, M., Im, S.-K., and Fang, S. (2018). Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. Lab Anim Res *34*, 147–159.
97. Yu, H.-M. I., Liu, B., Chiu, S.-Y., Costantini, F., and Hsu, W. (2005). Development of a unique system for spatiotemporal and lineage-specific gene expression in mice. Proc. Natl. Acad. Sci. USA *102*, 8615–8620.
98. Kent, R. S., Modrzynska, K. K., Cameron, R., Philip, N., Billker, O., and Waters, A. P. (2018). Inducible developmental reprogramming redefines commitment to sexual development in the malaria parasite Plasmodium berghei. Nat. Microbiol. *3*, 1206–1213.
99. Das, A. T., Tenenbaum, L., and Berkhout, B. (2016). Tet-On Systems For Doxycycline-inducible Gene Expression. Curr Gene Ther *16*, 156–167.
100. Baron, U., Freundlieb, S., Gossen, M., and Bujard, H. (1995). Co-regulation of two gene activities by tetracycline via a bidirectional promoter. Nucleic Acids Res. *23*, 3605–3606.
101. Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science *268*, 1766–1769.

102. Sun, Y., Chen, X., and Xiao, D. (2007). Tetracycline-inducible expression systems: new strategies and practices in the transgenic mouse modeling. Acta Biochim. Biophys. Sin. (Shanghai) *39*, 235–246.
103. Pham, D. H., Moretti, P. A. B., Goodall, G. J., and Pitson, S. M. (2008). Attenuation of leakiness in doxycycline-inducible expression via incorporation of 3' AU-rich mRNA destabilizing elements. BioTechniques *45*, 155–6, 158, 160 passim.
104. Costello, A., Lao, N. T., Gallagher, C., Capella Roca, B., Julius, L. A. N., Suda, S., Ducrée, J., King, D., Wagner, R., Barron, N., et al. (2019). Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance. Biotechnol. J. *14*, e1800219.
105. Morgens, D. W., Deans, R. M., Li, A., and Bassik, M. C. (2016). Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Nat. Biotechnol. *34*, 634–636.
106. Sun, N., Petiwala, S., Wang, R., Lu, C., Hu, M., Ghosh, S., Hao, Y., Miller, C. P., and Chung, N. (2019). Development of drug-inducible CRISPR-Cas9 systems for large-scale functional screening. BMC Genomics *20*, 225.
107. Lundin, A., Porritt, M. J., Jaiswal, H., Seeliger, F., Johansson, C., Bidar, A. W., Badertscher, L., Wimberger, S., Davies, E. J., Hardaker, E., et al. (2020). Development of an ObLiGaRe Doxycycline Inducible Cas9 system for pre-clinical cancer drug discovery. Nat. Commun. *11*, 4903.
108. Zetsche, B., Volz, S. E., and Zhang, F. (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. Nat. Biotechnol. *33*, 139–142.
109. Khajanchi, N., and Saha, K. (2022). Controlling CRISPR with small molecule regulation for somatic cell genome editing. Mol. Ther. *30*, 17–31.
110. Rose, J. C., Stephany, J. J., Valente, W. J., Trevillian, B. M., Dang, H. V., Bielas, J. H., Maly, D. J., and Fowler, D. M. (2017). Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. Nat. Methods *14*, 891–896.
111. Wei, C. T., Maly, D. J., and Fowler, D. M. (2020). Temporal and rheostatic control of genome editing with a chemically-inducible Cas9. Meth. Enzymol. *633*, 119–141.
112. Wei, C. T., Popp, N. A., Peleg, O., Powell, R. L., Borenstein, E., Maly, D. J., and Fowler, D. M. (2023). A chemically controlled Cas9 switch enables temporal modulation of diverse effectors. Nat. Chem. Biol. *19*, 981–991.
113. Meltzer, H., Marom, E., Alyagor, I., Mayseless, O., Berkun, V., Segal-Gilboa, N., Unger, T., Luginbuhl, D., and Schuldiner, O. (2019). Tissue-specific (ts)CRISPR as an efficient strategy for in vivo screening in Drosophila. Nat. Commun. *10*, 2113.

114. Rahman, F., Mishra, A., Gupta, A., and Sharma, R. (2022). Spatiotemporal regulation of crispr/cas9 enables efficient, precise, and heritable edits in plant genomes. Front. Genome Ed. *4*, 870108.
115. Ding, Y., Wang, L., Ji, W., Chen, Z., Wang, D., Chen, C., Tong, H., Han, Z., Niu, C., Chu, M., et al. (2020). Generation of a human induced pluripotent stem cell line with Cas9 driven by Tet-on operator via AAVS1 safe harbor gene-editing. Stem Cell Res. *49*, 102064.
116. Jurlina, S. L., Jones, M. K., Agarwal, D., De La Toba, D. V., Kambli, N., Su, F., Martin, H. M., Anderson, R., Wong, R. M., Seid, J., et al. (2022). A Tet-Inducible CRISPR Platform for High-Fidelity Editing of Human Pluripotent Stem Cells. Genes (Basel) *13*.
117. Hans, S., Zöller, D., Hammer, J., Stucke, J., Spieß, S., Kesavan, G., Kroehne, V., Eguiguren, J. S., Ezhkova, D., Petzold, A., et al. (2021). Cre-Controlled CRISPR mutagenesis provides fast and easy conditional gene inactivation in zebrafish. Nat. Commun. *12*, 1125.
118. Gendron, W. A. C., Rubin, J. D., Hansen, M. J., Nace, R. A., Simone, B. W., Ekker, S. C., and Barry, M. A. (2021). Unlocking loxP to Track Genome Editing In Vivo. genes *12*, 1204.
119. Le Roch, K. G., Chung, D. W. D., and Ponts, N. (2012). Genomics and integrated systems biology in Plasmodium falciparum: a path to malaria control and eradication. Parasite Immunol *34*, 50–60.
120. Talundzic, E., Ravishankar, S., Kelley, J., Patel, D., Plucinski, M., Schmedes, S., Ljolje, D., Clemons, B., Madison-Antenucci, S., Arguin, P. M., et al. (2018). Next-Generation Sequencing and Bioinformatics Protocol for Malaria Drug Resistance Marker Surveillance. Antimicrob. Agents Chemother. *62*.
121. Akash, S., Abdelkrim, G., Bayil, I., Hosen, M. E., Mukerjee, N., Shater, A. F., Saleh, F. M., Albadrani, G. M., Al-Ghadi, M. Q., Abdel-Daim, M. M., et al. (2023). Antimalarial drug discovery against malaria parasites through haplopine modification: An advanced computational approach. J. Cell Mol. Med. *27*, 3168–3188.
122. Phelan, J. E., Turkiewicz, A., Manko, E., Thorpe, J., Vanheer, L. N., van de Vegte-Bolmer, M., Ngoc, N. T. H., Binh, N. T. H., Thieu, N. Q., Gitaka, J., et al. (2023). Rapid profiling of Plasmodium parasites from genome sequences to assist malaria control. Genome Med. *15*, 96.

123. Dogga, S. K., Rop, J. C., Cudini, J., Farr, E., Dara, A., Ouologuem, D., Djimdé, A. A., Talman, A. M., and Lawniczak, M. K. N. (2024). A single cell atlas of sexual development in Plasmodium falciparum. Science *384*, eadj4088.
124. Volkman, S. K., Neafsey, D. E., Schaffner, S. F., Park, D. J., and Wirth, D. F. (2012). Harnessing genomics and genome biology to understand malaria biology. Nat. Rev. Genet. *13*, 315–328.
125. Su, X., and Wootton, J. C. (2004). Genetic mapping in the human malaria parasite Plasmodium falciparum. Mol. Microbiol. *53*, 1573–1582.
126. Sidhu, A. B. S., Verdier-Pinard, D., and Fidock, D. A. (2002). Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. Science *298*, 210–213.
127. Foote, S. J., Kyle, D. E., Martin, R. K., Oduola, A. M., Forsyth, K., Kemp, D. J., and Cowman, A. F. (1990). Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in Plasmodium falciparum. Nature *345*, 255–258.
128. Wellems, T. E., Panton, L. J., Gluzman, I. Y., do Rosario, V. E., Gwadz, R. W., Walker-Jonah, A., and Krogstad, D. J. (1990). Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. Nature *345*, 253–255.
129. Peterson, D. S., Walliker, D., and Wellems, T. E. (1988). Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. Proc. Natl. Acad. Sci. USA *85*, 9114–9118.
130. Su, X., Kirkman, L. A., Fujioka, H., and Wellems, T. E. (1997). Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant P. falciparum in Southeast Asia and Africa. Cell *91*, 593–603.
131. Takala-Harrison, S., Clark, T. G., Jacob, C. G., Cummings, M. P., Miotto, O., Dondorp, A. M., Fukuda, M. M., Nosten, F., Noedl, H., Imwong, M., et al. (2013). Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin treatment in Southeast Asia. Proc. Natl. Acad. Sci. USA *110*, 240–245.
132. Miotto, O., Amambua-Ngwa, A., Amenga-Etego, L. N., Abdel Hamid, M. M., Adam, I., Aninagyei, E., Apinjoh, T., Awandare, G. A., Bejon, P., Bertin, G. I., et al. (2024). Identification of complex Plasmodium falciparum genetic backgrounds circulating in Africa: a multicountry genomic epidemiology analysis. Lancet Microbe, 100941.
133. Miotto, O., Amato, R., Ashley, E. A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S. O., Dhorda, M., et al. (2015). Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat. Genet. *47*, 226–234.

134. Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., et al. (2010). New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol. Microbiol. *76*, 12–24.
135. López-Barragán, M. J., Lemieux, J., Quiñones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., and Su, X. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. BMC Genomics *12*, 587.
136. Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., et al. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Science *301*, 1503–1508.
137. Jiang, H., Patel, J. J., Yi, M., Mu, J., Ding, J., Stephens, R., Cooper, R. A., Ferdig, M. T., and Su, X. (2008). Genome-wide compensatory changes accompany drug-selected mutations in the Plasmodium falciparum crt gene. PLoS One *3*, e2484.
138. 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*, e1001165.
139. Chen, Y., Lun, A. T. L., and Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. [version 2; peer review: 5 approved]. F1000Res. *5*, 1438.
140. Costa-Silva, J., Domingues, D., and Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. PLoS One *12*, e0190152.
141. McDermaid, A., Monier, B., Zhao, J., Liu, B., and Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: review and integration. Brief. Bioinformatics *20*, 2044–2054.
142. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. *24*, 695–713.
143. Lim, J., Park, C., Kim, M., Kim, H., Kim, J., and Lee, D.-S. (2024). Advances in single-cell omics and multiomics for high-resolution molecular profiling. Exp Mol Med *56*, 515–526.
144. Howick, V. M., Russell, A. J. C., Andrews, T., Heaton, H., Reid, A. J., Natarajan, K., Butungi, H., Metcalf, T., Verzier, L. H., Rayner, J. C., et al. (2019). The Malaria Cell

Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. Science *365*.

145. Reid, A. J., Talman, A. M., Bennett, H. M., Gomes, A. R., Sanders, M. J., Illingworth, C. J. R., Billker, O., Berriman, M., and Lawniczak, M. K. (2018). Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. Elife *7*.
146. Real, E., Howick, V. M., Dahalan, F. A., Witmer, K., Cudini, J., Andradi-Brown, C., Blight, J., Davidson, M. S., Dogga, S. K., Reid, A. J., et al. (2021). A single-cell atlas of Plasmodium falciparum transmission through the mosquito. Nat. Commun. *12*, 3196.
147. Aravind, L., Iyer, L. M., Wellems, T. E., and Miller, L. H. (2003). Plasmodium biology: genomic gleanings. Cell *115*, 771–785.
148. Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., et al. (2003). PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. Nucleic Acids Res. *31*, 212–215.
149. Alvarez-Jarreta, J., Amos, B., Aurrecoechea, C., Bah, S., Barba, M., Barreto, A., Basenko, E. Y., Belnap, R., Blevins, A., Böhme, U., et al. (2024). VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center in 2023. Nucleic Acids Res. *52*, D808–D816.
150. Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S., and McFadden, G. I. (1998). Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum. Proc. Natl. Acad. Sci. USA *95*, 12352–12357.
151. Outlaw, D. C., and Ricklefs, R. E. (2011). Rerooting the evolutionary tree of malaria parasites. Proc. Natl. Acad. Sci. USA *108*, 13183–13187.
152. Pick, C., Ebersberger, I., Spielmann, T., Bruchhaus, I., and Burmester, T. (2011). Phylogenomic analyses of malaria parasites and evolution of their exported proteins. BMC Evol. Biol. *11*, 167.
153. Waters, A. P. (2002). Orthology between the genomes of Plasmodium falciparum and rodent malaria parasites: possible practical applications. Philos. Trans. R. Soc. Lond. B, Biol. Sci. *357*, 55–63.
154. Oberstaller, J., Otto, T. D., Rayner, J. C., and Adams, J. H. (2021). Essential genes of the parasitic apicomplexa. Trends Parasitol.
155. Mok, S., Ashley, E. A., Ferreira, P. E., Zhu, L., Lin, Z., Yeo, T., Chotivanich, K., Imwong, M., Pukrittayakamee, S., Dhorda, M., et al. (2015). Drug resistance.

Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. Science *347*, 431–435.

156. Mok, S., Stokes, B. H., Gnädig, N. F., Ross, L. S., Yeo, T., Amaratunga, C., Allman, E., Solyakov, L., Bottrill, A. R., Tripathi, J., et al. (2021). Artemisinin-resistant K13 mutations rewire Plasmodium falciparum's intra-erythrocytic metabolic program to enhance survival. Nat. Commun. *12*, 530.
157. Zhang, M., Wang, C., Oberstaller, J., Thomas, P., Otto, T. D., Casandra, D., Boyapalle, S., Adapa, S. R., Xu, S., Button-Simons, K., et al. (2021). The apicoplast link to fever-survival and artemisinin-resistance in the malaria parasite. Nat. Commun. *12*, 4563.
158. Pires, C. V., Oberstaller, J., Wang, C., Casandra, D., Zhang, M., Chawla, J., Adapa, S. R., Otto, T. D., Ferdig, M. T., Rayner, J. C., et al. (2023). Chemogenomic Profiling of a Plasmodium falciparum Transposon Mutant Library Reveals Shared Effects of Dihydroartemisinin and Bortezomib on Lipid Metabolism and Exported Proteins. Microbiol. Spectr. *11*, e0501422.
159. Neal, M. L., Wei, L., Peterson, E., Arrieta-Ortiz, M. L., Danziger, S. A., Baliga, N. S., Kaushansky, A., and Aitchison, J. D. (2021). A systems-level gene regulatory network model for Plasmodium falciparum. Nucleic Acids Res.
160. Zanghi, G., Patel, H., Camargo, N., Smith, J. L., Bae, Y., Flannery, E. L., Chuenchob, V., Fishbaugher, M. E., Mikolajczak, S. A., Roobsoong, W., et al. (2023). Global gene expression of human malaria parasite liver stages throughout intrahepatocytic development. BioRxiv.
161. Bártfai, R., Hoeijmakers, W. A. M., 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*, e1001223.
162. Josling, G. A., Petter, M., Oehring, S. C., Gupta, A. P., Dietz, O., Wilson, D. W., Schubert, T., Längst, G., Gilson, P. R., Crabb, B. S., et al. (2015). A plasmodium falciparum bromodomain protein regulates invasion gene expression. Cell Host Microbe *17*, 741–751.
163. Subudhi, A. K., O'Donnell, A. J., Ramaprasad, A., Abkallo, H. M., Kaushik, A., Ansari, H. R., Abdel-Haleem, A. M., Ben Rached, F., Kaneko, O., Culleton, R., et al. (2020). Malaria parasites regulate intra-erythrocytic development duration via serpentine receptor 10 to coordinate with host rhythms. Nat. Commun. *11*, 2763.

164. Duffy, M. F., Selvarajah, S. A., Josling, G. A., and Petter, M. (2012). The role of chromatin in Plasmodium gene expression. Cell Microbiol. *14*, 819–828.
165. Otto, T. D., Böhme, U., Jackson, A. P., Hunt, M., Franke-Fayard, B., Hoeijmakers, W. A. M., Religa, A. A., Robertson, L., Sanders, M., Ogun, S. A., et al. (2014). A comprehensive evaluation of rodent malaria parasite genomes and gene expression. BMC Biol. *12*, 86.
166. Siegel, T. N., Hon, C.-C., Zhang, Q., Lopez-Rubio, J.-J., Scheidig-Benatar, C., Martins, R. M., Sismeiro, O., Coppée, J.-Y., and Scherf, A. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum. BMC Genomics *15*, 150.
167. Vignali, M., Armour, C. D., Chen, J., Morrison, R., Castle, J. C., Biery, M. C., Bouzek, H., Moon, W., Babak, T., Fried, M., et al. (2011). NSR-seq transcriptional profiling enables identification of a gene signature of Plasmodium falciparum parasites infecting children. J. Clin. Invest. *121*, 1119–1129.
168. Kucharski, M., Tripathi, J., Nayak, S., Zhu, L., Wirjanata, G., van der Pluijm, R. W., Dhorda, M., Dondorp, A., and Bozdech, Z. (2020). A comprehensive RNA handling and transcriptomics guide for high-throughput processing of Plasmodium blood-stage samples. Malar. J. *19*, 363.
169. Guillochon, E., Fraering, J., Joste, V., Kamaliddin, C., Vianou, B., Houzé, L., Baudrin, L. G., Faucher, J. F., Aubouy, A., Houzé, S., et al. (2022). Transcriptome Analysis of Plasmodium falciparum Isolates From Benin Reveals Specific Gene Expression Associated With Cerebral Malaria. J. Infect. Dis. *225*, 2187–2196.
170. Ruiz, J. L., Yerbanga, R. S., Lefèvre, T., Ouedraogo, J. B., Corces, V. G., and Gómez-Díaz, E. (2019). Chromatin changes in Anopheles gambiae induced by Plasmodium falciparum infection. Epigenetics Chromatin *12*, 5.
171. Hoo, R., Zhu, L., Amaladoss, A., Mok, S., Natalang, O., Lapp, S. A., Hu, G., Liew, K., Galinski, M. R., Bozdech, Z., et al. (2016). Integrated analysis of the Plasmodium species transcriptome. EBioMedicine *7*, 255–266.
172. 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*.
173. Eappen, A. G., Li, T., Marquette, M., Chakravarty, S., Kc, N., Zanghi, G., Hoffman, B. U., Hettiarachchi, H., Patil, A., Abebe, Y., et al. (2022). In vitro production of infectious Plasmodium falciparum sporozoites. Nature *612*, 534–539.
174. Chêne, A., Vembar, S. S., Rivière, L., Lopez-Rubio, J. J., Claes, A., Siegel, T. N., Sakamoto, H., Scheidig-Benatar, C., Hernandez-Rivas, R., and Scherf, A. (2012).

PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. Nucleic Acids Res. *40*, 3066–3077.

175. Ponts, N., Fu, L., Harris, E. Y., Zhang, J., Chung, D.-W. D., Cervantes, M. C., Prudhomme, J., Atanasova-Penichon, V., Zehraoui, E., Bunnik, E. M., et al. (2013). Genome-wide mapping of DNA methylation in the human malaria parasite Plasmodium falciparum. Cell Host Microbe *14*, 696–706.
176. Pires, C. V., Cassandra, D., Xu, S., Laleu, B., Burrows, J. N., and Adams, J. H. (2024). Oxidative stress changes the effectiveness of artemisinin in Plasmodium falciparum. MBio *15*, e0316923.
177. 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*, 395.
178. Lasonder, E., Rijpma, S. R., van Schaijk, B. C. L., Hoeijmakers, W. A. M., Kensche, P. R., Gresnigt, M. S., Italiaander, A., Vos, M. W., Woestenenk, R., Bousema, T., et al. (2016). Integrated transcriptomic and proteomic analyses of P. falciparum gametocytes: molecular insight into sex-specific processes and translational repression. Nucleic Acids Res. *44*, 6087–6101.
179. Tan, Q. W., and Mutwil, M. (2020). Malaria.tools-comparative genomic and transcriptomic database for Plasmodium species. Nucleic Acids Res. *48*, D768–D775.
180. Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., and Tyler, J. K. (2015). The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. Mol. Cell. Biol. *36*, 662–667.
181. Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., et al. (2016). A survey of best practices for RNA-seq data analysis. Genome Biol. *17*, 13.
182. Hicks, S. C., Okrah, K., Paulson, J. N., Quackenbush, J., Irizarry, R. A., and Bravo, H. C. (2018). Smooth quantile normalization. Biostatistics *19*, 185–198.
183. Zhao, Y., Wong, L., and Goh, W. W. B. (2020). How to do quantile normalization correctly for gene expression data analyses. Sci. Rep. *10*, 15534.
184. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

185. Yu, M., Kumar, T. R. S., Nkrumah, L. J., Coppi, A., Retzlaff, S., Li, C. D., Kelly, B. J., Moura, P. A., Lakshmanan, V., Freundlich, J. S., et al. (2008). The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. Cell Host Microbe *4*, 567–578.
186. Vaughan, A. M., O'Neill, M. T., Tarun, A. S., Camargo, N., Phuong, T. M., Aly, A. S. I., Cowman, A. F., and Kappe, S. H. I. (2009). Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. Cell Microbiol. *11*, 506–520.
187. Shears, M. J., Botté, C. Y., and McFadden, G. I. (2015). Fatty acid metabolism in the Plasmodium apicoplast: Drugs, doubts and knockouts. Mol. Biochem. Parasitol. *199*, 34–50.
188. Evans, A. G., and Wellems, T. E. (2002). Coevolutionary genetics of Plasmodium malaria parasites and their human hosts. Integr. Comp. Biol. *42*, 401–407.
189. Loy, D. E., Liu, W., Li, Y., Learn, G. H., Plenderleith, L. J., Sundararaman, S. A., Sharp, P. M., and Hahn, B. H. (2017). Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. Int. J. Parasitol. *47*, 87–97.
190. Carrillo-Bustamante, P., Costa, G., Lampe, L., and Levashina, E. A. (2023). Evolutionary modelling indicates that mosquito metabolism shapes the life-history strategies of Plasmodium parasites. Nat. Commun. *14*, 8139.
191. Levine, M., and Davidson, E. H. (2005). Gene regulatory networks for development. Proc. Natl. Acad. Sci. USA *102*, 4936–4942.
192. Li, E., and Davidson, E. H. (2009). Building developmental gene regulatory networks. Birth Defects Res C Embryo Today *87*, 123–130.
193. Kim, D., Tran, A., Kim, H. J., Lin, Y., Yang, J. Y. H., and Yang, P. (2023). Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. NPJ Syst. Biol. Appl. *9*, 51.
194. Yuda, M., Kaneko, I., Murata, Y., Iwanaga, S., and Nishi, T. (2021). Mechanisms of triggering malaria gametocytogenesis by AP2-G. Parasitol. Int., 102403.
195. Josling, G. A., Williamson, K. C., and Llinás, M. (2018). Regulation of sexual commitment and gametocytogenesis in malaria parasites. Annu. Rev. Microbiol. *72*, 501–519.
196. Llorà-Batlle, O., Michel-Todó, L., Witmer, K., Toda, H., Fernández-Becerra, C., Baum, J., and Cortés, A. (2020). Conditional expression of PfAP2-G for controlled massive sexual conversion in Plasmodium falciparum. Sci. Adv. *6*, eaaz5057.

197. Kafsack, B. F. C., Rovira-Graells, N., Clark, T. G., Bancells, C., Crowley, V. M., Campino, S. G., Williams, A. E., Drought, L. G., Kwiatkowski, D. P., Baker, D. A., et al. (2014). A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature *507*, 248–252.
198. Bopp, S., Pasaje, C. F. A., Summers, R. L., Magistrado-Coxen, P., Schindler, K. A., Corpas-Lopez, V., Yeo, T., Mok, S., Dey, S., Smick, S., et al. (2023). Potent acyl-CoA synthetase 10 inhibitors kill Plasmodium falciparum by disrupting triglyceride formation. Nat. Commun. *14*, 1455.
199. Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.
200. Ge, S. X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics *36*, 2628–2629.
201. Garcia-Fossa, F., Cruz, M. C., Haghighi, M., de Jesus, M. B., Singh, S., Carpenter, A. E., and Cimini, B. A. (2023). Interpreting Image-based Profiles using Similarity Clustering and Single-Cell Visualization. Curr. Protoc. *3*, e713.
202. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software. SIGKDD Explor. Newsl. *11*, 10.
203. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., et al. (2012). Scikit-learn: Machine Learning in Python. arXiv.
204. Adjalley, S., and Lee, M. C. S. (2022). CRISPR/Cas9 Editing of the Plasmodium falciparum Genome. Methods Mol. Biol. *2470*, 221–239.
205. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., and Yang, S.-H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol. Ther. Nucleic Acids *4*, e264.
206. Guo, C., Ma, X., Gao, F., and Guo, Y. (2023). Off-target effects in CRISPR/Cas9 gene editing. Front. Bioeng. Biotechnol. *11*, 1143157.
207. Allemailem, K. S., Almatroodi, S. A., Almatroudi, A., Alrumaihi, F., Al Abdulmonem, W., Al-Megrin, W. A. I., Aljamaan, A. N., Rahmani, A. H., and Khan, A. A. (2023). Recent Advances in Genome-Editing Technology with CRISPR/Cas9 Variants and Stimuli-Responsive Targeting Approaches within Tumor Cells: A Future Perspective of Cancer Management. Int. J. Mol. Sci. *24*.
208. Liao, H., Wu, J., VanDusen, N. J., Li, Y., and Zheng, Y. (2024). CRISPR/Cas9-Mediated Homology-Directed Repair for Precise Gene Editing. Molecular Therapy - Nucleic Acids, 102344.

209. Xu, Z., Kuang, Y., Ren, B., Yan, D., Yan, F., Spetz, C., Sun, W., Wang, G., Zhou, X., and Zhou, H. (2021). SpRY greatly expands the genome editing scope in rice with highly flexible PAM recognition. Genome Biol. *22*, 6.
210. Walton, R. T., Christie, K. A., Whittaker, M. N., and Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science *368*, 290–296.
211. Hughes, R. A., and Ellington, A. D. (2017). Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. Cold Spring Harb. Perspect. Biol. *9*.
212. Skinner-Adams, T. S., Lawrie, P. M., Hawthorne, P. L., Gardiner, D. L., and Trenholme, K. R. (2003). Comparison of Plasmodium falciparum transfection methods. Malar. J. *2*, 19.
213. Prommana, P., Uthaipibull, C., Wongsombat, C., Kamchonwongpaisan, S., Yuthavong, Y., Knuepfer, E., Holder, A. A., and Shaw, P. J. (2013). Inducible knockdown of Plasmodium gene expression using the glmS ribozyme. PLoS One *8*, e73783.
214. 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*, 687–701.
215. 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.
216. Douglass, A. N., Kain, H. S., Abdullahi, M., Arang, N., Austin, L. S., Mikolajczak, S. A., Billman, Z. P., Hume, J. C. C., Murphy, S. C., Kappe, S. H. I., et al. (2015). Host-based prophylaxis successfully targets liver stage malaria parasites. Mol. Ther.
217. Arang, N., Kain, H. S., Glennon, E. K., Bello, T., Dudgeon, D. R., Walter, E. N. F., Gujral, T. S., and Kaushansky, A. (2017). Identifying host regulators and inhibitors of liver stage malaria infection using kinase activity profiles. Nat. Commun. *8*, 1232.
218. Boulet, C., Siddiqui, G., Gaynor, T. L., Doerig, C., Creek, D. J., and Carvalho, T. G. (2022). Red Blood Cell BCL-xL Is Required for Plasmodium falciparum Survival: Insights into Host-Directed Malaria Therapies. Microorganisms *10*.
219. 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*, e0278221.

220. O'Donnell, R. A., Preiser, P. R., Williamson, D. H., Moore, P. W., Cowman, A. F., and Crabb, B. S. (2001). An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected Plasmodium falciparum parasites. Nucleic Acids Res. *29*, 716–724.
221. Smith, T. A., Lopez-Perez, G. S., Herneisen, A. L., Shortt, E., and Lourido, S. (2022). Screening the Toxoplasma kinome with high-throughput tagging identifies a regulator of invasion and egress. Nat. Microbiol.
222. Laughery, M. F., and Wyrick, J. J. (2019). Simple CRISPR-Cas9 Genome Editing in Saccharomyces cerevisiae. Curr. Protoc. Mol. Biol. *129*, e110.
223. Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y.-W., et al. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol. *18*, 35.
224. Jonsdottir, T. K., Paoletta, M. S., Ishizaki, T., Hernandez, S., Ivanova, M., Herrera Curbelo, A., Saiki, P. A., Selinger, M., Das, D., Henriksson, J., et al. (2025). A scalable CRISPR-Cas9 gene editing system facilitates CRISPR screens in the malaria parasite Plasmodium berghei. Nucleic Acids Res. *53*.
225. Qian, P., Wang, X., Guan, C., Fang, X., Cai, M., Zhong, C.-Q., Cui, Y., Li, Y., Yao, L., Cui, H., et al. (2022). Apical anchorage and stabilization of subpellicular microtubules by apical polar ring ensures Plasmodium ookinete infection in mosquito. Nat. Commun. *13*, 7465.
226. Hliscs, M., Millet, C., Dixon, M. W., Siden-Kiamos, I., McMillan, P., and Tilley, L. (2015). Organization and function of an actin cytoskeleton in Plasmodium falciparum gametocytes. Cell Microbiol. *17*, 207–225.
227. Ferreira, J. L., Heincke, D., Wichers, J. S., Liffner, B., Wilson, D. W., and Gilberger, T.-W. (2020). The dynamic roles of the inner membrane complex in the multiple stages of the malaria parasite. Front. Cell Infect. Microbiol. *10*, 611801.
228. Ronda, C., Maury, J., Jakočiunas, T., Jacobsen, S. A. B., Germann, S. M., Harrison, S. J., Borodina, I., Keasling, J. D., Jensen, M. K., and Nielsen, A. T. (2015). CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae. Microb. Cell Fact. *14*, 97.
229. Zhao, Y., Coelho, C., Lauer, S., Majewski, M., Laurent, J. M., Brosh, R., and Boeke, J. D. (2023). CREEPY: CRISPR-mediated editing of synthetic episomes in yeast. Nucleic Acids Res.

230. Zhao, Y., Wang, F., Wang, C., Zhang, X., Jiang, C., Ding, F., Shen, L., and Zhang, Q. (2021). Optimization of CRISPR/Cas System for Improving Genome Editing Efficiency in Plasmodium falciparum. Front. Microbiol. *11*, 625862.
231. 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*, 18501.
232. 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*, 3881.
233. Tibúrcio, M., Yang, A. S. P., Yahata, K., Suárez-Cortés, P., Belda, H., Baumgarten, S., van de Vegte-Bolmer, M., van Gemert, G.-J., van Waardenburg, Y., Levashina, E. A., et al. (2019). A Novel Tool for the Generation of Conditional Knockouts To Study Gene Function across the Plasmodium falciparum Life Cycle. MBio *10*.
234. Bansal, A., Sharma, M., and Choudhury, H. (2024). Generation of a new DiCre expressing parasite strain for functional characterization of Plasmodium falciparum genes in blood stages. Sci. Rep. *14*, 24076.
235. Nkrumah, L. J., Muhle, R. A., Moura, P. A., Ghosh, P., Hatfull, G. F., Jacobs, W. R., and Fidock, D. A. (2006). Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase. Nat. Methods *3*, 615–621.
236. Fierro, M. A., Hussain, T., Campin, L. J., and Beck, J. R. (2023). Knock-sideways by inducible ER retrieval enables a unique approach for studying Plasmodium-secreted proteins. Proc. Natl. Acad. Sci. USA *120*, e2308676120.
237. Ramaprasad, A., and Blackman, M. J. (2024). A scaleable inducible knockout system for studying essential gene function in the malaria parasite. Nucleic Acids Res.
238. Kumar, S., Abatiyow, B. A., Haile, M. T., Oualim, K. M. Z., Leeb, A. S., Vaughan, A. M., and Kappe, S. H. I. (2022). A Putative Plasmodium RNA-Binding Protein Plays a Critical Role in Female Gamete Fertility and Parasite Transmission to the Mosquito Vector. Front. Cell Dev. Biol. *10*, 825247.
239. Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., and Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.
240. Trager, W., and Jensen, J. B. (1976). Human malaria parasites in continuous culture. Science *193*, 673–675.

241. Lambros, C., and Vanderberg, J. P. (1979). Synchronization of Plasmodium falciparum erythrocytic stages in culture. J. Parasitol. *65*, 418–420.
242. Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M., et al. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. Elife *5*.
243. Joung, J., Engreitz, J. M., Konermann, S., Abudayyeh, O. O., Verdine, V. K., Aguet, F., Gootenberg, J. S., Sanjana, N. E., Wright, J. B., Fulco, C. P., et al. (2017). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature *548*, 343–346.
244. Replogle, J. M., Bonnar, J. L., Pogson, A. N., Liem, C. R., Maier, N. K., Ding, Y., Russell, B. J., Wang, X., Leng, K., Guna, A., et al. (2022). Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. Elife *11*.
245. Datlinger, P., Rendeiro, A. F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L. C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301.
246. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.
247. Wong, N., Liu, W., and Wang, X. (2015). WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome Biol. *16*, 218.
248. Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F., et al. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. Nat. Commun. *10*, 4284.
249. Jensen, K. T., Fløe, L., Petersen, T. S., Huang, J., Xu, F., Bolund, L., Luo, Y., and Lin, L. (2017). Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. FEBS Lett. *591*, 1892–1901.
250. Wang, J., Zhang, X., Cheng, L., and Luo, Y. (2020). An overview and metanalysis of machine and deep learning-based CRISPR gRNA design tools. RNA Biol. *17*, 13–22.
251. Xiang, X., Corsi, G. I., Anthon, C., Qu, K., Pan, X., Liang, X., Han, P., Dong, Z., Liu, L., Zhong, J., et al. (2021). Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning. Nat. Commun. *12*, 3238.

252. Zhang, H., Yan, J., Lu, Z., Zhou, Y., Zhang, Q., Cui, T., Li, Y., Chen, H., and Ma, L. (2023). Deep sampling of gRNA in the human genome and deep-learning-informed prediction of gRNA activities. Cell Discov. *9*, 48.
253. Labun, K., Krause, M., Torres Cleuren, Y., and Valen, E. (2021). CRISPR genome editing made easy through the CHOPCHOP website. Curr. Protoc. *1*, e46.
254. Radzisheuskaya, A., Shlyueva, D., Müller, I., and Helin, K. (2016). Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. Nucleic Acids Res. *44*, e141.
255. Sanson, K. R., Hanna, R. E., Hegde, M., Donovan, K. F., Strand, C., Sullender, M. E., Vaimberg, E. W., Goodale, A., Root, D. E., Piccioni, F., et al. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat. Commun. *9*, 5416.
256. Doench, J. G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. Nat. Rev. Genet. *19*, 67–80.
257. Fan, Q., An, L., and Cui, L. (2004). Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling. Eukaryotic Cell *3*, 264–276.
258. Zhu, A. Y., Zhou, Y., Khan, S., Deitsch, K. W., Hao, Q., and Lin, H. (2012). Plasmodium falciparum Sir2A preferentially hydrolyzes medium and long chain fatty acyl lysine. ACS Chem. Biol. *7*, 155–159.
259. Liu, Y., Yu, C., Daley, T. P., Wang, F., Cao, W. S., Bhate, S., Lin, X., Still, C., Liu, H., Zhao, D., et al. (2018). CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. Cell Stem Cell *23*, 758–771.e8.
260. Shang, X., Wang, C., Fan, Y., Guo, G., Wang, F., Zhao, Y., Sheng, F., Tang, J., He, X., Yu, X., et al. (2022). Genome-wide landscape of ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network during Plasmodium falciparum blood-stage development. Nucleic Acids Res. *50*, 3413–3431.
261. Painter, H. J., Campbell, T. L., and Llinás, M. (2011). The Apicomplexan AP2 family: integral factors regulating Plasmodium development. Mol. Biochem. Parasitol. *176*, 1–7.
262. Matthews, K. A., Senagbe, K. M., Nötzel, C., Gonzales, C. A., Tong, X., Rijo-Ferreira, F., Bhanu, N. V., Miguel-Blanco, C., Lafuente-Monasterio, M. J., Garcia, B. A., et al. (2020). Disruption of the Plasmodium falciparum Life Cycle through Transcriptional Reprogramming by Inhibitors of Jumonji Demethylases. ACS Infect. Dis. *6*, 1058–1075.

263. Veening, J.-W., Smits, W. K., and Kuipers, O. P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. Annu. Rev. Microbiol. *62*, 193–210.
264. Levy, S. F., Ziv, N., and Siegal, M. L. (2012). Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS Biol. *10*, e1001325.
265. Starrfelt, J., and Kokko, H. (2012). Bet-hedging--a triple trade-off between means, variances and correlations. Biol. Rev. Camb. Philos. Soc. *87*, 742–755.
266. Seco-Hidalgo, V., Osuna, A., and Pablos, L. M. D. (2015). To bet or not to bet: deciphering cell to cell variation in protozoan infections. Trends Parasitol. *31*, 350–356.
267. Casci, T. (2012). Adaptation: Malarial bet hedging. Nat. Rev. Genet. *13*, 298–299.
268. Rovira-Graells, N., Gupta, A. P., Planet, E., Crowley, V. M., Mok, S., Ribas de Pouplana, L., Preiser, P. R., Bozdech, Z., and Cortés, A. (2012). Transcriptional variation in the malaria parasite Plasmodium falciparum. Genome Res. *22*, 925–938.
269. Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., Schwarzer, A., Mateo, J. L., Klusmann, J.-H., and Heckl, D. (2018). Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. Nucleic Acids Res. *46*, 1375–1385.
270. 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*, 10894.
271. Iwanaga, S., Kubota, R., Nishi, T., Kamchonwongpaisan, S., Srichairatanakool, S., Shinzawa, N., Syafruddin, D., Yuda, M., and Uthaipibull, C. (2022). Genome-wide functional screening of drug-resistance genes in Plasmodium falciparum. Nat. Commun. *13*, 6163.
272. Adjalley, S. H., Lee, M. C. S., and Fidock, D. A. (2010). A method for rapid genetic integration into Plasmodium falciparum utilizing mycobacteriophage Bxb1 integrase. Methods Mol. Biol. *634*, 87–100.
273. Miao, J., Wang, C., Lucky, A. B., Liang, X., Min, H., Adapa, S. R., Jiang, R., Kim, K., and Cui, L. (2021). A unique GCN5 histone acetyltransferase complex controls erythrocyte invasion and virulence in the malaria parasite Plasmodium falciparum. PLoS Pathog. *17*, e1009351.
274. Bonnell, V. A., Zhang, Y., Brown, A. S., Horton, J., Josling, G. A., Chiu, T.-P., Rohs, R., Mahony, S., Gordân, R., and Llinás, M. (2024). DNA sequence and chromatin

differentiate sequence-specific transcription factor binding in the human malaria parasite Plasmodium falciparum. Nucleic Acids Res. *52*, 10161–10179.
275. Cui, L., Fan, Q., Cui, L., and Miao, J. (2008). Histone lysine methyltransferases and demethylases in Plasmodium falciparum. Int. J. Parasitol. *38*, 1083–1097.
276. Tintó-Font, E., Michel-Todó, L., Russell, T. J., Casas-Vila, N., Conway, D. J., Bozdech, Z., Llinás, M., and Cortés, A. (2021). A heat-shock response regulated by the PfAP2-HS transcription factor protects human malaria parasites from febrile temperatures. Nat. Microbiol. *6*, 1163–1174.
277. Rawat, M., Srivastava, A., Johri, S., Gupta, I., and Karmodiya, K. (2021). Single-Cell RNA Sequencing Reveals Cellular Heterogeneity and Stage Transition under Temperature Stress in Synchronized Plasmodium falciparum Cells. Microbiol. Spectr. *9*, e0000821.
278. Chawla, J., Goldowitz, I., Oberstaller, J., Zhang, M., Pires, C. V., Navarro, F., Sollelis, L., Wang, C. C. Q., Seyfang, A., Dvorin, J., et al. (2023). Phenotypic Screens Identify Genetic Factors Associated with Gametocyte Development in the Human Malaria Parasite Plasmodium falciparum. Microbiol. Spectr. *11*, e0416422.
279. Parapini, S., Olliaro, P., Navaratnam, V., Taramelli, D., and Basilico, N. (2015). Stability of the antimalarial drug dihydroartemisinin under physiologically relevant conditions: implications for clinical treatment and pharmacokinetic and in vitro assays. Antimicrob. Agents Chemother. *59*, 4046–4052.
280. Kishore, S. P., Stiller, J. W., and Deitsch, K. W. (2013). Horizontal gene transfer of epigenetic machinery and evolution of parasitism in the malaria parasite Plasmodium falciparum and other apicomplexans. BMC Evol. Biol. *13*, 37.
281. 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., et al. (2018). Emerging Southeast Asian PfCRT mutations confer Plasmodium falciparum resistance to the first-line antimalarial piperaquine. Nat. Commun. *9*, 3314.
282. Kavishe, R. A., Koenderink, J. B., and Alifrangis, M. (2017). Oxidative stress in malaria and artemisinin combination therapy: Pros and Cons. FEBS J. *284*, 2579–2591.
283. Cui, L., Wang, Z., Miao, J., Miao, M., Chandra, R., Jiang, H., Su, X., and Cui, L. (2012). Mechanisms of in vitro resistance to dihydroartemisinin in Plasmodium falciparum. Mol. Microbiol. *86*, 111–128.

284. Ye, R., Tian, Y., Huang, Y., Zhang, Y., Wang, J., Sun, X., Zhou, H., Zhang, D., and Pan, W. (2019). Genome-Wide Analysis of Genetic Diversity in Plasmodium falciparum Isolates From China-Myanmar Border. Front. Genet. *10*, 1065.