UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Development of a conditional gene knockout system to investigate the functional importance of regulatory genes in *Plasmodium falciparum*

by

Christea van Zyl

Submitted in fulfilment of the requirements for the degree

Magister Scientiae in Biochemistry

in the Faculty of Natural and Agricultural Sciences

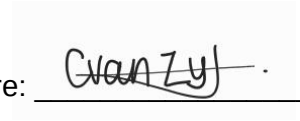
Department of Biochemistry, Genetics and Microbiology Division of Biochemistry

University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

SUBMISSION DECLARATION

I, Christea van Zyl, declare that the dissertation, which I hereby submit for the degree *Magister Scientiae* in the Department of Biochemistry, Genetics, and Microbiology at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: 

Date: **14 December 2023**

DECLARATION OF ORIGINALITY

University of Pretoria

Faculty of Natural and Agricultural Sciences

Department of Biochemistry, Genetics, and Microbiology

Full names of student: **Christea van Zyl**

Student number: **18001638**

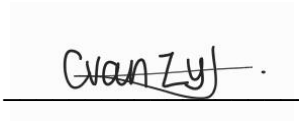
Declaration

1. I understand what plagiarism is and am aware of the University's policy in this regard.

2. I declare that this project proposal is my original work. Where other people's work has been used (either from a printed source, Internet, or any other source), this has been properly acknowledged and referenced in accordance with requirements.

3. I have not used work previously produced by another student or any other person to hand in as my own.

4. I have not allowed and will not allow anyone to copy my work with the intention of passing it off as his or her own.

Signature: 

Date: **14 December 2023**

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Acknowledgements

First, I would like to thank the most important force in my life, God, for giving me the strength I needed during the most challenging times of completing this dissertation, and for giving me the knowledge and perseverance to push through and finish this degree, which once did not seem possible to me.

Secondly, I would like to show my appreciation to my supervisor, the very knowledgeable and awe-inspiring, Prof Lyn-Marie Birkholtz, for allowing me to join this group and set forth in chasing my dreams. Without her belief in my abilities, I would never have realized how much I love science, even in the most difficult of circumstances. I also would like to thank my co-supervisor, Dr Jandeli Niemand, for her time and expertise, especially when I thought science wasn't on my side. I would also like to show my appreciation for Dr Elisha Mugo, who has since left the group, but without his invaluable insight into the field of cloning in *Plasmodium*, I would most certainly never have made a semi-success of this system.

Furthermore, I would like to thank my rock and solace, my mother (Christelle van Zyl) for her unwavering support, inspiring me every day to try my best and not ever back out. And to my dad (Ockert van Zyl), thank you for keeping me going with the endless coffees and egg on toast that got me through so many days. Also, to my brother (Lourens van Zyl), thank you for providing a comforting hug when it was oh-so-needed. Lastly, to my closest confidant and life partner, Ruan de Villiers, for you I will always be grateful as you never ceased to provide me with just the right words of comfort and always assured me that I can do anything, even though we both know it is not necessarily the case. Without all your love and support, I would never be where I am today.

Last, but certainly not least, I would like to thank all the members of the M_2PL group who helped me in every aspect of science, be it working with new equipment or simply relaxing lunches which turned out to be all I needed to tackle the rest of the day. Even more so, without the support and all of the gelado-runs with my dear friend, Dana Schultz, during the last and the roughest months of this degree, I would not have coped as well as I did. I would like to thank the PhDs (Daniel Opperman, Nicola Grundling, Savannah Moore, and Mariska Naude Hatting) for their reassuring words and assistance with even the smallest of things. I am also grateful for "The Triplets" who made the office environment just a little lighter and nicer to be in.

I want to acknowledge and give great thanks for my financial assistance from the National Research Foundation as an NRF-Grantholder linked recipient because, without this bursary, nothing laid out in this paper would have been possible.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABS	Asexual blood stage
ACT	Artemisinin-based combination therapy
Amp	Ampillicin
ApiAP2	*Apicomplexan* APETALA2 domain protein
ATc	Anhydrotetracycline
BSL2	Biosafety level 2
CRISPR-Cas9	Clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9s
coCas9	*P. falciparum* codon optimized Cas9
COVID-19	Coronavirus disease 2019
DD	Destabilization domain
DiCre	Dimerisable Cre recombinase
FCU	Cytosine deaminase/uridyl phosphoribosyl transferase
FIKK	Phenylalanine-Isoleucine-Lysine-Lysine motif
FKPB	FK506-binding protein
FRB	FKBP12-rapamycin-associated protein
gDNA	Genomic DNA
GFP	Green fluorescent protein
GlcN	Glucosamine
glmS	Glucosamine-6-phosphate riboswitch ribozyme
GOI	Gene of interest
HA	Haemagglutinin
hDHFR	Human dihydrofolate reductase
HEPEs	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hpi	Hours post-invasion
HR	Homology region
IDC	Intraerythrocytic developmental cycle
Kan	Kanamycin
KD	Knockdown
KO	Knockout
KS	Knock-sideways
LB	Luria Bertani broth
LB-Ampicillin	LB supplemented with 50 mg/mL ampicillin
Neo-R	Neomycin resistance marker
OD_{600}	Optical density at 600 nm
PAM	Protospacer-adjacent motif
PCR	Polymerase chain reaction
PlasmoDB	*Plasmodium* database
PTMs	Post-translational modifications
SARS-Cov-2	Severe acute respiratory syndrome coronavirus 2
SLI	Selection-linked integration
T2A	Thosea asigna virus 2A slip peptide
TAE	Tris-acetate-EDTA buffer
TetR-DOZI	Tet-repressor protein - Development of zygote inhibited
TFs	Transcription factors
UTR	Untranslated region
WHO	World Health Organization

Summary

Malaria cases have surged in recent years, with the World Health Organization reporting 619 000 deaths in 2021. A recently approved malaria vaccine has been recommended for children younger than 5 years; this vaccine, however, does not prevent cases amongst older children and adults. This stresses the necessity to allocate resources wisely for sustainable malaria control and treatment, emphasized by the ongoing threat of antimalarial resistance. Advancements in knowledge of the most lethal malaria-causing species, *Plasmodium falciparum*, offer hope for malaria elimination. *P. falciparum* genetic modification strategies have been a growing field, allowing for the targeting of specific components of this deadly parasite's genome. Controlled genetic interrogation at specific time-points within the highly dynamic and complex life cycle of the parasite has become an increasingly popular method for establishing the essentiality of genes in this parasite. However, non-inducible genetic knockout (KO) interrogation strategies do not facilitate the genetic probing of regulatory genes, due to the immediate death phenotype observed. The Dimerisable Cre recombinase (DiCre) system is a KO genetic interrogation tool used to flox any nucleotide sequence flanked with the cre-specific *loxP* sites, but only when induced with rapamycin. A drawback of this system is the current cloning approach employed for this system, which relies on continuous cloning of selection markers used for screening, and these critical *loxP* sites. This highlights the demand to bridge the gap and establish a more regulated cloning approach to generate transgenic parasite lines.

In this study, we developed a new DiCre cloning strategy. Specifically, we created a universal DiCre repair plasmid containing these *loxP* sites, together with a selection marker cassette and multiple cloning sites for homology regions and a recodonised gene insert. This will facilitate genomic integration through either the CRISPR-Cas9 or SLI genetic modification systems. The application of this technology was explored using two proxy regulatory genes, *gcn5* and *set7*, which play a role in histone post-translational modifications across the parasite's epigenome. This study presents a unique cloning approach for conditional gene knockout and supports future research to help expand our knowledge of gene essentiality in the *P. falciparum* parasite.

1. Literature review

Thousands of deadly microorganisms inhabit the modern world, adeptly infecting their host, evading its immune system, and then transmitting to the next host, perpetuating an unrelenting cycle. In livestock and humans, persistent and highly drug-resistant eukaryotic parasite infections have been documented for centuries, with some being among the most virulent and pathogenic parasitic species. The eradication of these parasites has been extremely difficult in recent years. A notable example is the pathogenic protozoan responsible for causing malaria.

1.1. Malaria: Understanding the global burden and pursuing solutions

Malaria is a global burden on public health and well-being. This intracellular protozoan has afflicted several countries, extending its impact beyond immediate health implications to hamper economic productivity and exacerbate social disparities. In 2021, there were over 619,000 reported deaths from malaria, with the majority of cases caused by two species of the malaria parasite, including *Plasmodium vivax* and *Plasmodium falciparum* [1]. The World Health Organization (WHO) reports the impact of the COVID-19 pandemic, particularly among low-income and developing countries such as those in Africa, where malaria control and treatment-related services were disrupted [1], exacerbating the socioeconomic stagnation in these malaria-endemic countries and setting back efforts to eradicate malaria over the last few years.

1.1.1. Control and preventative strategies

Effective strategies to combat this disease require a multifaceted approach that target the parasite during all stages of its life cycle. The intricate association of the human host and mosquito vector complicates control and preventative measures. Combined approaches such as effective antimalarial drugs, vector control strategies and the application of emergent vaccine technologies have proven to be most effective in alleviating malaria-related health burdens in affected countries. One such method is using mosquito nets treated with insecticides such as pyrethroid–piperonyl butoxide and indoor residual spraying. However, these vector control strategies have resulted in resistance amongst *Anopheles* mosquitos, posing a global threat to malaria control and eradication efforts [1].

Effective treatment and prevention of malaria in humans present significant challenges due to the complex involvement of epidemiological factors and the parasite's biology, coupled with logistical difficulties. This renders treatment of the disease particularly difficult due to the need to manage several factors, including treatment of symptoms, prevention of transmission, and mitigating the emergence of antimalarial drug resistance.

The WHO recently proposed using a new vaccine called R21/Matrix-M in October 2023 [2]. The vaccine is particularly recommended for children under five years in countries where malaria is prevalent. This vaccine is reported to aid in preventing malaria infections in children with an improved efficacy of 66 %

compared to the 29 % of the previous vaccine (RTS,S) [2-4]. This vaccine offers hope for improved malaria control but does not address the need for effective antimalarials that can be administered to infected children and adults.

Artemisinin and artemisinin-based combination therapies (ACTs) have become the gold-standard malaria drug treatment in many malaria-endemic countries. However, on a global scale, parasite populations are increasingly displaying both partial and full resistance to these ACTs, which renders the drugs ineffective. This stems from the parasite's intricate pathogenicity, enhancing its adeptness at evading the human immune system with great efficacy [1, 5]. Ultimately, discovering a novel biological mechanism can reveal the next target for therapeutic intervention in preventing and eradicating malaria.

1.1.2. The pathogenicity of the malaria-causing parasite

Human malaria can be caused by five different *Plasmodium* species. *Plasmodium malariae* and *Plasmodium ovale* present the mildest symptoms, while *P. vivax* presents mild to severe symptoms [1, 6]. Interestingly, *Plasmodium knowlesi*, another malaria-causing parasite with very mild symptoms, primarily infects macaque monkeys and contributes to zoonotic malaria in Southeast Asia [6-8]. *P. falciparum* is widely regarded as the most lethal malaria-causing parasite species, responsible for acute disease and most recorded malaria-related deaths. The difficulty in treating malaria is exacerbated by the non-specificity of symptoms reported by infected individuals, leading to misdiagnosis. These include arthralgia, fatigue, headaches, myalgia, or chest and stomach discomfort that could resemble a viral infection. Symptoms of cyclical fever result from repeated schizont eruptions and the perpetuation of the cycle of infection [9-11]. Malaria infections in pregnant women are particularly dangerous for both mother and child [12, 13], and therefore require safe and effective antimalarials.

The WHO further stresses malaria's impact on health and socioeconomic aspects that could hinder efforts towards elimination and eradication. The success of malaria control in countries from the Middle East is commendable, with some of these countries being certified malaria-free or with no reported cases in the past three years [1] (Figure 1.1). To make significant progress towards malaria elimination, it is essential to redirect focus from conventional control and treatment strategies towards investigating the parasite's biological processes, which can reveal novel critical targets.

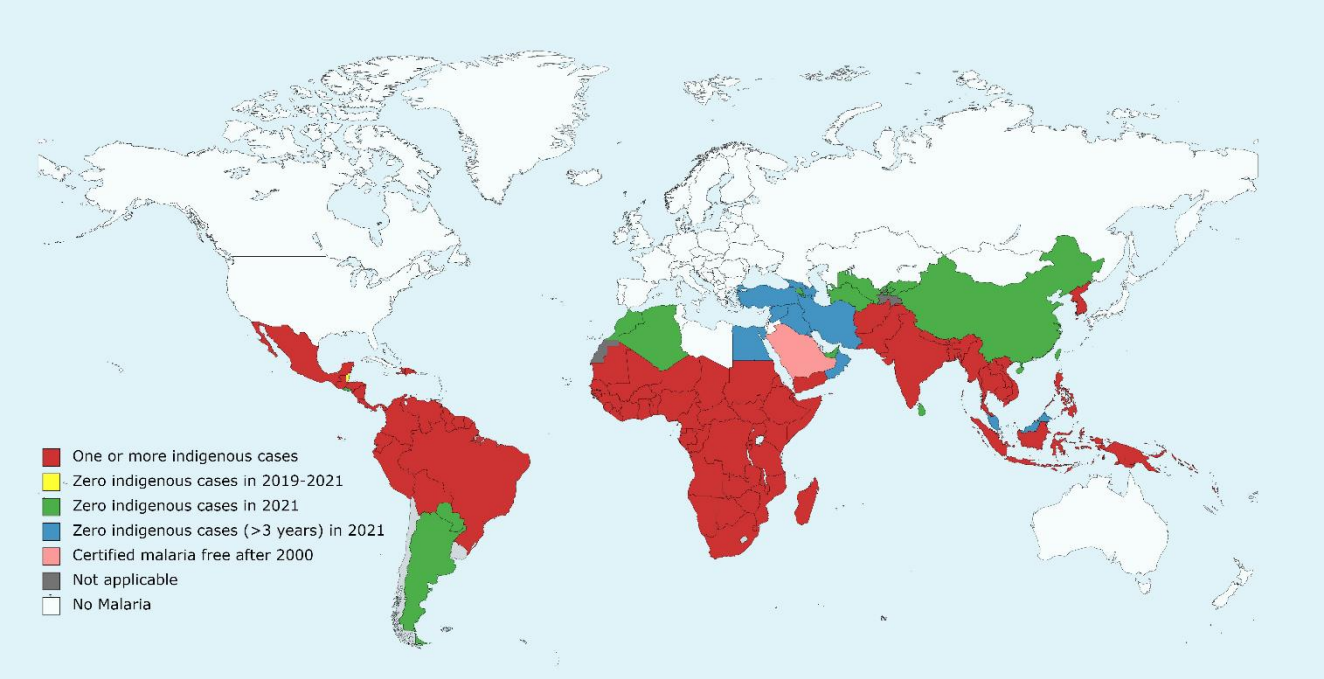


Figure 1.1: A representation of the burden of indigenous cases of malaria according to the World Malaria report 2022 [1]. Countries with at least one case of malaria in recent years are indicated in red and are malaria-endemic countries. Countries in yellow indicate those without any malaria cases between 2019 and 2021, while countries in green indicate a more recent period of zero malaria cases (as of 2021). Countries indicated in blue represent those that have been certified without cases for a period of three years as of 2021 with those in pink indicating certified malaria-free countries since 2000. (Created with mapchart.net)

1.2. *P. falciparum*'s lethal impact

P. falciparum is widely studied due to its global impact and fatality rate. The study of *P. falciparum*'s biological processes and transmission mechanisms is essential because it illustrates the intricate survival strategies employed by this malaria-causing parasite, providing critical insights. These insights can be applied to develop targeted intervention strategies, advance global health initiatives, and ultimately contribute to the elimination of this disease [6]. Characteristics of *P. falciparum*, such as its genetic diversity and resistance, facilitate evasion of complex immune systems and induce lasting effects in its host. This significantly threatens public health in countries where *P. falciparum* infections persist. The rise of infections by this species can be attributed to the growing cases of resistance, for example the upregulation of the genes associated with chloroquine and artemisinin resistance (multidrug resistance protein 1, MDR1) in parasite populations, noted since the early 1990s [14, 15]. A notable trait is the inherent complexity of the parasite's life cycle [11]. Understanding the nature of the parasite's life cycle allows the identification of mechanisms that can enhance susceptibility to therapeutic interventions.

1.2.1. *P. falciparum* parasite's life cycle

The *Plasmodium* parasite has one of the most complex life cycles among protists [6]. This is due to the association of two biological entities, where sexual proliferation occurs in the midgut of a female mosquito vector, followed by the infection of a human host in which asexual proliferation occurs [6, 16, 17]. The parasite's life cycle starts when it is transmitted in the form of sporozoites into the bloodstream of the

human through the blood feed of an infected female *Anopheles* mosquito [16, 18, 19], after which it invades hepatocytes (Figure 1.2A). Once within the hepatocyte, the sporozoites go through asexual proliferation to become hepatic schizonts (exo-erythrocytic schizogony). After 10 to 14 days, the hepatocytes lyse, releasing thousands of merozoites that enter the bloodstream. These merozoites infect erythrocytes and further develop into three distinct stages, each associated with a unique set of metabolic processes [16]. These stages encompass the ring stage, the metabolically active trophozoite stage, and the ultimate multi-DNA copy schizont stage. The schizont can release between 16 to 32 daughter cells (merozoites) upon erythrocyte rupture, perpetuating the cycle by infecting numerous cells [16, 20]. This is known as the asexual blood stages (ABS) of the parasite, where they undergo an intraerythrocytic developmental cycle (IDC) (Figure 1.2B).

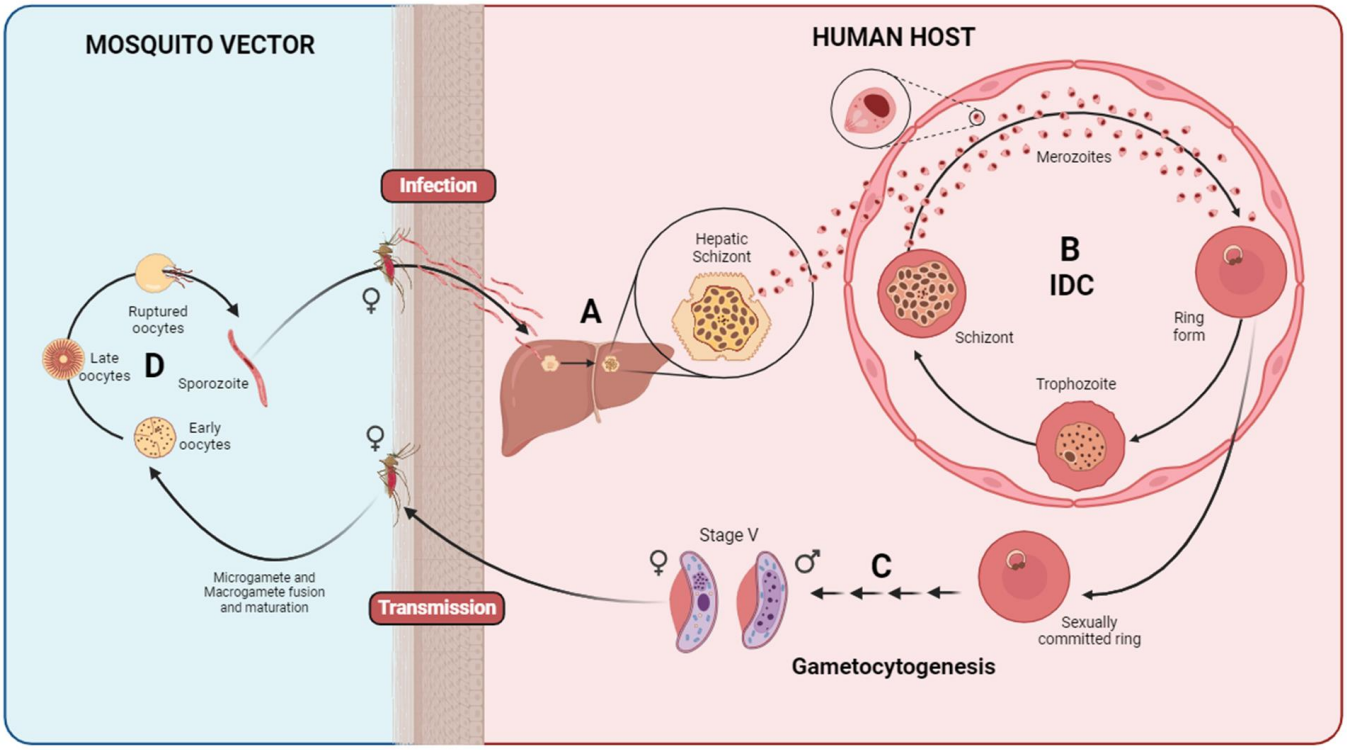


Figure 1.2: Life cycle of *P. falciparum* parasites. Vector-to-vertebrate transmission (mosquito-to-human), hepatocytes are infected with sporozoites (**A**: liver stage) which mature into hepatic schizonts that rupture to release merozoites. This is known as the exo-erythrocytic cycle (**B**: asexual proliferation or IDC, intraerythrocytic developmental cycle) which may continue indefinitely. Some merozoites infect erythrocytes and become part of the erythrocytic cycle which consists of three morphologically distinct stages, including the ring, trophozoite and schizont stage. During the latter stage, erythrocytes rupture to release dozens more merozoites. Some merozoites commit in a prior cycle to form a sexually committed ring which then later differentiates into (**C**) gametocytes over five stages with the last stage (V) being transmitted back to *Anopheles* mosquitos through a blood feed (transmission from human-to-mosquito). In the mosquito midgut, the male microgametes fuse with the female macrogamete to form a zygote. Zygotes enter the sporogonic cycle (**D**), which is also known as the sexual proliferation cycle. Zygotes become motile and elongate to become ookinetes which are then able to penetrate the wall of the mosquito's midgut where they further develop to become oocytes. The oocytes can rupture and release sporozoites which then deposit within the mosquito's salivary glands and may be transmitted back to the host and re-enter the human host cycle. Adapted from [19, 21]. Figure created with Biorender.com under the basic licence for educational purposes.

The IDC lasts about 48 hours and is associated with symptoms of malaria [16, 19, 22]. The erythrocyte is an optimal site for ABS development, given the abundance of essential nutrients and the absence of a nucleus. As a result, erythrocytes cannot express markers to initiate apoptosis through targeted cell lysis

by host immune system components. This allows the parasites to go undetected for the duration of their invasion of the cell [19, 23].

Some rings commit in a prior cycle to differentiate sexually and may enter the sexual erythrocytic stages within the spleen or bone marrow to become gametocytes [16] (Figure 1.2C). In *P. falciparum*, this process occurs over five distinct stages (I-V), each characterised by unique morphological features [11]. Stage V male and female gametocytes mature, are released into the blood circulation, and can thus be transmitted back to the mosquito during the next blood feed [19, 24]. The sexually dimorphic gametocytes generate a heterogeneous population, allowing the parasites to sexually reproduce within the vector and permitting the parasite to enter yet another life cycle [19, 24]. Only a small number of ABS parasites (<10 %) enter the sexual commitment [19, 22]. Nevertheless, only a small ratio of male and female stage V gametocytes (approximately 3 to 5 females:1 male) are required at any given time to sustain sexual proliferation within the vector [19, 24] (Figure 1.2D).

Importantly, the *Plasmodium* life cycle encompasses a multitude of developmental stages and cycles, highlighting the complexity associated with the various biological phenomena of each stage, which is also visible in the occurrence of precise transcriptional control of stage-specific gene sets [25, 26]. Given the parasite's complicated life cycle, multiple publications have revealed the targetability of parasite gene regulatory components [20, 25, 26].

1.2.2. Mechanisms of gene control

Tightly regulated biological processes and genes characterise the parasite’s control over its life cycle stages, including asexual proliferation and sexual differentiation [20, 25, 26]. Gene regulation is a crucial mechanism employed by all eukaryotic organisms, including *P. falciparum*. This mechanism distinctly controls the expression of genes in the parasite, enabling them to optimise their gene expression patterns [20]. By doing so, parasites conserve energy and resources by only expressing the genes necessary for each life cycle stage. This is characterised by the precise and timely activation and suppression of genes, essential for processes such as host cell invasion, immune evasion and adaptation to different host and vector environments [20, 25]. In a study focussed on the transcriptome of ABS parasites, authors revealed that during these stages, the parasite uses what is referred to as “just-in-time” gene expression [25, 27].

The regulatory processes include transcription factors (TFs), post-transcriptional modifiers, and epigenetic writers and erasers [26, 27]. In *Plasmodium* parasites, the sequence-specific *Apicomplexan* APETALA2 domain protein (apiAP2) DNA binding family of genes are the most critical transcriptional regulators influencing stage-specific transitions [28]. In a knockout (KO) study of the AP2 genes, the authors revealed the involvement of these various TFs on a large network of other gene expression regulatory genes [28]. Furthermore, control over the unique gene expression patterns displayed by the parasite is additionally modified with unique epigenetic control mechanisms [20, 26, 29].

1.3. Epigenetic gene regulation in *P. falciparum* parasites

Epigenetic gene regulation is an essential biochemical process in eukaryotic organisms, including humans, where gene expression is activated or silenced through a cascade of reversible genome modifications prompted by post-transcriptional modifications [30]. This allows the activation and suppression of a gene's transcription patterns without altering the organism's underlying DNA sequences [20]. In some cases, this may initiate commitment to a new stage of development and is also one of the main determinants of cellular phenotype, specifically in *P. falciparum* [26, 30, 31]. *P. falciparum* incorporates various epigenetic regulatory components, such as histone post-translational modifications (PTM), DNA methylation and the overall chromatin structure [20]. The parasite further employs epigenetic writers and erasers that act on histones to cause conformational changes in the overall DNA structure and subsequently facilitate or restrict accessibility to certain genes by TFs.

Histone PTMs are one of the most prominent epigenetic regulation approaches used by many organisms. Humans possess over 40 histone PTMs to facilitate a plethora of cellular processes [32], while in *P. falciparum* there are at least 50 with much higher levels of complexity and gene regulatory observations [33]. This is presented with a unique interconnected network of histone PTMs, affecting the deposition of subsequent PTMs, which further complicates studies into the biology of the parasite and the role of this vast histone PTM network in gene regulatory systems [26, 33]. Acetylation, methylation, phosphorylation, ubiquitylation, and SUMOylation of all conventional and variant histone proteins are identified as N-terminal tail PTMs in *P. falciparum* [34]. Of the several PTMs identified within the *Plasmodium* species, histone methylation and acetylation correlate highly with transcriptional reprogramming of the parasite [26, 35]. Histone modifications in the parasite’s epigenome can promote the extensively condensed heterochromatin or the euchromatin state. The latter facilitates gene activation, enabling stage-specific gene transcription [26, 31]. The genes encoding proteins implicated in these epigenetic modifications, including transcriptional writers such as histone methyltransferases (HMTs), must be investigated.

1.3.1. Histone methylation

Histone methylation patterns are characterised by depositing a methyl group on both lysine and arginine residues of histone N-terminals. Lysine methylation depends on the site, pattern and degree of methylation, translating into a unique role in gene expression [36]. This involves histone lysine residues' mono-, di- or tri-methylation. Histone methylation is associated with “transcriptional reprogramming” and subsequent gene silencing [32, 35, 37]. SET7 (PF3D7_1115200) is involved in sexual stage commitment and suppressing multi-copy virulence gene families [38]. SET7 has been described as binding and transferring the relevant groups to the chromatin and recruiting other interacting partners [33]. Furthermore, SET7 is an ideal candidate for genetic interrogation through a KO system due to its predicted involvement and significance in gametocytogenesis. This will determine the essentiality of SET7 and its role in recruiting all chromatin-interacting proteins.

1.3.2. Histone acetylation

Histone acetylation involves depositing an acetyl group on the N-terminal tails of various histone proteins [33]. Histone acetylation PTMs are linked to an euchromatic DNA state [39]. This strongly correlates with gene activation by enabling transcription factors to interact with the parasite's DNA [32]. Many genes within *P. falciparum* participate in this PTM, including GCN5 (PF3D7_0823300), which specifically acetylates residues H3K9 and H3K14 [33]. Although there is some reporting of the importance of GCN5 in *P. falciparum* under stress conditions [40], the importance of the GNAT domain or how the two domains interact to serve their function is unknown. The deposition of acetyl groups on histones is facilitated by GCN5's interacting partners, which form an intricate epigenetic network [33].

1.4. Gene manipulation and genetic modification approaches in *P. falciparum*

Genetically modified malaria parasites offer an effective way to investigate *Plasmodium* species' biology and reveal unique and essential pathways [41]. Discoveries made through this approach offer insight to identify potential targets for antimalarial interventions and serve as a platform for therapeutic development [42]. However, genetic manipulation in *Plasmodium* is complex and challenging [41], and stems from *P. falciparum*'s AT-rich genome, further exacerbated by limitations such as poor transfection efficiency and integration challenges for larger genes, accompanied by demanding KO procedures [43]. Due to these challenges, various genetic modification and gene manipulation techniques have been developed for use in *P. falciparum*.

1.4.1. Gene manipulation techniques

Genetic manipulation techniques involve three main methods: KO, knockdown (KD), and knock-sideways (KS). The KO method involves removing a gene entirely from the genome, while the KD method removes the gene product, i.e. either mRNA or protein. In contrast, the KS method involves removing the protein from its native cellular location and relocating it to another, non-functional location [41]. Lastly, mutations can be induced on the genome to modify a protein into a non-functional form [41]. In *P. falciparum*, these manipulations techniques have been used to influence a protein's native function and offer insight into the gene's biological significance. Furthermore, these systems can be conditional which defines a conditional means for the modification to occur, under certain conditions, such as the addition of an exogenous signal.

Researchers often require precise control over a target gene's expression, which makes these conditional systems ideal. This control can be mediated by these exogenous signals, such as for example rapamycin-mediated excision by the DiCre system. Most of the KO, KD and KS systems can be conditionally regulated. Secondly, in addition to being conditional some systems can be reversible, such as the KD systems where the exogenous signal can be added and removed as desired. However, this may require time to completely restore protein expression and for the parasite's biological processes to revert to

normal, if possible. In *P. falciparum*, examples of conditional, reversible systems are the *glmS* ribozyme, TetR-DOZI, and destabilising domain [44] (Table 1).

Table 1: Comparison of different systems that has been used to genetically interrogate *P. falciparum* genes.

Functional effect	System/s	Level	Advantage/s	Disadvantage/s	Efficiency	Reference/s
Knockout (KO)	Dimerisable Cre recombinase	DNA	▪ Conditional and rapid ▪ Complete/ incomplete gene disruption ▪ Single- or dual plasmid transfection	▪ Irreversible ▪ Requires DiCre-expressing parental line ▪ Inducer (rapamycin/ rapalog) may be toxic ▪ Size limitation (>5 kb)	~80-99 %	de Koning-Ward *et al.*, 2015 [41] Tibúrcio *et al.*, 2019 [45]
	CRISPR-Cas9 (traditional non-conditional)	DNA	▪ Complete gene disruption, mutation, or replacement ▪ Rapid	▪ Irreversible ▪ Unconditional ▪ Immediate, making timing of knock critical	~90-100 %	de Koning-Ward *et al.*, 2015 [41] Zhang *et al.*, 2014 [46]
Knock-down (KD)	*glmS* ribozyme	mRNA	▪ Conditional ▪ Reversible	▪ Inducer (glucosamine) may be cytotoxic	~50-90 (max) %	Prommana *et al.*, 2013 [47]
	TetR-DOZI aptamers	mRNA	▪ Conditional and rapid ▪ Reversible	▪ Not applicable to membrane proteins	~50-80 (max) %	Ganesan *et al.*, 2016 [48] Rajaram *et al.*, 2020 [49]
	FKBP12 Destabilising Domain (DD)	Protein	▪ Conditional and rapid ▪ Reversible	▪ Intolerance towards tag ▪ Inducer (Shield1) may be toxic ▪ Not applicable to secreted proteins	~60-80 (max) %	Banaszynski *et al.*, 2006 [50] Armstrong and Goldberg, 2007 [51]
Knock-sideways (KS)	Dimerisable Cre recombinase	Protein	▪ Conditional ▪ Variety of retrieval sequences, i.e. ER (KDEL) or nuclear (NLS)	▪ Irreversible ▪ Not applicable to membrane proteins ▪ Requires DiCre-expressing parental line	~60-90 %	Birnbaum *et al.*, 2017 [52] Fierro *et al.*, 2023 [53]

Table 1 summarises the genetic manipulation techniques applied in *P. falciparum*, with the comparable advantages and disadvantages indicated here. Some systems prove to be at a disadvantage given the requirement for a parent line contributing to the application of these systems, a notable example is that of the DiCre system. In contrast to this, the DiCre and CRISPR-Cas9 systems are more efficient than KD (*glmS* ribozyme, TetR-DOZI aptamers, and destabilising domain) and KS systems [41], which is characterised by an incomplete knock by these systems [47-53]. Although KD and KS systems are preferred for regulatory gene interrogations over the irreversible KO systems, their notably lower efficiency may affect conclusions made if only a partial knock is achieved [41].

1.4.1.1. Gene knockout systems

KO approaches rely on altering or eliminating the entire gene sequence to the extent that the expression is prohibited or an altered gene product is produced, resulting in a complete loss of its native function. Until

recently, robust tools to genetically interrogate regulatory genes were not readily available for *P. falciparum*. Given the haploid nature of the parasite during the asexual stages, which is also when genetic alterations are commonly performed [41], KO of an essential gene at any of these stages will either cause major parasite growth defects or death. This highlights the necessity for gene manipulation systems that can be applied on a conditional level, by inducing the knock only when required.

The DiCre and CRISPR-Cas9 systems have accelerated gene deletion and disruption studies in *P. falciparum*. Even though the CRISPR-Cas9 demonstrates the highest efficiency in *P. falciparum* parasites (Table 1), this system does not facilitate a more controlled gene KO [54]. The conditional Dimerisable Cre recombinase (DiCre) system, which is induced by an effector like rapamycin and permits genome-level gene alteration and deletion only upon induction, offers a more nuanced understanding of the gene's cellular impact [41, 45, 55].

In *P. falciparum*, the conditional Cre recombinase system has led to rapid advancements and great success in evaluating the function of target genes [45, 52, 56, 57]. In a previous study applied with this system where 3' UTR excision-based deletion of some genes was performed, there was no notable decline in protein expression, therefore, excision of the full gene should yield a greater phenotypic effect for some genes and is especially useful in studies involving the essentiality of a target gene [41, 58, 59]. This system relies on the activity of functional Cre recombinase fragments to allow conditional control of Cre recombinase activity, as the dimerisation of this protein can be manipulated. Cre monomers are fused to proteins immunophilin FK506-binding protein (FKPB) and FKBP12-rapamycin-associated protein (FRB), respectively, by a linker and contain the Cre N- and C-termini (containing the active domain) [55]. Rapamycin, a small ligand inducer, promotes the non-covalent binding of FKPB and FRB, facilitating the heterodimerisation and activation of Cre-fragments in closer proximity (Figure 1.3) [55]. The active Cre enzyme can selectively bind to and cleave the target sequence of interest, known as *lox*. This sequence is artificially introduced to flank any gene in the parasite's genome, allowing for highly specific enzymatic gene alterations. Different variants of the lox sequence (*loxN*, *lox2272* and *loxP*) can be applied simultaneously to yield multiple genetic alterations on the same gene [60].

This system has been adapted and successfully applied in studies on *P. falciparum* genes, as previously described [45] (Figure 1.3). The two Cre moieties are constitutively expressed in a parasite line NF54::DiCre created by Tibúrcio *et al.*, 2021, specific to two *loxP* sequences, which is included to flank any sequence for the target gene [45, 55]. These are the sites that the active DiCre enzyme will recognise and excise from the DNA, including the region between these two sites, which should contain the gene of interest (GOI). Critically the directionality of these *loxP* sequences relative to each other may also result in different genetic alterations, including excision (sequences in the same orientation), inversion of the flanked sequence (sequences in opposite orientations) or even transposition (sequences in the same orientation on different locations of the genome) [60].

Given the conditional nature of DiCre gene manipulation, this system holds promise to explore genes and their functions, including the effect of eliminating specific domains or sequences with unknown functions.

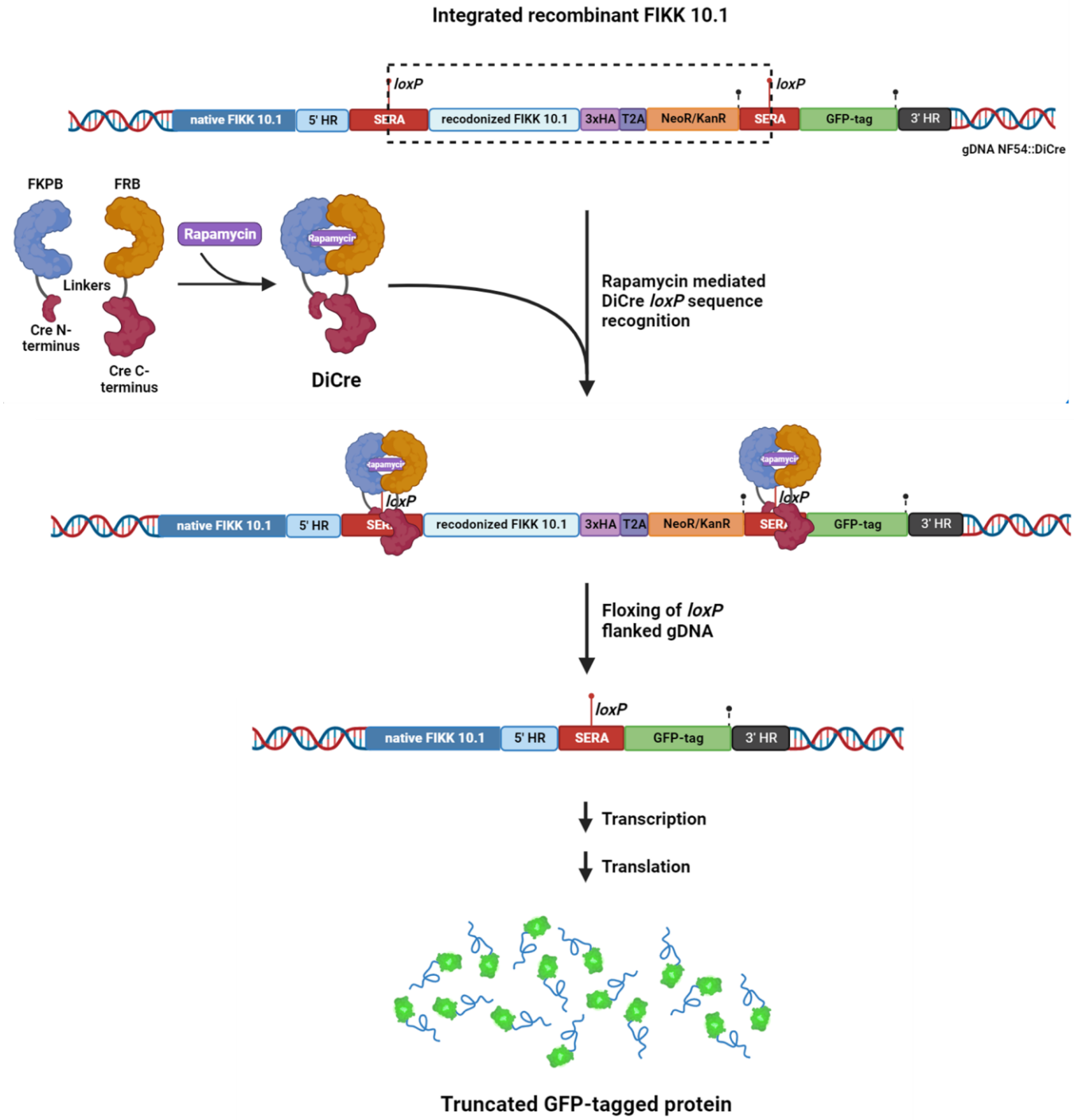


Figure 1.3: DiCre-mediated excision of *loxP* intervening sequence of FIKK 10.1. Repair section for FIKK 10.1 at the native locus of *P. falciparum* NF54 cre expressing parasites, with heterodimerization of the cre-dimers induced by rapamycin, followed by site-specific excision of the *loxP* intervening sequence. The blue (FKPB) and orange (FRB) protein are linked to the two enzymatically inactive cre-dimers through linkers. Enzymatic function induced with rapamycin, causing the dimers to come into closer proximity to serve their function. The result is a truncated protein product tagged with GFP given the out-of-frame nature of the cloning of the recodonised region and the GFP-tag downstream. Schematic is not to scale. Adapted from [45, 52, 55, 56]. Figure created with Biorender.com under the basic licence for educational purposes.

1.4.1.2. Gene knockdown systems

The KD approaches provide versatility given that a knock can be accomplished on either an RNA or protein level, depending on the GOI [41]. In some instances, KD of RNA is preferred when the protein of interest is observed to be promiscuous or involved in several pathways within the parasite's biological processes. Therefore, a more precise control will encompass RNA-level modulation.

A conditional system that relies on the regulation of the gene product on a transcription level is that of the *glmS* ribozyme system (Figure 1.5A). This system is based on tagging the target mRNA of the gene on the 3' terminal of the gene with an inactive *glmS* ribozyme [47]. Once transcribed, the gene's mRNA can be conditionally targeted for degradation by activating the ribozyme autocatalytic function with the inducer glucosamine. This renders the mRNA molecule without the required stabilisation to exit the nucleus (polyA-tail). The TetR (Tet-repressor protein) DOZI (development of zygote inhibited) aptamer system is another RNA KD approach where gene translation is suppressed under normal conditions by the binding of the TetR-DOZI complex to the aptamer present on the mRNA sequence (Figure 1.5B) [49, 61]. Translation can easily be activated when induced with anhydrotetracycline (aTc), which selectively binds to the TetR component and renders it unable to bind to the aptamer sequence [49, 61].

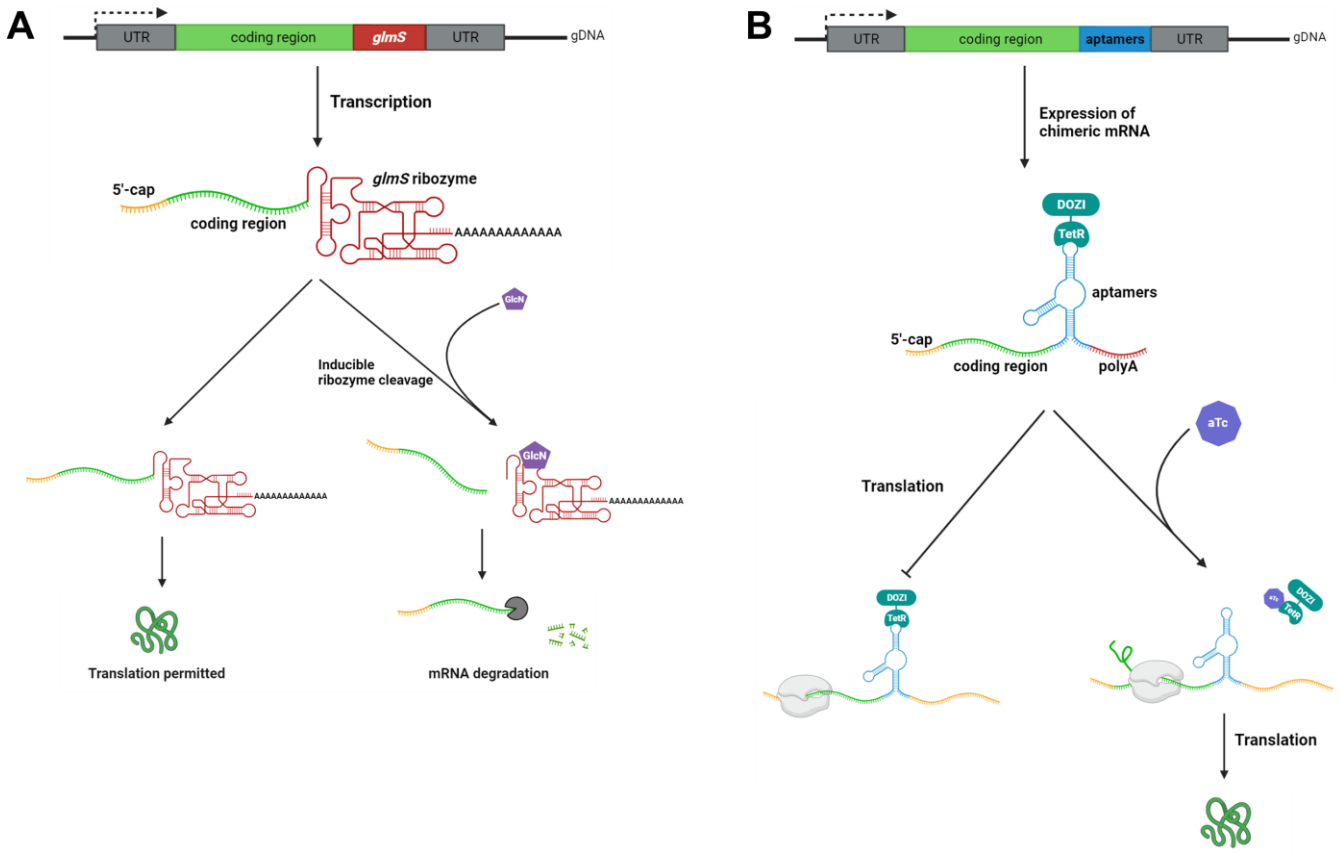


Figure 1.4: Principle of the knockdown systems (A) *glmS* ribozyme and (B) TetR-DOZI aptamers. The *glmS* ribozyme system relies on mRNA degradation through induced cleavage by inducer glucosamine (GlcN). The TetR-DOZI system relies on suppression of protein expression and is mediated by the inducer anhydrotetracycline (aTc). The destabilising domain system relies on stabilising the protein of interest by Shield1 (Shld1). Adapted from [47-51]. Figure created with Biorender.com under the basic licence for educational purposes.

Some other systems rely on regulating a gene product on a translational level, achieved by a destabilising domain (DD) on either the N- or C-terminus of the protein [50, 51]. DDs cause protein instability and recruit degradation machinery, however, some proteins are intolerant to DD tagging, which is therefore unsuitable for some genetic interrogation approaches [41].

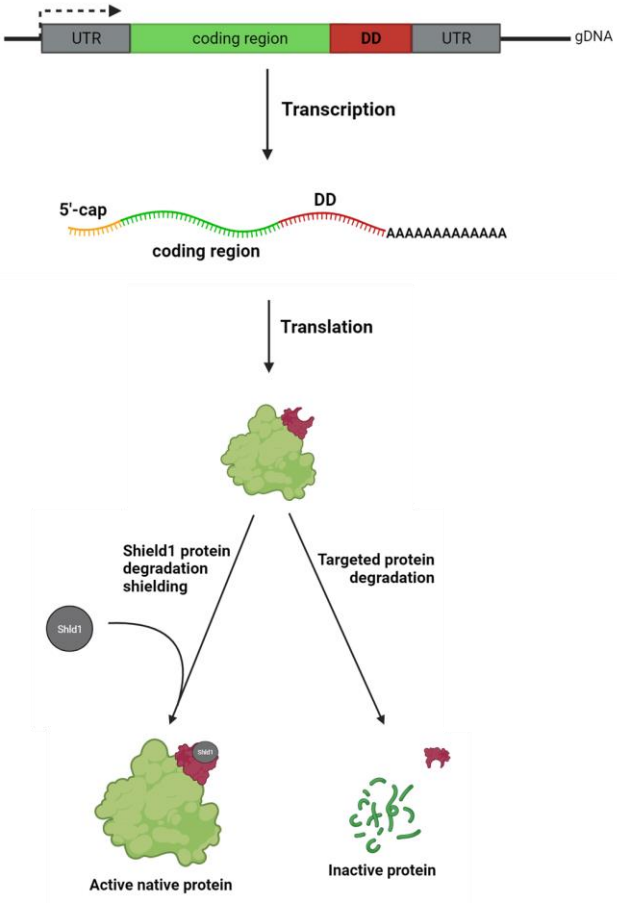


Figure 1.5: Principle of the destabilising domain reverse genetic tool. The *glmS* ribozyme system relies on mRNA degradation through induced cleavage by inducer glucosamine (GlcN). The TetR-DOZI system relies on suppression of protein expression and is mediated by the inducer anhydrotetracycline (aTc). The destabilising domain system relies on stabilisation of the protein of interest by Shield1 (Shld1). Adapted from [47-51]. Figure created with Biorender.com under the basic licence for educational purposes.

1.4.1.3. Gene knock-sideways systems

In the approaches described above, the basic principle involves disrupting the native gene or gene product. As an alternative, a KS approach achieves protein mislocalisation by tagging the protein of interest with retrieval sequences, such as endoplasmic reticulum (ER) (e.g., KDEL) or nucleus sequestration (e.g., NLS) [52, 53]. The process of sequestration is typically performed to relocate the protein of interest away from its native location of action, rendering the processes in which it participates incomplete or ineffective. This system necessitates a genome-level alteration to specifically induce expression of the retrieval sequence at either the N- or C-terminus of the protein of interest [52, 53] (Figure 1.6). More recent studies have employed ‘floxing’-based excision to induce the KS for the GOI.

The appropriate gene manipulation system depends on the scientific question being addressed in the study. It is crucial to introduce genetic manipulation systems either at the native gene site to enable genetic interrogation or an alternative site on the genome, depending on the desired level of expression of the target gene.

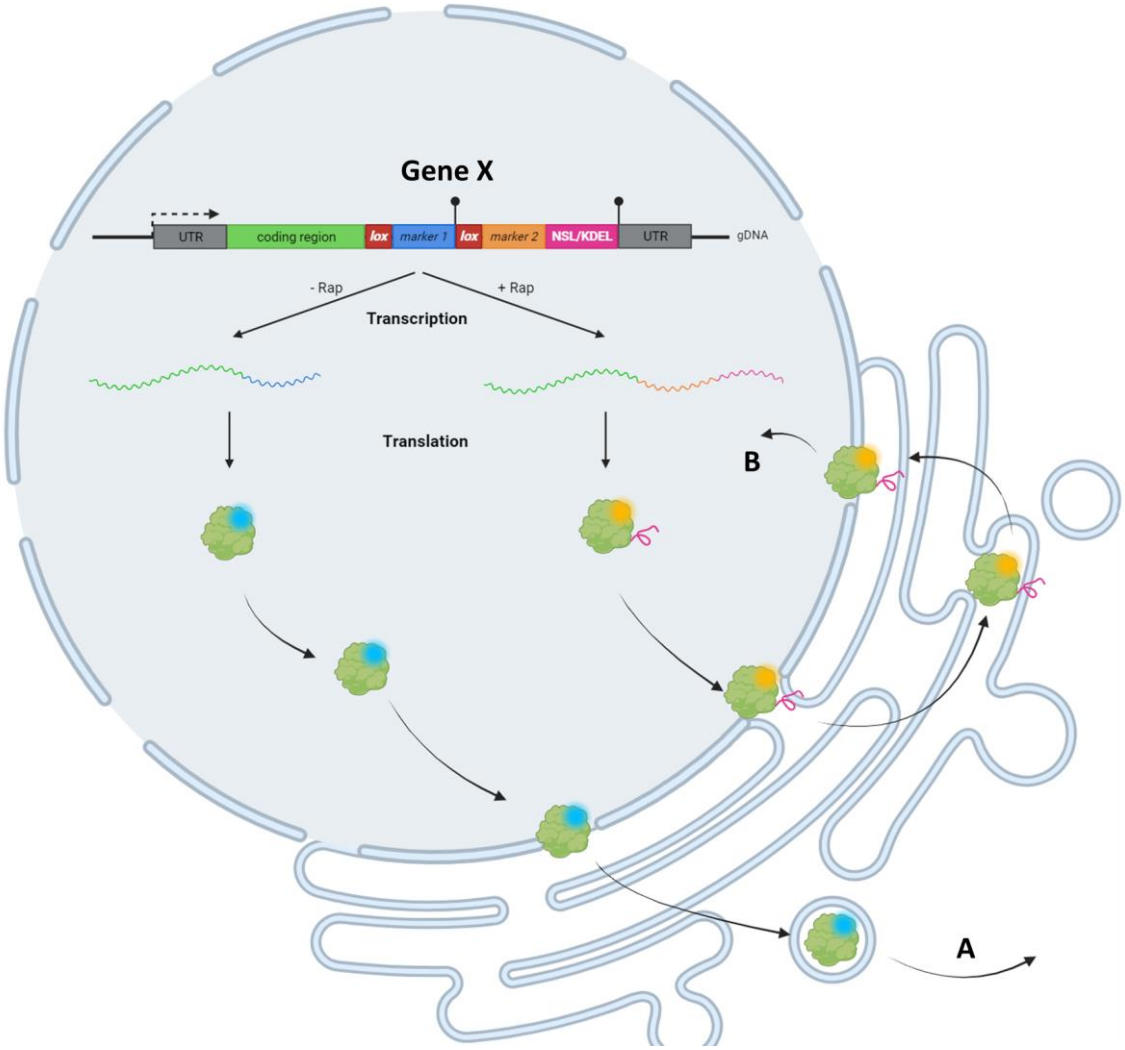


Figure 1.6: Principle of the gene knock-sideways flox-mediated approach. The knock-sideways effect of gene X is seen in the transcription process before and after floxing (DiCre-mediated), with unique coloured markers (blue and orange, for visual purposes) included. These markers represent the region where protein X will be directed within the parasite. Site A indicates the protein's native functional pathway of localization, whereas pathway B shows the induced site for protein relocation (in this case the nucleus) following the floxing of the sequestering sequence (indicated in pink) when knocked into frame. Adapted from [52, 53]. Figure created with Biorender.com under the basic licence for educational purposes.

1.4.2. Genetic modification approaches

To investigate a gene on a molecular level through gene manipulation systems such as the *glmS* ribozyme and DiCre approaches, changes in DNA must first be accomplished. This is necessary given that the native sequence for the GOI lacks the genetic information required for the intended system utilised in gene interrogation. As a result, the system with the relevant nucleotide sequences must be introduced either at the native gene locus or off-site depending on the system being used. Genetic modification approaches in *P. falciparum* include selection-linked integration (SLI), BxB1 integrase and CRISPR-Cas9 systems.

1.4.2.1. Selection linked integration genetic modification

This system has been devised to select transgenic lines more efficiently through a dual-selection approach. This relies on the repair plasmid, containing several selection markers, including the human dihydrofolate reductase gene (hDHFR) which is initially expressed. These markers can be selected for in parasites which contain the plasmid episomally, by the addition of the antifolate, WR99210. In the event of a stochastic break in the genomic DNA, the repair plasmid can act as a donor DNA to integrate into the genome. Homologous recombination allows for the integration of the construct at the 5'- or 3'-end, depending on the system (Figure 1.7) [62, 63]. The recombinant parasites are then selected for by neomycin treatment, given the presence and integration of the neomycin resistance cassette (Neo-R) in this system, which is expressed under the native promoter at the integration site. One drawback of this system is that recombination occurs spontaneously and relies on a double-stranded DNA break that happens infrequently [21, 63].

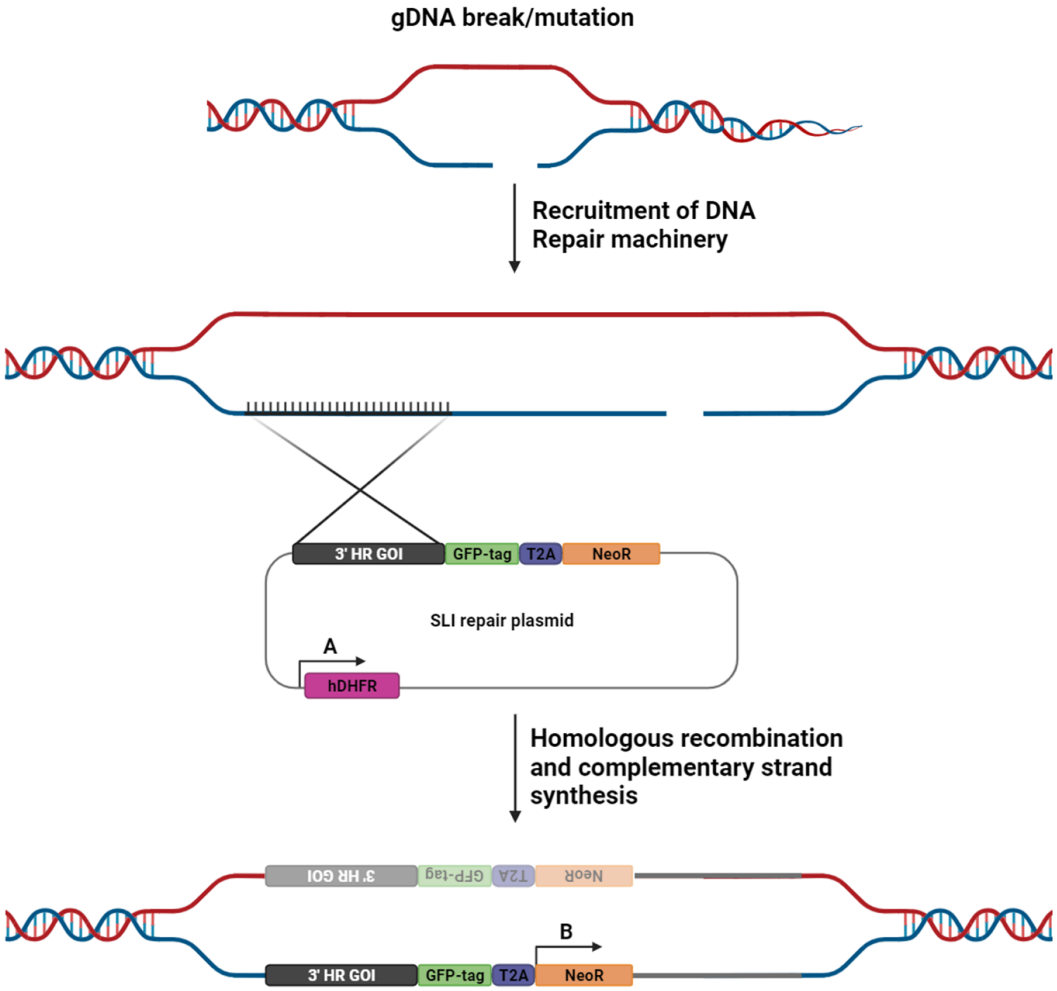


Figure 1.7: Principle of the Selection Linked Integration (SLI) genetic modification approach. A break or change in the DNA triggers the DNA repair mechanism, which uses donor DNA (in the form of a repair plasmid) to align itself with the 3' homologous region (HR) that is typically located near the site of the break or change. The donor DNA is then incorporated into the parasite's genome and includes any changes that were present on the plasmid. Initial screening for transgenic parasites **(A)** is completed by way of the episomally expressed hDHFR, and then once integration occurs, parasites are selected for **(B)** under NeoR treatment, expressed under the integrated site promoter. Adapted from [47, 64]. Figure created with Biorender.com under the basic licence for educational purposes.

1.4.2.2. BxB1 integrase genetic manipulation

This system relies on the mycobacteriophage Bxb1 integrase expressed on a separate plasmid to allow homology-specific recombination between the repair plasmid containing the GOI and the *attB* sequence within the endogenous, non-essential *cg6* gene, on the *P. falciparum* genome [65]. The *attB* sequence is specific for the *attP* sequence on the incoming repair plasmid. This system causes complete integration of the repair plasmid and inversion of the *attP* sequence to form the selectable *attL* and *attR* sequences. This system allows for optimisation of the expression levels for the GOI, by way of a different promoter included on the repair plasmid. An example is the calmodulin (cam) promoter, which allows for ectopic and constitutive expression of the GOI above basal levels [66]. The authors reported successful integrants within 2 – 4 weeks after transfecting [65]. The drawback of this system is the reliance on a parent line with the required *attB* locus, which has been previously established.

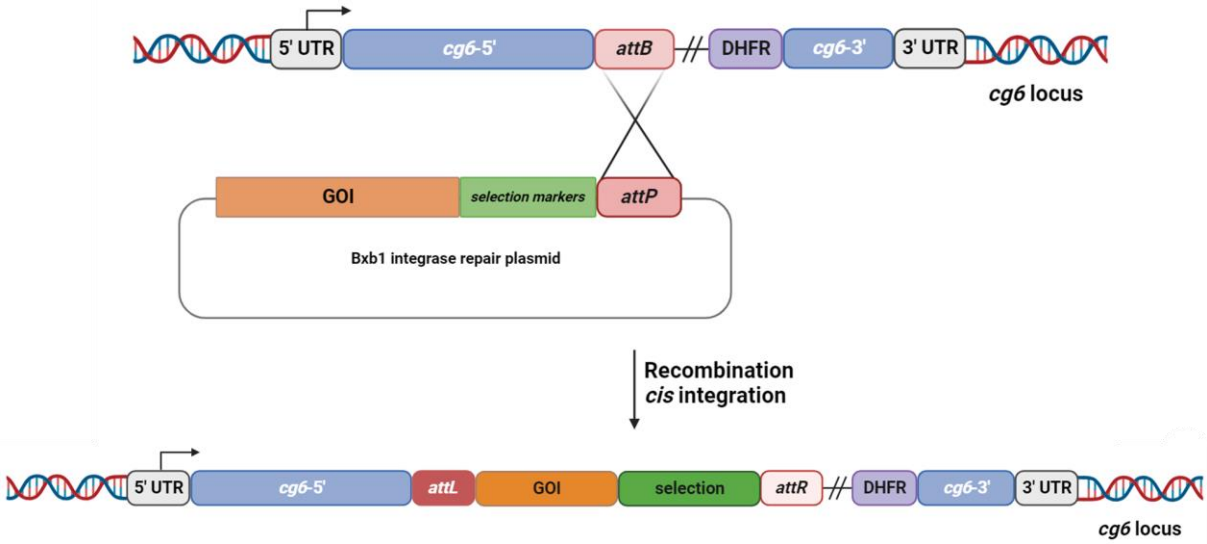


Figure 1.8: Principle of the BxB1 integrase genetic modification system. Integrase (expressed episomally on a different plasmid, not shown) catalyses the recombination between *attB* and *attP* on the same location (cis) of the DNA to produce the unique *attL* and *attR* sequences. These unique sequences are selected for to confirm integration. Adapted from [47, 64]. Figure created with Biorender.com under the basic licence for educational purposes.

1.4.2.3. CRIPSR-Cas9 genetic manipulation

The CRISPR-Cas9 editing system functions together with a gene-specific gRNA as part of the complex to induce a double-stranded break that aids in recruiting DNA repair mechanisms [21, 64]. This system is ideal for quick and highly specific genome editing [64]. For CRISPR-Cas9-based genome editing, a 5'- and 3'-homology region is normally required. This ensures correct alignment of the repair plasmid to the native locus and minimizes off-target recombination, making this system very precise and even more efficient [64]. For many other recombination systems, HRs are typically 1 kb in length [21], but because CRISPR-Cas9 causes dsDNA breaks close to the site of recombination, a smaller homology region is normally sufficient. Several publications report successful integrants within one month of transfecting [41, 46, 64, 67]. This makes CRISPR-Cas9 gene editing approaches ideal for integration at the native locus on the genome and even more useful for genetic alterations to the native gene. The recently developed

codon-optimized Cas9 enzyme is ideal for use in *P. falciparum* given the modifications applied for increased expression efficiency [68]. Similarly, a Cas9 codon-optimized (coCas9) plasmid pDC2-coCas9-gRNA-hDHFR, contains the positive selectable marker (hDHFR) and the negative selectable marker cytosine deaminase/uridyl phosphoribosyl transferase (FCU) [69, 70]. The hDHFR marker enables the selection of parasites that have incorporated the plasmid in the genome following transfection, using WR99210 [70, 71], while the FCU-marker enables the selection of parasites containing the plasmid episomally, using 5-fluorocytosine [72].

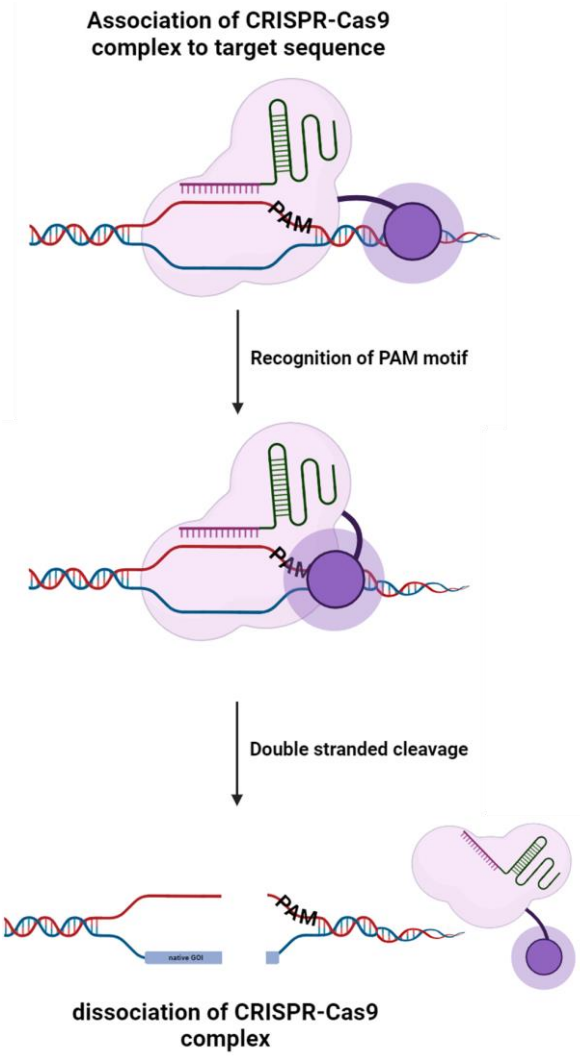


Figure 1.9: Principle of the CRISPR-Cas9 mediated genetic modification approach. The association of the CRIPSR-Cas9 is mediated by the gRNA recognition of the target site on the genome. The Protospacer-adjacent (PAM) motif mediates the enzymatic double-stranded cleavage of the genomic DNA followed by the dissociation of the CRISPR-Cas9 complex. Adapted from [45, 62]. Figure created with Biorender.com under the basic licence for educational purposes.

1.4.3 Combining the power of CRISPR-Cas9 genetic manipulation with DiCre gene modification

To successfully apply the conditional DiCre system, two *loxP* recognition sequences needs to be introduced at the GOI to flank this region at the native site of the gene locus. This typically entails a complicated cloning approach to introduce the required gene segments and can use various integration

systems. A universal repair plasmid incorporating two *loxP* sequences would greatly enhance the cloning efficiency of the subsequent required gene inserts. This, coupled with the integration efficiency and specificity of the CRISPR-Cas9 system, would be a highly beneficial combined system. Therefore, this study aimed to generate a universal DiCre plasmid called pCS22_DiCre to simplify gene editing processes through optimized cloning and efficient CRISPR-Cas9/SLI-mediated gene editing. Figure 1.10 illustrates the proposed integration of CRISPR-Cas9-mediated gene editing with applying the pCS22_DiCre plasmid. The design would allow cut-and-paste cloning of any GOI into the pCS22_DiCre plasmid to facilitate gene KO. Parallel to this, guide RNAs are cloned into the pDC2-coCas9-gRNA-hDHFR plasmid to allow for CRISPR-Cas9 genomic integration. Once co-transfected into the parasite, the gRNA will allow CRISPR-Cas9 localisation and genomic disruption at the site of interest, followed by integration of the *loxP* flanked version of the gene through homologous alignment and repair from the pCS22_DiCre plasmid. Once integrated, the area on the genome can then be ‘floxed’ (rapamycin-mediated excision of *lox*-flanked sequences), removed by the DiCre action of recognition the spanning *loxP* sites to efficiently delete the GOI (Figure 1.10).

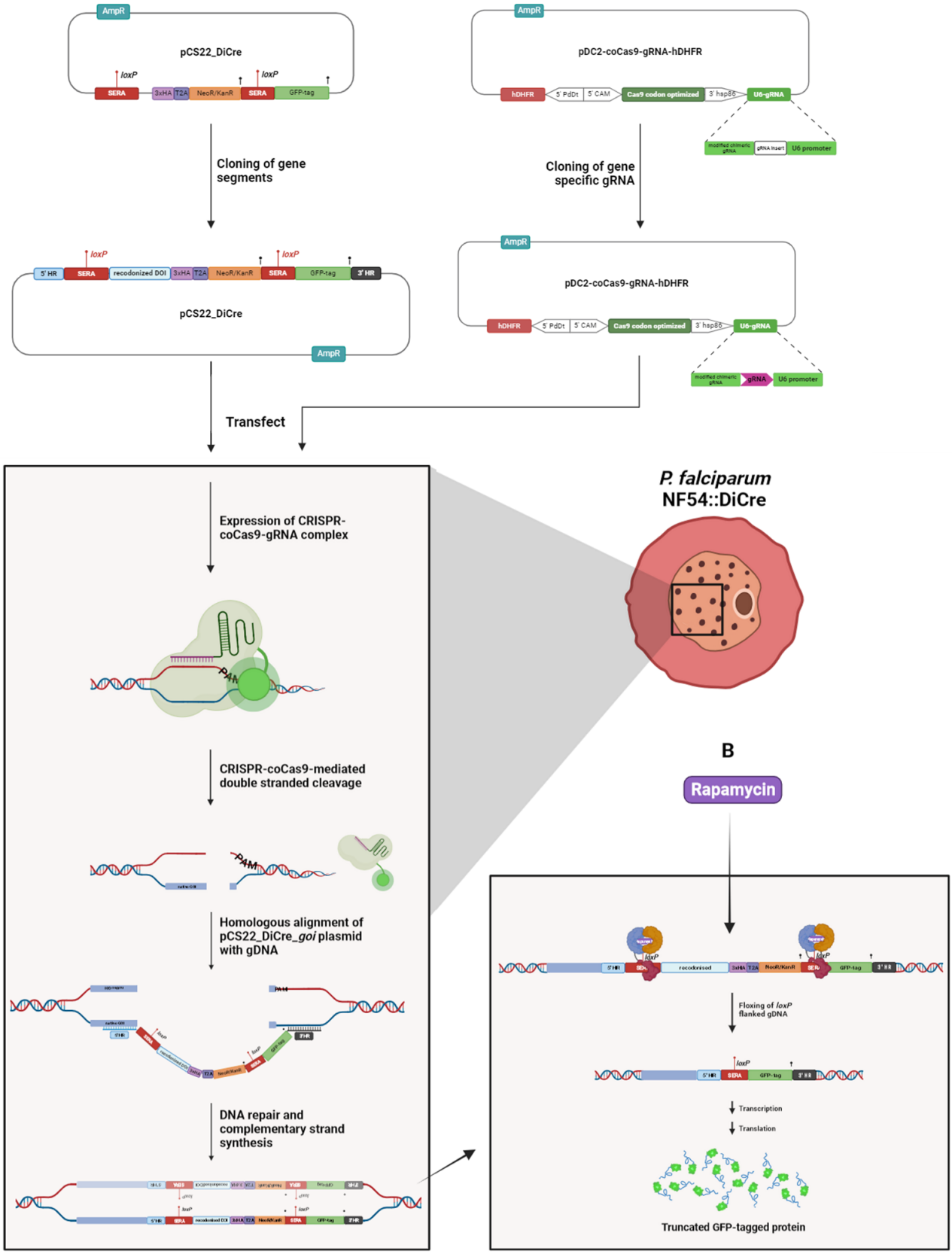


Figure 1.10: Illustration of CRISPR-coCas9-mediated genomic integration of DiCre universal plasmid, containing the respective gene segments. Cloning of the relevant gene inserts into pCS22_DiCre through restriction cut site ligation, thus generating the gene-specific repair plasmid (pCS22_DiCre_*goi*). Cloning of gene-specific guide RNA (gRNA) into pDC2-coCas9-gRNA-hDHFR. This can then be co-transfected (**A**) into asexual *P. falciparum* NF54::DiCre parasites to allow integration at the locus of interest. Initial screening for transgenic parasites can be completed by way of the episomall expression of the hDHFR located on the CRISPR-coCas9 plasmid, and then once integration occurs, parasites can be selected for under NeoR treatment, expressed under the integrated site promoter. **(B)** Rapamycin can be added to the *P. falciparum* NF54::DiCre culture to mediate *loxP* site excision. Figure created with Biorender.com under the basic licence for educational purposes.

Hypothesis

A unique cloning system can be generated for use on any *P. falciparum* GOI, to generate a repair plasmid for the DiCre system.

Aim and Objectives

The aim of this project was to generate an optimized unique cloning system for efficient cloning of various *P. falciparum* genes for CRISPR-Cas9 or SLI mediated integration for DiCre-mediated conditional KO to investigate gene function in *P. falciparum*.

Objectives include:

a) *In silico* design of the cloning strategy for the DiCre recombinase system.

b) Generate a generic repair plasmid for the DiCre system.

c) Generate a repair plasmid for the gene candidates, *gcn5* and *set7*.

d) Generate a gRNA containing plasmids for the gene candidates, *gcn5* and *set7*.

Research outputs

van Zyl, C., Niemand, J., Birkholtz, LM. Genetic interrogation of GCN5 and SET7 as essential regulators of histone post-translational modifications in the malaria parasite, *Plasmodium falciparum*. *8th South African Malaria Research Conference*. Poster presentation. Pretoria. August 2023

2. Materials and Methods

2.1. *In silico* design of the universal DiCre repair plasmid and cloning system

A cloning strategy was devised to bypass the need for continuous cloning of the selection marker cassette described in the approach by Tibúrcio *et al.* [45], and generate a universal DiCre repair plasmid, hereafter referred to as pCS22_DiCre (Figure 2.1). Additionally, the cloning approach described here would allow either genomic integration through the CRISPR-Cas9 or SLI genetic modification systems. An advantage of this system is the availability of many cloning sites to insert the 5' HR and 3' HRs. However, the *Sna*BI site is an exception which is used in all cloning instances. Given that it is right at the start of the first SERA2 intron and it is critical that no nucleotides are deleted, as it may affect the splicing of this intron (Figure 2.1). This is also true for the *Bsm*I site located at the very end of the first SERA2 intron, which is lost together with the 3' AG- overhang, once digested. To accommodate for this, the AG overhang is therefore incorporated in the recodonised insert on the forward primer at the 5'-end.

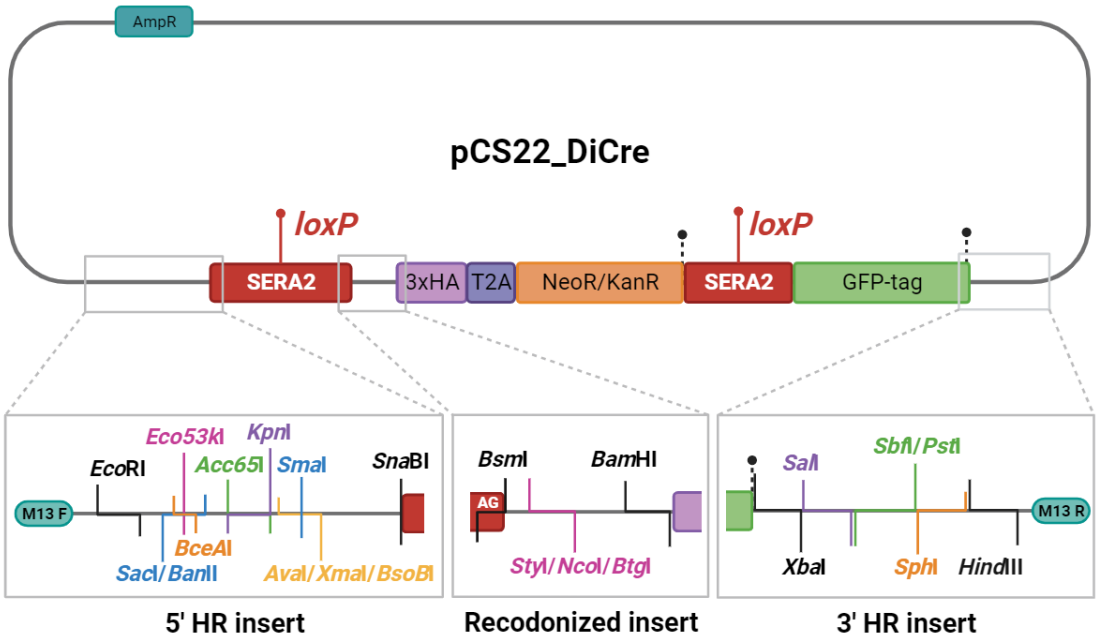


Figure 2.1: Schematic of the cloning sites of gene segments on the pCS22_DiCre plasmid. Names of enzymes indicated in black are those applied specifically in this study. Alternative enzymes are indicated with their corresponding colours. The relevant sites for cloning the required 5' HR, recodonised region and 3' HR, are also indicated. Stop codons are indicated with dotted vertical lines. **M13 F:** forward plasmid backbone primer. **SERA2:** SERA intron-containing ***loxP*** sequence. **AmpR:** Ampicillin resistance selection marker. **KanR:** Kanamycin resistance selection marker. **HA:** Haemagglutinin. **T2A:** skip peptide. **NeoR:** Neomycin resistance selection marker. **M13 R:** reverse plasmid backbone primer. Figure created with Biorender.com under the basic licence for educational purposes.

The pCS22_DiCre plasmid was strategically designed to contain two *loxP* sequences, along with a selection marker cassette and restriction cut sites for cloning purposes (Figure 2.1). One *loxP* sequence and several restriction cut sites were identified on the pSN28 pUC19 loxPint plasmid (a kind gift from the Moritz Treeck lab, Francis Crick Institute) through Benchling (https://www.benchling.com/), and facilitated the generation of the pCS22_DiCre plasmid. The pSN28 plasmid was therefore used as backbone for the

entire construction of pCS22_DiCre. The pSN28 plasmid has the advantage of several different restriction cut sites that were thereby retained for downstream cloning of the relevant gene inserts required to generate the final gene repair plasmid. (Figure 2.1) summarises the restriction cut sites proposed for the cloning strategy described here, all of which span six nucleotides in length. Additionally, the necessary selection marker cassette containing the second *loxP* sequence, was sourced from the pH33 FIKK10.1cKOGFP_Cas9 [56]. The selection marker cassette was cloned downstream of the *loxP* site on the pSN28 plasmid, through restriction enzyme overhang for *Bam*HI and *Xba*I sequences, to generate the pCS22_DiCre.

Table 2: Restriction enzymes optimized for use in this cloning strategy and the desired location on the respective primers.

Restriction enzyme	Sequence (5' – 3')	Located on	Alternative enzyme on pSN28/pCS22_DiCre
***Eco*RI**	GAATTC	5' HR forward primer	**Eco53k*I, *Bce*AI, #*Ban*II↔*Sac*I, *Acc65*I, #*Ava*I↔*Xma*I↔*BsoB*I↔*TspM*I, **Sma*I, *Kpn*I
***Sna*BI**/*Bsa*AI	TACGTA	None	None
***Bsm*I**	GAATGCN	None	None
***Bam*HI**	GGATCC	Recodonised reverse Selection cassette forward	*Sty*I↔ #*Nco*I↔#*Fat*I↔*Btg*I, *Xba*I, #*Sal*I, *Acc*I, *Hinc*II, *Sbf*I, #*Pst*I, *BspM*I, *BfuA*I, #*Sph*I, *Hind*III
***Xba*I**	TCTAGA	Selection cassette reverse 3' HR forward	
***Hind*III**	AAGCTT	3' HR reverse	

*not recommended due to blunt cut; #not recommended due to the presence of the same cut site located in the selection marker cassette; ↔ enzymes with the same recognition sequence

The designed system should then allow efficient, rapid and accurate cloning of any GOI, as depicted in Figure 2.3. The recombinant plasmid is generated by amplifying (or purchasing synthetic versions) of three gene fragments: the 5' HR, the recodonised region and the 3' HR with primers containing the required overhangs. For the 5' HR, the PCR product is blunted with T4 DNA Polymerase, digested with the corresponding restriction enzymes (*Eco*RI), and then ligated into the pCS22_DiCre plasmid with complementary overhangs (sticky *Eco*RI and blunt *Sna*BI site), as indicated. For the recodonised gene fragment, the nucleotide sequence is obtained from a purchased pUC18 plasmid, amplifying from the plasmid using primers with the desired overhangs ("AG" for 5' and *Bam*HI for 3'); the plasmid is blunted after digestion with the *Bsm*I enzyme, followed by the digestion with *Bam*HI. For the 3' HR, the PCR product is only digested with the required enzymes (*Xba*I and *Hind*III).

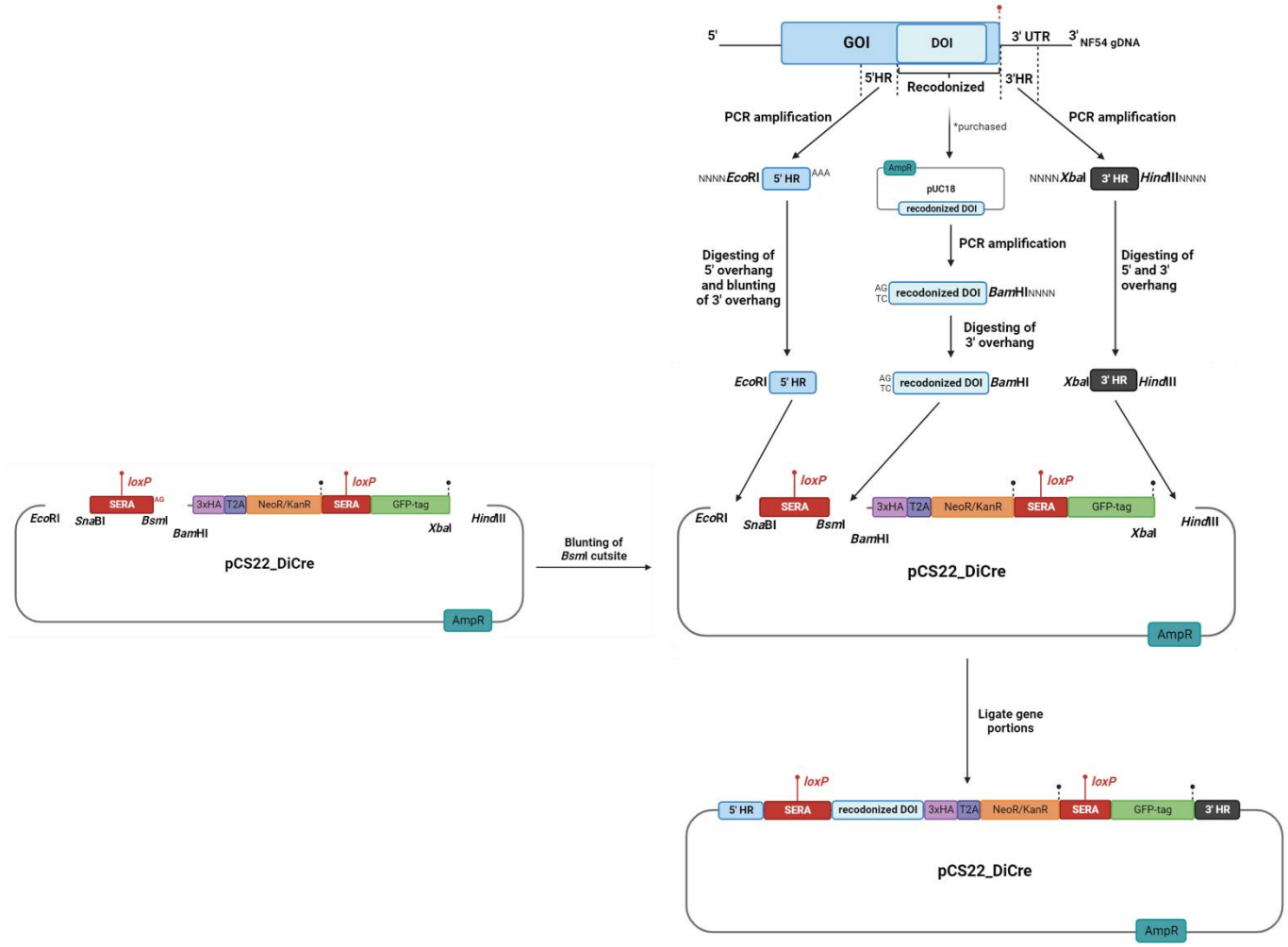


Figure 2.2: Cloning strategy to generate a recombinant pCS22_DiCre plasmid containing the gene fragments of interest. Stop codons are indicated as dashed vertical lines. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

2.2. Cloning of the pCS22_DiCre plasmid

The cloning strategy, as depicted in Figure 2.3, outlines the steps that were used to generate the pCS22_DiCre plasmid. The selection marker cassette was amplified from the pH33 FIKK10.1cKOGFP_Cas9 (a kind gift from the Moritz Treeck lab, Francis Crick Institute) [56]. Benchling was used to create primers designed for the amplification of this cassette, considering the unique features required for the cloning of this cassette, including the desired complementary restriction overhang sequences. These sequences were artificially introduced by PCR of the cassette with primers containing restriction recognition sequences for *Bam*HI (5' G/GATCC 3') and *Xba*I (5' T/CTAGA 3') to the forward (SLC_F) and reverse primers (SLC_R), respectively (Figure 2.3 – Step 1). A 4-nucleotide overhang was also introduced on the 5' end of each primer to allow for the binding and cleavage of the sequence by the restriction enzyme. This selection cassette with *Bam*HI and *Xba*I sites then contained the desired overhang sequence required to generate pCS22_DiCre, through complementary sticky end ligation using the restriction sites available on pSN28 pUC19 loxPint plasmid.

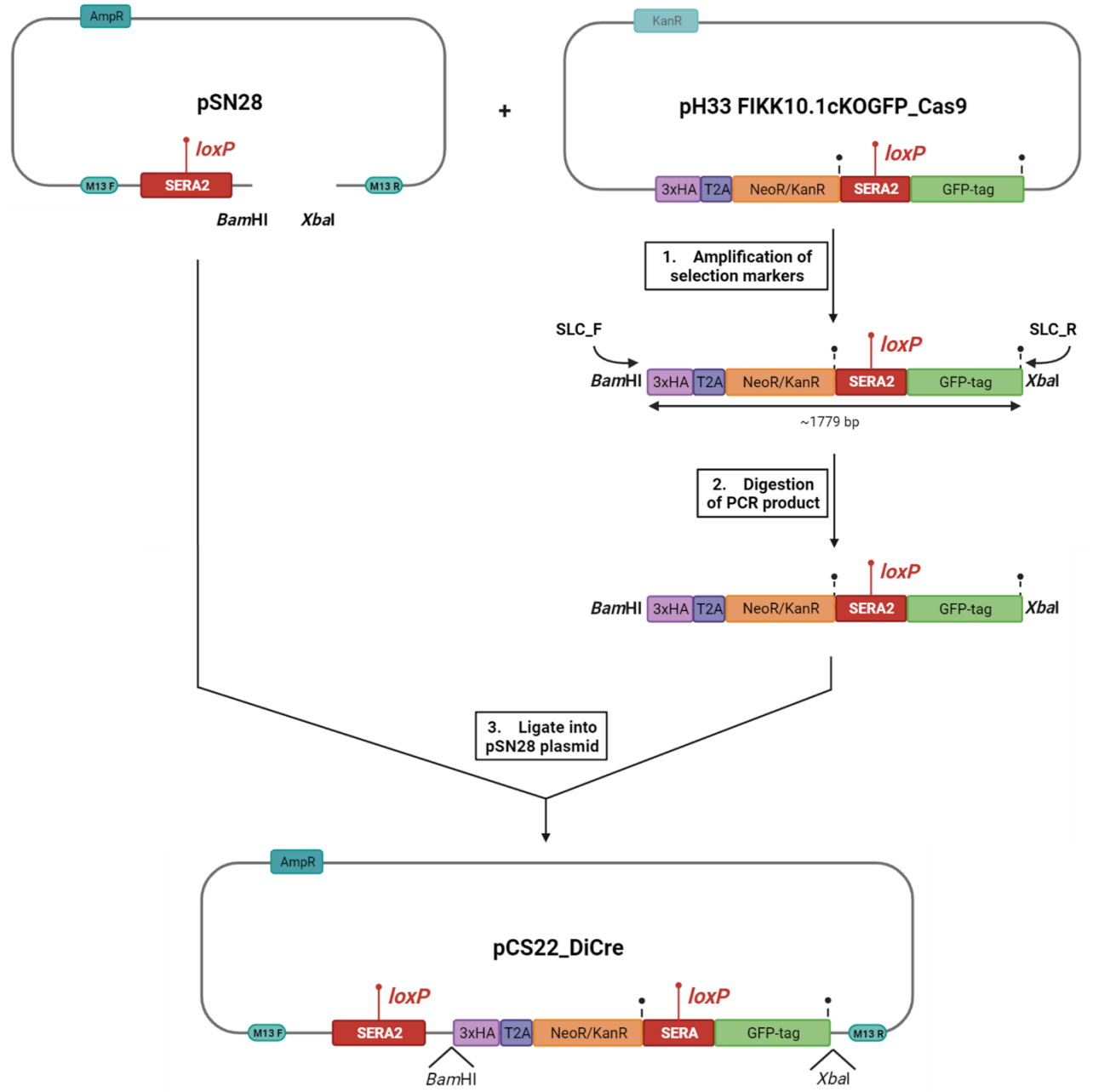


Figure 2.3: Schematic representation of the cloning strategy employed to generate the pCS22_DiCre plasmid. The universal cut-and-paste DiCre plasmid was generated by **(1)** amplifying the selection cassette from the pH33 plasmid, with cassette-specific primers listed in **Table 2**, containing the required overhangs. The PCR product was **(2)** digested with the corresponding restriction enzymes (*Bam*HI and *Xba*I), and then **(3)** ligated into the pSN28 plasmid with complementary overhangs, as indicated. Stop codons are indicated as dashed vertical lines. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

The primer sequences (Table 3) were optimized for the appropriate melting temperatures, adapting for larger ΔG values to ensure a smaller chance of monomerization of the primers and the formation of secondary structures, including hetero- or homodimers. Primers were manufactured by Inqaba Biotechnical Industries (Pty) Ltd (South Africa) and dissolved in sterile $dddH_2O$ according to the instructions provided. Primer primary stocks and aliquots were stored at - 20 ºC.

Table 3: Primer sequences used to amplify the selection cassette. Nucleotide overhangs are indicated in lowercase, while the cassette-specific sequence is all uppercase. The bold and underlined sequence represents the restriction enzyme recognition sequences.

Primer code	Oligo primer sequence (5' – 3')		Length of amplicon
SLC_F	F	cgct**<u>GGATCC</u>**TACCCATATGATGTACCGGATTACG	~1799 bp
SLC_R	R	cgct**<u>TCTAGA</u>**TTATTTGTATAGTTCATCCATGC	

To introduce the cassette of selection marker into the pSN28 plasmid, the PCR product was digested with *Bam*HI and *Xba*I, and then ligated with the pSN28 plasmid digested with the same restriction enzymes, which allowed for directional cloning of the cassette into the now universal plasmid, pCS22_DiCre, that is used in downstream cloning steps.

2.3. Generation of pCS22_DiCre plasmid

2.3.1. Preparation of $CaCl_2$ heatshock-competent bacterial cells

$CaCl_2$-heatshock competent bacterial cells were prepared as per the method published by Inoue *et al.* [73]. Competent *E. coli* cells were prepared from a saturated *E. coli* DH5α culture, grown overnight in 1× LB (Luria-Bertani broth [0.5 % (w/v) yeast extract, 1 % (w/v) tryptone, 171.23 mM NaCl, pH 7.5; Glentham Life Sciences, UK) at 37 ºC, shaking at 180 rpm. From this, a secondary culture was inoculated (diluted 1/50 in 1× LB media, 37 ºC and shaking at 180 rpm) until an OD_{600} reading of ~0.4 (log phase of bacterial growth) was obtained. This culture was then chilled on ice and then spun down at 1865 ×*g* for 30 min (SL 8R Centrifuge, Thermo Scientific, ZA), after which the pelleted cells were resuspended and washed two times in ice-cold 0.1 M $CaCl_2$-solution (Merck, ZA). The final round of suspended cells was then stored with a 70 % glycerol, by way of snap freezing at -80 ºC.

2.3.2. Transformation of plasmid DNA

The pH33 and pSN28 plasmids were transformed, separately, with $CaCl_2$-heatshock competent bacteria, previously prepared (section 2.2.1.), by incubating 10 ng of the respective plasmids with 100 µL competent bacterial cells thawed on ice, for 30 min. Bacteria cells were heat shocked at 42 ºC for 90 s and then placed back on ice for 2 min. Immediately after this, 900 µL pre-warmed LB-glucose (1/100 dilution of 2 M glucose in 1× LB media) was added to the transformation reaction and incubated at 37 ºC (180 rpm) for 1 h. A total of 100 µL of the transformed bacteria transformation was then plated on 1 % (w/v) LB-agar plates supplemented with ampicillin (100 µg/mL final concentration, LB-agar-amp) to select for bacteria containing the pSN28 plasmid, while the pH33 plasmid transformation was plated on 1 % (w/v) LB-agar plates supplemented with kanamycin (100 µg/mL final concentration). The transformants were left to grow overnight at 37 ºC.

Individually picked colonies were grown overnight in 10 mL of the same 1× LB-ampicillin (LB-amp) or LB-kanamycin media, respectively, after which the plasmid was isolated from the overnight culture using the plasmid DNA isolation kit NucleoSpin Plasmid (Machery Nagel GmbH & Co.KG, Germany) according to the manufacturer's instructions. Isolated DNA was quantified using a NanoDrop One^C Spectrophotometer

(Fisher Scientific, USA) which gives a reading of the concentration and purity, which is based on the maximal absorbance of pyrimidine and purine bases (at a wavelength of 260 nm) that is present in the sample, together with the maximal absorbance of proteins (or aromatic amino acids at 280 nm). Additionally, an absorbance reading of 230 nm represents the maximal absorbance of salts and phenolic compounds. DNA samples with readings of 1.7 – 2.2 for both the A_{260}/A_{280} and A_{260}/A_{230} ratios, were regarded as pure and used in downstream cloning steps. Samples were stored at -20 °C until used to validate through restriction enzyme (RE) mapping.

2.3.3. Restriction enzyme mapping

Restriction enzyme digestion was used to further confirm the identity of the isolated plasmid DNA by digesting 0.5 µg of each plasmid (pSN28 and pH33) with 10 U each of *Eco*RI-HF, *Hind*III-HF and *Bam*HI-HF in different combinations, including *Eco*RI only, *Eco*RI and *Hind*III, and *Eco*RI and *Bam*HI. The digestions were incubated with 1x 2.1 buffer (10 mM magnesium chloride, 10 mM Tris-hydrochloride, 50 mM sodium chloride and 100 µg/mL Bovine Serum Albumin (BSA), pH 7.9; New England Biolabs, UK), for 3 h at 37 ºC using 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). The digested plasmid was analysed on a 0.8 % (w/v) agarose gel (prepared with Tris-Acetate-EDTA (TAE) [1 mM EDTA, 20 mM acetic acid and 20 mM Tris, pH 8.0, supplemented with 1 µg/mL Ethidium Bromide (EtBr, Invitrogen, USA)], loading the samples and a 1 kb DNA ladder (Thermo Fisher, USA) using 6× TriTrack DNA Loading Dye (10 mM Tris-HCl (pH 7.6), 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, 0.15 % orange G, 60 % glycerol, and 60 mM EDTA) to identify plasmid fragments with corresponding sizes of the backbone. Gel images were visualized with the Gel Doc XR Imaging system (Bio-Rad, USA), using the corresponding imaging software, Image Lab (Version 6.1, USA)

2.3.4. PCR amplification

A PCR product was obtained by setting up an amplification reaction using ~50 ng of previously isolated pH33 plasmid, 5 pmol final concentration of each primer (SLC_F and SLC_R, Table 3) and polymerase 1× KAPA Taq Ready Mix [specifications are KAPA Taq DNA Polymerase (0.5 U/25 µL), proprietary KAPA Taq buffer supplemented with $MgCl_2$ (1.5 mM), dNTPs (0.2 mM each) and proprietary stabilizers; KAPA Biosystems, Wilmington, USA], in a 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). Previously optimized thermal cycling steps included an initial denaturation step (95 ºC, 3 min), 30 - 35 cycles including denaturation (95 ºC, 30 s), annealing (optimized at 58 ºC, 30 s), extension (68 ºC, 2 min), with a final extension step of (68 ºC, 5 min). The PCR reaction was loaded on a 0.8 % (w/v) agarose/TAE gel (prepared as before), loading a 100 bp DNA ladder (Thermo Fisher, USA) with 6× TriTrack DNA Loading and the sample with 6× purple loading dye (0.02 % proprietary Dye 1, 0.001 % proprietary Dye 2, 2.5 % (w/v) Ficoll 400, 3.3 mM Tris-HCl, 10 mM EDTA, and 0.08 % SDS). DNA fragments of the desired size were analysed by UV visualization, using the imaging systems and corresponding imaging software as previously stated (section 2.1.1.3). The amplified DNA was then isolated using the NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer's instructions. The

bands of the desired size were excised from the agarose gel and mixed with the binding buffer (30-45 % guanidinium thiocyanate, proprietary mix). The solution was heated to dissolve the agarose, while the chaotropic salt from the buffer allowed the binding of the DNA to the silica membrane. Contaminants were then washed out with the use of an ethanol-containing wash buffer, after which an elution buffer (5 mM Trish/HCl, pH 8.5) was used to elute the DNA sample, quantified spectrophotometrically, and stored at -20 °C, as previously stated.

2.3.5. Restriction enzyme digestion

Next, both the isolated PCR product and ~1 µg of previously isolated pSN28 plasmid were digested with 10 U of both *Bam*HI-HF and *Xba*I (New England Biolabs, UK), including 1× CutSmart buffer (10 mM $Mg(CH_3COO)_2$, 20 mM Tris-acetate, 50 mM CH_3CO_2K and 100 µg/mL BSA; New England Biolabs, UK), for 3 h at 37 ºC using a 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). The PCR product and backbone was then purified from the solution using the NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer's instructions, and stored at -20 ºC or immediately ligated as described below.

2.3.6. Ligation

The digested selection markers cassette PCR product (~10 ng) was ligated with ~50 ng of the complementary sticky ends digested pSN28 (*Bam*HI and *Xba*I) using 400 U of T4 DNA Ligase (New England Biolabs, UK) and 1× T4 ligation buffer (1 mM ATP, 10 mM Dithiothreitol (DTT), 10 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5; New England Biolabs, UK). To determine the amount of vector and insert that were required to ligate the vector and insert, a 3:1 ligation reaction of insert to vector molar ratio was used in Equation 1 to calculate this amount (in ng). The ligation reaction was incubated at room temperature overnight and then placed at 4 ºC until 10 µL of the reaction could be transformed with 100 µL heatshock competent *E. coli* DH5α cells, previously prepared, and then plated onto LB-agar-amp plates and incubated overnight at 37 ºC.

Equation 1: Formula to calculate the amount of insert needed (ng) for ligation with backbone.

$$\frac{\text{ng of vector (50 ng)} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert:vector molar ratio} = \textbf{ng of insert}$$

2.3.7. Screening of positive colonies containing selection cassette insert

Individually picked bacterial colonies, from the transformed and plated *E. coli* DH5α cells with recombinant pCS22_DiCre plasmids, were inoculated into 100 µL of LB-amp and incubated with shaking at ~180 rpm for ≤3 h at 37 °C in an MRC LM-570 incubator (MRC, Israel). Following the incubation, a PCR screening reaction was set up with 1× KAPA Taq Ready Mix (KAPA Biosystems, Wilmington, USA), in the 2720 Thermo Cycler (Applied Biosystems, Foster City, USA), 5 pmol of each of the backbone or gene-specific primer, and 1 µL of the singular incubated colony. The primer pair for this PCR screening included a plasmid backbone specific M13 forward (M13 F) and selection cassette-specific reverse primer (SLC_R),

refer to Tables 3 & 4. Previously optimized thermal cycling steps included an initial denaturation step (95 ºC, 5 min), 30 - 35 cycles including denaturation (95 ºC, 30 s), annealing (optimized at 58 ºC, 30 s), extension (68 ºC, 1 min), with a final extension step of (68 ºC, 2 min). The PCR product was then visualized on a 0.8 % (w/v) agarose/TAE gel (prepared as stated above).

Positive colonies (band with size ~1800 bp) were inoculated into 10 mL 1× LB-amp media and incubated overnight with shaking, whereafter the plasmid was isolated using the plasmid DNA isolation kit NucleoSpin Plasmid (Machery Nagel GmbH & Co.KG, Germany) according to the manufacturer's instructions, quantified spectrophotometrically, and stored at -20 °C until use.

Table 4: Primer sequences used to select for and sequence the selection cassette. Nucleotide overhangs are indicated in lowercase, while cassette specific sequence is all uppercase.

Primer code	Purpose	Oligo primer sequence (5' – 3')	
SP1	Screen/Sequence 5' HR and/or recodonised region	F	CGGGCCTCTTCGCTATTACG
M13 forward	Screen/Sequence 5' HR and/or recodonised region	F	TGTAAAACGACGGCCAGT
SP2	Screen/Sequence Recodonised region	R	ACGACCGCATAATCCGGTAC
SP3	Screen/Sequence for selection markers cassette	F	GGCCGCTTTTCTGGATTCATCG
SP4	Screen/Sequence for selection markers cassette	R	CAAGAATTGGGACAACTCCAGTG
SP5	Screen/Sequence 3' HR	F	GCCCTTTCGAAAGATCCCAACG
M13 reverse	Screen/Sequence 3' HR and/or selection markers	R	CAGGAAACAGCTATGACCATG
SP6	Screen/Sequence 3' HR and/or selection makers	R	ACCCCAGGCTTTACACTTTATGC

2.3.8. Sanger sequencing of insert

Isolated plasmids were then sequenced using the Sanger sequencing system. This system is based on the incorporation of dideoxynucleotides labelled fluorescently, where every other base incorporated causes termination of the chain of nucleotides being amplified. These signals are detected and read as a single sequence. The sequencing reaction (20 µL) was set up with components from the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City) which includes 1× BigDye (proprietary mix, 2 µL) and 2× BigDye sequencing buffer (proprietary mix, 4 µL), together with 300 ng of plasmid DNA, 5 pmol of primer M13 F, SP3, SP4 and M13 R, in different reactions (Figure 2.4). The thermal cycling setup was previously optimized and included an initial denaturation step (95 ºC, 1 min) and a repeat of 25 cycles that include a denaturation step (95 ºC), annealing step (58 ºC, 5 s) and final extension step (60 ºC, 4 min).

Each 20 µL reaction was the cleaned by using ethanol precipitation steps. This included mixing the reaction with ice-cold 100 % absolute EtOH (50 µL) with 3 M CH_3COONa (pH 5.2, 2 µL) and chilling on ice for 15 min. The precipitate reaction was then centrifuged at 12 500 ×*g* for 35 min (4 ºC; 5415 R Eppendorf Centrifuge, Hamburg, Germany). The pellet was washed by carefully adding 70 % (v/v) ethanol (250 µL)

and then immediately centrifuging the reaction for 15 min (as above). The supernatant was carefully removed, and the pellet dried by evaporating residual ethanol using an *in vacuo* system (Bachofer, BA-VC-300H vacuum concentrator). The purified cycle sequencing PCR extension products were submitted for sequencing to the Sanger Sequencing facilities at the Centre for Bioinformatics and Computational Biology which forms part of the Forestry and Agricultural Biotechnology Institute (FABI, University of Pretoria), with an ABI3500xL Genetic Analyser (Applied Biosystems, Thermo Fischer Scientific, USA). Alignment and analysis of sequencing reactions were carried out with the alignment tool on Benchling.

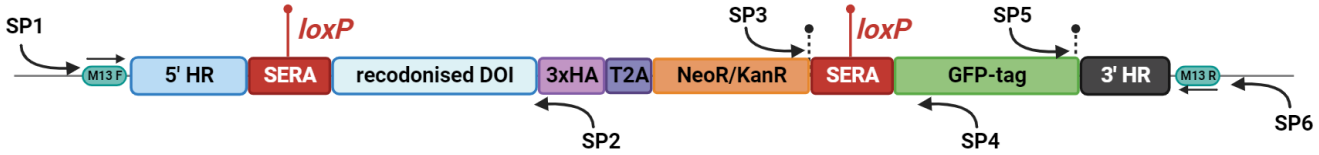


Figure 2.4: Schematic of the various sequencing primers used for pCS22_DiCre. Arrows indicated present relative position and orientation of screening and sequencing primers for the pCS22_DiCre plasmid, i.e. forward primers are arrows from left to right and reverse primers are arrows from right to left. Figure created with Biorender.com under the basic licence for educational purposes.

2.4. *In silico* analysis for cloning of proxy genes

As proof of concept for the useful application of the design strategy described in this study, two gene candidates were chosen: *gcn5* (PF3D7_0823300) and *set7* (PF3D7_1115200), given their proposed regulatory nature that could be addressed using this system. Benchling was used throughout for *in silico* analysis.

The two domains of *gcn5* (GNAT and Bromodomain) and the singular domain found in *set7* (SET), were chosen to be cloned. The genomic DNA nucleotide sequence as it reflects the well-documented reference strain, 3D7, was obtained for each gene from the online *Plasmodium* database (PlasmoDB: https://plasmodb.org/plasmo/app/), to allow the correct placement of the *loxP* sequences to flank the domain/s of interest and ensure the correct 'floxing' of the gene. The relevant domain/s amino acid sequence was retrieved from InterPro (https://www.ebi.ac.uk/interpro/) using the nucleotide sequence obtained from PlasmoDB to correctly define the domain areas. Once the sequence for the domains were aligned to the genomic sequence, the domain and the remaining downstream nucleotides were chosen as part of the recodonised gene region. This region was designed to be genetically distinct from the genomic DNA. This distinction prevents the region from aligning with the genomic DNA, ensuring the accurate alignment of the 5' and 3' homology region (cloned downstream of GFP), during DNA repair. Recodonisation enhances the application of the CRISPR-Cas9-mediated genetic editing by facilitating the mutation of the PAM motif on this DNA sequence, preventing continuous cleavage of the genomic DNA after integration. Furthermore, the recodonisation region is designed to exclude any introns to allow for optimal gene expression and does not include the native stop codon. Additionally, this specific region will be synthesized by a manufacturer, and cannot be amplified from genomic DNA.

The native stop codon is not included in the recodonisation gene insert given the presence of the artificial stop codon at the end of the Neo-R cassette (Figure 2.5). This stop codon functions in expression pre-floxing with rapamycin, while the stop codon after the GFP functions for post-floxing with rapamycin. The GFP is originally cloned out-of-frame in the cassette and therefore requires that post-floxing that it come into frame. This serves as a proxy for correct expression of the gene once floxed.

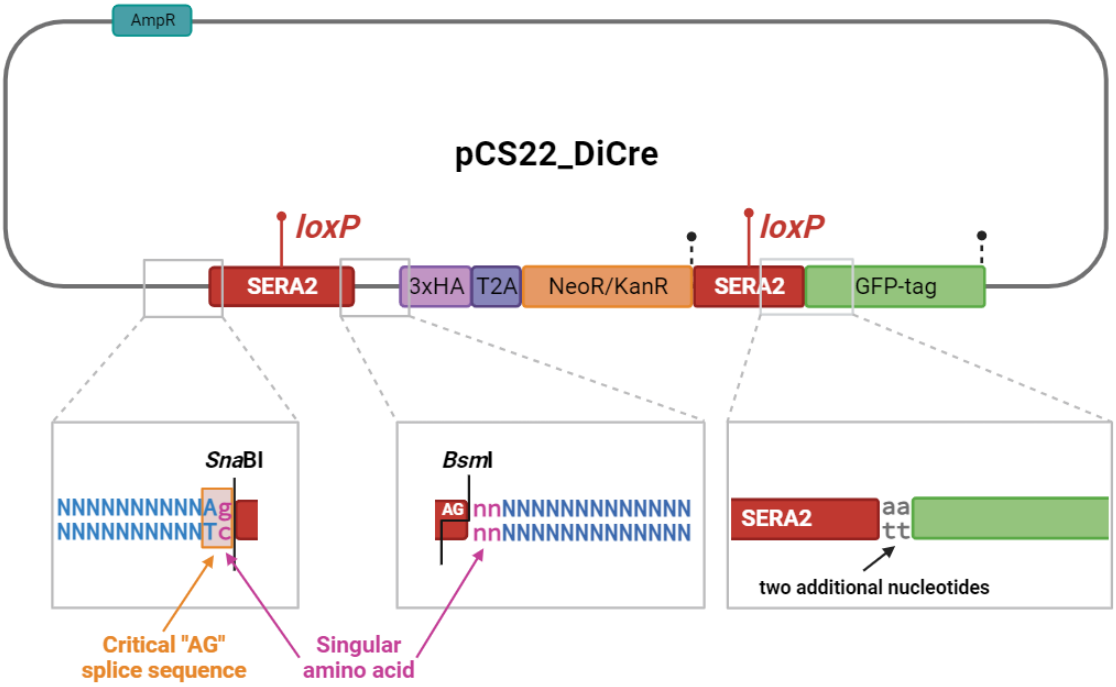


Figure 2.5: Schematic depicting how an "out-of-frame" AG sequence needs to be incorporated for correct incorporation of the GFP post-floxing. The light blue sequence on the left represents the 5' HR sequence chosen, while the AG encircled in orange represents the "out-of-frame" splicing sequence. The sequence in pink are the three nucleotides that together code for the same amino acid on the native sequence, split between the first SERA2 intron to accommodate for the out-of-frame GFP downstream. Figure created with Biorender.com under the basic licence for educational purposes.

Given the presence of the *loxP* sequences on the pCS22_DiCre plasmid, cloning was accomplished by including complementary restriction cut sites on primers as overhangs when amplifying the relevant gene segments, and cloning was completed by way of complementary overhang ligation. For this system, the splice sequence "AG" (Figure 2.5) was chosen to form part of the 5' HR sequence. This is based on the normally described splice sequences in *P. falciparum* [74] and was used to aid in the successful splicing of the SERA2 introns, which already contain the "AG" splice sequence. For this cloning system, the 5' HR (minimum 500 bps in length, 3n+1) is cloned upstream of the first *loxP* at the *Sna*BI cut site on pCS22_DiCre. However, to accommodate for the downstream out-of-frame GFP on nucleotide from the first amino acid to form part of the recodonised gene segment will need to be included for the 5' HR with corresponding primers (Figure 2.5). The "out-of-frame" nature of the floxing of GFP refers to the amino acid sequence chosen for the 5' HR. Under rapamycin-mediated excision of the *loxP* sequence, with the remaining sequence spliced with the SERA2 intron, this will allow the downstream GFP to come into frame [56], hence the additional nucleotide at the end of the 5' HR will splice with the "aa" nucleotides (Figure 2.5). The presence of the GFP facilitates the detection of correct floxing of the domain/GOI.

The recodonised region is cloned between the first *loxP* sequence and HA-tag, using the *Bsm*I and *Bam*HI clone sites. This region's first amino acid sequence will therefore only encompass the two remaining nucleotides (Figure 2.5). Upon cloning, this region is flanked by the *loxP* sequences.

Finally, the 3' UTRs act as the 3' HR (minimum 300 bps in length) and are amplified from genomic DNA with corresponding primers. This region was cloned by way of the *Xba*I and *Hind*III cloning sites on pCS22_DiCre.

The cloning approach described here is used to accommodate the use of CRISPR-Cas9-mediated genetic modification, furthermore, by including homology regions this can accommodate SLI-mediated genomic modification. Guide RNA (gRNA) candidates were searched for based on the criteria mentioned in section 2.6.1, and specifically chosen to fall within this recodonised gene portion as previously discussed.

2.5. Generating gene-specific pCS22_DiCre plasmid

Benchling was utilized to create primers designed for the amplification of the relevant gene inserts, considering the unique features required for the cloning of the respective gene inserts, including the desired complementary restriction overhang sequences and several preparation steps (Figure 2.5).

These complementary sequences were artificially introduced by PCR with primers containing restriction recognition sequences for the 5' HR (*Eco*RI, 5' G/AATTC 3'), recodonised gene insert (*Bam*HI 5' G/GATCC 3') and 3' HR (*Xba*I, 5' T/CTAGA 3' and *Hind*III, 5' A/AGCTT 3') (Figure 2.5). A 4-nucleotide overhang was also introduced on the 5' end of each primer to allow for the binding and cleavage of the sequence by the restriction enzyme. The respective gene inserts with the desired overhang sequence required to generate gene-specific pCS22_DiCre repair plasmids, were cloned through complementary sticky end ligation using the restriction sites available on this plasmid.

2.5.1. *In vitro P. falciparum* parasite cultivation

2.5.1.1. Ethical clearance statement

This study was performed and approved under the Faculty of Natural and Agricultural Sciences Ethics Committee (ethics approval number: 297/2022). All experiments were carried out at the Malaria Parasite Molecular Laboratory (M2PL) biosafety level 2 facility (BSL2; registration number: 39.2/University of Pretoria-19/160), in accordance with the relevant guidelines and regulations. The *in vitro* cultivation of malaria-causing parasites, *P. falciparum*, in human erythrocytes was approved by umbrella ethics for the SARChI program under Prof. Birkholtz by the Faculty of Natural and Agricultural Sciences Ethics Committee (ethics approval number: 180000094), and additionally by the Faculty of Health Sciences Ethics Committee (ethics approval number: 506/2018).

2.5.1.2. *In vitro* asexual parasite cultivation

Asexual drug-sensitive *P. falciparum* (NF54; MRA-1000, BEI resources) parasites were cultured and maintained at ~5 % haematocrit of human erythrocytes (varying blood types), with daily replenishment of complete culture media (RPMI-1640 culture medium (Sigma-Aldrich, USA); supplemented with 0.024 mg/mL gentamycin (HyClone, USA), with an additional 0.2 % (w/v) glucose (Merck, Germany), 200 µM hypoxanthine (Sigma-Aldrich, USA), 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES; Sigma-Aldrich, USA), 23.81 mM sodium bicarbonate (Sigma-Aldrich), and then completed with 5 g/L Albumax II (a human serum albumin replacement, Life technologies, USA) [75]. Intraerythrocytic *P. falciparum* cultures were incubated at 37 °C with shaking on a rotary platform at 60 rpm to ensure maximal merozoite invasion, maintaining the parasites in a hypoxic environment consisting of 5 % O_2, 5 % CO_2 and 90 % N_2 (Afrox, South Africa). The percentage of parasite-infected erythrocytes (parasitaemia) were determined once daily through optical microscopy (1000 x magnification; Nikon, Japan) of Rapi-Diff-stained (Merck, South Africa) blood smears. A parasitaemia of ~4 % was maintained until the desired stage was established to isolate genomic DNA for downstream cloning uses.

2.5.1.3. Total *P. falciparum* DNA isolation

The commercially available kit Quick DNA Miniprep Kit (Zymo Research, USA) including its proprietary buffers, was used to isolate total DNA from *P. falciparum* NF54 asexual parasite cultures (>5 % parasitaemia, >50 % schizont population, see section 2.5.1.2). The buffers supplied with this kit contain high concentrations of chaotropic salts, together with a silica-based membrane, that collectively allows the selective binding of the whole DNA from a packed parasite-infected blood sample (pelleted through centrifugation at 3200 x *g* for 3 min), using the kit as per the manufacturer's instructions. Isolated DNA was quantified using a NanoDrop One^C Spectrophotometer as stated before. DNA samples with readings of 1.8 – 2.1 for both the A_{260}/A_{280} and A_{260}/A_{230} ratios, were regarded as pure and used in downstream cloning steps. All samples were stored at -20 ºC. This isolated DNA proves critical to perform downstream cloning steps, acted as template for PCRs amplification of the 5' and 3' homology regions (HRs) for *gcn5* and *set7* (section 2.5.2).

2.5.2. Optimisation and amplification of gene segment inserts

Sequences for all primers used in this section are listed under Supplementary Table 1. All primers were manufactured by Inqaba Biotechnical Industries (Pty) Ltd (South Africa) and dissolved in sterile dddH_2O according to the instructions provided. Primer primary stocks and aliquots were stored at -20 ºC.

2.5.2.1. 5' Homology region amplification for *gcn5* and *set7*

Cloning of the 5' HRs (~524 bp and ~416 bp for *gcn5* and *set7*, respectively) was completed by way of complementary restriction cut site overhangs and subsequent directional ligation with the universal

plasmid (pCS22_DiCre) (Figure 2.5). Forward and reverse primers (see Supplementary Table 1) used to amplify these gene regions, were designed using Benchling. The same software was also applied to artificially introduce the desired complementary restriction cut site overhang, which includes *Eco*RI (5' G/AATTC 3') to the forward including a 4-nucleotide overhang to allow for the binding of the restriction enzyme to the PCR product. The primer sequences were optimized for the melting temperatures, adapting it for larger ΔG values to avoid monomerization and/or the formation of secondary structures including hetero- or homodimers. Critically, no overhangs were added to the reverse primers which is crucial for downstream blunt-sticky end ligation of the PCR with the target vector, pCS22_DiCre. Importantly, only the pCS22_DiCre plasmid was digested with both *Eco*RI and *Sna*BI (5' T/CTAGA 3') as depicted in (Figure 2.5).

PCR products were obtained by setting up an amplification reaction using 30 – 100 ng of isolated DNA, 5 pmol final concentration of each primer (forward and reverse) and finally the polymerase 1× KAPA Taq Ready Mix [specifications are KAPA Taq DNA Polymerase (0.5 U/25 µL), proprietary KAPA Taq buffer supplemented with $MgCl_2$ (1.5 mM), dNTPs (0.2 mM each) and proprietary stabilizers; KAPA Biosystems, Wilmington, USA], in the 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). Previously optimized thermal cycling steps included an initial denaturation step (95 ºC, 3 min), 30 - 35 cycles including denaturation (95 ºC, 30 s), annealing (optimized at 48 ºC and 50 ºC for *gcn5* and *set7* respectively, 30 s), extension (68 ºC, 1 min), with a final extension step of (68 ºC, 2 min).

PCR reactions were loaded on a 1.2 - 1.5 % (w/v) agarose/TAE gel (prepared as before), loading a 100 bp DNA ladder (Thermo Fisher, USA) with 6× TriTrack DNA Loading and the samples with 6× purple loading dye (as before). Bands of the desired size were visualized by UV visualization, using the Gel Doc XR Imaging system (Bio-Rad, USA) and corresponding imaging software as previously stated.

Bands of the desired size were excised from the agarose gel and subsequently purified with the NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer instructions, and spectrophotometrically quantified, as previously stated.

Next, the PCR product was blunted using T4 DNA Polymerase (New England Biolabs, UK) with the required 1× CutSmart buffer (New England Biolabs, UK) and supplementing the reaction with 1 mM dNTPs, as per the manufacturer's instructions. The blunting reaction was incubated at 12 ºC for 15 min, and then immediately column purified (NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer's instructions). In parallel, ~1 µg of pCS22_DiCre plasmid was digested with 10 U of *Eco*RI-HF and *Sna*BI (New England Biolabs, UK), including 1× CutSmart buffer (New England Biolabs, UK), for 3 h at 37 ºC using 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). The PCR product and backbone was then column purified from the solution (NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer's instructions), and stored at -20 ºC or immediately ligated as described in section 2.5.3 below.

2.5.2.2. Recodonised gene segment amplification for *gcn5* and *set7*

Cloning of the recodonised gene portions (~1031 bp and ~1262 bp for *gcn5* and *set7*, respectively) relied on purchasing the segment of nucleotide sequences (GeneUniversal, USA) including the domain/s of interest (i.e. GNAT and bromodomain for *gcn5*, and SET domain for *set7*). This stretch of nucleotide sequences was recodonised by excluding any introns and the native stop codon, and mutating the PAM motif where possible, by using the free recodonisation services available on GeneScript's website (https://www.genscript.com/tools/gensmart-codon-optimization).

This stretch of recodonised nucleotides will not completely match the native genomic DNA sequence, thus preventing the continuous cleaving of the dsDNA at the site of recognition by the gRNA. As seen in Figure 2.5, the cloning was completed by way of complementary restriction cut site overhangs and subsequent directional ligation. Forward and reverse primers (see Supplementary Table 1) used to amplify these gene regions from the pUC18 plasmid, were designed using Benchling and used to artificially introduce the desired complementary restriction cut site overhang, which includes *Bam*HI (5' G/AATTC 3') to the forward including a 4-nucleotide overhang to allow for the binding of the restriction enzyme to the PCR product, critically no overhangs are added to the reverse primers. Importantly, only the pCS22_DiCre plasmid was digested with both *Bsm*I and *Bam*HI (5' GAATGCN/ 3' and 3' C/GATTC 5') as depicted in Figure 2.3. The primer sequence was optimized as previously stated.

Recodonised gene segments were manufactured by Gene Universal Inc (USA) (pUC18_*gcn5* and pUC18_*set7*) and dissolved in sterile dddH_2O, according to the instructions provided. Primary stocks and aliquots were stored at -20 ºC. Some of the plasmid (10 ng) was transformed with previously prepared $CaCl_2$-heatshock competent cells, as previously stated. Colonies were screened with M13 F and recodonised specific reverse primers. Positive colonies were grown overnight in 5 mL of 1 × LB-amp medium; 2 mL of cells from the overnight culture, containing the plasmid DNA, were snap-frozen in cells with 70 % (v/v) glycerol. The remaining 3 mL of bacterial culture were used to isolate plasmid using the plasmid DNA isolation kit NucleoSpin Plasmid (Machery Nagel GmbH & Co.KG, Germany) according to the manufacturer's instructions, quantified spectrophotometrically, and stored at -20 °C until use.

PCR products were obtained by setting up an amplification reaction using 50 ng of isolated plasmid, 5 pmol final concentration of each primer and finally the polymerase mixture 2× Phusion® High-Fidelity PCR Master Mix with HF Buffer [specifications are Phusion DNA Polymerase (0.4 U/20 µL), proprietary 5× Phusion HF Buffer supplemented with $MgCl_2$ (1.5 mM), dNTPs (0.2 mM each) and proprietary stabilizers; New England Biolabs, UK], in the 2720 Thermo Cycler. Thermal cycling steps were optimized for the following steps: initial denaturation step (98 ºC, 30 s), 25 - 30 cycles including denaturation (98 ºC, 10 s), annealing (optimized at 58 ºC for both *gcn5* and *set7*, 5 s), extension (68 ºC, 1 min), with a final extension step (68 ºC, 5 min).

PCR reactions were visualized on a 0.8 % (w/v) agarose/TAE gel (prepared as before). Bands of the desired sizes identified, as previously described in section above, were excised from the agarose gel and column purified with the NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer instructions, and eluted samples were quantified by spectrophotometry and stored at -20 ºC.

Next, ~1 µg the backbone of pCS22_DiCre plasmid was digested with 10 U of *Bsm*I (New England Biolabs, UK), including 1 × r2.1 buffer (10 mM magnesium chloride, 10 mM Tris-hydrochloride, 50 mM sodium chloride and 100 µg/mL Recombinant Albumin, pH 7.9; New England Biolabs, UK), for 3 h at 65 ºC using 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). Next, the *Bsm*I digested plasmid was immediately blunted using T4 DNA Polymerase (New England Biolabs, UK). To the digestion reaction (*Bsm*I), an additional buffer (1× rCutsmart; New England Biolabs) was added to accommodate the manufacturer's instructions for T4 DNA Polymerase, supplementing the reaction with 1 mM dNTPs. The blunting reaction was incubated at 12 ºC for 15 min, and then immediately column purified (NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer's instructions). The eluted sample was then incubated with 2 U of Shrimp Alkaline Phosphatase (rSAP, New England Biolabs, UK) and the rCutSmart buffer (New England Biolabs) at 37 ºC for 30 min, as per the manufacturer's instructions. This prevents the re-circularisation of the digested vector backbone by way of dephosphorylation of the 5' termini of both ends of the vector. The reaction was heat inactivated at 80 ºC for 5 min, and immediately column purified, as per usual. The eluted sample was then immediately digested with the final digestion enzyme, 10 U *Bam*HI-HF (New England Biolabs) by adding rCutSmart buffer to the sample and incubating at 37 ºC for 3 h. This was done in parallel with the digestion of the recodonised region PCR product with 10 U *Bam*HI, under the same conditions. Both digestion reactions were column purified as usual, and the final concentrations were determined spectrophotometrically.

2.5.2.3. 3' Homology region amplification for *gcn5* and *set7*

Cloning of the 3' HR gene portions (~604 bp and ~577 bp for *gcn5* and *set7* respectively) was completed by way of complementary restriction cut site overhangs and subsequent directional ligation with the universal plasmid (pCS22_DiCre) (Figure 2.5). Forward and reverse primers used to amplify these gene regions, were designed using Benchling. The same software was also applied to artificially introduce the desired complementary restriction cut site overhang, which includes *Xba*I (5' G/AATTC 3') to the forward and *Hind*III (5' A/AGCTT 3') to the reverse primer, including a 4-nucleotide overhang and optimizing the primer sequence as previously discussed.

PCR products were obtained in a reaction using 50 – 200 ng of isolated DNA, 5 - 10 pmol final concentration of each primer (forward and reverse, see Supplementary Table 1) and Takara Ex Taq® Polymerase (Taq Polymerase (5 U/µL; Takara Bio Inc.), including the proprietary 10× ExTaq Buffer (supplemented with Mg^{2+} (1.5 mM each) Takara Bio Inc.), and further supplementing the reaction with dNTPs (2.5 mM; Takara Bio Inc). The reaction was run in the 2720 Thermo Cycler (Applied Biosystems,

Foster City, USA). The thermal cycling steps were optimized due to continuous low band intensity, which included an initial denaturation step (95 ºC, 3 min), 30 - 35 cycles including denaturation (95 ºC, 30 s), annealing (optimized at 50 ºC for both *gcn5* and *set7*, 30 s), extension (68 ºC, 1 min), with a final extension step (68 ºC, 2 min).

PCR reactions were loaded on a 1.2 – 1.5 % (w/v) agarose/TAE gel (prepared as before) using the setup and analysing steps as mentioned above. Bands of the desired size were excised from the agarose gel and purified as previously stated.

Next, both the isolated PCR product and ~1 µg of pCS22_DiCre plasmid were digested with 10 U of both *Xba*I *and Hind*III-HF (New England Biolabs, UK) in 1× r2.1 buffer (New England Biolabs, UK), for 3 h at 37 ºC using 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). The PCR product and backbone was then isolated from the solution, using the NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer instructions, and eluted samples were quantified by spectrophotometry and stored at -20 ºC or immediately ligated.

2.5.3. Ligation of gene inserts

All ligation reactions were set up with a standard 10 µL, including 400 U T4 DNA Ligase and 1× T4 DNA Ligase buffer and using 50 ng of the pCS22_DiCre plasmid, digested with the respective restriction enzymes. These enzymes include 10 U of both *Eco*RI and *Sna*BI for the 5’ HR, 10 U of both *Bsm*I and *Bam*HI for the recodonised gene insert and 10 U of both *Xba*I and *Hind*III for the 3’ HR insert. The respective PCR products (as prepared in sections 2.4.3.1, 2.4.3.2 and 2.4.3.3, respectively) were used in each of the respective ligation reaction. Using the steps previously described in section 2.1.2 and the formula in section 2.1.2 (Equation 1), the amount of insert of each gene segment was determined for a 3:1 ligation reaction. Competent *E. coli* DH5α cells were transformed with the respective ligation reactions, as previously stated in section 2.1.2.

2.5.4. Screening of positive colonies containing desired insert

Screening of positive colonies for the relevant gene inserts were completed by way of PCR screening, isolation of plasmid and subsequent sequencing, as described in section 2.5.2. Positive colonies were screened by way of PCR using Kapa Taq Polymerase, with conditions as previously described. The primer pair for the PCR screenings are summarised in Supplementary Table 2. A different annealing temperature for the thermal cycling was optimized at 58 ºC for all reactions, while the extension step was set at 68 ºC for 2 min, and final extension at 68 ºC for 5 min for all the reactions. Positive colonies were inoculated overnight in 10 mL 1× LB-amp media with shaking, whereafter the plasmid was isolated using the plasmid DNA isolation kit NucleoSpin Plasmid (Machery Nagel GmbH & Co.KG, Germany) according to the manufacturer’s instructions, quantified spectrophotometrically, and stored at -20 °C until use.

Isolated plasmids were then sequenced using the Sanger sequencing system. The sequencing reaction was setup as mentioned in section 2.2.8. Primers used include a combination of SP1, SP2, SP5, and SP6

(5 pmol) and gene specific primers (see Supplementary Table 1 and 2). The thermal cycling conditions were set up as before, with annealing temperatures all at 58 ºC for 5 s. Reactions were ethanol precipitated and submitted for sequencing to the Sanger Sequencing facilities at the Centre for Bioinformatics and Computational. The results were alignment and analysis of sequencing reactions were carried out with the alignment tool on Benchling.

2.6. Cloning strategy to generate Cas9-gRNA plasmid construct

Cloning of the double-stranded gRNAs, as depicted in Figure 2.6, was completed by way of complementary restriction cut site overhangs and subsequent directional ligation. To introduce these gRNAs into the pDC2-coCas9-gRNA-hDHFR plasmid, the annealed complementary gRNAs were then ligated with the digested (*Bbs*I) pDC2-coCas9-gRNA-hDHFR plasmid, which allowed for the directional cloning of the gRNA into the now final gRNA containing Cas9 plasmid.

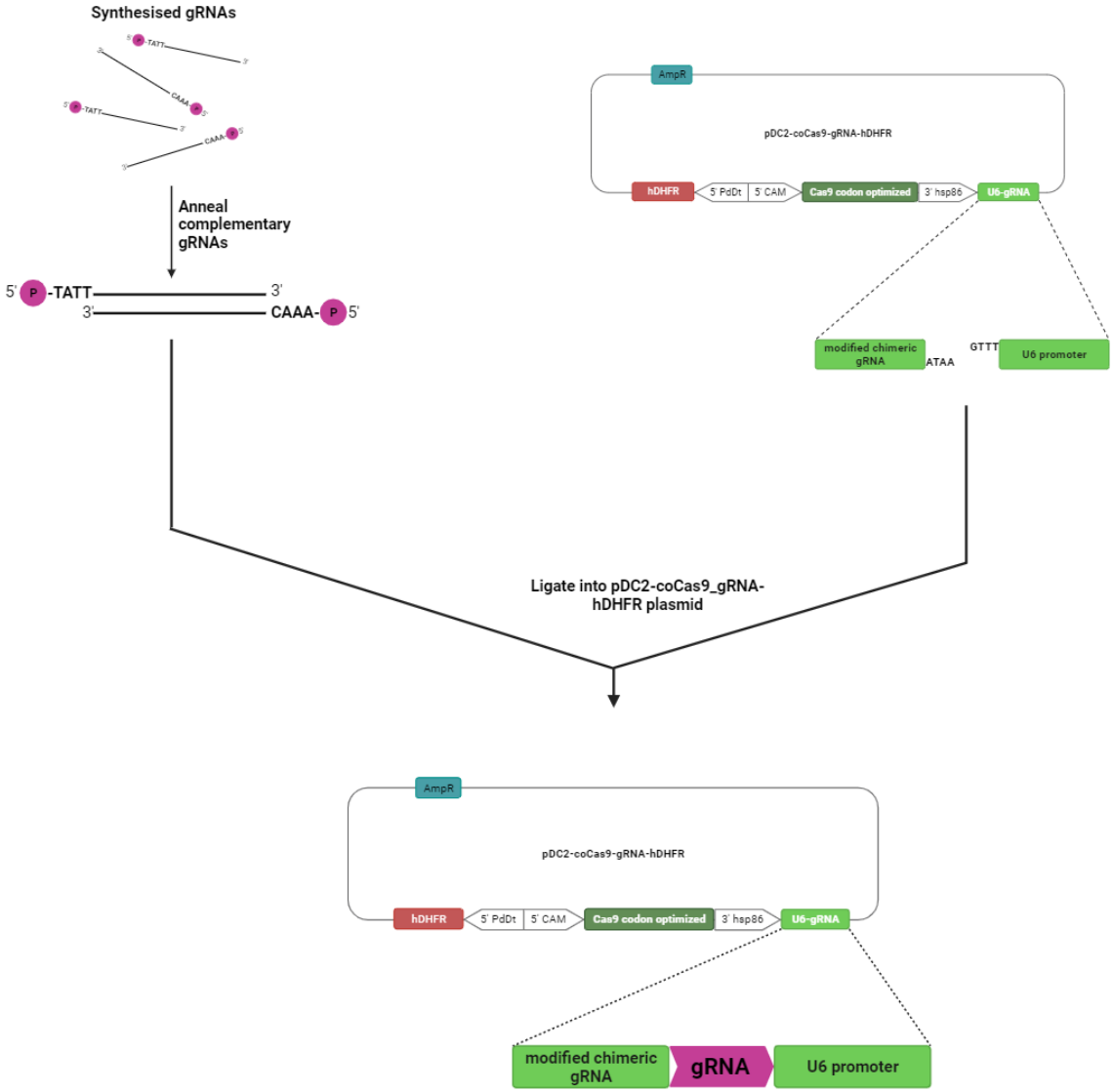


Figure 2.6: Schematic representation of the cloning strategy employed to generate the pDC2-coCas9-gRNA-hDHFR plasmid. The plasmid containing the double stranded gRNA insert for each of the gene candidates (*gcn5* and *set7*) was generated by annealing two complimentary pre-phosphorylated gRNAs, with the desired *Bbs*I cut site overhangs, then ligated into the pDC2-coCas9-gRNA-hDHFR plasmid with complementary overhangs, as indicated. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

2.6.1. Design of gRNAs

Possible sequences that may act as the guide for the CRISPR-Cas9 complex *in vitro,* was generated using Benchling. Parameters were set for single guide RNA (gRNA) with a base length of ~20 bp, for the codon optimized Cas9, a protospacer-adjacent motif (PAM) sequence of 5'-NGG 3' [64] and using the nucleotide sequences (available on PlasmoDB) for both *gcn5* and *set7* respectively.

Candidate gRNAs were filtered based on their proximity (within 200 bps) to either the 5' or 3' ends of the recodonised gene portion (described in section 2.2), the relative on- and off-target score (50≤x≤100) which represents the efficiency of Cas9 cleavage and the inverse probability of Cas9 binding off target [76], excluding any with the presence of A and T stretches (<5) which could lead to off-target binding, and the percentage of G+C nucleotides (~40 %) within this 20 bp nucleotide sequence. Importantly, the final exclusion category for each gRNA candidate was to evaluate for any non-specific target recognition and binding, by interrogating the e-value of each through a BLAST search function (under "Tools" on PlasmoDB) with the multi-query sequence alignment selecting for specificity to the *Plasmodium falciparum* NF54 reference genome. Any gRNA candidate with an e-value greater than 1 (>85 % similarity) to any off-target sequence were not used.

Two gRNA candidates were chosen per gene. In the case that there are no guanine nucleotide at the start of the ~20 bp gRNA hit sequence, one was added. Although not distinctively proven in *P. falciparum* it has been successfully applied in other species due to the preference of RNA Polymerase III during transcription initiation [64].

No gRNA sequences were excluded based on their directionality (sense or antisense strand hits) but was rather adapted to include the relative restriction cut site (*Bbs*I) overhang complementary to those generated on the pDC2-coCas9co-Cas9-gRNA-hDHFR plasmid (see Figure 2.6), with the forward overhang sequence (5' TATT- 3') allocated to the hit gRNA sequence (as provided by Benchling) and the reverse complement sequence of the hit gRNA was allocated the reverse overhang sequence (5' AAAC-3'). This is critical as feasibility of the gRNA is dependent on the directionality relative to the U6 promoter. These overhangs were artificially introduced on Benchling after the two gRNA candidates were chosen.

Single stranded phosphorylated gRNAs (oligos) were manufactured by Inqaba Biotechnical Industries (Pty) Ltd (South Africa) and dissolved in sterile dddH_2O according to the instructions provided. The primary stock and aliquots were stored at -20 ºC.

2.6.2. Cloning of gRNAs into pDC2-coCas9-gRNA-hDHFR

Restriction enzyme digestion was used to further confirm the identity of the isolated plasmid DNA by digesting 0.5 µg of isolated plasmid (pDC2-coCas9-gRNA-hDHFR) with 10 U each of *Eco*RI-HF, *Hind*III-

HF and *Bam*HI-HF in different combinations, including *Eco*RI only, *Eco*RI and *Hind*III, and *Eco*RI and *Bam*HI (as previously stated in section 2.1.3). The digestions were visualised on a 0.8 % (w/v) agarose/TAE gel (prepared as stated before) to identify plasmid fragments with corresponding sizes of the backbone.

Restriction enzyme digestion was used to further confirm the identity of the isolated plasmid DNA by digesting 0.5 µg of isolated plasmid (pDC2-coCas9-gRNA-hDHFR) with 10 U each of *Eco*RI-HF, *Hind*III-HF and *Bam*HI-HF in different combinations, including *Eco*RI only, *Eco*RI and *Hind*III, and *Eco*RI and *Bam*HI. The digestions were incubated with 1x 2.1 buffer, for 3 h at 37 ºC using 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). The digested plasmid was analysed on a 0.8 % (w/v) agarose/TAE gel (prepared as stated before), loading the samples and a 1 kb DNA ladder (Thermo Fisher, USA) using 6× TriTrack DNA Loading Dye to identify plasmid fragments with corresponding sizes of the backbone.

The complementary oligos was annealed after denaturing any secondary structures of the reconstituted gRNA oligos (100 µM of each) on a heating block set to 95 ºC for 5 min, thereafter allowing the oligos to anneal to their complementary oligo by gradually cooling the sample to room temperature overnight. The annealed oligos were then stored at -20 ºC.

Next, the gRNA plasmid was prepared by digesting ~5 µg of pDC2-coCas9-gRNA-hDHFR, purified as previously stated from transformed $CaCl_2$-heatshock competent *E. coli* DH5α bacteria cells, with 6 U of *Bbs*I (New England Biolabs, UK) with and 1× r2.1 buffer (10 mM $MgCl_2$, 10 mM Tris-Hydrochloride, 50 mM NaCl and 100 µg/mL Recombinant Albumin, pH 7.9; New England Biolabs, UK), for 3 h at 37 ºC using 2720 Thermo Cycler (Applied Biosystems, Foster City, USA). The backbone was then isolated using the NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel GmbH & Co.KG, Germany) as per the manufacturer, and stored at -20 ºC.

Ligation was accomplished by incubating ~50 ng digested vector backbone and 1 µL of previously annealed gRNA oligos (diluted 1:100 in $dddH_2O$), with T4 DNA Ligase and 10× T4 ligation buffer (1 mM ATP, 10 mM Dithiothreitol, 10 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5; New England Biolabs, UK). The reaction was incubated at room temperature overnight and then placed at 4 ºC until the reaction could be transformed with heatshock competent *E. coli* DH5α cells. The gRNA constructs were then transformed with $CaCl_2$-heatshock competent bacteria, previously prepared, as previously discussed in section 2.1.2.

2.6.3. Screening of positive clones containing gRNA insert

Positive colonies were screened using PCR, with conditions as previously described. PCR reactions were visualized on a 1.5 % (w/v) agarose/TAE gel (prepared as before). The primer pair for this PCR screening included a plasmid backbone specific p35 reverse and gene specific gRNA primer (in Table 5 below).

Table 5: Primers used to screen for positive bacterial colonies containing gRNA insert. Nucleotide overhangs are indicated in lowercase, while gene specific sequences are all uppercase. The bold nucleotide references the added Guanine nucleotide. The screening and sequencing primer is backbone specific.

Purpose	Primer code	Oligo sequence (5' – 3')		Length of amplicon
Colony screening forward	*gcn5*_gRNA1	F	tattTGAATGTATAACGAACGATA	~110 bp
	*gcn5*_gRNA2	F	tatt**G**ttagGGTGATTATAAAACGA	
	*set7*_gRNA1	F	tattGCTGCATAATGCCATTTAAG	
	*set7*_gRNA2	F	tattCTCAATTTGAGGAGCATCAG	
Colony screening & sequencing reverse	p35	R	AAGCACCGACTCGGTGCCAC	

A different annealing temperature for the thermal cycling was optimized at 50 ºC, while the extension step was set at 68 ºC for 1 min, and final extension at 68 ºC for 2 min. Positive colonies (band with size ~112 bp) were grown overnight in 10 mL 1× LB-amp media whereafter the plasmid was isolated using the plasmid DNA isolation kit NucleoSpin Plasmid (Machery Nagel GmbH & Co.KG, Germany) according to the manufacturer's instructions, quantified spectrophotometrically, and stored at -20 °C until use.

Isolated plasmids were then sequenced using the Sanger sequencing system as described in section 2.1.3. using the reverse primer (p35) specific to the pDC2-coCas9 backbone, with annealing temperature optimized at 58 ºC. The EtOH precipitated samples were then submitted for sequencing reactions performed by Sanger Sequencing facilities at the Centre for Bioinformatics and Computational Biology, as previously stated. Alignment and analysis of sequencing reactions were carried out on the pDC2-coCas9-gRNA-hDHFR containing the sequence of the respective gRNA sequences, i.e. gRNA1 and 2 for *gcn5* and gRNA1 and 2 for *set7*, with the alignment tool on Benchling.

3. Results

3.1. Cloning of pCS22_DiCre plasmid

The pCS22_DiCre plasmid was created from the pSN28 pUC19 loxPint plasmid as core and pH33 FIKK10.1cKOGFP_Cas9 as donor of the selection cassette. Both these plasmids were successfully isolated and purified from competent *E. coli* DH5α and characterised by restriction mapping (Figure 3.1). The relevant concentrations for the isolated plasmids after purification ranged from 313 - 330 ng/µL while purity was in the range of 1.91 – 1.99 and 1.70 – 2.11 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, indicating pure DNA suitable for downstream use.

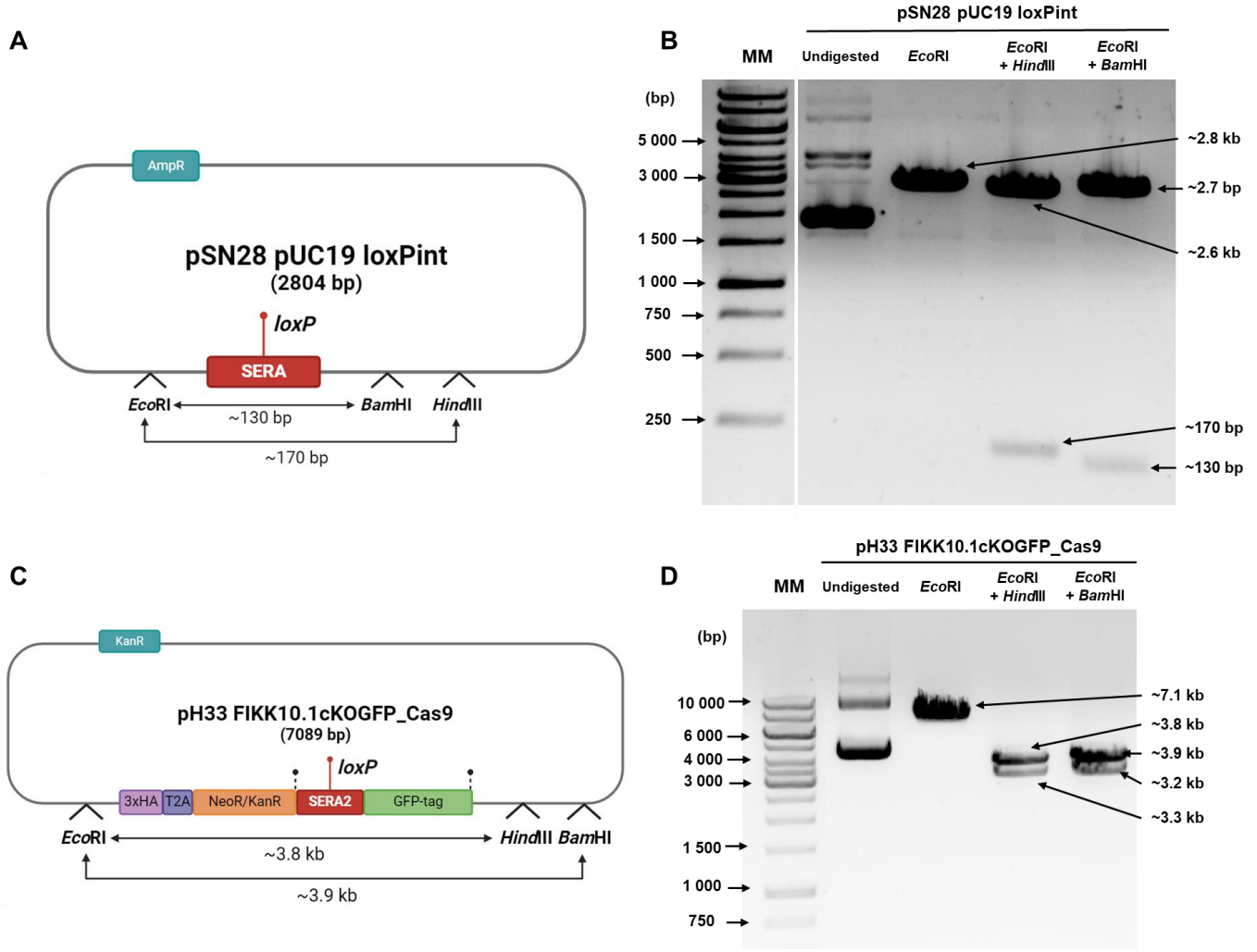


Figure 3.1: Restriction enzyme digestion of the starting plasmids for pCS22_DiCre. (A) Schematic representation of the expected band sizes for the digestion of the pSN28 pUC19 loxPint plasmid, and (B) validation of successful isolation of the plasmid through restriction enzyme mapping. (C) Schematic representation of the expected band sizes for the digestion of the pH33 FIKK10.1cKOGFP_Cas9 and (D) validation of successful isolation through restriction enzyme mapping. Restriction enzyme mapping performed indicates the separation of plasmid DNA and 1 kb ladder (Lane MM) on a 1.0 and 0.8 % (w/v) agarose/TAE gel, respectively, and visualized with EtBr (1 µg/mL) pre-staining. Expected band sizes are indicated, no Rf-values were obtained. Digestion of three combinations of enzymes (*Eco*RI, *Hind*III and *Bam*HI) indicated the presence of the selection markers and the intact nature of at least three of the desired downstream clone sites (*Eco*RI, *Bam*HI and *Hind*III). Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

pSN28 was linearized with *Eco*RI-HF as expected, as a ~2.8 kb band was observed (Figure 3.1B). Furthermore, the bands observed for RE mapping with *Eco*RI-HF and *Hind*III-HF (~2.6 kb and ~170 bp) and *Eco*RI-HF and *Bam*HI-HF confirmed the expected sizes (~2.7 kb and ~130 bp) of the plasmid. This also further confirms the presence of the three cloning sites on the pSN28 plasmid for subsequent cloning steps. Similarly, the pH33 plasmid was linearized with *Eco*RI-HF, with the expected band size ~7.1 kb (Figure 3.1C) and was further characterised with the same RE mapping combination as for pSN28. The bands of expected sizes for *Eco*RI and *Hind*III (~3.8 and ~3.3 kb) and *Eco*RI and *Bam*HI (~3.9 and ~3.2 kb) were observed. This therefore confirmed the successful isolation of the pSN28 and pH33 plasmids, and pH33 were thus used for the cloning of the selection marker cassette. Undigested plasmids indicate the three natures in which plasmids are found, namely open circular (top), linear (middle) and supercoiled (bottom). Faint bands at ~1.5 kb across all digested plasmid samples could presents a contaminant or the partial digestion of the pSN28 plasmid. The amplification of this selection marker cassette was performed using Phusion® Polymerase with sequence-specific primers with the relevant restriction recognition site overhangs (Figure 3.2).

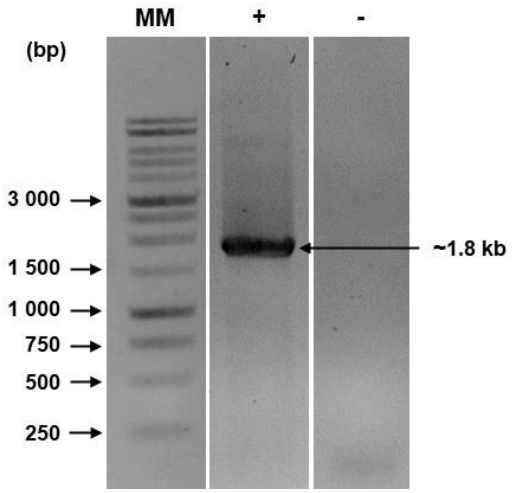


Figure 3.2: PCR amplification of the selection marker cassette from the pH33 plasmid. The cassette was amplified at an annealing temperature of 58 °C. The amplified DNA and 1 kb DNA ladder (Lane MM) were separated on a 1.0 % (w/v) agarose/TAE gel and visualized with EtBr (1 µg/mL) pre-staining. Expected band sizes are indicated, at different intensities for different amounts of DNA template. No Rf-values were obtained to define the magnitude of difference. The negative control (-) included the same amount of primer as per the normal setup, but without template.

The amplification product of ~1.8 kb indicates the identity of the amplicon, which was purified from the gel. The relevant concentrations for the PCR product before and after purification ranged from 43 – 222 ng/µL while purity was in the range of 1.87 – 1.97 and 1.86 – 2.23 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, which indicated pure, good quality DNA. Previously purified pSN28 plasmid and selection marker cassette PCR product was digested with *Bam*HI and *Xba*I, column-purified from the reaction solution and then ligated.

Following the transformation of the ligation reaction into competent *E. coli* DH5α cells, colony screening PCR was performed, using selection marker cassette specific reverse primers (SLC_R) and pCS22_DiCre backbone specific forward primer (M13 F), to identify bacterial clones containing recombinant pSN28, now

pCS22_DiCre, plasmid. Colonies that were screened identified only one positive clone (colony 5, Figure 3.3), containing the recombinant pSN28 plasmid.

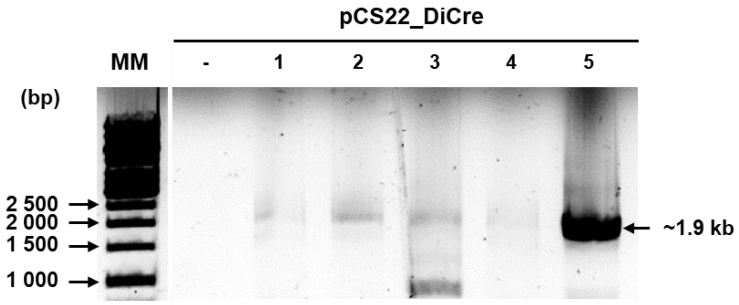


Figure 3.3: PCR analysis of screening of bacterial colonies positive for the selection marker cassette insert in pSN28, or pCS22_DiCre. Using selection marker specific primers and the M13 forward (pCS22_DiCre backbone-specific primer), amplified DNA and 1 kb ladder (Lane MM) was separated with a 0.8 % w/v agarose/TAE gel that contained EtBr (1 µg/mL) for visualization under UV light. Expected band sizes are indicated, no Rf-values were obtained. The negative control (-), primers with no template is indicated.

The positive clone was grown overnight in LB-amp media, followed by the isolation of the plasmid, obtaining a concentration of 392 ng/µL with purity ratios of 1.84 and 1.63 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, indicating good quality DNA. The isolated plasmid was then characterised through RE mapping (Figure 3.4).

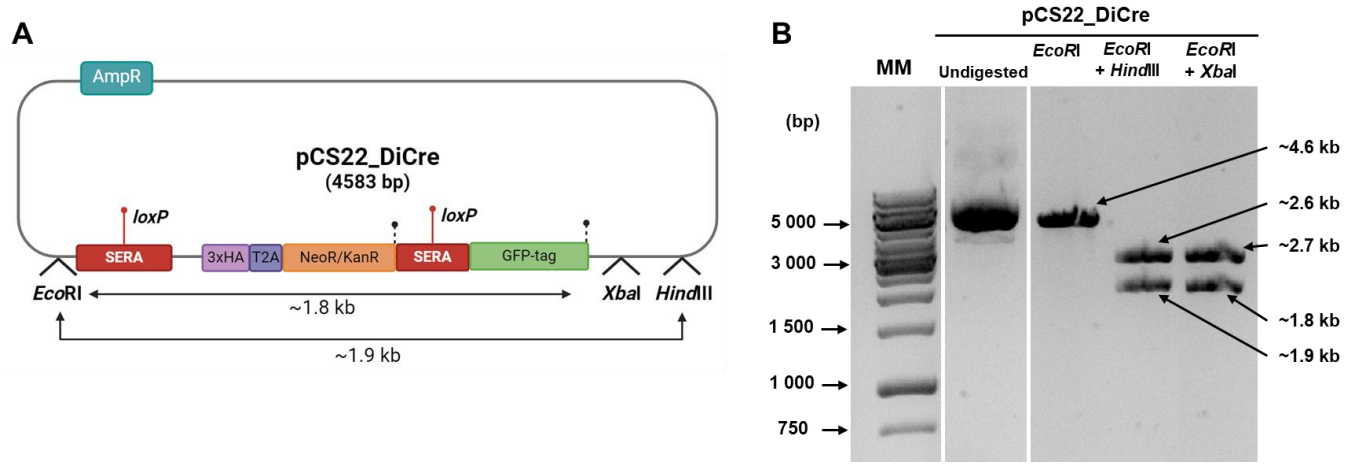


Figure 3.4: Validation of successful isolation of pCS22_DiCre. Restriction enzyme mapping performed indicates the separation of plasmid DNA and 1 kb ladder (Lane MM) on a 1.0 % (w/v) agarose/TAE gel, respectively, and visualized with EtBr (1 µg/mL) pre-staining. Expected band sizes are indicated, no Rf-values were obtained. Digestion of three combinations of enzymes (*Eco*RI, *Hind*III and *Xba*I) indicated the presence of the selection markers and the intact nature of at least three of the desired downstream clone sites (*Eco*RI, *Bam*HI and *Xba*I). Figure created with Biorender.com under the basic licence for educational purposes.

pCS22_DiCre plasmid was linearized with *Eco*RI-HF as expected, as a ~4.6 kb band was observed (Figure 3.4). Furthermore, the bands observed for RE mapping with *Eco*RI-HF and *Hind*III-HF (~2.6 kb and ~1.9 kb) and *Eco*RI-HF and *Xba*I-HF further confirmed the expected sizes (~2.7 kb and ~1.8 kb) of the plasmid. This provided further confirmation that the selection marker cassette (with a size of ~1.8 kb) was successfully cloned and indicated the presence of at least three of the cloning sites on the pCS22_DiCre plasmid required for subsequent cloning steps.

A final verification of the correct construction of the pCS22_DiCre plasmid was performed with Sanger dideoxy sequencing. The sequencing alignment in the chromatogram below (Figure 3.5), confirmed that the selection marker cassette was successfully cloned into pSN28 without any alterations to generate pCS22_DiCre, and further confirmed the retention and presence of the restriction enzyme recognition sites (*Bsm*I, *Bam*HI and *Xba*I and *Hind*III).

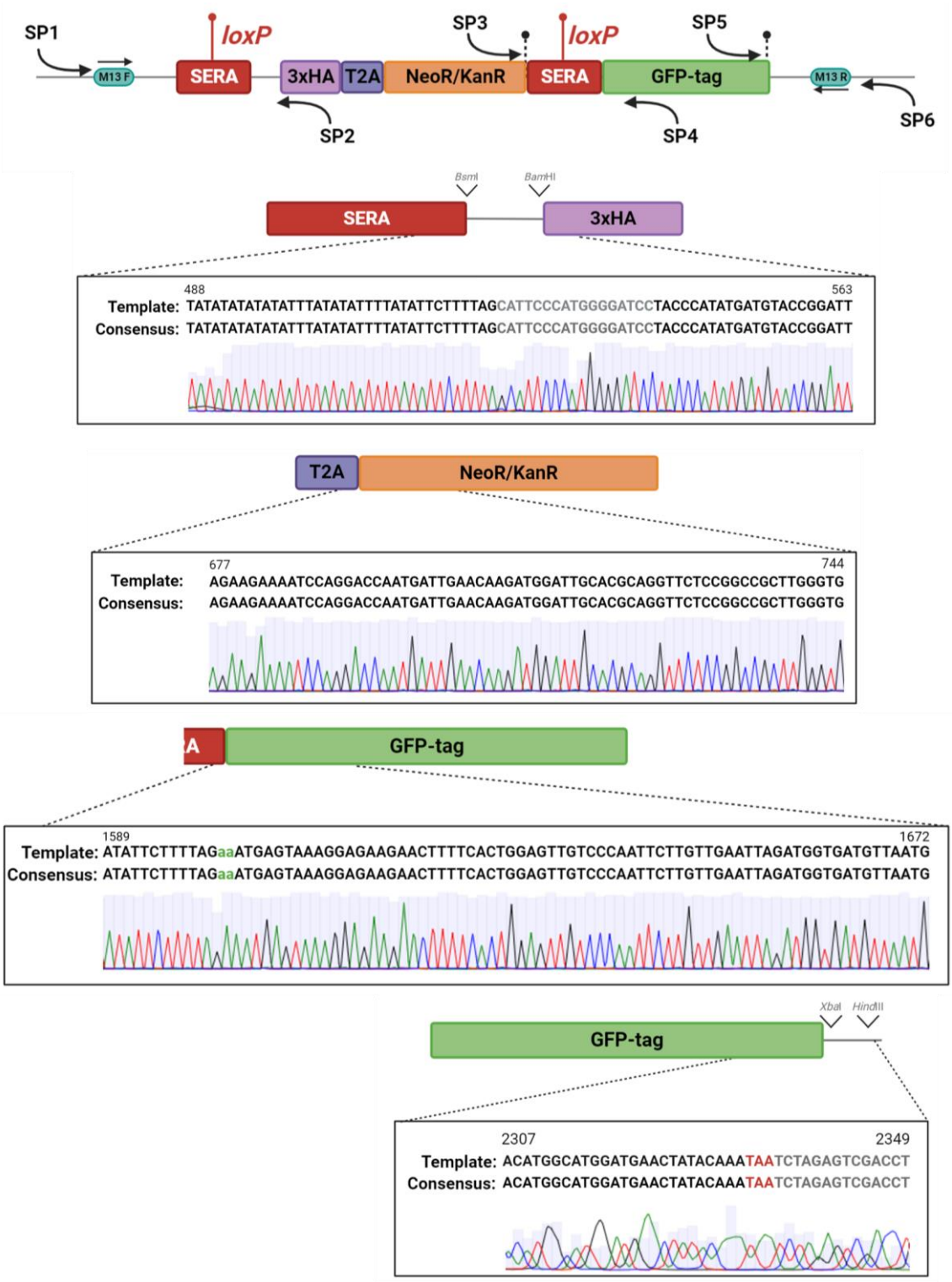


Figure 3.5: Sanger sequencing of selection cassette from pH33 cloned into pSN28, now pCS22_DiCre. Alignment of the consensus sequence to the expected reference sequence. The consensus sequence was obtained from analysis of a sequencing reactions obtained from either pCS22_DiCre backbone specific primers (M13 forward) or selection cassette-specific reverse primers indicated at the top of the figure. Restriction cut-sites are indicated in grey. Stop codons are indicated in red. A chromatogram of the forward or reverse trace data is represented with the

alignment sequences. Grey bars represent the confidence score out of 50 visually represented by the size of the bars, generated by the Benchling programme as a representation of basecalling. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

Notably, there are areas with lower confidence scores, such as the 3'-end of GFP at the *Xba*I/*Hind*III sites (bases 2307 to 2349), which is due to the proximity of the sequencing primers to this region. This is further supported by the much taller grey bars for the more upstream sequences for the GFP-tag. The read length for all the chromatograms here were 1000 bps or more, which could also contribute to the low resolution of the last bases for the GFP-tag since the overlap of sequences could not be better amplified with the available primers. Based on these results, the pCS22_DiCre construct was used in subsequent cloning. For the full consensus Sanger sequence alignment of pCS22_DiCre, see supplementary data 1.

3.2. *In vitro*, asexual *P. falciparum* parasite cultivation

P. falciparum NF54 asexual parasites were cultured, and the genomic DNA was isolated to use as template for downstream cloning of homology regions. Over 48 h (one developmental cycle), parasites were observed as per the normal stages (Figure 3.6) as seen in literature [19], starting with the merozoites (0 hours post infection, hpi) that bursts from the erythrocyte, subsequently infecting the next available erythrocyte, followed by a ring stage (from ~10 - 17 hpi), early trophozoite stage (~18 hpi) with the formation of the prominent hemozoin crystal being prominent and increasing in size as the parasite progressed through to the late-stage trophozoite (~30 - 37 hpi). Following the trophozoite stage, the parasite enters the final stage of the life cycle, the schizont stage, which occurs up until 48 hpi, after which the newly formed merozoites burst from the infected erythrocyte, for the cycle of infection to continue [19].

Figure 3.6: Asexual developmental stages of *P. falciparum* parasites, *in vitro*. Representation of the morphology of the three distinct stages including ring, trophozoite and schizont, of *P. falciparum* parasites. Rapi-Diff-stained thin blood smears were used to visualize parasites with optical microscopy at 1000x magnification. Hemozoin crystal is indicated with the arrow and individual merozoite within schizont indicated with arrowhead.

3.3. Total DNA isolation from asexual *P. falciparum* parasites

The total genomic DNA isolated from asexual *P. falciparum* parasites that were cultured *in vitro* until a parasitaemia of ~8 - 12 % was obtained (>50 % schizont population), were used in downstream PCR amplifications, with the DNA acting as the template to amplify the required gene segments. The concentrations and purity of each DNA sample was determined by spectrophotometric analysis, obtaining samples with concentrations ranging from 27 ng/µL – 114 ng/µL, while the purity was in the range of 1.67 – 2.06 and 0.91 – 1.95 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively.

3.4. Design and construction of transgenic repair plasmids

3.4.1. *In silico* design of repair plasmid for transgenic repair constructs

To obtain a clearer indication of the relative size of the genomic DNA sequence of *gcn5* (PF3D7_0823300) and *set7* (PF3D7_1115200) and the extent to which the gene sequence can be applied in this cloning system, an *in silico* analysis of the genes were performed (Figure 3.7). This included obtaining the nucleotide and amino acid sequence from PlasmoDB and using the amino acid sequence to evaluate the identity and to confirm the sequence of the domains, using InterPro. This identified the exact sequences for the two domains for *gcn5* (GNAT- and Bromodomain) and one for *set7* (SET domain) previously described in literature [38, 77-79]. However, the GNAT domain for *gcn5* has not been well-defined, which could make this domain an interesting domain to study. The relative proximity of the respective domains was near the C-terminus on the genomic sequence. This supported the use of all of the domains, as presented in Figure 3.7. This depicts the genomic sequence of the *gcn5* and *set7*, with the relevant exons, introns, and domains indicated. Given the relatively close proximity of the two domains of *gcn5*, it creates an ideal setting to apply the cloning strategy described in here. For *set7*, the proximity of the SET domain to the 3' UTR should allow effective cloning of the required gene inserts. Also presented in this figure is how the "out-of-frame" AG nucleotide pair was chosen to accommodate for the out of frame GFP downstream on the pCS22_DiCre plasmid, as applied by Davies *et al.* in their study [56].

The gene-specific recodonised region was synthesized by Gene Universal Inc (USA) and then successfully amplified from the supplied plasmid (pUC18_*gcn5* and pUC18_*set7*, respectively). This gene segment was recodonised from the native sequence to mutate the critical PAM motif or gRNA recognition sequence where possible, and to exclude any introns and the native stop codon.

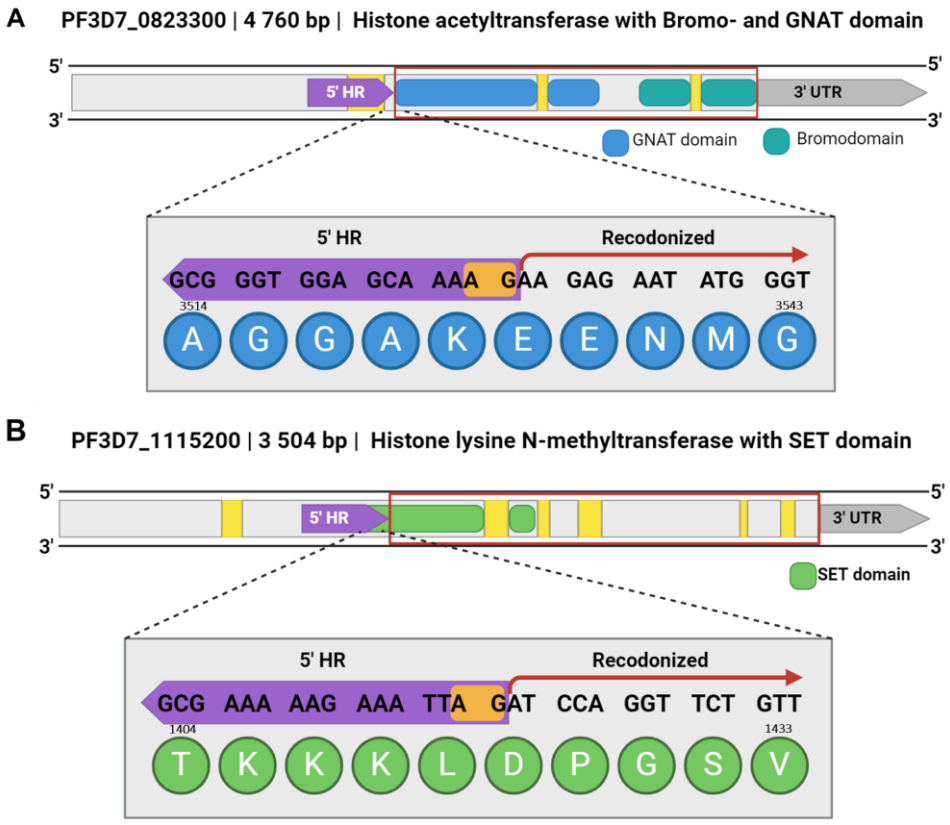


Figure 3.7: Depiction of the domains of (A) *gcn5* and (B) *set7* according to InterPro, and how the respective HRs were chosen accordingly. Lighter grey blocks indicate exons, with yellow blocks in between indicating introns. Domains for the respective genes are indicated with the relevant colours. Homology regions are indicated (5' HR - purple, and 3'HR – grey). Blocks in red indicate the recodonised gene region, chosen around the "out-of-frame" AG, indicated in yellow. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

Denaturation, annealing, and extension times and temperatures, as well as DNA template/primer concentration ratios, were optimized for PCR amplifications. PCR optimisation was performed for different polymerases due to low or no amplification, in instances such as the 3' HRs. After some optimisation, PCR products with the expected relative sizes were obtained, but there were also several faint bands noted for the 3' HR of *gcn5*, indicating non-specific amplification. To ensure no non-specific amplification PCR products were used in downstream cloning steps, all bands of the expected size were purified through gel extraction, with concentrations ranging from 3.0 – 86 ng/µL while purity was in the range of 1.49 – 2.15 and 0.99 – 2.01 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively. Given the continuous low yield of PCR products, specifically for the 5' HR of *set7* and 3' HR of *gcn5*, alternative amplification approaches were tested, as mentioned above.

3.4.2. Generating a recombinant pCS22_DiCre plasmid

3.4.2.1. Cloning a 5' gene fragment of *gcn5* and *set7* into pCS22_DiCre

The 5' gene fragment (or 5' HR) of both gene fragments were amplified from isolated genomic DNA from *in vitro* cultured *P. falciparum* NF54 parasites (as per section 3.3). The PCR conditions were investigated for amplification using gene-specific primers (Supplementary Table 1), aiming to identify the optimal

annealing temperature for maximum product amplification, based on the visual intensity after EtBr staining (Figure 3.8).

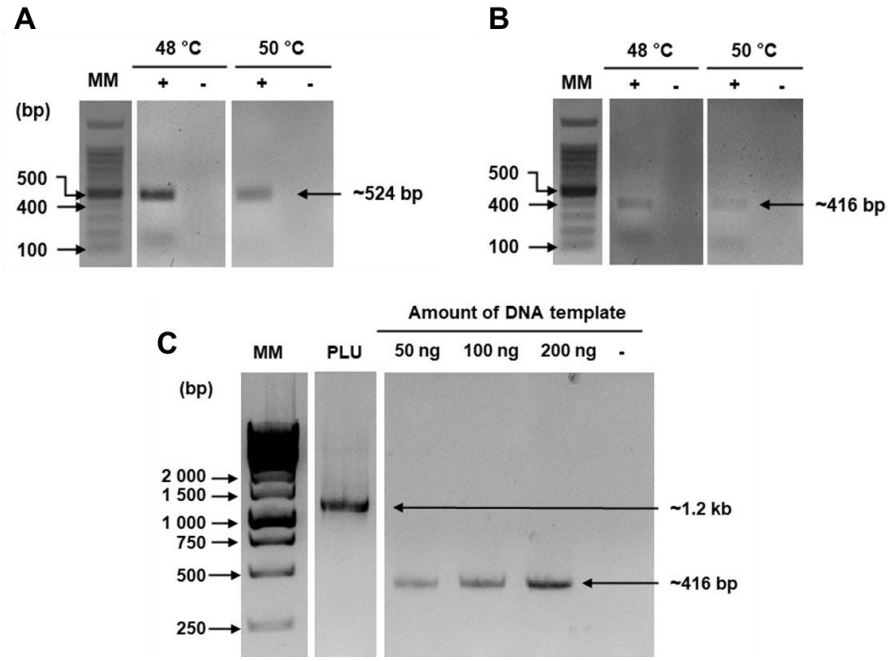


Figure 3.8: PCR amplification of the 5' HR of (A) *gcn5* and (B) *set7* at different annealing temperatures (48 °C and 50 °C), and (C) *set7* under different DNA template amounts. The amplified DNA and 100 bp and 1 kb DNA ladder (Lane MM) were separated with a 1.5 % (w/v) agarose/TAE gel and visualized with EtBr (1 µg/mL) pre-staining. Expected band sizes are indicated, at different intensities for different amount of DNA template. No Rf-values were obtained to define the magnitude in difference. All amplifications were performed with KAPA Taq Polymerase using the same gDNA sample. The positive control amplification is that of the standard 18S RNA (PLU, ~1.2 kb) using primers PLU5 and PLU6 [80], while the negative control (-) included the same amount of primer as per the normal setup, but without template.

The 5' HR for *gcn5* was successfully amplified as indicated by the presence of the expected ~542 bp product (Figure 3.8A). Low amplification of *set7* was obtained at 48 °C and 50 °C (Figure 3.8B), which necessitated additional optimisation, including different template concentrations (Figure 3.8C). Following this optimisation, PCR amplicons of the anticipated fragment sizes for the 5' HR were successfully obtained for *set7* at ~416 bp. The PCR amplicons corresponding to the respective gene fragments of interest were excised and purified from the gel for further downstream experiments. The purified PCR products were analysed spectrophotometrically to have a DNA concentration of 69 - 96 ng/µL, with A_{260}/A_{280} and A_{260}/A_{230} purity ratios of 1.77 - 1.88 and 1.40 – 1.61, respectively.

Cloning of the 5' HRs required a sticky *Eco*RI 5'-overhang and a blunt 3'-end. Since Taq polymerase adds additional adenosine nucleotides to the 3'-ends, the 5' HR PCR products needed to be blunted prior to *Eco*RI digestion to allow for sticky-blunt ligation with the *Eco*RI and *Sna*BI digested pCS22_DiCre plasmid. The purified PCR products were blunted using T4 DNA polymerase, known 3' – 5' exonuclease activity, and then digested after which the final reaction was column purified yielding concentrations 48 - 60 ng/µL, with A_{260}/A_{280} and A_{260}/A_{230} purity ratios of 1.80 - 1.91 and 1.59 – 1.79. The purified, digested PCR products and pCS22_DiCre plasmid (digested with *Eco*RI and *Sna*BI) were ligated. Competent *E. coli* DH5α cells

were transformed with the ligated repair plasmids. Colony PCR was preformed to confirm the presence of recombinant clones (Figure 3.9).

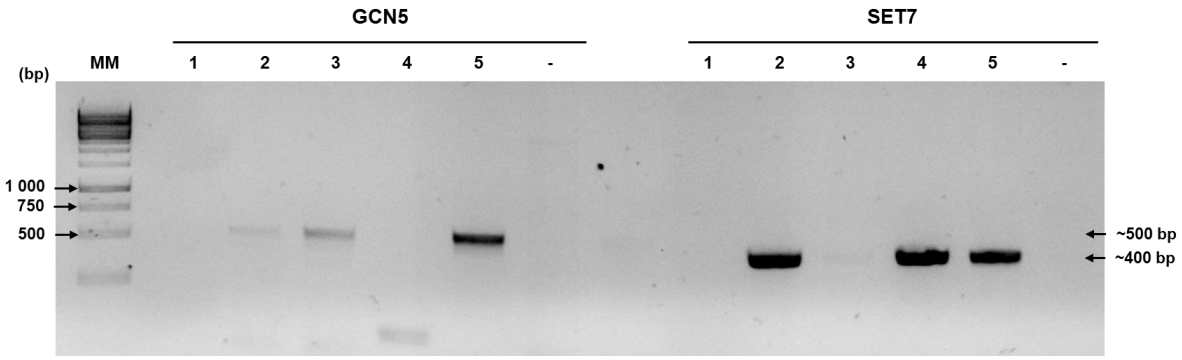


Figure 3.9: PCR analysis of screening of bacterial colonies positive for the 5' HR insert in pCS22_DiCre. Using gene-specific primers and the M13 forward (pCS22_DiCre backbone-specific primer), amplified DNA and 1 kb ladder (Lane MM) was separated with a 1.2 % (w/v) agarose/TAE gel that contained EtBr (1 µg/mL) for visualization under UV light. Expected band sizes are indicated, no Rf-values were obtained. The negative control (-), primers with no template are indicated.

Colonies that were screened identified only one positive clone for *gcn5* (colony 5, Figure 3.9) contained the insert in pCS22_DiCre. However, three clones were positive for the insert of *set7* (colony 2, 4 and 5, Figure 3.9), containing the recombinant pCS22_DiCre plasmid. The positive clones were grown overnight in LB-amp media, followed by the isolation of the plasmid, obtaining a concentration of 210 – 288 ng/µL with purity ratios of 1.60 – 1.70 and 1.55 – 1.63 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, indicating good quality DNA. A final verification of the identity of the cloned recombinant clones, included using Sanger dideoxy sequencing to screen for any inserts, deletions, or frameshift mutations. The sequencing alignment in the chromatogram (Figure 3.10), confirmed that the 5' HRs of *gcn5* and *set7* were successfully cloned into pCS22_DiCre without any alterations. The critical "AG" splice site at the 3'-end of both genes were detected and therefore supported the application of the pCS22_DiCre_*gcn5* and pCS22_DiCre_*set7* construct for downstream cloning. As predicted, the *Eco*RI cut site was retained for both pCS22_DiCre_*gcn5* and pCS22_DiCre_*set7*, respectively.

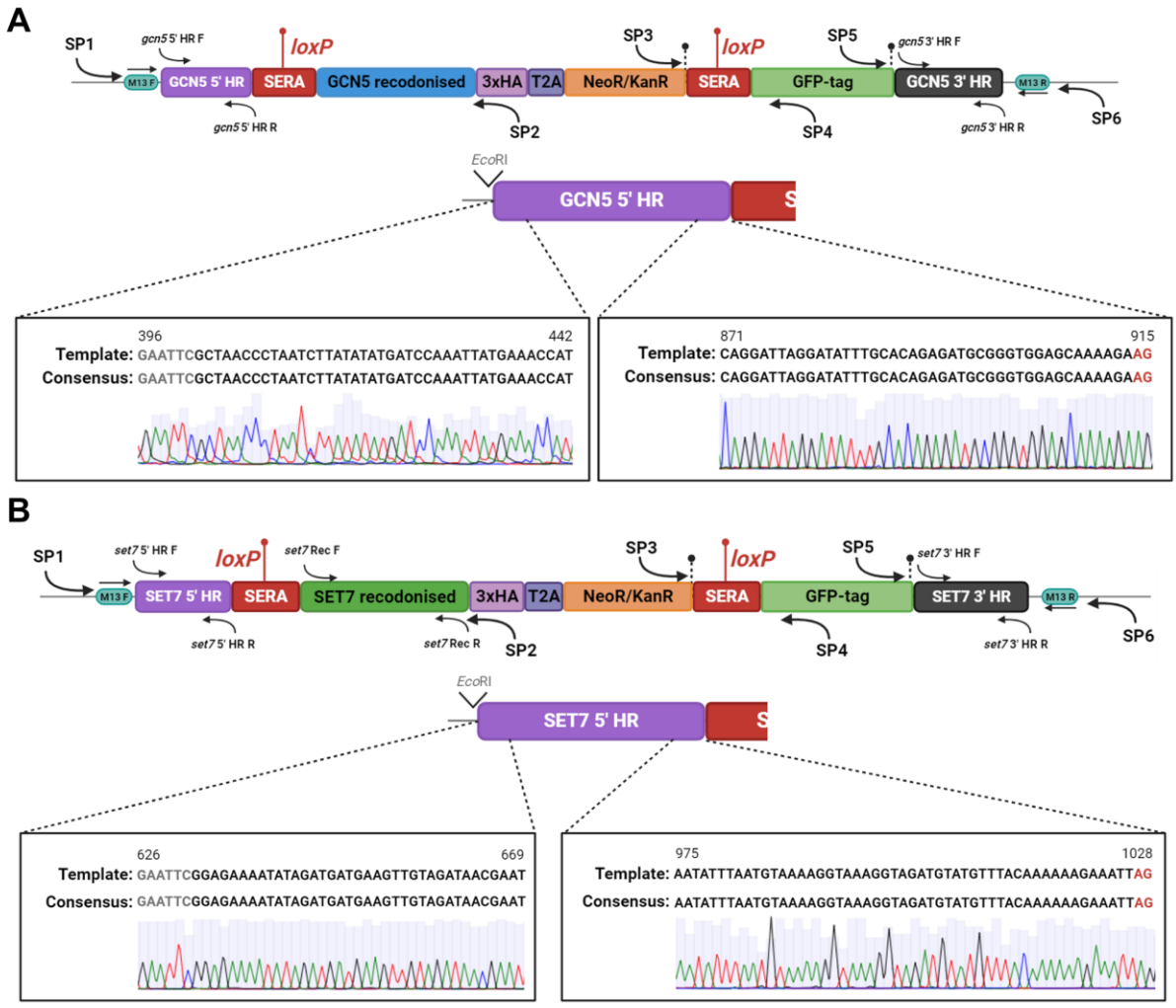


Figure 3.10: Sanger sequencing 5'HR inserts for (A) *gcn5* and (B) *set7*, cloned into pCS22_DiCre. Alignment of the consensus sequence to the expected reference sequence. The consensus sequence was obtained from analysis of a combination of two sequencing reactions obtained from either pCS22_DiCre backbone-specific primers (M13 forward or SP1) and 5' gene-specific reverse primers for both genes. A chromatogram of the forward or reverse trace data is represented with the alignment sequences. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

For the full consensus Sanger sequence alignment of the gene inserts in pCS22_DiCre, see supplementary data 2 and 3 respectively.

3.4.2.2. Cloning a recodonised gene fragment of *gcn5* and *set7* into pCS22_DiCre

The recodonised fragment was cloned downstream from the first *loxP* sequence and upstream of the HA-tag on pCS22_DiCre by complementary restriction cut site (blunt-sticky) ligation. These gene fragments were amplified from the pUC18_*gcn5* and pUC18_*set7* plasmids received from the sequence manufacturer GeneUniversal, following amplification optimisation. The PCR conditions were optimised for different annealing temperatures (not shown here) using gene-specific primers (Supplementary Table 1). The maximum product amplification was based on the visual intensity after EtBr staining (Figure 3.11). PCR amplicons of the anticipated fragment sizes for the recodonised regions were successfully obtained for both *gcn5* and *set7*, ~1029 bp and ~1260 bp, respectively.

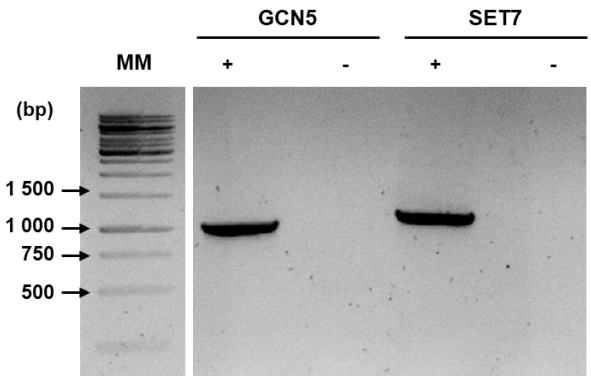


Figure 3.11: PCR amplification of the recodonised region of *gcn5* and *set7*. The amplified DNA and 1 kb DNA ladder (Lane MM) were separated with a 0.8 % (w/v) agarose/TAE gel and visualized with EtBr (1 µg/mL) pre-staining. Expected band sizes for *gcn5* (~1 kb) and set7 (~1.2 kb) are indicated. No Rf-values were obtained to define the magnitude in difference. All amplifications were performed with Phusion® Polymerase using the plasmid of origin as template. This also serves as conformation of presence of recodonised gene insert on pUC18 plasmid. The negative control (-) included the same amount of primer as per the normal setup, but without template.

The PCR amplicons of the corresponding sizes of interest were excised and purified from the gel for further downstream experiments. The purified PCR products were analysed spectrophotometrically to have a DNA concentration of 12 – 34 ng/µL, with A_{260}/A_{280} and A_{260}/A_{230} purity ratios of 1.51 – 1.78 and 0.28 – 0.30, respectively. Although low quality, these ratios were considered pure enough for downstream use.

Cloning of the recodonised regions required a blunt 5'-end and a sticky (*Bam*HI) 3'-overhang. Given the use of Phusion® Polymerase, which possess 3' – 5' exonuclease activity, blunting was not required. Following this, the PCR reaction was column purified and analysed spectrophotometrically to have a DNA concentration of 28 ng/µL for both pCS22_DiCre gene-specific plasmids, with A_{260}/A_{280} and A_{260}/A_{230} purity ratios of 1.78 and 0.26 – 0.41, respectively. The purified, digested PCR products and pCS22_DiCre plasmid were ligated and transformed into competent *E. coli* DH5α cells. Colony PCR was performed to confirm the presence of recombinant clones.

No positive clone for *gcn5* (Figure 3.12A) was obtained but two clones were positive for the insert (~2.0 kb) of *set7* (colony 1 and 9, Figure 3.12B). The positive clones were grown overnight in LB-amp media, followed by the isolation of the plasmid, obtaining a concentration of 210 – 288 ng/µL with purity ratios of 1.60 – 1.70 and 1.55 – 1.63 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, indicating good quality DNA. Given the presence of two positive clones for *set7* and the lack of a positive clone for *gcn5* suggests that the cloning of the *gcn5* recodonised regions were unsuccessful. This supports the idea that the cause could be attributed to the different gene characteristics. Furthermore, the presence of the 1 kb band across all colonies but one for *gcn5* indicates non-specific amplification by the primers used here, and therefore could point towards self-ligation of the plasmid backbone.

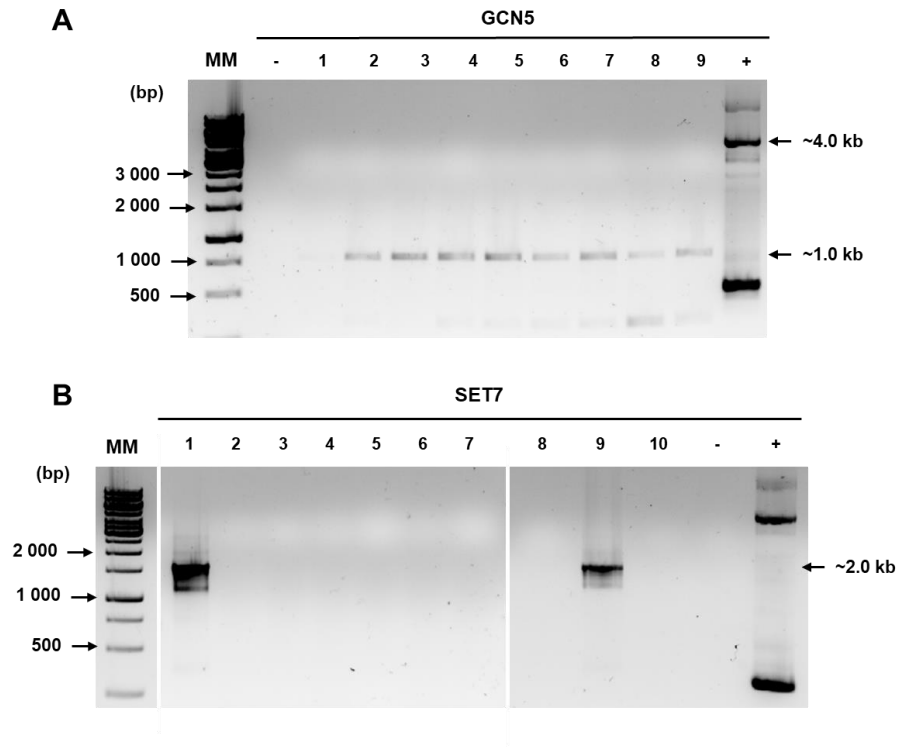


Figure 3.12: PCR analysis of screening of bacterial colonies positive for the recodonised insert in pCS22_DiCre. Using gene-specific primers and the M13 forward (pCS22_DiCre backbone-specific primer), amplified DNA and 1 kb ladder (Lane MM) was separated with a 0.8 % (w/v) agarose/TAE gel that contained EtBr (1 µg/mL) for visualization under UV light. Expected band sizes are indicated, no Rf-values were obtained. The positive control (+), primers specific for the selection marker cassette on pCS22_DiCre, and negative control (-), primers with no template is indicated.

The isolated plasmid was then characterised through Sanger dideoxy sequencing, which confirmed the identity of the isolated plasmid and the presence of the recodonised insert for *set7* only, pCS22_DiCre_*set7*, given that no positive clone for *gcn5* was obtained, even after optimisation (Figure 3.13). The critical "AG" splice site at the 3'-end of the SERA2 intron was detected and therefore supported the use of the pCS22_DiCre_*gcn5* and pCS22_DiCre_*set7* construct for downstream cloning. As predicted, the *Bam*HI cut site was retained.

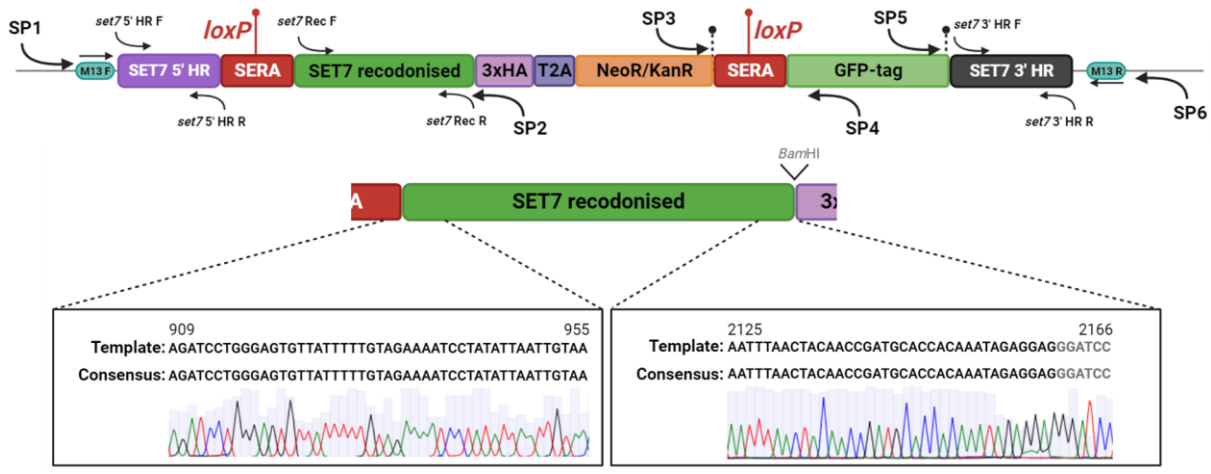


Figure 3.13: Sanger sequencing of recodonised gene insert for *set7*, cloned into pCS22_DiCre. Alignment of the consensus sequence to the expected reference sequence. The consensus sequence was obtained from analysis of a combination of two sequencing reactions obtained from either pCS22_DiCre backbone-specific primers (M13 F

and SP2) or gene-specific primers (*set7* 5'HR F or *set7* Rec F). Notably, the AG splice site on the 5'-end of the sequence is detected. A chromatogram of the forward or reverse trace data is represented with the alignment sequences. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

For the full consensus Sanger sequence alignment of the gene inserts in pCS22_DiCre, see supplementary data 2 and 3 respectively.

3.4.2.3. Cloning a 3' gene fragment of *gcn5* and *set7* into pCS22_DiCre

The 3' gene fragment (or 3' HR) was cloned upstream of the *loxP* sequence on pCS22_DiCre_*gcn5* and pCS22_DiCre_*set7* by complementary restriction cut site (sticky-sticky) ligation. These gene fragments were amplified from isolated genomic DNA following amplification optimisation, from *in vitro* cultured *P. falciparum* NF54 parasites (as per section 3.3). ExTaq DNA Polymerase was able to efficiently amplify the 3' HRs of both *gcn5* and *set7* with better yield after purification, ranging from 5 – 23 ng/µL with A_{260}/A_{280} and A_{260}/A_{230} purity ratios of 1.62 – 1.69 and 0.31 – 0.55, respectively (Figure 3.14). This successful amplification allowed for the gene segment to be cloned into the pCS22_DiCre_ plasmid.

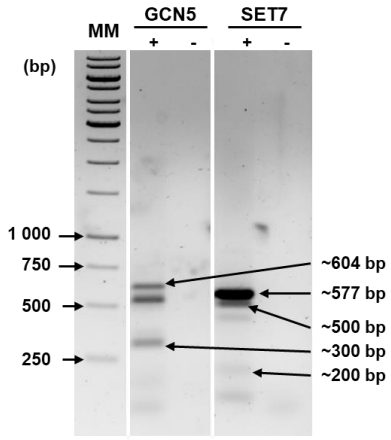


Figure 3.14: PCR amplification of the 3' HR's of *gcn5* and *set7*. DNA was separated on a 1.2 % (w/v) agarose/TAE gel and visualized with EtBr (1 µg/mL) pre-staining. All amplifications were performed with the same *P. falciparum* genomic DNA sample using ExTaq polymerase. Expected band sizes are indicated, at different intensities, however, no Rf-values were obtained. Both amplifications were performed from the same gDNA sample. The negative control (-) included the same amount of primer as per the gene amplification but without template. Non-specific amplification (~200 bp, ~300 bp and ~500 bp) was easily discernible from the desired band size.

The PCR amplicons of the corresponding sizes of interest were excised and purified from the gel for further downstream experiments. The purified PCR products were analysed spectrophotometrically with a DNA concentration of 9 - 25 ng/µL, with A_{260}/A_{280} and A_{260}/A_{230} purity ratios of 1.51 - 1.78 and 0.98 – 1.12, respectively.

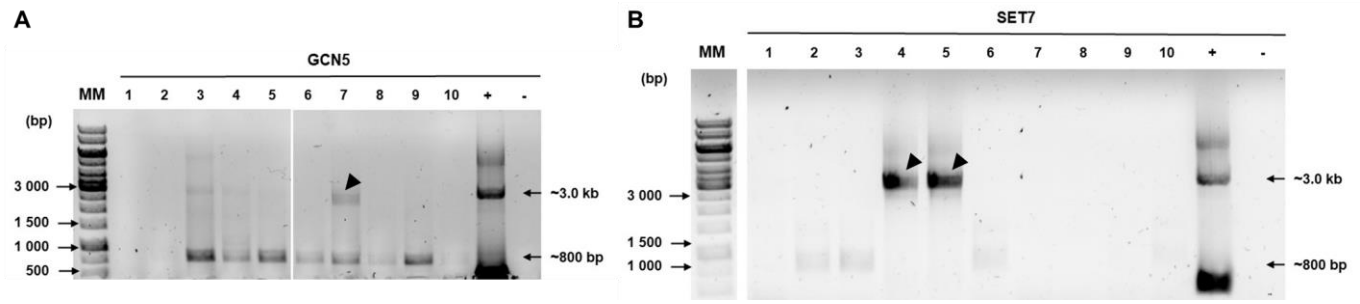


Figure 3.15: PCR analysis of screening of bacterial colonies positive for the 3' HR insert in pCS22_DiCre. Using gene-primers and the M13 forward (pCS22_DiCre backbone-specific primer), amplified DNA and 1 kb ladder (Lane MM) was separated with a 0.6 % (w/v) agarose/TAE gel that contained EtBr (1 µg/mL) for visualization under UV light. Expected band sizes are indicated, no Rf-values were obtained. The expected bands are indicated with arrow heads with the approximate sizes indicated with the corresponding arrow. The positive control (+) is primers specific for the selection marker cassette on pCS22_DiCre, and negative control (-), gene-specific primers with no template is indicated. Arrow heads

Colonies that were screened identified only one positive clone for *gcn5* (colony 7, Figure 3.15) contained the insert in pCS22_DiCre. However, three clones were positive for the insert of *set7* (colony 4 and 5, Figure 3.15), containing the recombinant pCS22_DiCre_*gcn5* plasmid. The positive clones were grown overnight in LB-amp media, followed by the isolation of the plasmid, obtaining a concentration of 210 – 288 ng/µL with purity ratios of 1.60 – 1.70 and 1.55 - 1.63 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, indicating good quality DNA. The isolated plasmid was then characterised through Sanger dideoxy sequencing, which confirmed the identity of the isolated plasmid and the presence of the 3' HR insert for both pCS22_DiCre_*gcn5* and pCS22_DiCre_*set7* (Figure 3.16).

As predicted, the restriction cut sites, *Xba*I and *Hind*III, was retained in both pCS22_DiCre_*gcn5* and pCS22_DiCre_*set7*, respectively. A complete final repair plasmid for *set7* was obtained. This plasmid, pCS22_DiCre_*set7*, may be prepared for transfection with the *P. falciparum* NF54::DiCre parasites.

For the full consensus Sanger sequence alignment of the gene inserts in pCS22_DiCre, see supplementary data 2 and 3 respectively.

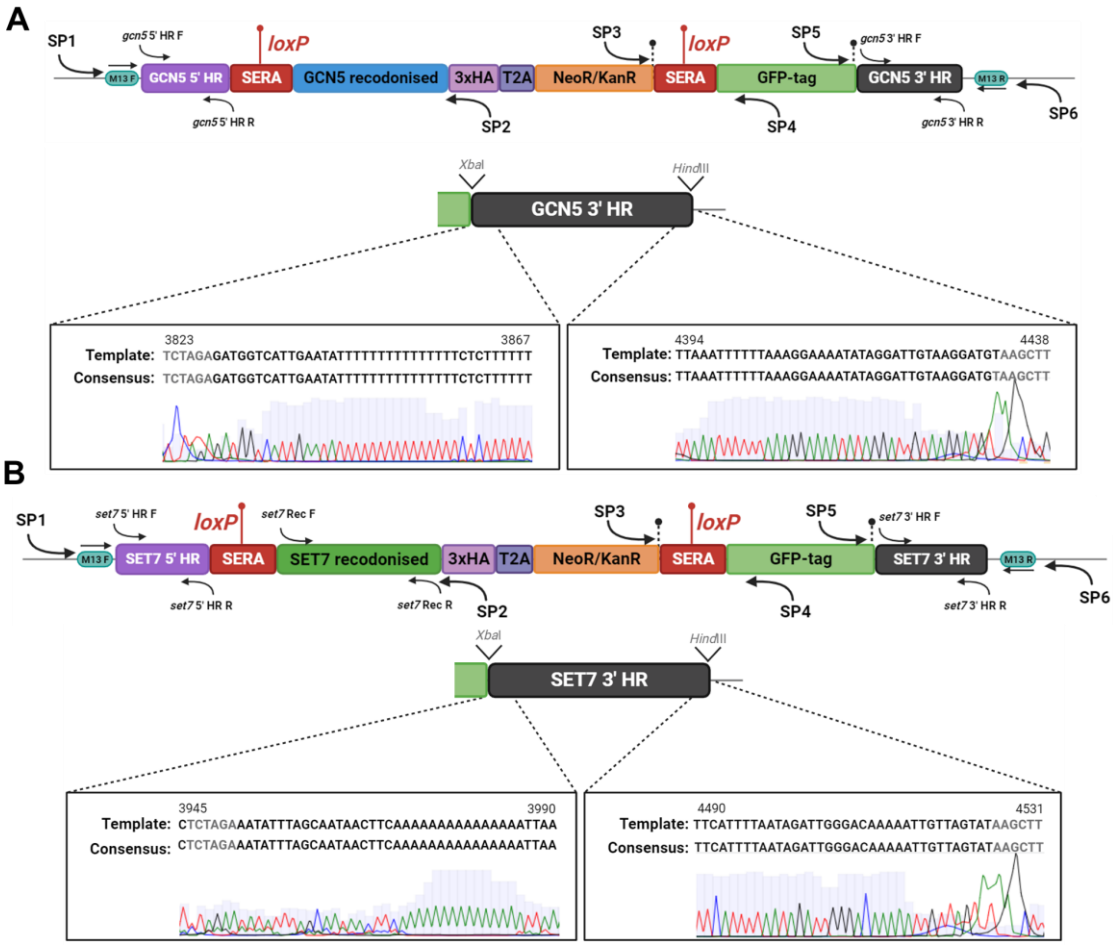


Figure 3.16: Sanger sequencing of 3' HR gene inserts for (A) *gcn5* and (B) *set7*, cloned into pCS22_DiCre. Alignment of the consensus sequence to the expected reference sequence. The consensus sequence was obtained from analysis of a combination of two sequencing reactions obtained from either pCS22_DiCre backbone-specific primers (M13 reverse or SP6) or gene-specific forward primers for both *gcn5* and *set7*. A chromatogram of the forward or reverse trace data is represented with the alignment sequences. Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

3.5. Generating a Cas9-gRNA plasmid construct

3.5.1. Design of guide RNAs for targeted gene dsDNA modification

Using the CRISPR-coCas9 system, a gRNA is required to align and induce a dsDNA break through the work of the coCas9 enzyme. Two gRNA per gene were designed and selected, and the efficiency would be compared once a KO is induced. Each gRNA was evaluated for homology in the genome in the *P. falciparum* genome to establish similarity to any off-target sequences through the e-value score, using the BLAST search tool. Any guides that met the required criteria discussed in section 2.6.1., were chosen if found in the recodonised region (i.e. between homology regions). In the instances where no gRNA candidates were identified in the indicated region, the 5' HR was moved upstream, and the search continued. Guides were specifically designed to be in close proximity to either one of the HRs. Hit gRNA oligonucleotides (Table 6) were synthesized and then annealed with their complementary oligo for subsequent cloning steps. The annealed gRNAs were then ligated with the *Bbs*I digested pDC2-coCas9-gRNA-hDHFR.

Table 6: Guide RNA sequences chosen for each gene candidate. Nucleotide overhangs are indicated in lowercase, while hit gRNA specific sequences in all uppercase. The addition of a G, where applicable, is indicated in bold. The e-value is indicated for the Blast score.

Sequence code	Final oligo for hit sequences (5' – 3')		e-value
*gcn5*_gRNA1	+1	tattTGAATGTATAACGAACGATA	3.00×10^{-4}
*gcn5*_gRNA2	+1	tatt**G**TTAGGGTGATTATAAAACGA	0.044
*set7*_gRNA1	-1	tattGCTGCATAATGCCATTTAAG	5.00×10^{-4}
*set7*_gRNA2	-1	tattCTCAATTTGAGGAGCATCAG	5.00×10^{-4}

Generating the Cas9-gRNA plasmid construct was initiated from the pDC2-coCas9-gRNA-hDHFR plasmid. The plasmid was successfully isolated from competent *E. coli* DH5α and characterised (Figure 3.17). The relevant concentrations for the isolated plasmids after purification ranged from 455 ng/µL while purity was in the range of 1.55 and 1.77 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, indicating pure DNA suitable for downstream use.

pDC2-coCas9-gRNA-hDHFR linearized with *Eco*RI-HF as expected, as a ~11.6 kb band was observed (Figure 3.17). Furthermore, the bands observed for RE mapping with *Eco*RI-HF and *Hind*III-HF (~7 kb, ~2.6 kb and ~1.9 kb) and *Eco*RI-HF and *Bam*HI-HF further confirmed the expected sizes (~8.1 kb, ~2.7 kb and ~700 bp) of the plasmid. Therefore, this confirmed the successful isolation of the pDC2-coCas9-gRNA-hDHFR plasmid which could be used for subsequent cloning of the gene-specific gRNAs.

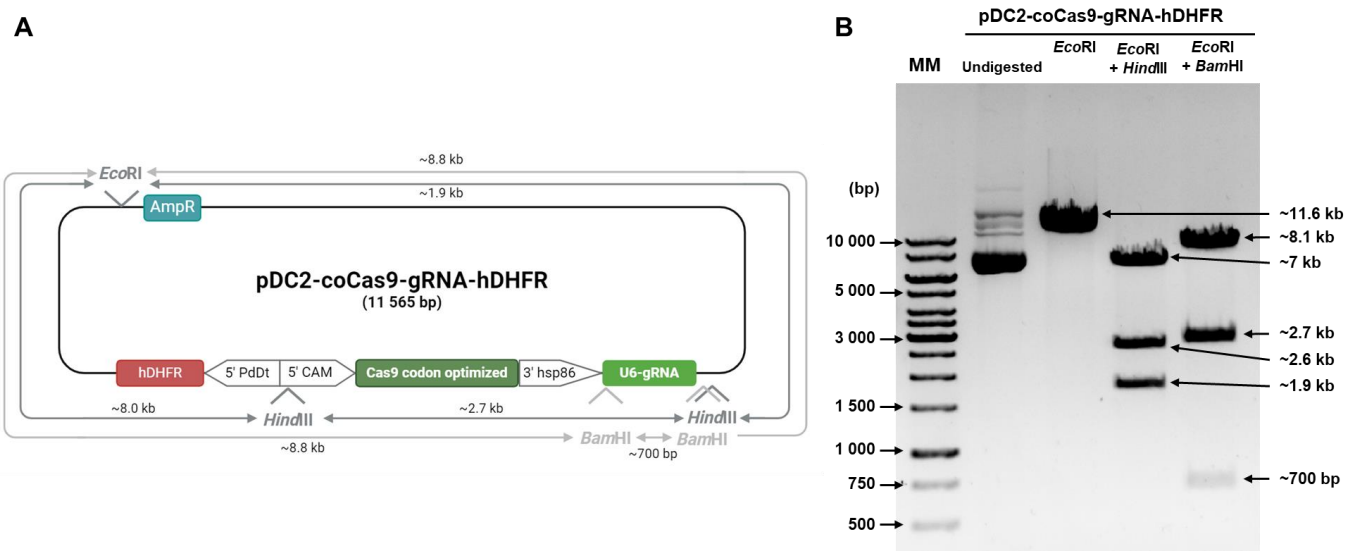


Figure 3.17: Restriction enzyme digestion of the pDC2-coCas9-gRNA-hDHFR plasmid. (A) Schematic representation of the expected band sizes for the digestion of the plasmid, and (B) validation of successful isolation of the plasmid through restriction enzyme mapping. Restriction enzyme mapping preformed indicates the separation of plasmid DNA and 1 kb ladder (Lane MM) on a 1.0 % (w/v) agarose/TAE gel and visualized with EtBr (1 µg/mL) pre-staining. Expected band sizes are indicated, no Rf-values were obtained. Digestion of three combinations of enzymes (*Eco*RI, *Hind*III and *Xba*I). Schematic is not to scale. Figure created with Biorender.com under the basic licence for educational purposes.

3.5.2. Cloning of gRNA of *gcn5* and *set7* into pDC2-coCas9-gRNA-hDHFR plasmid

The pDC2-coCas9-gRNA-hDHFR plasmid was digested with *Bbs*I restriction enzyme, after which it was column purified yielding a concentration 233 ng/µL while purity was in the range of 1.82 and 1.66 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively. Guide oligos were purchased in a 5' phosphorylated state, annealed to the respective complementary strands and then ligated with the digested plasmid. The ligation reaction was transformed with competent *E. coli* DH5α cells.

Several positive clones were obtained for *gcn5* (Figure 3.18 A) which contained the gRNA insert, 1 and 2, in pDC2-coCas9 plasmid. Furthermore, several positive clones for the gRNA insert, 1 and 2, of *set7* (Figure 3.18 B), containing the recombinant pDC2-coCas9 plasmid were detected. The positive clones were grown overnight in LB-amp media, followed by the isolation of the plasmid, obtaining a concentration of 210 – 288 ng/µL with purity ratios of 1.60 – 1.70 and 1.55 - 1.63 for A_{260}/A_{280} and A_{260}/A_{230} ratios, respectively, indicating good quality DNA.

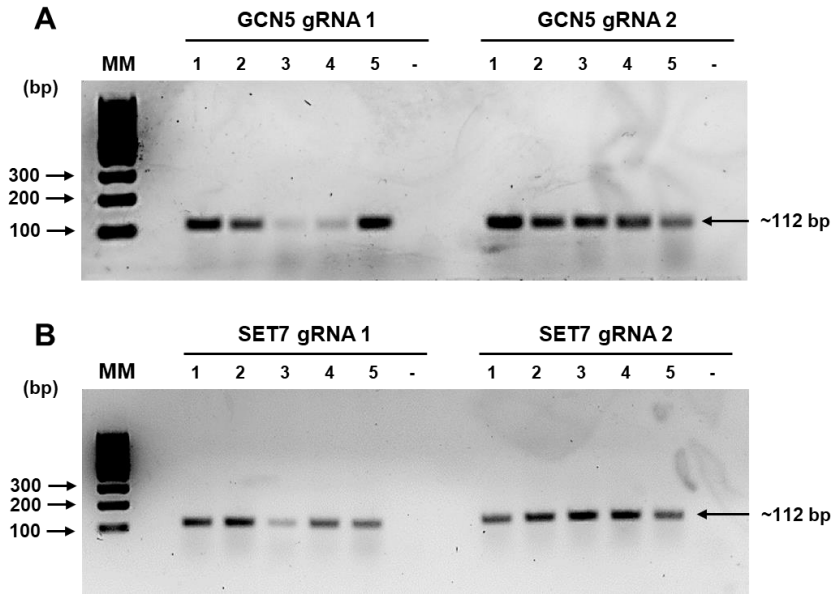


Figure 3.18: PCR analysis of screening of bacterial colonies with the gRNA insert for (A) *gcn5* and (B) *set7*. Using gRNA-specific primers and the p35 (pDC2-coCas9 backbone-specific primer), amplified DNA was separated with a 2.0 % w/v agarose/TAE gel that contained EtBr (1 µg/mL) for visualization under UV light. Expected band sizes are indicated, no Rf-values were obtained. The negative control (-), primers with no template is indicated.

The isolated plasmid was then characterised through Sanger dideoxy sequencing, which confirmed the identity of the isolated plasmid and the presence of the respective gRNA inserts for both *gcn5* and *set7* in pDC2-coCas9, respectively (Figure 3.19 A and B).

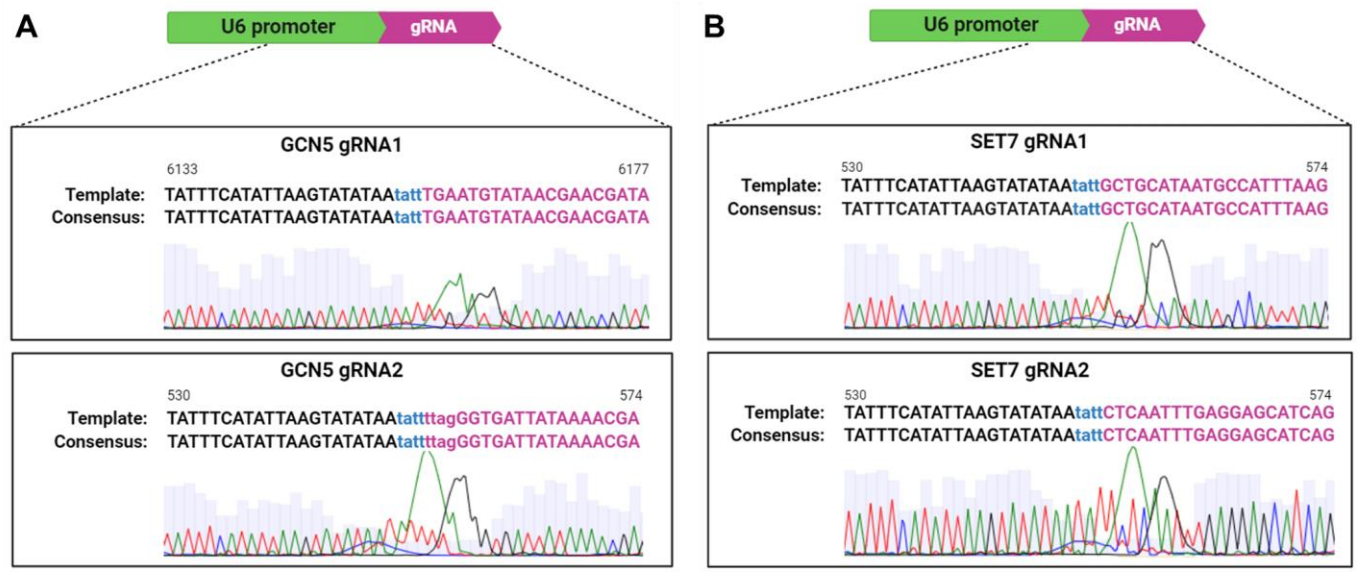


Figure 3.19: Sanger sequencing of gRNAs of (A) *gcn5* and (B) *set7* cloned into pDC2-coCas9co-Cas9-gRNA-hDHFR. Alignment of the consensus sequence to the expected reference sequence. The consensus sequence was obtained from analysis of a singular sequencing reaction, including the pDC2-coCas9-gRNA-hDHFR backbone-specific primers (p35). The chromatogram of the forward or reverse trace data is represented with the alignment sequences, with no deletions or mutations noted across the entirety of the forward or reverse alignment. Schematic of the sequenced gene segment is not drawn to scale. Figure created with Biorender.com under the basic licence for educational purposes.

The respective gRNAs were cloned into the pDC2-coCas9-gRNA-hDHFR vector. Therefore, in this study one of the two required components for the CRISPR-coCas9 mediated gene editing approach for the DiCre system was generated and is ready for transfection.

4. Discussion

Epigenetic modifiers are arguably one of the most elusive regulatory gene members involved in the complex epigenetic landscape of *P. falciparum*. Epigenetic regulation is of critical importance in the timely expression of genes in the parasite [30, 33]. Given the apparent lack of transcription factors in the parasite, epigenetic regulators are thought to be of major importance [26, 30, 31]. The conditional DiCre gene KO system is expected to enhance our understanding of genes that play a crucial role in specifically regulatory pathways, which were previously not accessible at a genetic level [45, 69]. This study focussed on optimising the conditional DiCre system and validate the application of this system by genetically investigating two regulatory genes associated with the epigenome of *P. falciparum* [38, 45, 79]. We produce a novel cloning method to simplify the process for the DiCre system and create a universal repair plasmid applicable to any GOI.

The optimised cloning system of the new pCS22_DiCre plasmid should have several advantages over the current cloning strategy that is dependent on Gibson assembly [45]. Firstly, the selection marker cassette was cloned into a plasmid carrying a SERA/*loxP* sequence, preparing it for cloning of the additional gene inserts. The technique critically relies on accurate nucleotide amplification and cloning, with no deletions in either the 5' HR or the recodonised inserts as these modifications may influence the original protein. This cloning approach carefully incorporates several validation steps to confirm successful cloning, including Sanger Sequencing. Secondly, the pSN28 plasmid was used to facilitate a “cut-and-paste” approach by designing the cloning system to use all possible restriction cloning sites available on the plasmid.

A potential pitfall for this system is identifying the required "out-of-frame" AG nucleotide sequence for the 5’ HR, which can be resolved by using alternative splice sequences [74]. Nevertheless, for the proxy genes used here, the out-of-frame AG was easily identified. Therefore, this is not considered a limitation, since the "cut-and-paste" approach can still be applied effectively. Another potential pitfall of this system is the requirement to clone a sticky-blunt ended 5’ HR and the requirement to blunt the *Bsm*I site overhang. This therefore requires a 3’ - 5’ exonuclease with highly specific activity to blunt any overhangs on the 3’-end and 5’ end, respectively. This can be mediated by applying a different cloning approach for the 5’ HR, and recodonised region. This entails synthesizing the 5' HR, SERA2/*loxP* sequence and recodonised gene insert as one stretch of DNA to clone into pCS22_DiCre. Alternatively, this can be mediated by synthezing the recodonised insert with a 5’-upstream SERA2/*loxP* sequence, including a *Sna*BI cut site on the 5’-end to clone into the pCS22_DiCre plasmid. These approaches were not feasible specifically for these two proxy genes. An alternative is to use a Polymerase with 3’ – 5’ exonuclease activity to amplify the 5’ HR.

A complete repair plasmid was generated for only one of the proxy genes, *set7*, using the cloning strategy laid out in this study as proof of concept of the system described here. Given the inability to generate the repair plasmid for the other proxy gene, *gcn5*, this system possibly reflects some cloning restrictions that could point towards specific characteristics of the GOI. Given the inability to predict the secondary

structures that PCR products possibly could form *in vitro*, literature describes the difficulties applying such a PCR product in downstream cloning [8]. The GC-content of *gcn5*'s recodonised insert consisted of ~30 %, compared to ~29 % for *set7*. While this seems low, in the event that GC-dense areas on the PCR product comes in close proximity to another GC-dense area, the nucleotides can form cross bonds which leads to these secondary structures, thus prohibiting downstream use. It is therefore plausible that this contributed to unsuccessful cloning of the *gcn5* recodonised insert. Alternatively, for a gene such as *gcn5*, it can be considered that the chosen area for recodonisation was not ideal for cloning and possibly influenced the clonability of this insert. To bypass this, the region can be altered by choosing a different size or GC-rich section to clone.

Furthermore, a notable non-specific deletion was observed on the 3'-end of *set7*'s 5' HR (see supplementary data 4), after attempting to clone this section at the *Eco*RI and *Sna*BI sites. This suggests that T4 DNA Polymerase exhibits non-specific activity. This is supported by the lack of a similar deletion for the cloning of *gcn5*'s 5' HR. These difficulties are substantiated by the findings is in a study on T4 DNA Polymerase, where the authors noted high rates and non-uniform digestion of DNA in conditions where the enzyme concentration is much lower than that of the DNA [81]. The findings here further support the notion that T4 DNA Polymerase exhibit non-specific activity and supports the idea that the *Bsm*I site was blunted non-specifically before attempting to clone *gcn5*'s recodonised insert, possibly owing to the unsuccessful cloning. This is further supported by the successful cloning of *set7*'s recodonised region, which did not indicate any non-specific deletions. To resolve this the straight-forward resolution is to amplify the 5' HR using a Polymerase containing 3' – 5' exonuclease activity, such as Phusion®, however, this does not address the need for the blunting of the *Bsm*I site. Thus, an alternative approach to address this would include to use a completely different 3' – 5' exonuclease and test the error rate of T4 DNA polymerase, which can then be applied as an alternate for this cloning strategy. Cloning the 5' HR and recodonised gene insert proves to be another pitfall given the required use of the *Sna*BI and *Bsm*I cut sites. This can be mitigated by synthesizing this entire region as one stretch of nucleotides and then cloning using the *Eco*RI and *Bam*HI cut sites, as previously described. This makes the cloning system very versatile for the cloning of the 5' HR and recodonised gene segments. However, under certain conditions, manufacturers are unable to synthesise the 5' HR of *P. falciparum* due to the high AT-richness of the parasite's genome and coding sequences; this is typically seen when the 5' HR includes intron sequences or forms part of the 5' UTR. Hence, regardless of the gene insert to be synthesized, this cloning strategy is optimized for any outcome, and cloning should still be accomplishable.

Generating gene-specific gRNA-plasmid construct is supported by reports of a two-plasmid approach for CRISPR-Cas9. This approach works ideally when linearized before transfection, decreasing the time to generate mutant parasites by as much as a week [82]. *P. falciparum* parasites have also been recorded to readily take up plasmids from erythrocytes, thus rendering the use of a two-plasmid approach even more viable [82]. Given the size of the selection marker cassette applied here, together with the desired gene fragments, i.e. 5' HR, recodonised and 3' HR, the repair plasmid was set to span a minimum of

3.5 kb. Therefore, together with the final recombinant plasmid size, in this cloning system a two-plasmid approach is more feasible. As a result, one plasmid codes for the CRISPR-coCas9-complex and gRNA specific to the GOI, while the repair plasmid contains the section that will be integrated into the genomic DNA with the desired gene segments, as previously stated.

Control over genetic interrogation at specific times within an organism's life cycle, particularly in the highly complex *P. falciparum*, has become an increasingly popular method for establishing the essentiality of any gene/gene product in a target organism. Using such a system can offer insight on the influence of a gene at any point in the parasite's life cycle. In this study, we derived and assessed a new cut-and-paste cloning strategy employing DiCre with a CRISPR-Cas9/SLI approach to enable conditional KO of target genes *gcn5* and *set7*, which will enable future studies exploring the roles of the proteins encoded by these genes in regulating the parasite life cycle. The results generated here support the applicability of the cloning system for genes with more challenging characteristics as observed for *set7* and *gcn5.* For genes that proves less challenging to clone, such as *myb1*, this cloning approach was successfully applied to generate a DiCre *myb1* transgenic line (Sizwe Tshabalala, MSc Dissertation, University of Pretoria). In conclusion, the cloning system described here have minimal drawbacks and should be easy to apply for any GOI.

5. Conclusion

The investigation of crucial regulatory genes within *P. falciparum* could be a major contributing factor prolonging the stagnation in malaria elimination research. The development and implementation of a novel cloning approach for the DiCre gene KO system in this study signifies a leap forward in the field of genetic research. By overcoming previous cloning limitations and streamlining the process of genetic manipulation, this innovative approach not only enhances the precision and efficiency of gene editing techniques but also opens avenues for broader applications in diverse biological systems. Here we provide evidence that a universal DiCre plasmid was generated and applied to genetically alter one proxy gene as proof of concept for the application for any GOI. The newfound adaptability and simplicity of this cloning approach not only enhances the DiCre system's usability but also holds promise for accelerating the identification of essential genes in *P. falciparum*. This will facilitate effective DiCre-mediated gene deletion and pave the way for expanded gene KO studies while insights gained using this cloning system are expected to contribute to malaria elimination efforts.

References

1. WHO, *World malaria report 2022*, in *Geneva: World Health Organization*. 2022.
2. World Health Organization. *WHO recommends R21/Matrix-M vaccine for malaria prevention in updated advice on immunization*. 2023; Available from: https://www.who.int/news/item/02-10-2023-who-recommends-r21-matrix-m-vaccine-for-malaria-prevention-in-updated-advice-on-immunization.
3. Nadeem, A.Y., et al., *Mosquirix™ RTS, S/AS01 Vaccine Development, Immunogenicity, and Efficacy.* Vaccines (Basel), 2022. **10**(5).
4. Stanisic, D.I. and M.F. Good, *Malaria Vaccines: Progress to Date.* BioDrugs, 2023. **37**(6): p. 737-756.
5. Bozdech, Z., et al., *Population genomics and transcriptomics of Plasmodium falciparum uncover components of the artemisinin resistance genetic background.* Research Square [Preprint], 2023.
6. Tuteja, R., *Malaria − an overview.* FEBS Journal, 2007. **274**(18): p. 4670-4679.
7. Ansari, H.R., et al., *Genome-scale comparison of expanded gene families in Plasmodium ovale wallikeri and Plasmodium ovale curtisi with Plasmodium malariae and with other Plasmodium species.* International Journal for Parasitology, 2016. **46**(11): p. 685-696.
8. Fredman, D., et al., *DFold: PCR design that minimizes secondary structure and optimizes downstream genotyping applications.* Human Mutatant, 2004. **24**(1): p. 1-8.
9. Moon, S., et al., *An image analysis algorithm for malaria parasite stage classification and viability quantification.* PLOS ONE, 2013. **8**(4): p. e61812.
10. Milner, D.A., *Malaria Pathogenesis.* Cold Spring Harbor Perspectives in Medicine, 2018. **8**(1): p. a025569.
11. Florens, L., et al., *A proteomic view of the Plasmodium falciparum life cycle.* Nature, 2002. **419**(6906): p. 520-526.
12. Basu, S. and P.K. Sahi, *Malaria: an update.* The Indian Journal of Pediatrics, 2017. **84**: p. 521-528.
13. Del Prado, G.R.L., et al., *Malaria in developing countries.* The Journal of Infection in Developing Countries, 2014. **8**(01): p. 001-004.
14. Si, K., et al., *The structure of Plasmodium falciparum multidrug resistance protein 1 reveals an N-terminal regulatory domain.* Proceedings of the National Academy of Sciences, 2023. **120**(32): p. e2219905120.
15. Alves, E., et al., *Combining IP3 affinity chromatography and bioinformatics reveals a novel protein-IP3 binding site on Plasmodium falciparum MDR1 transporter.* Current Research in Microbial Sciences, 2023. **4**: p. 100179.
16. Sherman, I.W., *Biochemistry of Plasmodium (malarial parasites).* Microbiological Reviews, 1979. **43**(4): p. 453-495.
17. Sato, S., *Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology.* Journal of Physiological Anthropology, 2021. **40**(1).
18. Frischknecht, F. and K. Matuschewski, *Plasmodium Sporozoite Biology.* Cold Spring Harbor Perspectives in Medicine, 2017. **7**(5): p. a025478.
19. van der Watt, M.E., J. Reader, and L.-M. Birkholtz, *Adapt or Die: Targeting Unique Transmission-Stage Biology for Malaria Elimination.* Frontiers in Cellular and Infection Microbiology, 2022. **12**: p. 901971.
20. Hollin, T. and K.G. Le Roch, *From Genes to Transcripts, a Tightly Regulated Journey in Plasmodium.* Frontiers in Cellular and Infection Microbiology, 2020. **10**.
21. Lee, A.H., L.S. Symington, and D.A. Fidock, *DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum.* Microbiology and Molecular Biology Reviews, 2014. **78**(3): p. 469-486.
22. Van Biljon, R., et al., *Hierarchical transcriptional control regulates Plasmodium falciparum sexual differentiation.* BMC Genomics, 2019. **20**(1).
23. Cowman, A.F. and B.S. Crabb, *Invasion of Red Blood Cells by Malaria Parasites.* Cell, 2006. **124**(4): p. 755-766.
24. Delves, M.J., et al., *Male and Female Plasmodium falciparum Mature Gametocytes Show Different Responses to Antimalarial Drugs.* Antimicrobial Agents and Chemotherapy, 2013. **57**(7): p. 3268-3274.
25. Bozdech, Z., et al., *The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum.* PLOS Biology, 2003. **1**(1): p. e5.
26. Ay, F., et al., *Multiple dimensions of epigenetic gene regulation in the malaria parasite Plasmodium falciparum.* BioEssays, 2015. **37**(2): p. 182-194.
27. Li, S.-J., Z. Huang, and L. Jiang, *Molecular Basis of Malaria Pathogenesis*, in *Malaria Control and Elimination in China: A successful Guide from Bench to Bedside*. 2023, Springer. p. 211-223.
28. Modrzynska, K., et al., *A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the Plasmodium Life Cycle.* Cell Host & Microbe, 2017. **21**(1): p. 11-22.
29. Duffy, M.F., et al., *Epigenetic regulation of the Plasmodium falciparum genome.* Briefings in Functional Genomics, 2014. **13**(3): p. 203-216.
30. Park, M., A.J. Keung, and A.S. Khalil, *The epigenome: the next substrate for engineering.* Genome Biology, 2016. **17**(1): p. 1-17.
31. Latchman, D., *Gene control*. 2 ed. 2015: Garland Science. 79-89.
32. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nature Genetics, 2008. **40**(7): p. 897-903.

33. Connacher, J., H. Von Grüning, and L. Birkholtz, *Histone modification landscapes as a roadmap for malaria parasite development.* Frontiers in Cell and Developmental Biology, 2022. **10**: p. 848797.
34. Coetzee, N., et al., *Quantitative chromatin proteomics reveals a dynamic histone post-translational modification landscape that defines asexual and sexual Plasmodium falciparum parasites.* Scientific Reports, 2017. **7**(1): p. 1-12.
35. Connacher, J., *Epigenetic repression of gene expression controls the initial phases of gametocytogenesis in the malaria parasite, Plasmodium falciparum*. 2021: Pretoria. p. 29-30, 33-80.
36. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Research, 2011. **21**(3): p. 381-395.
37. Chookajorn, T., et al., *Epigenetic memory at malaria virulence genes.* Proceedings of the National Academy of Sciences, 2007. **104**(3): p. 899-902.
38. Chen, P.B., et al., *Plasmodium falciparum PfSET7: enzymatic characterization and cellular localization of a novel protein methyltransferase in sporozoite, liver and erythrocytic stage parasites.* Scientific Reports, 2016. **6**(1): p. 1-14.
39. Miao, J., et al., *The malaria parasite Plasmodium falciparum histones: Organization, expression, and acetylation.* Gene, 2006. **369**: p. 53-65.
40. Lucky, A.B., et al., *Plasmodium falciparum GCN5 plays a key role in regulating artemisinin resistance–related stress responses.* bioRxiv [Preprint], 2023: p. 2023.01. 11.523703.
41. de Koning-Ward, T.F., P.R. Gilson, and B.S. Crabb, *Advances in molecular genetic systems in malaria.* Nature Reviews Microbiology, 2015. **13**(6): p. 373-387.
42. Coetzee, N., et al., *Epigenetic inhibitors target multiple stages of Plasmodium falciparum parasites.* Scientific Reports, 2020. **10**(1): p. 2355.
43. Gardner, M.J., et al., *Genome sequence of the human malaria parasite Plasmodium falciparum.* Nature, 2002. **419**(6906): p. 498-511.
44. Kudyba, H.M., et al., *Some conditions apply: Systems for studying Plasmodium falciparum protein function.* PLOS Pathogens, 2021. **17**(4): p. e1009442.
45. Tibúrcio, M. and et al., *A Novel Tool for the Generation of Conditional Knockouts To Study Gene Function across the Plasmodium falciparum Life Cycle.* mBio, 2019. **10**(5).
46. Zhang, C., et al., *Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System.* mBio, 2014. **5**(4): p. 10.1128/mbio.01414-14.
47. Prommana, P., et al., *Inducible Knockdown of Plasmodium Gene Expression Using the glmS Ribozyme.* PLOS ONE, 2013. **8**(8): p. e73783.
48. Ganesan, S.M., et al., *Synthetic RNA–protein modules integrated with native translation mechanisms to control gene expression in malaria parasites.* Nature Communications, 2016. **7**(1): p. 10727.
49. Rajaram, K., H.B. Liu, and S.T. Prigge, *Redesigned TetR-Aptamer System To Control Gene Expression in Plasmodium falciparum.* mSphere, 2020. **5**(4).
50. Banaszynski, L.A., et al., *A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules.* Cell, 2006. **126**(5): p. 995-1004.
51. Armstrong, C.M. and D.E. Goldberg, *An FKBP destabilization domain modulates protein levels in Plasmodium falciparum.* Nature Methods, 2007. **4**(12): p. 1007-1009.
52. Birnbaum, J. and et al., *A genetic system to study Plasmodium falciparum protein function.* Nature Methods, 2017. **14**(4): p. 450.
53. Fierro, M.A., et al., *Knock-sideways by inducible ER retrieval enables a unique approach for studying Plasmodium-secreted proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2023. **120**(33): p. e2308676120.
54. Liang, X., et al., *A leak-free inducible CRISPRi/a system for gene functional studies in Plasmodium falciparum.* Microbiology Spectrum, 2022. **10**(3): p. e02782-21.
55. Jullien, N., et al., *Regulation of Cre recombinase by ligand-induced complementation of inactive fragments.* Nucleic Acids Research, 2003. **31**(21): p. e131-e131.
56. Davies, H., et al., *An exported kinase family mediates species-specific erythrocyte remodelling and virulence in human malaria.* Nature Microbiology, 2020. **5**(6): p. 848-863.
57. Robert-Paganin, J., et al., *Plasmodium myosin A drives parasite invasion by an atypical force generating mechanism.* Nature Communications, 2019. **10**(1): p. 1-12.
58. Collins, C.R., et al., *Robust inducible cre recombinase activity in the human malaria parasite Plasmodium falciparum enables efficient gene deletion within a single asexual erythrocytic growth cycle.* Molecular Microbiology, 2013. **88**(4): p. 687-701.
59. Ecker, A., et al., *Tricks in Plasmodium's molecular repertoire–Escaping 3' UTR excision-based conditional silencing of the chloroquine resistance transporter gene.* International Journal for Parasitology, 2012. **42**(11): p. 969-974.
60. Kratochwil, C.F. and F.M. Rijli, *The Cre/Lox system to assess the development of the mouse brain.* Brain Development: Methods and Protocols, 2014: p. 295-313.
61. Pino, P., et al., *A tetracycline-repressible transactivator system to study essential genes in malaria parasites.* Cell Host & Microbe, 2012. **12**(6): p. 824-834.

62. Wagner, J.C., et al., *An integrated strategy for efficient vector construction and multi-gene expression in Plasmodium falciparum.* Malaria Journal, 2013. **12**: p. 1-13.
63. Crabb, B.S. and A.F. Cowman, *Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodium falciparum.* Proceedings of the National Academy of Sciences, 1996. **93**(14): p. 7289-7294.
64. Ghorbal, M., et al., *Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system.* Nature Biotechnology, 2014. **32**(8): p. 819-821.
65. Nkrumah, L.J., et al., *Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase.* Nature Methods, 2006. **3**(8): p. 615-621.
66. Rodríguez-Trelles, F., R. Tarrío, and F.J. Ayala, *Is ectopic expression caused by deregulatory mutations or due to gene-regulation leaks with evolutionary potential?* BioEssays, 2005. **27**(6): p. 592-601.
67. Lee, M.C.S., et al., *Cutting back malaria: CRISPR/Cas9 genome editing of Plasmodium.* Briefings in Functional Genomics, 2019. **18**(5): p. 281-289.
68. Adjalley, S. and M.C.S. Lee, *CRISPR/Cas9 editing of the Plasmodium falciparum genome*, in *Malaria Immunology: Targeting the Surface of Infected Erythrocytes*, A.T.R. Jensen and L. Hviid, Editors. 2022, Springer. p. 221-239.
69. Knuepfer, E., et al., *Generating conditional gene knockouts in Plasmodium–a toolkit to produce stable DiCre recombinase-expressing parasite lines using CRISPR/Cas9.* Scientific Reports, 2017. **7**(1): p. 1-12.
70. Summers, R.L., et al., *Chemogenomics identifies acetyl-coenzyme A synthetase as a target for malaria treatment and prevention.* Cell Chemical Biology, 2022. **29**(2): p. 191-201. e8.
71. Fidock, D.A. and T.E. Wellems, *Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil.* Proceedings of the National Academy of Sciences, 1997. **94**(20): p. 10931-10936.
72. Maier, A.G., et al., *Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in Plasmodium falciparum for targeted gene deletion by double crossover recombination.* Molecular and Biochemical Parasitology, 2006. **150**(1): p. 118.
73. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-28.
74. Zhang, X., et al., *Branch Point Identification and Sequence Requirements for Intron Splicing in Plasmodium falciparum.* Eukaryotic Cell, 2011. **10**(11): p. 1422-1428.
75. Trager, W. and J. Jensen, *Human malaria parasites in continuous culture.* Science, 1976. **193**(4254): p. 673.
76. Doench, J.G., et al., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.* Nature Biotechnology, 2016. **34**(2): p. 184-191.
77. Kaushik, M., et al., *Genomic Glimpse of the Chromatin Modifier SET Domain family in Plasmodium falciparum.* Journal of Communicable Diseases (E-ISSN: 2581-351X & P-ISSN: 0019-5138), 2019. **51**(4): p. 29-40.
78. Cui, L., et al., *Histone lysine methyltransferases and demethylases in Plasmodium falciparum.* International Journal for Parasitology, 2008. **38**(10): p. 1083-1097.
79. Fan, Q., L. An, and L. Cui, *Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling.* Eukaryotic Cell, 2004. **3**(2): p. 264-276.
80. Harrison, G.F., et al., *Plasmodium-specific molecular assays produce uninterpretable results and non-Plasmodium spp. sequences in field-collected Anopheles vectors.* The American Journal of Tropical Medicine and Hygiene, 2013. **89**(6): p. 1117.
81. Qi, R. and G. Otting, *Mutant T4 DNA polymerase for easy cloning and mutagenesis.* PLOS ONE, 2019. **14**(1): p. e0211065.
82. Deitsch, K.W., C.L. Driskill, and T.E. Wellems, *Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes.* Nucleic Acids Research, 2001. **29**(3): p. 850-853.

Supplementary information

Supplementary Table 1: Primer sequences used to amplify the all the required gene segments of both *gcn5* and *set7*, to generate the respective DiCre repair plasmids. Nucleotide overhangs are indicated in lowercase, while gene specific sequences are all uppercase. The underlined and bold sequence represents the restriction enzyme recognition sequences. No stops codons present.

	Primer code		Primer sequence (5' – 3')	Length of amplicon
gcn5	**gcn5_5'HR_FWD**	F	atta**GAATTC**GCTAACCCTAATCTTATATATGATCCAAATTATG	524 bp
	gcn5_5'HR_REV	R	CTTCTTTTGCTCCACCCGCATC	
	gcn5_Rec_FWD	F	agAAAATATGGGTATTATAACCTTTGAATG	1031 bp
	gcn5_Rec_REV	R	tggc**GGATCC**TTTTGCTGTATCTGTTATCGC	
	gcn5_3'HR FWD	F	ttaa**TCTAGA**GATGGTCATTGAATATTTTTTTTT	624 bp
	gcn5_3'HR REV	R	ttaa**AAGCTT**ACATCCTTACAATCCTATATTTTCC	
set7	**set7_5'HR_FWD**	F	ttaa**GAATTC**GGAGAAAATATAGATGATGAAGTTGTAG	416 bp
	set7_5'HR_REV	R	CTAATTTCTTTTTTGTAAACATACATCTACC	
	set7_Rec_FWD	F	agATCCTGGGAGTGTTATTTTTGTAG	1262 bp
	set7_Rec_REV	R	atat**GGATCC**CTCCTCTATTTGTGGTGCATC	
	set7_3'HR_FWD	F	aggc**TCTAGA**ATATATTTAGCAATAACTTCAAAAA	597 bp
	set7_3'HR_REV	R	tgtt**AAGCTT**ATACTAACAATTTTTGTCCC	

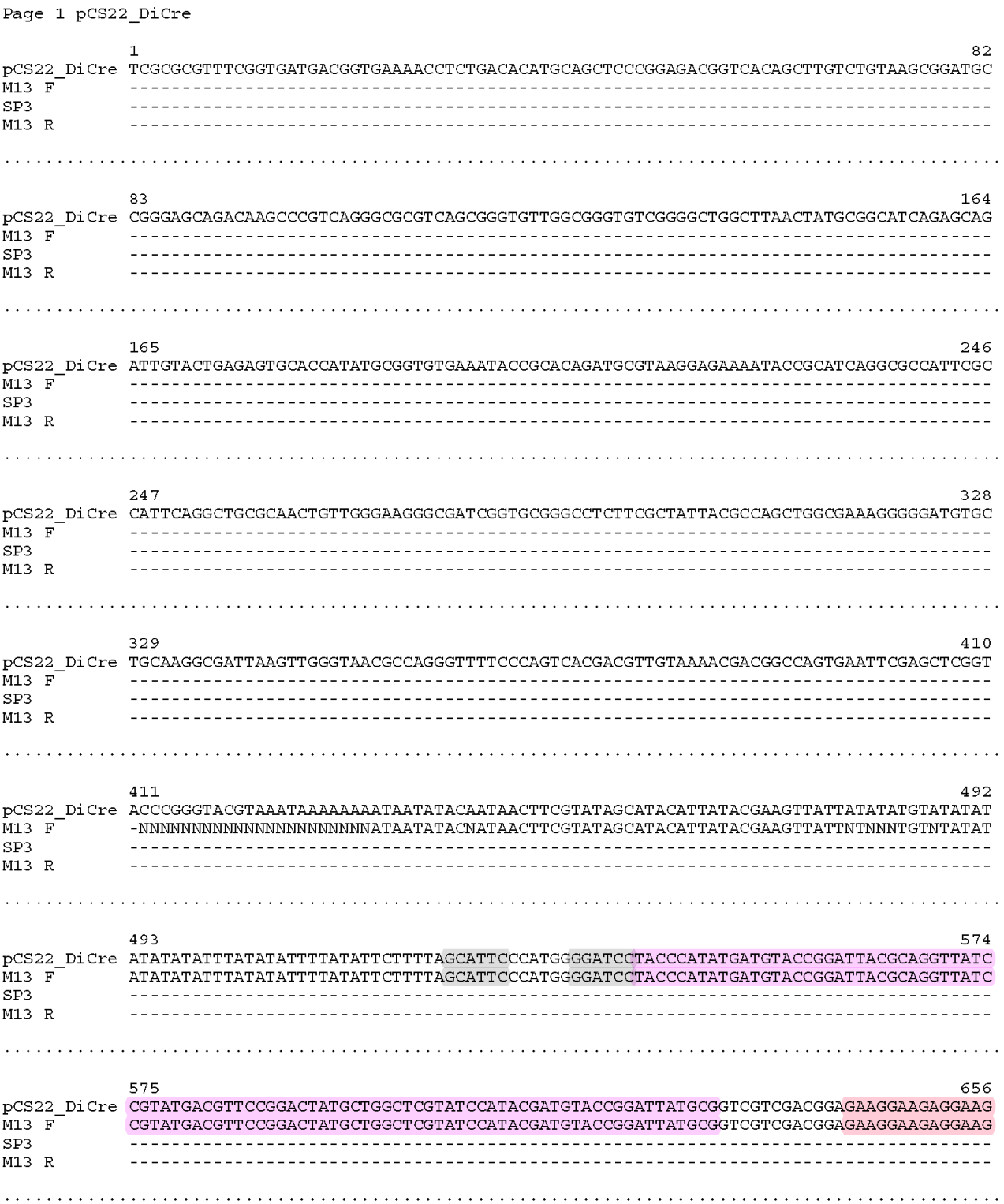
Supplementary Table 2: Primers used to screen for positive bacterial colonies and sequence positive integrants, in the process of generating the recombinant repair plasmid. Nucleotide overhangs are indicated in lowercase, while gene specific sequences are all uppercase. Restriction enzyme recognition sequences are indicated in bold, only for reference. Primers referred to as pCS22_DiCre is backbone specific, some located in the selection markers cassette as indicated in the "Location" column. The "amplifying/sequencing" column groups the primers that can be used for screening or sequencing the indicated region; this also depends on the cloning order. No stops codons present.

Purpose	Amplifying/ sequencing	Primer name	Oligo primer sequence (5' – 3')	Location
Colony screening and/or sequencing FORWARD	**5' HR + Recodonised**	M13 FWD	TGTAAAACGACGGCCAGT	pCS22_DiCre backbone
		gcn5 5'HR FWD	atta**GAATTC**GCTAACCCTAATCTTATATATGATCCAAATTATG	In gene sequence
		set7 5'HR FWD	ttaa**GAATTC**GGAGAAAATATAGATGATGAAGTTGTAG	In gene sequence
	Recodonised (C)	*gcn5* Rec FWD	agAAAATATGGGTATTATAACCTTTGAATG	In gene sequence
		set7 Rec FWD	agATCCTGGGAGTGTTATTTTTGTAG	In gene sequence
	3' HR	pCS22_DiCre 3'HR FWD	GCCCTTTCGAAAGATCCCAACG	In selection markers cassette (GFP-tag)
		gcn5 5'HR FWD	ttaa**TCTAGA**GATGGTCATTGAATATTTTTTTTT	In gene sequence
		set7 5'HR FWD	aggc**TCTAGA**ATATATTTAGCAATAACTTCAAAAA	In gene sequence
Colony screening and/or sequencing REVERSE	**5' HR**	*gcn5* 5'HR REV	CTTCTTTTGCTCCACCCGCATC	In gene sequence
		set7 5'HR REV	CTAATTTCTTTTTTGTAAACATACATCTACC	In gene sequence
	Recodonised (C)	*gcn5* Rec REV	tggc**GGATCC**TTTTGCTGTATCTGTTATCGC	In gene sequence
		set7 Rec REV	atat**GGATCC**CTCCTCTATTTGTGGTGCATC	In gene sequence
		pCS22 Rec REV	ACGACCGCATAATCCGGTAC	In selection markers cassette (HA-tag)
	3' HR	*gcn5* 3'HR REV	ttaa**AAGCTT**ACATCCTTACAATCCTATATTTTCC	In gene sequence
		set7 3'HR REV	tgtt**AAGCTT**ATACTAACAATTTTTGTCCC	In gene sequence
		M13 REV	CAGGAAACAGCTATGACCATG	pCS22_DiCre backbone

		pCS22 3'HR REV	ACCCCAGGCTTTACACTTTATGC	pCS22_DiCre backbone

Codes: Subscript numerals (**1/2/3)** – number of ways that desired region may be screened via PCR or sequencing; 5' – 5' region of gene; C – Recodonised gene region; **3'** – 3' region of gene; F – forward; R - reverse

Supplementary Data 1: Full consensus Sanger sequence alignment of pCS22_DiCre



Sequences highlighted in pink represent the HA-tag. Sequences highlighted in pale pink represent the T2A slip peptide.

```
Page 2 pCS22_DiCre

            657                                                                    738
pCS22_DiCre TTTATTAACATGTGGAGATGTAGAAGAAAATCCAGGACCAATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCT
M13 F       TTTATTAACATGTGGAGATGTAGAAGAAAATCCAGGACCAATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCT
SP3         ------------------------------------------------------------------------------
M13 R       ------------------------------------------------------------------------------

..............................................................................................

            739                                                                    820
pCS22_DiCre TGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGC
M13 F       TGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGC
SP3         ------------------------------------------------------------------------------
M13 R       ------------------------------------------------------------------------------

..............................................................................................

            821                                                                    902
pCS22_DiCre AGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTG
M13 F       AGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTG
SP3         ------------------------------------------------------------------------------
M13 R       ------------------------------------------------------------------------------

..............................................................................................

            903                                                                    984
pCS22_DiCre GCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAA
M13 F       GCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAA
SP3         ------------------------------------------------------------------------------
M13 R       ------------------------------------------------------------------------------

..............................................................................................

            985                                                                   1066
pCS22_DiCre GTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGC
M13 F       GTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGC
SP3         ------------------------------------------------------------------------------
M13 R       ------------------------------------------------------------------------------

..............................................................................................

            1067                                                                  1148
pCS22_DiCre ATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGG
M13 F       ATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGG
SP3         ------------------------------------------------------------------------------
M13 R       ------------------------------------------------------------------------------

..............................................................................................

            1149                                                                  1230
pCS22_DiCre TCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATG
M13 F       TCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATG
SP3         ------------------------------------------------------------------------------
M13 R       --------------GNNNNNNNNNTNNTGNNNNNANNNNNNTNNNNNNNNNNNNNNNNCNNNNNNNTNNNNNNNNNNNCNN

..............................................................................................

            1231                                                                  1312
pCS22_DiCre CCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGAT
M13 F       CCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGAT
SP3         ------------------------------------------------------------------------------
M13 R       NTNNCNNNNNNNGNNNNNGTCNNNNCCNNTNNNGATNCCTGCTTGNCGANNATNNTGGTGGAAAATNGCNNNTTTTNTGNNT

..............................................................................................
```

Sequences highlighted in pale pink represent the T2A slip peptide. Sequences highlighted in blue represent the neomycin resistance cassette.

```
Page 3 pCS22_DiCre

            1313                                                                1394
pCS22_DiCre TCATCGACTGTGGCCGGCTGGGTGTGGCGGAcCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGG
M13 F       TCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCNGNACATAGCGTTGGCTACCCGTGATATTGCTGAAGANCTTGG
SP3         ---------------NNNNNNNNNNNNNGCNNNNNCTATCNGGNNATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGG
M13 R       NCATCGNCTNNNGNCNGNTGGNNNNNGNGGACNGCTATCNGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGG

.....

            1395                                                                1476
pCS22_DiCre CGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTT
M13 F       CGGCNAATGGGCTGACCGCTTCCNCGNGCTTTANGGTATNNCCGCTCCCGATTCNNANCGCATCNCNTTNTATCNCNTTNNN
SP3         CGNNNAANGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTT
M13 R       CGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCNCNTTCTT

.....

            1477                                                                1558
pCS22_DiCre GACGAGTTCTTCTAActcgagGTAAATAAAAAAAATAATATACAATAACTTCGTATAGCATACATTATACGAAGTTATTATA
M13 F       GNCNNNNNNNTCTNNNCNAGNAATNNNNNNANNNNTANNNNNNNNNNNNNNAGCATACNTNANNCNNNNNNNNNNANNNNNN
SP3         GACGAGTTCTTCTAACTCGAGGTAAATAAAAAAAATAATATACAATAACTTCGTATAGCATACATTATACGAAGTTATTATA
M13 R       GACGAGTTCTTCTAACTCGAGGTAAATAAAAAAAATAATATACAATAACTTCGTATAGCATACATTATACGAAGTTATTATA

.....

            1559                                                                1640
pCS22_DiCre TATGTATATATATATATATTTATATATTTTATATTCTTTTAGaaATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCC
M13 F       NNNNNNNNNNTTTNNN----------------------------------------------------------------
SP3         TATGTATATATATATATATTTATATATTTTATATTCTTTTAGAAATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCC
M13 R       TATGTATATATATATATATTTATATATTTTATATTCTTTTAGAAATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCC

.....

            1641                                                                1722
pCS22_DiCre AATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGA
M13 F       ----------------------------------------------------------------------------
SP3         AATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGA
M13 R       AATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGA

.....

            1723                                                                1804
pCS22_DiCre AAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCAGTTCcatggccaacacttgtcactactttcgcgtatggtc
M13 F       ----------------------------------------------------------------------------
SP3         AAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCAGTTCCATGGCCAACACTTGTCACTACTTTCGCGTATGGTC
M13 R       AAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCAGTTCCATGGCCAACACTTGTCACTACTTTCGCGTATGGTC

.....

            1805                                                                1886
pCS22_DiCre ttcaatgctttgcgagatacccagatcatatgaaacagcatgacttttttcaagagtgccatgcccgaaggttatgtacagga
M13 F       ----------------------------------------------------------------------------
SP3         TTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGA
M13 R       TTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGA

.....

            1887                                                                1968
pCS22_DiCre aagaactatatttttcaaagatgacgggaactacaagacacgtgctgaagtcaagtttgaaggtgataccccttgttaataga
M13 F       ----------------------------------------------------------------------------
SP3         AAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGA
M13 R       AAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGA

.....
```

Sequences highlighted in blue represent the neomycin resistance cassette. Sequences highlighted in red represent the SERA2/*loxP* site. Sequences highlighted in green represent the GFP-tag.

```
Page 4 pCS22_DiCre

            1969                                                                    2050
pCS22_DiCre atcgagttaaaaggtattgattttaaagaagatggaaacattcttggacacaaattggaatacaactataactcacacaatg
M13 F       ------------------------------------------------------------------------------------
SP3         ATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATG
M13 R       ATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATG

.....

            2051                                                                    2132
pCS22_DiCre tatacatcatggcagacaaacaaaagaatggaatcaaagttaacttcaaaattagacacaacattgaagatggaagcgttca
M13 F       ------------------------------------------------------------------------------------
SP3         TATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCA
M13 R       TATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCA

.....

            2133                                                                    2214
pCS22_DiCre actagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtccacacaa
M13 F       ------------------------------------------------------------------------------------
SP3         ACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAA
M13 R       ACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAA

.....

            2215                                                                    2296
pCS22_DiCre tctgccctttcgaaagatcccaacgaaaagagagaccacatggtccttcttgagtttgtaacagctgctgggattacacatg
M13 F       ------------------------------------------------------------------------------------
SP3         TCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGANCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATG
M13 R       TCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGNNNNNNNNAGCTGCTGGGATTACACATG

.....

            2297                                                                    2378
pCS22_DiCre gcatggatgaActAtacaaaTAATCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTG
M13 F       ------------------------------------------------------------------------------------
SP3         GNATGGNNGAACTATACAAATAATCTANAGTCGANCNGCNGNNATGCAAGCTTNGNNTAANCNTGGNCATAGCTGTTTNCNG
M13 R       GCATGGATGAACTATACAAATAATCTAGANNCGNCNNNNNGNNNNNNNNNN----------------------------

.....

            2379                                                                    2460
pCS22_DiCre TGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGT
M13 F       ------------------------------------------------------------------------------------
SP3         NGNGAANNNNNNNTCCGCNNNNNATTCCNCACANNATACNANNNNNNNNATNANGNGNNNNNNNNNGNNNNNNNNNNNNANC
M13 R       ------------------------------------------------------------------------------------

.....

            2461                                                                    2542
pCS22_DiCre GAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATC
M13 F       ------------------------------------------------------------------------------------
SP3         TANNNNNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAANNNNNN--------------------
M13 R       ------------------------------------------------------------------------------------

.....

            2543                                                                    2624
pCS22_DiCre GGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC
M13 F       ------------------------------------------------------------------------------------
SP3         ------------------------------------------------------------------------------------
M13 R       ------------------------------------------------------------------------------------

.....
```

Sequences highlighted in green represent the GFP-tag.

Supplementary Data 2: Full consensus Sanger sequence alignment of pCS22_DiCre_*gcn5*

```
Page 1 pCS22_DiCre_gcn5

             1                                                                                82
GCN5 Fina... --------------------------------------------------------------------------------
SP1          --------------------------------------------------------------------------------
5' HR R      NNNNNNNNNNNNNNNNTNNNCNNNAANNNNNNNCNGNNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAAANNNNNGNNNNN
SP5          --------------------------------------------------------------------------------
SP6          --------------------------------------------------------------------------------

................................................................................................

             83                                                                              164
GCN5 Fina... ----TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGG
SP1          --------------------------------------------------------------------------------
5' HR R      NGNNNNNNNNNNNCGNNNNTTNNGNGATGNNNNNNANNNNTGACACATGCAGNNNCNNAGNNNTCNCAGNNNTNNNTNNNCNN
SP5          --------------------------------------------------------------------------------
SP6          --------------------------------------------------------------------------------

................................................................................................

             165                                                                             246
GCN5 Fina... ATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGA
SP1          --------------------------------------------------------------------------------
5' HR R      ATNCNGGNANCAGNCNAGCCCGTCNGGGCGNNTCAGCGGGTGNNNNCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGA
SP5          --------------------------------------------------------------------------------
SP6          --------------------------------------------------------------------------------

................................................................................................

             247                                                                             328
GCN5 Fina... GCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCAT
SP1          --------------------------------------------------------------------------------
5' HR R      GCAGATTGTACTGAGAGTGCNCCATATGCGGTGTGAAATNCCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCNT
SP5          --------------------------------------------------------------------------------
SP6          --------------------------------------------------------------------------------

................................................................................................

             329                                                                             410
GCN5 Fina... TCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGAT
SP1          ------------------------------------------------------------------------NNNNNNNNNNN
5' HR R      TCGCCNTTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTNTTCGCTATTACGCCAGCTGGCGAAAGGGGGAT
SP5          --------------------------------------------------------------------------------
SP6          --------------------------------------------------------------------------------

................................................................................................

             411                                                                             492
GCN5 Fina... GTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGCTAA
SP1          NNGCNGCNNGGCGANTAAGTTGGGTAACGCNNGGGTTTTCCCAGTCACGACGTTGTAAAA-NNNNNNNNNNGNATTCGCTAA
5' HR R      GTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGCTAA
SP5          --------------------------------------------------------------------------------
SP6          --------------------------------------------------------------------------------

................................................................................................

             493                                                                             574
GCN5 Fina... CCCTAATCTTATATATGATCCAAATTATGAAACCATATATTCTAAATGGAAAACATTTTTAAGAAAAGAGCAATCAAGTGGA
SP1          CCCTAATCTTATATATGATCCAAATTNTGAAACCATATATTCTAAATGGAAAACATTTTTAAGAAAAGAGCAATCAAGTGGA
5' HR R      CCCTAATCTTATATATGATCCAAATTATGAAACCATATATTNTAAATGGAAAACATTTTTAAGAAAAGAGCAATCAAGTGGA
SP5          --------------------------------------------------------------------------------
SP6          --------------------------------------------------------------------------------

................................................................................................
```

Sequences highlighted in blue represent the 5' HR of *gcn5*.

Page 2 pCS22_DiCre_gcn5

```
              575                                                                    656
GCN5 Fina...  AATTTAATTAGTAATTGTTTTTCACGAGATTTTTTACATACTATTTTGTTATGTAATTATGTAACTATAATTGAAGATTTAA
SP1           AATTTAATTAGTAATTGTTTTTCACGAGATTTTTTACATACTATTTTGTTATGTAATTATGTAACTATAATTGAAGATTTAA
5' HR R       AATTTAATTAGTAATTGTTTTTCACGAGATTTTTTACATANTATTTTGTTATGTAATTATGTAACTATAATTGAAGATTTAA
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

..............................................................................................

              657                                                                    738
GCN5 Fina...  GAAAAACGGCCGTTAAGAAAAAGTTGAAATATTTTTTCTTACACTTATGTTTGGAATCGGGCATATCGATTAACGTTGCATT
SP1           GAAAAACGGCCGTTAAGAAAAAGTTGAAATATTTTTTCTTACCCTTATGTTTGGAATCGGGCATATCGATTAACGTTGCATT
5' HR R       GAAAAACGGCCGTTAAGAAAAAGTTGAAATATTTTTTNTTACCCTTATGTTTGGAATCGGGCATATCGATTAACGTTGCATT
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

..............................................................................................

              739                                                                    820
GCN5 Fina...  GATGCTGTTTATAAATGCCACGAAGCAGAGCGATAAGTTACAGgtattaagagcatgttggttcataattatatgaacgata
SP1           GATGCTGTTTATAAATGCCACGAAGCAGAGCGATAAGTTACAGGTATTAAGAGCATGTTGGTTCATAATTATATGAACGATA
5' HR R       GATGCTGTTTATAAATGCCNCGAAGCAGAGCGATAAGTTACAGGTATTAAGAGCATGTTGGTTCATAATTATATGAACGATA
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

..............................................................................................

              821                                                                    902
GCN5 Fina...  ttgtatgtatgaatgtgtatatatgaataaatagataaataattaaatatatatatatatatatatatatgaatatgtgtat
SP1           TTGTATGTATGAATGTGTATATATGAATAAATAGATAAATAATTAAATATATATATATATATATATATGAATATGTGTATAT
5' HR R       TTGTATGTATGAATGTGTATATATGAATAAATAGATAAATAATTAAATATATATATATATATATATNTGAATATGTGTATNT
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

..............................................................................................

              903                                                                    984
GCN5 Fina...  attttttttttttttttttctttttatgtgtaaagTCTCTACTACCCTCCGAAACAGGATTAGGATATTTGCACAGAGATGC
SP1           -TTTTTTTTTTTTTTTTTTCTTTTTATGTGTAAAGTCTCTACTACCCTCCNAAACAGGATTAGGATATTTGCACAGAGATGC
5' HR R       -TTTTTTTTTTTTTTTTTTCTTTTTATGTGTAAAGTCTCTACTACCCTCCGAAACAGGATTAGGATATTTGCACAGAGATGC
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

..............................................................................................

              985                                                                   1066
GCN5 Fina...  GGGTGGAGCAAAAGAAGGTAAATAAAAAAAATAATATACAATAACTTCGTATAGCATACATTATACGAAGTTATTATATATG
SP1           GGGNGGAGCAAAANAAGGTAAATAAAAAAAATAATATACAATAACTTCGTATAGCATACATTATACNAAGTTATTATATATG
5' HR R       GGGTGGAGCAAAAGAAGGTAAATAAAAAAAATAATATACAATAACTTCGTATAGCATACATTATACGAAGTTATTATATATG
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

..............................................................................................

              1067                                                                  1148
GCN5 Fina...  TATATATATATATATTTATATATTTTATATTCTTTTAGAA--AATATGGGTATTATAACCTTTGAATGTATAACAAATGAcA
SP1           TATATATATATATATTTATATATTTTATATTCTTTTAGCATTCCCATGGGGATCCTACCCATATGATGTACCGGATTACGCA
5' HR R       TATATATATATATATTTATATATTTTATATTCTTTTAGCATNNNCNNGGGGATCCTACCCATATGATGTACCGGATTACGCA
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

..............................................................................................
```

Sequences highlighted in red represent the SERA2/*loxP* site. Sequences highlighted in orange represent the recodonised gene insert site, however, no insert was detected.

```
Page 3 pCS22_DiCre_gcn5

              1149                                                                1230
GCN5 Fina...  GaGAACCAGATCATTTAATCAAATTAATTACGCTGAAAAATAT-----------TTTTTCCAGACAATTACCCAAAATGCC
SP1           GGTTATCCGTATGACGTTCCGGACTATGCTGGCTCGTATCCATACGATGTACCGGATTATGCGGTCGTCGACGGAGAAGGAA
5' HR R       GGTTATCCG-----------------------------------------------------------TANGACGTC
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

................................................................................................

              1231                                                                1312
GCN5 Fina...  GAGAGAGTATATT-GTACGTTTAGTTTTTGATAGAAATCATTATACATTTTGTTTGTTAAAGAAAAATACTGTTATAGGTGG
SP1           GAGGAAGTTTATTAACATGTGGAGATGTAGAAGAAAATC-------------------------------------------
5' HR R       CGGNNNNTGCNGGNNNNNN--------------------------------------------------------------
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

................................................................................................

              1313                                                                1394
GCN5 Fina...  TGTATGTTTTAGACCATATTTTGAACAAAAATTCGCAGAAATAGCTTTCTTAGCTGTGACATCTACAGAGCAAGTAAAGGGG
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

................................................................................................

              1395                                                                1476
GCN5 Fina...  TATGGAACTAGGTTGATGAATCACCTTAAGGAACATGTAAAAAAATTTGGCATTGAATATTTTTTAACATATGCTGATAATT
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

................................................................................................

              1477                                                                1558
GCN5 Fina...  TTGCTATAGGGTACTTCAGGAAGCAAGGATTTTCACAAAAAATTAGTATGCCTAAAGAAAGATGGTTTGGTTATATCAAAGA
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

................................................................................................

              1559                                                                1640
GCN5 Fina...  CTATGACGGAGGAACTCTAATGGAATGTTATATATTTCCTAATATAAATTATTTACGGCTAAGTGAAATGTTATATGAACAG
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

................................................................................................

              1641                                                                1722
GCN5 Fina...  AAAAAAGCCGTTAAGAAAGCTATACATTTTATTAAACCTCAGGTTATTTACAAGGGTATAAACTATTTTGCAGATAATAAAG
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
SP5           --------------------------------------------------------------------------------
SP6           --------------------------------------------------------------------------------

................................................................................................
```

Sequences highlighted in orange represent the recodonised gene insert site, however, no insert was detected. The full sequence is not shown.

```
Page 7 pCS22_DiCre_gcn5

              3445                                                                                3526
GCN5 Fina...  agcatgacttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatatttttcaaagatgacgggaactacaa
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           ----------------------------------------------------------------------------------------
SP6           NCNNNNNNNNNNNNNNNNNNNNNNCNNNNGAANNNNNNNNNTNTTTTNNNNANNNNGNNNNNGNNNNNNNNNNANNCNNGNN

..............................................................................................................

              3527                                                                                3608
GCN5 Fina...  gacacgtgctgaagtcaagtttgaaggtgatacccttgttaatagaatcgagttaaaaggtattgattttaaagaagatgga
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           ----------------------------------------------------------------------------------------
SP6           NNANNNNNNNNANNNNNNNNNNGNNNANCCNNNTNNNNNNNANNNNNGNNNNNNANNNNNTNNNNNTTNNANNANNNNNNNN

..............................................................................................................

              3609                                                                                3690
GCN5 Fina...  aacattcttggacacaaattggaatacaactataactcacacaatgtatacatcatggcagacaaacaaaagaatggaatca
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           ----------------------------------------------------------------------------------------
SP6           NNNNNNNTNNNNNACANNNNNNNNNNNNANNNNTANNTCTCACNNNNTNTANANNNCNNCNNACANNNANAAGANNNNNNNNN

..............................................................................................................

              3691                                                                                3772
GCN5 Fina...  aagttaacttcaaaattagacacaacattgaagatggaagcgttcaactagcagaccattatcaacaaaatactccaattgg
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           ----------------------------------------------------------------------------------------
SP6           NNNNNNNNTNNTNNANNNNACACANNNNNNNNGAGNNNAGNNNTNTNNNAGCNNNNNNCNNTCTNNNNAATANNNNNNNNNN

..............................................................................................................

              3773                                                                                3854
GCN5 Fina...  cgatggccctgtccttttaccagacaaccattacctgtccacacaatctgcccttcgaaagatcccaacgaaaagagagac
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           ----------------------------------------------------------------------------------------N
SP6           NGNGGNNNNNNTNNNNTTNNNNCACANNNNCNNNNNNTNNNCACACTNTNNNNNNNNNNNNANANNNNNNNNNNNAGAGAGAG

..............................................................................................................

              3855                                                                                3936
GCN5 Fina...  cacatggtccttcttgagtttgtaacagctgctgggattacacatggcatggatgaActAtacaaaTAATCTAGAgatggtc
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           NNNNNNNNNNNNCTTGANTTTGTAACNGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAANCNNNAgAtGGTC
SP6           CACACNNNCTCTNTNGNGTNTNTNNNNCNNNNNGNGNNNNCNCANNNCNCNNNNNNNNTATACNNNNANTNTNGAGANNNNN

..............................................................................................................

              3937                                                                                4018
GCN5 Fina...  attgaatatttttttttttttttctctttttcgtaaaggaaaatgtatacatatatatatatatatatatatatatatatata
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           ATTGAATATTTTTTTTTTTTTTTCtCTTTTTTCGTAAAGGAAAATGTATACATATATATATATATATATATATATATATN
SP6           NNNNNNNATNTTTTTTTTTTTTTTNTCTNTTNTNNNNNNNNNNNANNTNNNNANATATATATATATATATATATATATA

..............................................................................................................
```

Sequences highlighted in grey represent the *Xba*I site. Sequences highlighted in green represent the 3' HR for *gcn5*.

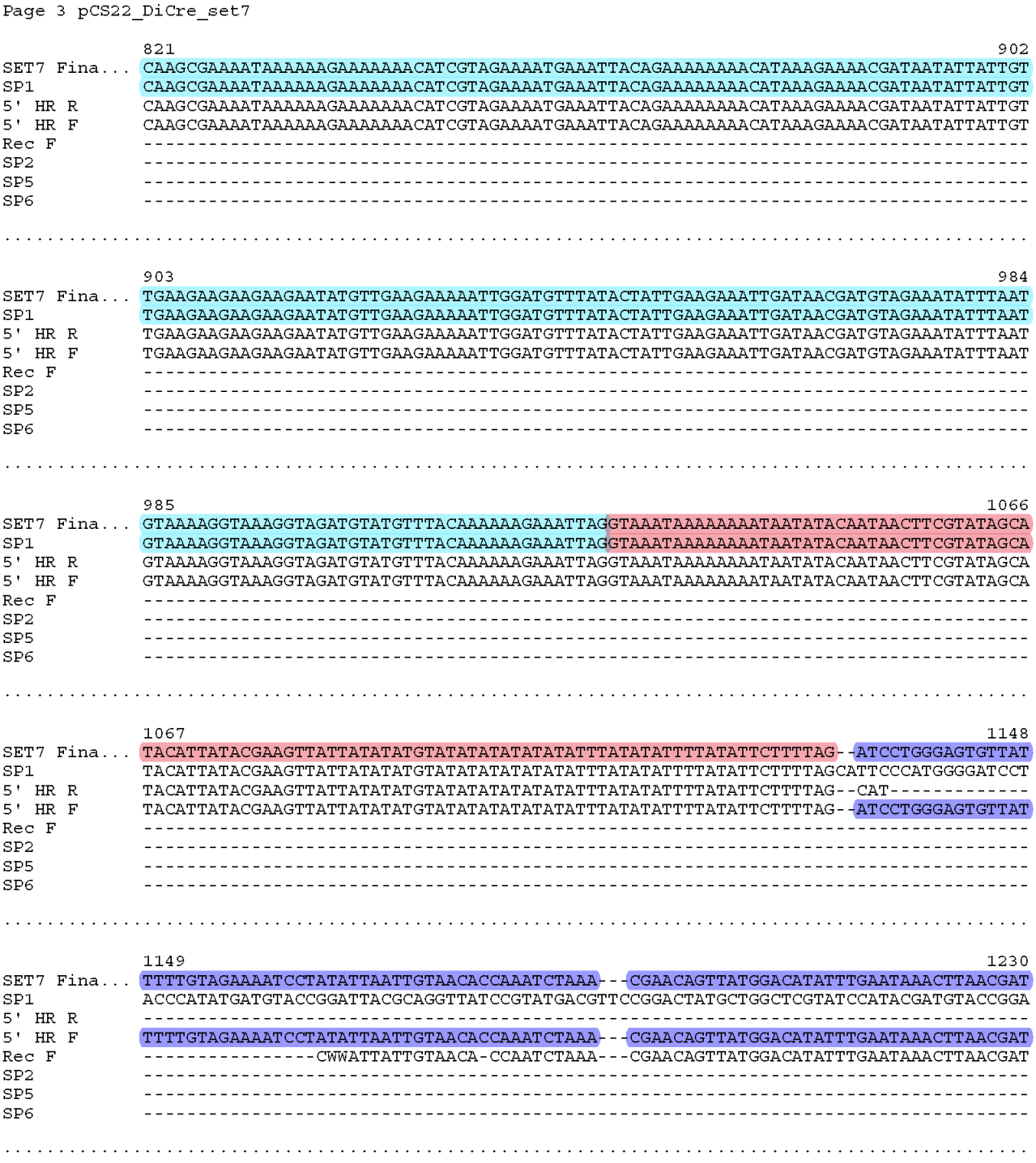
```
Page 8 pCS22_DiCre_gcn5

              4019                                                                                4100
GCN5 Fina...  ttaataaaaattaccatatacataacatatatcattatttttttatttgatccaattaataatacactccttttttatttttt
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           NNNANAAANNNCCCCATANACANNACATATNTCNNTTTTTTTTTNTNNNNCCCNNNNNANAANACNCNNNTTTTTTNTTTGN
SP6           TTAATAAAAATTACCATATACATAACATATATCATTATTTTTTTATTTGATCCAATTAATAATACACTCCTTTTTTATTTTT

..........................................................................................................

              4101                                                                                4182
GCN5 Fina...  gggaaaataaaaaagatgtatatatatttttttatatttgtataccagataagcatatggttttattatacagtgaatttttt
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           GNNAAAANAAAANANNNNNATATATNTTTTTTTATNNNNNNNNNNNNANANNNNNNNNNNNNNNNNNNNNNNNNNNNNTNN
SP6           GGGAAAATAAAAAAGATGTATATATATTTTTTTATATTTGTATACCAGATAAGCATATGGTTTTATTATACAGTGAATTTTT

..........................................................................................................

              4183                                                                                4264
GCN5 Fina...  tattggtcatttgtttatctcaatacatttaatcatttacatatatatatatatatatatatatatatatatatatttcttttttat
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           NNNNNNNNNNNNNNNNTNNNNNNNNNNNNTNNNNNNNTNNANATNTATATATATATATATATATATATATATNNNNNNNNTNTNNNN
SP6           TATTGGTCATTTGTTTATCTCAATACATTTAATCATTTACATATATATATATATATATATATATATATATATTCTTTTTAT

..........................................................................................................

              4265                                                                                4346
GCN5 Fina...  tgatgtgattatatgtaatgaatacaccctatatacatttattccgaaataatttttttttattcttttttttatattatttttt
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNNAANNNTTTTNNNNNNNNNNTTTNTATNNNNNNTTT
SP6           TGATGTGATTATATGTAATGAATACACCCTATATACATTTATTCCGAAATAATTTTTTTTTATTCTTTTTTTTATATTATTTTT

..........................................................................................................

              4347                                                                                4428
GCN5 Fina...  tttaatttttatttttttttgtccgtttgaaaaagacgcgattaaaaaataatttatgcatcaacttatatttttttttaactt
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           TNNNNTTNNNNNNTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTNTTTTTTNNNNNNN
SP6           TTTAATTTTTATTTTTTTTTGTCCGTTTGAAAAAGACGCGATTAAAAAATAATTTATGCATCAACTTATATTTTTTTTAACTT

..........................................................................................................

              4429                                                                                4510
GCN5 Fina...  atttttataatatttttttaaagaatatcaatgaatttgtaaatttttttaaagcaaacagatatgcttaaatttttttaaagg
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           NNNNNNNNNNNNTANNTTNNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
SP6           ATTTTTTATAATATTTTTTTAAAGAATATCAATGAATTTGTAAATTTTTTTAAAGCAAACAGATATGcTTAAATTTTTTAAAGG

..........................................................................................................

              4511                                                                                4592
GCN5 Fina...  aaaatataggattgtaaggatgtAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAA
SP1           ----------------------------------------------------------------------------------------
5' HR R       ----------------------------------------------------------------------------------------
SP5           ANNNNNNNNANNNNNNNNNNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGNNNNNNNNNNNNNNNNNNNNN
SP6           AAAATATAGGATTGTAAGGATGtAaGNTTNGCGTAATCATGGTCATAGCTGTTTCCTGTGNGAAATTGTTNTCCGCTCACAN

..........................................................................................................
```

Sequences highlighted in green represent the 3' HR for *gcn5*. Sequences highlighted in grey represent the *Hind*III site.

Supplementary Data 3: Full consensus Sanger sequence alignment of pCS22_DiCre_*set7*

```
Page 2 pCS22_DiCre_set7

              411                                                                 492
SET7 Fina...  GTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAA
SP1           ------------------------------------------------------------------------------
5' HR R       GTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAA
5' HR F       ------------------------------------------------------------------------------
Rec F         ------------------------------------------------------------------------------
SP2           ------------------------------------------------------------------------------
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              493                                                                 574
SET7 Fina...  GGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC
SP1           ----------------------------------------NNNNNNNNNNNNNNGCTGC-AGGCGATTAAGTTGGGTAACGC
5' HR R       GGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC
5' HR F       ------------------------------------------------------------------------------
Rec F         ------------------------------------------------------------------------------
SP2           ------------------------------------------------------------------------------
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              575                                                                 656
SET7 Fina...  CAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGGAGAAAATATAGATGATGAAGTTGTAGATAACGA
SP1           CAGGGTTTTCCCAGTCACGACGTTGTAAAANNNNNNNNNGNGAATTCGGAGAAAATATAGATGATGAAGTTGTAGATAACGA
5' HR R       CAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGGAGAAAATATAGATGATGAAGTTGTAGATAACGA
5' HR F       ------------------------------------------------------------------------NNNNN
Rec F         ------------------------------------------------------------------------------
SP2           ------------------------------------------------------------------------------
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              657                                                                 738
SET7 Fina...  ATTTGTAGATAACGAATTTGTAGATAATGAAAATGCTGATATGGAAGAGGTGAATATGGAAAATATCGATGTGCAATATGTT
SP1           ATTTGTAGATAACGAATTTGTAGATAATGAAAATGCTGATATGGAAGAGGTGAATATGGAAAATATCGATGTGCAATATGTT
5' HR R       ATTTGTAGATAACGAATTTGTAGATAATGAAAATGCTGATATGGAAGAGGTGAATATGGAAAATATCGATGTGCAATATGTT
5' HR F       NNNNNANNNNAACGAATTTGTAGANAANGAAANTGCTGATATGGAAGAGGTGAATANNNNNNNTANNGATGTGCAATATGTT
Rec F         ------------------------------------------------------------------------------
SP2           ------------------------------------------------------------------------------
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              739                                                                 820
SET7 Fina...  GATGAGGAAAATATGTATGCAGAAGAACTAGAGAAGAAATATCCTAATAATGAAAATATGAACGATGTAAAAGTTACATTAC
SP1           GATGAGGAAAATATGTATGCAGAAGAACTAGAGAAGAAATATCCTAATAATGAAAATATGAACGATGTAAAAGTTACATTAC
5' HR R       GATGAGGAAAATATGTATGCAGAAGAACTAGAGAAGAAATATCCTAATAATGAAAATATGAACGATGTAAAAGTTACATTAC
5' HR F       GATGAGGAAAATATGTATGCAGAAGAACTAGAGAAGAAATATCCTAATAATGAAAATATGAACGATGTAAAAGTTACATTAC
Rec F         ------------------------------------------------------------------------------
SP2           ------------------------------------------------------------------------------
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................
```

Sequences highlighted in grey represent the *Eco*RI site. Sequences highlighted in blue represent the 5'HR insert.



Sequences highlighted in blue represent the 5' HR insert. Sequences highlighted in red represent the SERA2/*loxP* site. Sequences highlighted in blue represent the recodonised gene insert site.

```
Page 4 pCS22_DiCre_set7

                1231                                                                1312
SET7 Fina...    GAACAAAATTTTGAATTACCATTAAAgTGGCAcTAcGCTGCGTTATGTAGTATCACAATGTTAAATGATTTTAATTATAAAG
SP1             TTATGCGGTCGTCGACGGAGAAGGAAGAGGAAGTTTATTAACATGTGGAGATGTAGAAGAAAATCCAGGACCAATGATTGAA
5' HR R         ------------------------------------------------------NNNCNNGNNGATCCTACCCATATGA
5' HR F         GAACAAAATTTTGAATTACCATTAAAGTGGCACTACGCTGCGTTATGTAGTATCACAATGTTAAATGATTTTAATTATAAAG
Rec F           GAACAAAATTTTGAATTACCATTAAAGTGGCACTACGCTGCGTTATGTAGTATCACAATGTTAAATGATTTTAATTATAAAG
SP2             ------------------------------------------------------------------------------------
SP5             ------------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------------

....................................................................................................

                1313                                                                1394
SET7 Fina...    CATGTTTAG--ATAAGTGGGTTCCTGAACCTGATAAAGAACCAGATAATGATATATACAATGTCCTAGACAAAGTATGTGAA
SP1             CAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTAT-------------------------------
5' HR R         TGTACCGGA--TTACGCAGGTT---------------------------ATCCGTANGACGTCCGGNNNNTGCNNNNNNNN
5' HR F         CATGTTTAG--ATAAGTGGGTTCCTGAACCTGATAAAGAACCAGATAATGATATATACNATGTCCTAGACAAAGTATGTGAA
Rec F           CATGTTTAG--ATAAGTGGGTTCCTGAACCTGATAAAGAACCAGATAATGATATATACAATGTCCTAGACAAAGTATGTGAA
SP2             ------------------------------------------------NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
SP5             ------------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------------

....................................................................................................

                1395                                                                1476
SET7 Fina...    AAAACATCATTCGTAAATGGAAATAAGTATTACTACTATAAAAATAAATTAATAGATCCAAAAATTTATAGTAGGATAATTC
SP1             ------------------------------------------------------------------------------------
5' HR R         NN----------------------------------------------------------------------------------
5' HR F         AAAACATCATTCGTAANTGGNANTNNNTATTACTACTATAAA---ATAATTANNGATCC-NAAANTTATAGTNGGANNATTC
Rec F           AAAACATCATTCGTAAATGGAAATAAGTATTACTACTATAAAAATAAATTAATAGATCCAAAAATTTATAGTAGGATAATTC
SP2             NNNNNNNNNNANNNNNNTNGNNNNNNNNNNNNNNTATTNNNNNNANNAANNNNNNNNNNGATNCNANNNTTNNTNNNNGNNNN
SP5             ------------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------------

....................................................................................................

                1477                                                                1558
SET7 Fina...    AAGTATGGCATTATAATGCTTTTGGACATCATACTGACAATGAAGGATTAGTGTTATATAATAGAATCAGTATGTTGGCTCA
SP1             ------------------------------------------------------------------------------------
5' HR R         ------------------------------------------------------------------------------------
5' HR F         ANTATGGCATTNNANGCTTTTNGNCNTCNTACTGANNNNNNNNNTNNNGNNNNNNNNNNNANNANCAGTNTGNTNGCNCNTNAT
Rec F           AAGTATGGCATTATAATGCTTTTGGACATCATACTGACAATGAAGGATTAGTGTTATATAATAGAATCAGTATGTTGGCTCA
SP2             NNANNTNNNNNNNTNATGCTTTNNNACATCATNCTGNCNANNNNGGANNANNNNNNNNNANNNNNNTCAGNNNGNNGGCNCA
SP5             ------------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------------

....................................................................................................

                1559                                                                1640
SET7 Fina...    TTCATGTATTAGTACAGCCTGTTGGCACTATGGTGAAAACGATTCATTCGTTTTACGTGCTCGTATAAACTTAAATCCGGGC
SP1             ------------------------------------------------------------------------------------
5' HR R         ------------------------------------------------------------------------------------
5' HR F         NNNNTNNNACANNNNNNNNNNANGNNNAAANNATNNTNNTTTNNNNNNNCNNNNNNNNANNNNNNNNATGANNNNCNNNNNN
Rec F           TTCATGTATTAGTACAGCCTGTTGGCACTATGGTGAAAACGATTCATTCGTTTTACGTGCTCGTATAAACTTAAATCCGGGC
SP2             TTCANGTANTAGTNCAGCCNNNTGGNNNNNANGNNGAAAACGNTNCATTCGTTTTACGTGCTNGTATAAACTNAAANCNGGGC
SP5             ------------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------------

....................................................................................................
```

Sequences highlighted in blue represent the recodonised gene insert site.

```
Page 5 pCS22_DiCre_set7

                1641                                                                1722
SET7 Fina...    GATGAGATAACTATATCATATTTAGGAGATGATGATTTGTATAAAAGCTCCAATATAAGAAGAGAAAAATTGACTAATTGGT
SP1             ------------------------------------------------------------------------------
5' HR R         ------------------------------------------------------------------------------
5' HR F         NNTANNNANNNNNNNNNNAANNNNNNTTNNNNNNNAANNNNNNNNNNNNTNNNNNNCNNNNNNNNNNNNNNNNNN--------
Rec F           GATGAGATAACTATATCATATTTAGGAGATGATGATTTGTATAAAAGCTCCAATATAAGAAGAGAAAAATTGACTAATTGGT
SP2             GATGAGATNANTATATCATATTTAGGAGATGATGATTTGTATAAAAGNNCCAATATAAGAAGAGAAAAATTGACTAATTGGT
SP5             ------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------

..............................................................................................

                1723                                                                1804
SET7 Fina...    TATTTGTTTGTATGTGTTCGAGATGTACACATCCTGTTGATAATTGTAGAGGTTTTAGATGCAGTAGTTGTGGTATAGGTAC
SP1             ------------------------------------------------------------------------------
5' HR R         ------------------------------------------------------------------------------
5' HR F         ------------------------------------------------------------------------------
Rec F           TATTTGTTTGTATGTGTTCGAGATGTACACATCCTGTTGATAATTGTAGAGGTTTTAGATGCAGTAGTTGTGGTATAGGTAC
SP2             TATTTGTTTGTATGTGTTCGAGATGTACACATCCTGTTGATAATTGTAGAGGTTTTAGATGCAGTAGTTGTGGTATAGGTAC
SP5             ------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------

..............................................................................................

                1805                                                                1886
SET7 Fina...    TTTTTTTATAAAATCTGAATATCACGATGATATTCCCATTATTTCTAAATGTAATGTGTGCTTGTGTGAAATTACGGAAAGT
SP1             ------------------------------------------------------------------------------
5' HR R         ------------------------------------------------------------------------------
5' HR F         ------------------------------------------------------------------------------
Rec F           TTTTTTTATAAAATCTGAATATCACGATGATATTCCCATTATTTCTAAATGTAATGTGTGCTTGTGTGAAATTACGGAAAGT
SP2             TTTTTTTATAAAATCTGAATATCACGATGATATTCCCATTATTTCTAAATGTAATGTGTGCTTGTGTGAAATTACGGAAAGT
SP5             ------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------

..............................................................................................

                1887                                                                1968
SET7 Fina...    GCAGCATATGAATATATTGAATATGAAAATAGTTATATAGAAAGGTTACAAGAAACAGATAAAAATGATTTATCTGATGCTT
SP1             ------------------------------------------------------------------------------
5' HR R         ------------------------------------------------------------------------------
5' HR F         ------------------------------------------------------------------------------
Rec F           GCAGCATATGAATATATTGAATATGAAAATAGTTATATAGAAAGGTTACAAGAAACAGATAAAAATGATTTATCTGATGCTT
SP2             GCAGCATATGAATATATTGAATATGAAAATAGTTATATAGAAAGGTTACAAGAAACAGATAAAAATGATTTATCTGATGCTT
SP5             ------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------

..............................................................................................

                1969                                                                2050
SET7 Fina...    TAGCTGTTTTTGTACAAGCAGATAAAATATTTACGCAACATTGGATAATGTTTCAATTATATACAATTCTTTTCGAAGGTTA
SP1             ------------------------------------------------------------------------------
5' HR R         ------------------------------------------------------------------------------
5' HR F         ------------------------------------------------------------------------------
Rec F           TAGCTGTTTTTGTACAAGCAGATAAAATATTTACGCAACATTGGATAATGTTTCAATTATATACAATTCTTTTCGAAGGTTA
SP2             TAGCTGTTTTTGTACAAGCAGATAAAATATTTACGCAACATTGGATAATGTTTCAATTATATACAATTCTTTTCGAAGGTTA
SP5             ------------------------------------------------------------------------------
SP6             ------------------------------------------------------------------------------

..............................................................................................
```

Sequences highlighted in blue represent the recodonised gene insert site.

```
Page 6 pCS22_DiCre_set7

              2051                                                                2132
SET7 Fina...  TAGGGATACATCTCAATGGGAAAAGGCTATATATTATCAAATGCAAAGAATTAAATATGCAATTGATGTAATACCAAGAGCA
SP1           ------------------------------------------------------------------------------
5' HR R       ------------------------------------------------------------------------------
5' HR F       ------------------------------------------------------------------------------
Rec F         TAGGGATACATCTCAATGGGAAAAGGCTATATATTATCAAATGCAAAGAATTAAATATGCAATTGATGTAATACCAAGAGCA
SP2           TAGGGATACATCTCAATGGGAAAAGGCTATATATTATCAAATGCAAAGAATTAAATATGCAATTGATGTAATACCAAGAGCA
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              2133                                                                2214
SET7 Fina...  AATTATGTTCTAGCATGGTTATATGAAGAGCTTGGAGAAATACATGCCAATTCAATATCCACGGATATATTGAGCACAGAAA
SP1           ------------------------------------------------------------------------------
5' HR R       ------------------------------------------------------------------------------
5' HR F       ------------------------------------------------------------------------------
Rec F         AATTATGTTCTAGCATGGTTATATGAAGAGCTTGGAGAAATACATGCCAATTCAATATCCACGATATATTGAGCACAGAAAA
SP2           AATTATGTTCTAGCATGGTTATATGAAGAGCTTGGAGAAATACATGCCAATTCAATATCCACGGATATATTGAGCACAGAAA
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              2215                                                                2296
SET7 Fina...  ATGACTTTACCATAACATTTGAAGAAAAAAAAAAGAATTTGTTCTCATTTTTTAAAGTCTATTCATTTATTAGAAATTTTATG
SP1           ------------------------------------------------------------------------------
5' HR R       ------------------------------------------------------------------------------
5' HR F       ------------------------------------------------------------------------------
Rec F         TGACTTTACCATAACATTTGAAGAAAAAAAAAAAGAATTTGTTCTCATTTTTTAAGTCTTATTCATTTATTAGAAATTTTATG
SP2           ATGACTTTACCATAACATTTGAAGAAAAAAAAAAGAATTTGTTCTCATTTTTTAAAGTCTATTCATTTATTAGAAATTTTATG
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              2297                                                                2378
SET7 Fina...  TGGATATTCACATGACTATTTAAGAGACTCACTTAATAAATATTATCGAATCGATAATTTAACTACAACCGATGCACCACAA
SP1           ------------------------------------------------------------------------------
5' HR R       ------------------------------------------------------------------------------
5' HR F       ------------------------------------------------------------------------------
Rec F         TGAWTCMATGACTATTAAGARACTCACTAATAAATATTTATCGGAATTCGAWTAACTCAC----------------------
SP2           TGGATATTCACATGACTATTTAAGAGACTCACTTAATAAATATTATCGAATCGATAATTTAACTACAACCGATGCACCACAA
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................

              2379                                                                2460
SET7 Fina...  ATAGAGGAGGGATCCTACCCATATGATGTACCGGATTACGCAGGTTATCCGTATGACGTTCCGGACTATGCTGGCTCGTATC
SP1           ------------------------------------------------------------------------------
5' HR R       ------------------------------------------------------------------------------
5' HR F       ------------------------------------------------------------------------------
Rec F         ------------------------------------------------------------------------------
SP2           ATAGAGGAGGGATCCTACCCATATGATGTACCGGATTACGCAGGTTANCCGTATGACGTCCNNNNTNNNNNNNNNNN-----
SP5           ------------------------------------------------------------------------------
SP6           ------------------------------------------------------------------------------

..............................................................................................
```

Sequences highlighted in blue represent the recodonised gene insert site. Sequences highlighted in grey represent the *Xba*I site.

```
Page 11 pCS22_DiCre_set7

              4101                                                                                4182
SET7 Fina...  gaccacatggtccttcttgagtttgtaacagctgctgggattacacatggcatggatgaActAtacaaaTAATCTAGATAAa
SP1           ---------------------------------------------------------------------------------
5' HR R       ---------------------------------------------------------------------------------
5' HR F       ---------------------------------------------------------------------------------
Rec F         ---------------------------------------------------------------------------------
SP2           ---------------------------------------------------------------------------------
SP5           --NNNNNNNNNNNNNNTTGANTTTGTAACAGCTGCTGGGANTACACATGGCATGGATGAACTATACAAATAANC---NNNNN
SP6           GACCNCATGGTCCTTNTTGAGTTTGTAACAGCTGCTGGGATTACNCATGGCANGGATGAACTATACAAATAATNTAG---AA

..............................................................................................

              4183                                                                                4264
SET7 Fina...  tatatttagcaataacttcaaaaaaaaaaaaaaattaaaaaccacaaaataataaaatcataatttttattatttacgagta
SP1           ---------------------------------------------------------------------------------
5' HR R       ---------------------------------------------------------------------------------
5' HR F       ---------------------------------------------------------------------------------
Rec F         ---------------------------------------------------------------------------------
SP2           ---------------------------------------------------------------------------------
SP5           NATATTTAGCAATAACTTCAAAAAAAAAAAAAAATTAAAAACCACAAAATAATAAAATCATAATTTTTATTATTTACGAGTA
SP6           TATATTTAGCAANAACTTCAAAAAAAAAAAAAAATTAAAAACCNCAAAATAATAAAATNATAATTTTTNTTATTTNNGAGTA

..............................................................................................

              4265                                                                                4346
SET7 Fina...  tattatcatattgtgtaaatgtataatatatatatatatatatatatatttttttttttcttattcatatttttttttttt
SP1           ---------------------------------------------------------------------------------
5' HR R       ---------------------------------------------------------------------------------
5' HR F       ---------------------------------------------------------------------------------
Rec F         ---------------------------------------------------------------------------------
SP2           ---------------------------------------------------------------------------------
SP5           TATTATCATATTGTGTAAATGTATAATATATATATATATATATATA-TTTTTTTTTTCTTATTCATATTTTTTTTTTTT
SP6           TATTATCATATTGNGTAAATGTATAANANANANANANANANANATA-TTTTTTTTTTNTTATTCANATTTTTTTTTTTT

..............................................................................................

              4347                                                                                4428
SET7 Fina...  ttttacaataaaatgattatcattttgttttttcatgtgaaatattataaaaatcttgttgtttttaaaatattgttagaac
SP1           ---------------------------------------------------------------------------------
5' HR R       ---------------------------------------------------------------------------------
5' HR F       ---------------------------------------------------------------------------------
Rec F         ---------------------------------------------------------------------------------
SP2           ---------------------------------------------------------------------------------
SP5           TTTTACAATAAAATGATTATCATTTTGTTTTTTCATGTGAAATATTATAAAAATCTTGTTGTTTTTAAAATATTGTTANAAC
SP6           TTTTACAATAAAATGATTNTCATTTTGTTTTTTCATGTGAAATATTATAAAAATCTTGTTGTTTTTAAAATATTGTTAGAAC

..............................................................................................

              4429                                                                                4510
SET7 Fina...  attttacatatgttcttaaaataatatattaaaatttccataaggtggaaaatgaaattaaagaatttcaaatcaagtccac
SP1           ---------------------------------------------------------------------------------
5' HR R       ---------------------------------------------------------------------------------
5' HR F       ---------------------------------------------------------------------------------
Rec F         ---------------------------------------------------------------------------------
SP2           ---------------------------------------------------------------------------------
SP5           ATTTTACATATGTTCTTAAAATAATATATTAAAATTTCCATAAGGNGGAAAATGAAATTAAANAATTTCAAATCAAGTCCAC
SP6           ATTTTACATATGTTNTTAAAATAATATATTAAAATTTCCATAAGGTGGAAAATGAAATTAAAGAATTTCAAATCAAGTCCNC

..............................................................................................
```

Sequences highlighted in grey represent the *Bam*HI site. Sequences highlighted in green represent the 3' HR gene insert site.

Page 12 pCS22_DiCre_set7

```
              4511                                                                4592
SET7 Fina...  atcaaaaaaaaaaaaaaAAAAAAAACACTAGATAATAAAAATGAATATAAGAATTAATGGATATGTAGAATATATAAAACTT
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
5' HR F       --------------------------------------------------------------------------------
Rec F         --------------------------------------------------------------------------------
SP2           --------------------------------------------------------------------------------
SP5           ATC-AAAAAAAAAAAAAAAAAAAAACNCTANATAATAAAAATGAATATAANAATTAANGGATATGTANAATATATAAAACTT
SP6           ATC-AAAAAAAAAAAAAAAAAAAAACACTAGATAATAAAAATGAATATAAGAATTAATGGATATGTAGAATATATAAAACTT

................................................................................................

              4593                                                                4674
SET7 Fina...  GTACCTATTTGTTCTAAGGTATCTCTTAAGATGAACATCCATATATTTCAATGTTCATTTTTACTGGAAATACACATAACAT
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
5' HR F       --------------------------------------------------------------------------------
Rec F         --------------------------------------------------------------------------------
SP2           --------------------------------------------------------------------------------
SP5           GNNCCTNTTTGTTCTAAGGNATCTCTTAANATGAACATCNATATNTTTNANNGNTCATTTTTACNGGAAATACACATAACAT
SP6           GTACCTATTTGTTCTAAGGTATCTCTTAAGATGAACATCCATATATTTCAATGTTCATTTTTACTGGAAATACACATAACAT

................................................................................................

              4675                                                                4756
SET7 Fina...  TCTTTTCTGTTTTTACATATATTTTTTTTATCTTTACTCATTTTTGTTCATTTTAATAGATTGGGACAAAAATTGTTAGTATA
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
5' HR F       --------------------------------------------------------------------------------
Rec F         --------------------------------------------------------------------------------
SP2           --------------------------------------------------------------------------------
SP5           TCTTTTCNGTTTTTACNTATATTTTTTTTATCTNNACNCNTTTTTGTTCATTTTAATANATNGGGACAAAAATTGTTAGTATA
SP6           TCTTTTCTGTTTTTACATATATTTTTTTTATCTTTACTCATTTTTGTTCATTTTAATAGATTGGGACAAAAATTGTTANNNNN

................................................................................................

              4757                                                                4838
SET7 Fina...  AGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
5' HR F       --------------------------------------------------------------------------------
Rec F         --------------------------------------------------------------------------------
SP2           --------------------------------------------------------------------------------
SP5           ANCTNGGCGNAATCATGGTCATANCTGTTTCCTGNGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGNAA
SP6           NNNTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGNGAAATTGTTNNCCGCTCNNNNNNCNNNNNNNNNNNN--------

................................................................................................

              4839                                                                4920
SET7 Fina...  GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
SP1           --------------------------------------------------------------------------------
5' HR R       --------------------------------------------------------------------------------
5' HR F       --------------------------------------------------------------------------------
Rec F         --------------------------------------------------------------------------------
SP2           --------------------------------------------------------------------------------
SP5           GCATAAAGTGTAAAGCCTGGGGNGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
SP6           --------------------------------------------------------------------------------

................................................................................................
```

Sequences highlighted in green represent the 3' HR gene insert site. Sequences highlighted in grey represent the *Hind*III site.

Supplementary Data 4: Full consensus Sanger sequence alignment of deletion in 5' HR for *set7*

```
Page 2 G07_19_pCS22_7_FWD_2023-05-02_17A_18929_2023-05-02-10-17-14 (G07_19_pCS22_7_FWD_2023-...

              903                                                                        984
set7_5'HR... TCTTTTAGCATTCCCATGGGGATCCTACCCATATGATGTACCGGATTACGCAGGTTATCCGTATGACGTTCCGGACTATGCT
G07_19_pC... TCTTTTAGCATTCCCATGGGGATCCTACCCATATGATGTACCGGATTACGCAGGTTATCCGTATGACGTTCCGGACTATGCT

..................................................................................................

              985                                                                       1066
set7_5'HR... GGCTCGTATCCATACGATGTACCGGATTATGCGGTCGTCGACGGAGAAGGAAGAGGAAGTTTATTAACATGTGGAGATGTAG
G07_19_pC... GGCTCGTATCCATACGATGTACCGGATTATGCGGTCGTCGACGGAGAANGAAGAGGAAGTTTATTAACATGTGGAGATGTAG

..................................................................................................

              1067                                                                      1148
set7_5'HR... AAGAAAATCCAGGACCAATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTA
G07_19_pC... AAGAAAATCCAGGANCAATGANTGAACNAGATGGATTGNACGC-NNNTCTCCNGNCGCTTGGGTGGAGAGGCTATTNNNCTA

..................................................................................................

              1149                                                                      1230
set7_5'HR... TGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTC
G07_19_pC... TGACTGGGNANANAGANANNNNNTGCTNNNATNNNNNNNNNNNCGGCNNNCANNNNNNNNNNNNNNNNTTTTNNNANANCNN

..................................................................................................

              1231                                                                      1312
set7_5'HR... AAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTT
G07_19_pC... NCCNNNNNCNGGNNNNNCCNNNNN----------------------------------------------------------

..................................................................................................

              1313                                                                      1394
set7_5'HR... GCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTC
G07_19_pC... ----------------------------------------------------------------------------------

..................................................................................................

              1395                                                                      1476
set7_5'HR... ATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGC
G07_19_pC... ----------------------------------------------------------------------------------

..................................................................................................

              1477                                                                      1558
set7_5'HR... CCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGG
G07_19_pC... ----------------------------------------------------------------------------------

..................................................................................................

              1559                                                                      1640
set7_5'HR... ACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGT
G07_19_pC... ----------------------------------------------------------------------------------

..................................................................................................

              1641                                                                      1722
set7_5'HR... GACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGT
G07_19_pC... ----------------------------------------------------------------------------------

..................................................................................................

              1723                                                                      1804
set7_5'HR... GTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCC
G07_19_pC... ----------------------------------------------------------------------------------

..................................................................................................
```

Sequences highlighted in blue represent the 5' HR insert. Sequences highlighted in red represent the SERA2/*loxP* site. Deletion is indicated in yellow.