Experimentally Inducible Gametocytogenesis: A new tool to uncover the early stages of commitment in *P. berghei*

Robyn Kent

M.Sc, M.Res, B.Sc

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

September 2016

Institute of Infection, Immunity & Inflammation
College of Medicine, Veterinary Medicine & Life Sciences
Wellcome Trust Centre for Molecular Parasitology
University of Glasgow, Glasgow, U.K.

Abstract

Plasmodium, the causative agent of malaria, has a complex life cycle requiring a mammalian host and a mosquito vector. Its cyclic infection of red blood cells gives rise to the characteristic fevers and resulting anaemia associated with the disease. Many drug intervention strategies target this, the asexual blood cycle. Little is known about the underlying processes governing the differentiation from this cycle to the sexual stage, the gametocyte, which is uniquely able to transmit through the mosquito to new hosts.

In this work we have developed a novel conditional system to build on existing knowledge about the trigger for sexual commitment (*ap2*-g) to control and expand commitment to gametocytogenesis in the rodent malaria model, *Plasmodium berghei*. We have characterised the effect of controlling and increasing expression of this initiating transcription factor on asexual and gametocyte development in an effort to obtain an over representation of biologically relevant gametocytes.

Using this novel system we initiated an untargeted transcriptomics study to uncover novel factors involved in the process of commitment and other gametocyte specific roles such as gender assignation, sex specific components and overall gametocyte development. From the pilot data obtained from this transcriptomics we identified and screened 40 candidates potentially specifically involved in gametocyte biology. Five of these candidates have been investigated further to uncover novel roles in the early stages of commitment or development (2 gametocyte non-producers), male specific development (1 male non-producer and 1 which seems to effect development) and potentially female specific development (1 *in vivo* female non-producer).

Using the data generated in this study we hope further work can be completed to characterise many aspects of commitment to gametocytogenesis and the processes involved in downstream events required for successful transmission.

Author's Declaration

I, Robyn Kent hereby declare than I am the sole author of this thesis and I have performed all the work presented here, with the following exceptions:

1. Several vectors used in this study were generated by other lab members and repurposed. These are acknowledges throughout the work. Rachael Cameron (RC) generated the GIMO vectors and DiCre vectors. Katarzyna Modrzynska (KM) generated the conditional *ap2-g* overexpressor vector. Katie Hughes (KH) generated the constitutive CFP and GFP vectors used. Rebecca Lee (RL) generated the early gametocyte RFP vector.

2. RNAseq library preparation and running of the samples was completed at the Sanger institute by Katarzyna Modrzynska (KM).

3. The CFP promoter lines used in this study were generated by Katie Hughes (KH) for analysis (Sinha *et al.*, 2014).

Acknowledgements

Wow, I thought writing a thesis was hard. Nobody warns you about the acknowledgements.

Starting with the most obvious, I would like to thank Prof. Andy Waters for being such a good supervisor. For allowing me the freedom to drive and love this project, the incredible insight to make it possible and the foresight to control me in order to keep it on track- it wouldn't have been the PhD it was without you. In a similar fashion I would like to thank my second supervisor, Prof. Markus Meissner for providing me with a great start in the PhD and my assessors Prof. Sylke Muller, Dr. Richard McCulloch and Dr.Karl Burgess for all their advice on how to complete a PhD; I am confident without them I would still be in the lab desperately seeking that final great experiment.

Next to all the amazing people in the Waters and Meissner groups; you're a nurturing, supportive, fun and above all knowledgeable bunch! Particularly I want to thank Brett Roberts for the unending patience for my inability to stay in the past tense and desire to make the tiniest figures imaginable; I am 100% sure without you I'd be laughed out of a Viva. I also thank Nisha, Katie, Julie, Nick and Mallory for unending patience with the line "can I just ask you". I am sure having someone in the office writing a thesis was not easy but you guys made it not only possible but... let's go with tolerable. Lastly, at work, though not least of course I want to thank Jamie, Fernanda and Pieter for being fantastic friends. Between you you've solved every problem I've come to you with. Whether it needed an experiment, a de-stressing run, a few beers or an excuse to jump around punching strangers (aka dancing) you guys always helped. Finally, to the best housemate, Carmen, thanks for being as emotionally available as I am and knowing the best way to deal with my insanity. I can only hope to be as accommodating when you are in the same position.

Back to the project, this wasn't a solo-endeavour but a massive team effort. Before me came the illustrious work carried out by Abhinav, Rachael, Katie and Anne. Thank you. Building on your foundations and having the faith in everything you did made everything simpler. To my current collaborator Kasia, all I can say is thank God you were always so patient. I can only apologise for the stupid

questions and hope that my excitement made up for it a little bit. Finally, though sometimes I think you guys should be the foremost thank in every thesis, every paper, and every grant ever written, the "support" staff of Glasgow University. Particularly, I want to thank everyone at the JRF and in the flow suite. I shudder to think what this project would have been like without Colin.

Now to the bit I am rubbish at...

People who've met me will know I am the most stable person in the world who never falters in her self-belief and confidence. Those who know me will know this is utter... nonsense. To those people; Stoowert, my ever-more-insightful little sister Smelly, the Stokies and many more, thanks for seeing through the front I put on. Your support has helped me get through not just the past 4 years of PhD but, probably every minute leading up to it.

Now to my mum. You are the strongest, most determined person I know. You have an unwavering belief in me that has made this whole endeavour possible. There is nothing I could do to thank you for everything you've ever done and all you've helped me achieve. I'll just end with the soppiest thing, I love you bigger than the world.

To everyone else I have probably forgotten, don't take it personally. I appreciate everything everyone does, I just have an awful memory and have exhausted the pool of emotion I have. When all this is over, I'll shower you with cakes, biscuits and bread.

Table of Contents

List of Figures

List of Tables

Abbreviations

°C	Degrees Celcius
μ	Micro
μg	Microgram
μl	Microliter
μM	Micromolar
μm	Micrometer
μmol	Micromoles
3' UTR	Three Prime Untranslated Region
5' UTR	Five Prime Untranslated Region
5-FC	5-Fluorocytosine
AID	Auxin inducible degron
ATc	Anhydrotetracycline
bp	Base Pairs Pairs
BSA	Bovine Serum Albumin
DD	Destabilisation domain
ddH2O	Double Distilled Water
DE	Differentially expressed
DHFR/TS	Dihydrofolate Reductase Synthase Reductase-Thymidylate Synthase
DiCre	Dimerisable Cre
DMSO	Dimethyl Sulphoxide Sulphoxide
DNA	Deoxyribonucleic Acid Acid
dNTP	Deoxynucleotide Triphosphate Triphosphate
DTT	Dithiothreitol
dUTP	Deoxyuridine Triphosphate Triphosphate
ECL	Enhanced Chemiluminescence Chemiluminescence
EDTA	Ethylene Diamine Tetraacetic Acid
EG	Early Gametocyte
ePBS	Enriched Phosphate Buffered Saline
FACS	Fluorescence-Activated Cell Sorting
FBS	Foetal Bovine Serum
g	Grams
g	Relative Centrifugal Force
gDNA	Genomic Deoxyribonucleic Acid
GFP	Green Fluorescent Protein
GNP	Gametocyte Non-Producer
goi	Gene Of Interest
GWS	Genome Wide Sequencing
hpi	hours post induction

hdfr	Human Dihydrofolate Reductase
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HRP	Horseradish Peroxidase
kb	Kilobases
kDa	Kilodaltons
ko	Knockout
KS	Knock Sideways
m	Meter
m	Milli
M	Molar
mg	Milligram
min	Minute
ml	Milliliter
mM	Millimolar
n	Nano
NaCl	Sodium Chloride
ng	Nanogram
nm	Nanometer
nM	Nanomolar
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
pmol	Picomole
POI	Protein of Interest
RBC	Red Blood Cell
RNA	Ribonucleic Acid
SDS	Sodium Dodecyl Sulphate
sec	Seconds
SM	Selectable Marker
SSR	Site specific recombination
TBE	Tris-Borate Containing EDTA
TE	Tris Containing EDTA
TetO	Tetracycline Operon
TetR	Tetracycline Repressor
V	Volt
WHO	World Health Organization
WT	Wild-Type
yfcu	Uridyl Phosphoribosyl Transferase

Chapter 1   Introduction

1.1 **Malaria**

It is estimated that around half the world's population, 3.3 billion people, are at risk of contracting human malaria (Figure 1.1 (WHO, 2014)). Reported cases of malaria in 2015, caused by one of the five species known to infect humans, was 215 million resulting in almost half a million deaths (WHO, 2015).
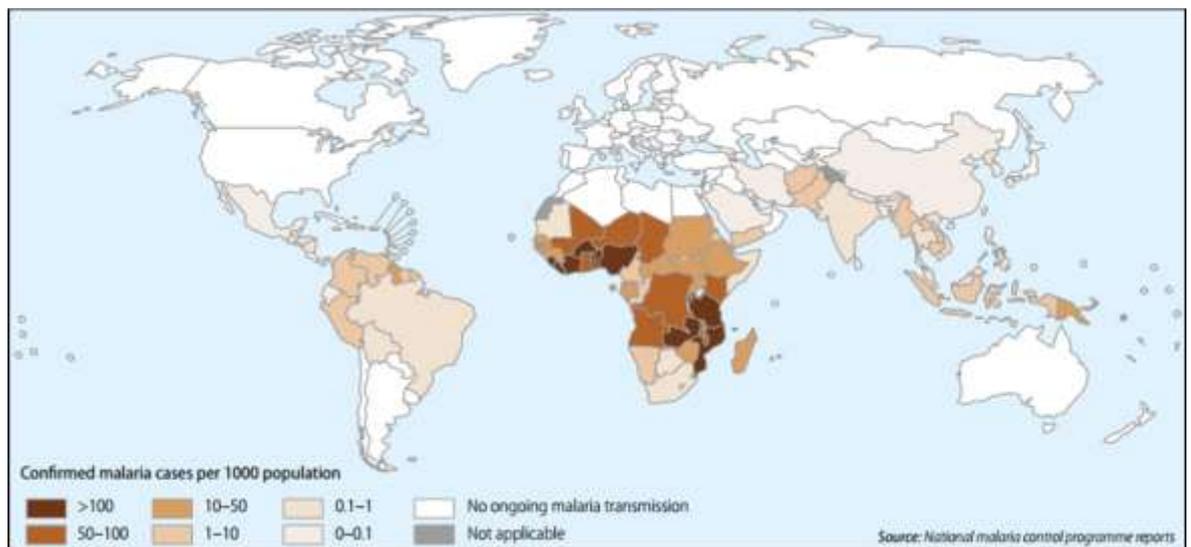
Figure 1.1 Global malaria distribution

Confirmed malaria cases per 1000 across the globe (WHO, 2014).

The majority of deaths in 2015 occurred in the WHO African Region (90%), followed by the WHO South-East Asia Region (7%) and the WHO Eastern Mediterranean Region (2%). Approximately 70% of all deaths were children under the age of 5 years; however malaria was no longer the leading cause of death among children in Sub-Saharan Africa, moving to the fourth biggest killer accounting for around 10% of deaths. The leading cause of death in children under 5 years old in Africa is acute respiratory tract infections such as pneumonia (WHO, 2015; WHO, 2016).

The number of reported cases of malaria and therefore attributed deaths is considered a vast underestimate due to reliance on passive detection and national reporting for disease estimates (Snow *et al.*, 2005). Furthermore, the actual burden of malaria is underappreciated when considering the vast swathes of the globe exhibiting resistance to drug treatments. This includes both the

economic cost to the elimination programs and public health (Talisuna *et al.*, 2004). In Senegal the incidents of malaria-specific mortality increased 2 – 11-fold in children under 5 years old in association with chloroquine (CQ) resistance. The economic cost of resistance is further exacerbated as the most heavily affected regions are also the poorest. An adult course of CQ treatment is around 20 cents whereas alternatives, after emergence of resistance, range from more than $2 (10-fold increase) to $80 (400-fold increase) per treatment. In terms of control programs this shift makes the majority of strategies unfeasible (Talisuna *et al.*, 2004).

Current front-line treatments utilise Artemesinin Combination Therapies (ACTs) and treatments of malaria cases in Africa with these compounds has risen from < 1% (2005) to up to 22% (2014). As a treatment ACTs are highly successful and have been shown to reduce malaria mortality in children under 5 years by 97% - 99%. However, as of 2015 artemesinin resistance has been reported in five countries in the Greater Mekong subregion (Cambodia, Lao People's Democratic Republic, Myanmar, Thailand and Vietnam). This resistance is attributed to a delay in parasite clearance. Although this delay currently does not prevent parasite clearance it increases the proportion of parasites killed by the partner drug and may increase the likelihood of resistance developing. Furthermore, ACT are the sole treatment for severe malaria, due to the fast action of the artemisinin compounds, if this activity is no longer active then more severe malaria cases will result in mortality (WHO, 2014; WHO, 2015)
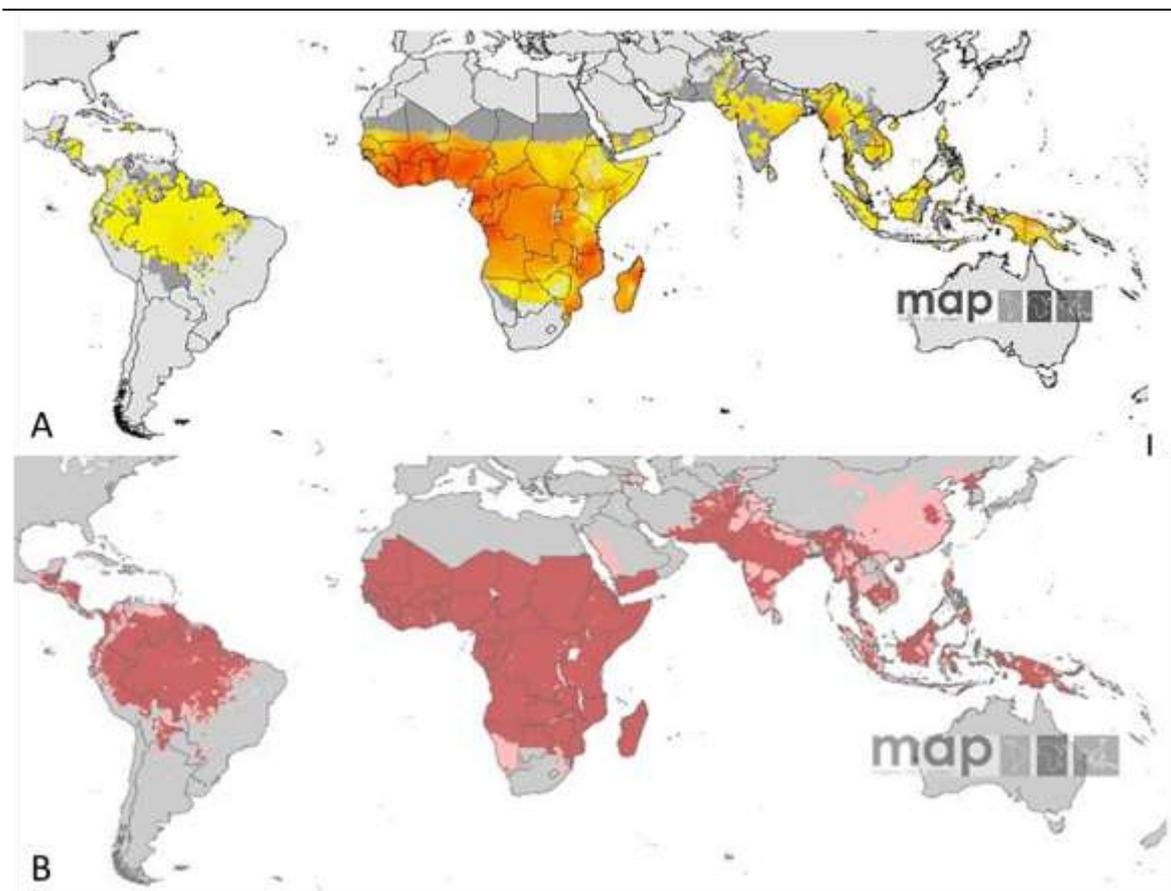
Figure 1.2 Distributions of Plasmodium falciparum (A) and Plasmodium vivax (B)

Global endemic distributions of Plasmodium falciparum (A) and Plasmodium vivax (B) mapped by
www.map.ox.ac.uk (Autino *et al.*, 2012; Bhatt *et al.*, 2015).

While *Plasmodium falciparum* is the most deadly human infective species, *Plasmodium vivax* is now regarded as a contributor to disease burden and mortality (Mendis *et al.*, 2001). The distribution of these two species overlaps considerably (Figure 1.2).

1.2 The Lifecycle of *Plasmodium spp.*

The lifecycle of all *Plasmodium spp.* is complex, requiring a mammalian host and an insect vector (Figure 1.3 (MALWEST, 2016)). During blood feeding by a female Anopheles mosquito, sporozoites are deposited in the host's skin. From the skin these sporozoites make their way to the host's liver where they traverse multiple hepatocytes before invading and replicating forming a hepatic schizont. The merozoites formed in the hepatic schizont emerge in merosomes and enter the blood stream. In the blood stream merozoites invade red blood cells and mature through the trophozoite and begin replicating in a schizont. After

maturation the schizont ruptures releasing merozoites that invade new red blood cells. This forms the asexual cycle responsible for increasing parasitaemia and the symptoms of malaria. A proportion of merozoites in each cycle however do not go on to replicate but differentiate into the sexual forms, gametocytes. Mature gametocytes circulate in the blood and are taken up during the blood meal of Anopheles mosquitoes. Once in the midgut of the mosquito the male gametes exflagellate and the female gametes activate. Fertilisation of the female by the male forms a zygote which matures into an ookinete that can traverse the midgut of the mosquito and form an oocyst in which thousands of sporozoites develop and mature. After the oocyst ruptures the sporozoites migrate to the salivary gland of the mosquito where they can be deposited into a new host (Menard, 2005).
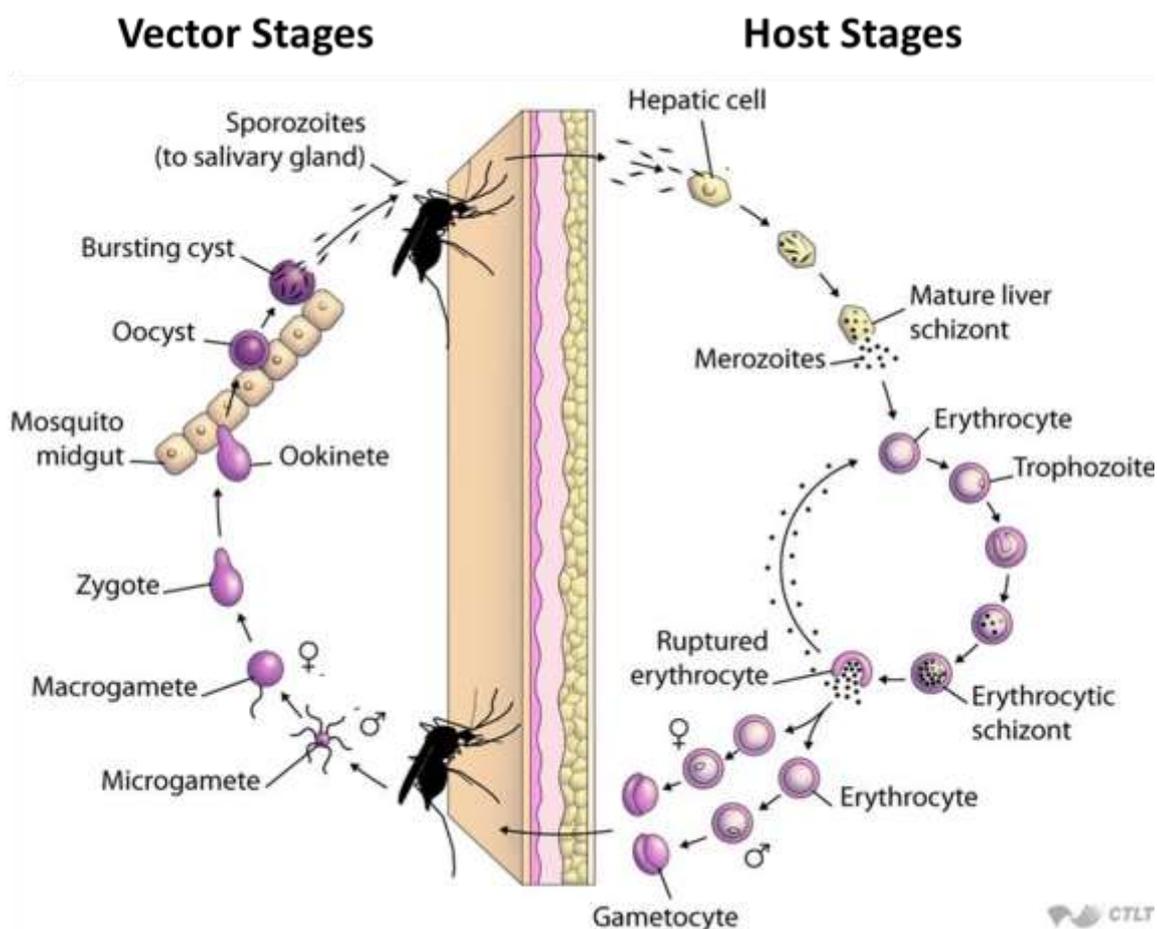
Figure 1.3 Lifecycle of *Plasmodium spp* (adapted from (MALWEST, 2016)).

1.2.1 Parasite transmission to the host

Initial deposition of *Plasmodium* into the host occurs when a blood meal is taken by a female anopheles mosquito. In experimental conditions it was estimated around 100 sporozoites are deposited by one mosquito, however in nature bites were estimated to deliver fewer than 50 *P. falciparum* sporozoites per bite (Sidjanski & Vanderberg, 1997). Based on studied in *P. berghei,* the sporozoites are deposited in the extracellular matrix of the skin where around 50% move, by gliding motility, from the bite site to vessels in the skin (Amino *et al.*, 2006). When the sporozoites encounter a vessel they invaded the blood or lymphatic vessel (Amino *et al.*, 2006; Menard *et al.*, 2013). Sporozoites that enter lymphatic vessels are drained to the proximal lymph node where most are degraded but some partially develop into pre-erythrocytic stages (Amino *et al.*, 2006). Approximately 70% of the sporozoites that left the bite site invaded blood vessels and continued their journey to the liver (Amino *et al.*, 2006).

1.2.2 Pre-erythrocytic development

Sporozoites that successfully enter the blood stream reach the liver where they cross the lining of the liver, the sinusoidal barrier, by traversing through Kuppfer cells (in approximately 70% of cases) or endothelial cells (approximately 30% of cases). Often the sporozoites interact with multiple cells of the two types before traversal (Tavares *et al.*, 2013). Sporozoites, now in the liver often traverse through multiple hepatocytes before invading and establishing themselves within a parasitophorous vacuole in the final hepatocyte. Two mechanisms by which cell traversal can occur have been demonstrated. Firstly, parasites were shown to breach the cell membrane of the hepatocyte and glide through the cytoplasm emerging from the hepatocyte by breaching another membrane (Mota *et al.*, 2001). Secondly, parasites invaded the hepatocyte by the formation of a transient vacuole which does not require discharge of secretory organelles (for details see section 1.3.1), called rhoptries, which are associated with the formation of a moving junction. This invading parasite did not pass through a moving junction formed between the parasite and the host cell in this process but relied upon vigorous motility (Risco-Castillo *et al.*, 2015). After hepatocyte migration, the sporozoite invades a hepatocyte forming a parasitophorous vacuole. Invasion in this way discharges the parasites secretory organelles and

always forms a parasitophorous vacuole, in which the parasite resides. This vacuole is composed of parasite derived material and host cell membrane (Mota *et al.*, 2001; Prudencio *et al.*, 2006). The sporozoite now undergoes rounds of DNA replication and in *P. berghei* within 2 – 3 days begins the process of merozoite release into the blood stream. This process is achieved by the delivery of membrane bound vesicles, called merosomes, filled with merozoites. This process is thought to improve parasite delivery and survival rates. Merozoites directly released into the liver sinusoids would be recognised and potentially phagocytosed by the Kuppfer cells. After replication, hepatocytes infected with merozoite-filled schizonts were shown to give off signals pertaining to cell-death which would ultimately lead to their clearance (Sturm *et al.*, 2006).

1.2.3 The asexual cycle

Once in the bloodstream, merozoites rapidly invade erythrocytes. In infections with the rodent malaria, *P. berghei,* merozoites show a strong preference for the invasion of young erythrocytes called reticulocytes (Janse *et al.*, 1989a). The invasion process in *P. falciparum* is made up of three distinct phases, preinvasion, active invasion and echinocytosis and takes less than two minutes. The first two of these phases are conserved across *Plasmodium* species. Initial contact results in low-affinity interactions with the red blood cell and often significant deformation of the surface to achieve reorientation of the apical end of the merozoites (for details see section 1.3.1). Irreversible attachment marks the beginning of active invasion and occurs by the formation of the tight junction. This junction comprises parasite proteins that tightly interact with blood cell proteins forming a ring through which the parasite invades. Invasion through this tight junction, where the merozoites penetrates the red blood cell membrane, results in the parasite encasing itself in a parasitophorous vacuole where it resides and develops. After entering the red blood cell the volume of the cell reduces and then recovers its usual biconcave shape (Gilson & Crabb, 2009). Inside the red blood cell the parasite begins to consume the haemoglobin and grow within the vacuole maturing to a trophozoite. The digestion of the cells haemoglobin and subsequent detoxification of haem results in the

formation of hemozoin granules. This process of haemoglobin digest and detoxification is the target for many anti-malarials (Bakar *et al.*, 2010).

In *P. berghei a*pproximately 16 hours post invasion (hpi) of the red blood cell, the trophozoite is mature and schizogony begins. Firstly, the trophozoite begins replicating its DNA and undergoes nuclear divisions. The last round of this mitosis is synchronous and at the same time as this final replication is ongoing, daughter cell buds begin to form at the periphery of the parasite (Absalon *et al.*, 2016; Francia & Striepen, 2014). Schizogony is completed approximately 24 hours post merozoite invasion and results in the formation of 8 – 24 daughter cells. It is uncommon for schizonts to be found in the peripheral blood as the stage sequesters in the spleen, lungs and adipose tissue of the rodent host (Franke-Fayard *et al.*, 2005). In all species of malaria, mature budded daughter merozoites egress from the red blood cell in a two-step process. Firstly, the parasitophorous vacuole membrane, in which replication has occurred, ruptures releasing the merozoites into the cytosol of the host cell. Then the red blood cell membrane ruptures releasing the merozoites into the blood stream to go on and invade new reticulocytes (Blackman & Carruthers, 2013; Das *et al.*, 2015). This process of invasion, growth, replication and release continues in a cyclic manner termed the asexual blood cycle and increases parasitaemia in the host.

1.2.4 Sexual differentiation and gametocytogenesis

While proliferation within the host is achieved by the asexual cycle, transmission to new hosts requires terminal differentiation to a sexually dimorphic cell. Within each asexual cycle a proportion of the parasites differentiate into gametocytes. In the rodent species *P. berghei* 5 – 15% of the parasites develop into sexually distinct cells, gametocytes. *P. berghei* gametocytes have been estimated to be mature from approximately 26 hours post merozoite invasion (25 – 33hpi) demonstrated by the onset of exflagellation of male gametocytes. In *P. falciparum* gametocyte maturation is a lengthier process taking 10 – 12 days to reach maturity (see section 1.3.2 and Figure 1.5 (Josling & Llinas, 2015)). In *P. berghei* gametocytes are indistinguishable from asexual parasites until 20hpi when they become recognisable from the schizont stages. However, morphological gender identification is not reliable until gametocytes are apparently mature at around 26hpi (see section 1.3.2 and Figure 1.6 (Mons *et*

al., 1985)). Mature gametocytes are morphologically distinguishable from asexual parasites by their encompassment of the entire red blood cell, the presence of a single, large nucleus and pigmentation (Figure 1.6). It is notable that not only do gametocytes develop from within the asexual proliferative cycle but also directly from the pre-erythrocytic stages in the liver showing that transmission is possible from the onset of blood stage infection (Suhrbier *et al.*, 1987).

Two models of sexual differentiation have been proposed for *P. falciparum*. The first proposes that immediately after invasion the early blood stage parasite is uncommitted and retains the potential to become either a gametocyte or remain in the asexual cycle. The second suggests that invading merozoites are pre-committed to either the asexual cycle or commitment to differentiation into a gametocyte. Work from several groups has led to the favouring of the second model (Bruce *et al.*, 1990; Carter & Miller, 1979). Bruce 1989, demonstrated that after the seeding of single infected schizonts on monolayers of red blood cells and the subsequent analysis of plaques generated from rupture and invasion the majority of these plaques contained either uniquely asexual parasites or gametocytes. The presence of a few mixed plaques was attributed to the presence of multiply infected red blood cells that had developed to generate two schizonts, one producing only asexual parasites and the other producing only gametocytes. This result led to the identification of late schizogony as the stage of commitment where parasites make the decision to commit to gametocytogenesis or to remain in the asexual proliferative stage. Further study into this phenomenon was initially hampered by the inability to fully mature *P. falciparum* gametocytes in monolayer culture which would allow sexing of the individual cells. However, the ability to maintain *P. falciparum* in static culture to allow gametocyte maturation was developed and gender identification in individual plaques of gametocytes became possible. In all unique gametocyte plaques it was found that the gametocytes either stained with a male specific or a female specific gametocyte markers, a cohort of gametocytes that contained both male and female gametocytes was not identified. This added to the previous assignation of schizogony as the point of commitment to gametocytogenesis as also the point of gender determination (Silvestrini *et al.*, 2000). While this model has been widely adopted, an alternate

model was proposed in *P. berghei* suggesting invading merozoites are not pre-committed. The initiation of *in vitro* cultures from synchronous *in vivo* infections at different times post invasion led to variability in the proportion of parasites that commit to gametocytogenesis without reinvasion. This led to the identification of early trophozoite (12 – 16hpi) as the stage at which gametocyte commitment becomes fixed. From 12hpi when *in vitro* and *in vivo* infections are compared for gametocyte commitment they are equal. Before this time the removal of the infection from the *in vivo* setting reduced gametocyte conversion rates (Mons, 1986b; Mons *et al.*, 1985).

While these works aimed to identify the point at which commitment occurred, they did not provide the molecular basis underlying the differentiation process. In *P. berghei* and *P. falciparum* the transcription factor responsible for the onset of gametocytaemia was identified as *ap2-g* (see 1.5.2 for full details). Absence of *ap2-g* expression results in no commitment to gametocytogenesis and all parasites remain in the asexual proliferative cycle (Kafsack *et al.*, 2014; Sinha *et al.*, 2014).

1.2.5 Mosquito stages

During mosquito blood feeding all blood stage parasites are ingested. This includes asexual stages, immature gametocytes and mature gametocytes. Only mature gametocytes develop further in the mosquito midgut. Inside the midgut of the mosquito conditions are very different to those in the mammalian host, in particular the temperature and pH change (lower temperature and slightly higher pH). In addition the gametocytes now encounter a multitude of new mosquito derived stimuli. One key mosquito derived factor is xanthurenic acid which was shown to provide key developmental cues (Billker *et al.*, 1998). Inside the midgut, male gametes are stimulated to form 8 microgametes through three rounds of nuclear division and female gametocytes activate forming one large macrogamete. The developmental cues result in exit from the blood cells encasing the gametocyte allowing fusion of the micro and macrogametes. This fusion forms the diploid zygote which in *P. berghei* undergoes meiosis within four hours of the infected blood meal (Billker *et al.*, 1997; Janse *et al.*, 1986). Within 24 hours, the invasive and polarised ookinete is formed from this fusion. To stimulate activation of gametocytes and ookinete formation *in vitro* the pH increase from 7.4 in mouse blood must be increased from 7.5 – 7.6 in the midgut of mosquitoes to 8 in the presence of xanthurenic acid (Billker *et al.*, 2000).

Migration of the ookinete through the mosquito midgut is a damaging process often destroying multiple cells during the traversal process. The ookinete is destined for the basal lamina of the mosquito's midgut where it forms an oocyst. Inside this oocyst thousands of sporozoites are formed (Sinden & Billingsley, 2001). Mature sporozoites are released from the oocyst where they must migrate to the salivary gland of the mosquito for successful delivery into the new host. This process has been shown to be inefficient and sporozoites are effectively

cleared by the mosquito, however the migration of sporozoites has also been shown to be facilitated by the flow of the haemolymph in the mosquito (Hillyer *et al.*, 2007). Once in the salivary glands of the mosquito sporozoites are ready to be deposited in a new host, beginning the life cycle anew.

1.3 **Morphology and structure of *Plasmodium***

The invasive stages of *Plasmodium* (sporozoites, merozoites and ookinete), collectively known as the zoite stages, share overall structural similarities and contain some key organelles that facilitate their motility and invasive ability. Meanwhile, p*lasmodium* gametocytes have a very different morphology to the asexual blood stages. As this work utilised the *P. berghei* rodent model these stages will be briefly described.

1.3.1 Morphology of the zoite

The invasive stages share a common set of organelles involved in invasion and an overall structural organisation (Figure 1.4). These stages are all polar in organisation with apical polar rings (APR) denoting the apical end of the parasite. The zoites contain a collection of secretory organelles unique to the apicomplexan; the rhoptries (merozoites and sporozoites), the micronemes (all) and the dense granules (merozoites and sporozoites). Upon contact with the host blood cell to be invaded the merozoites reorients to position the apical end of the parasite in contact with the host cell. From there is forms a tight junction made up of host cell receptors and proteins secreted from the rhoptries. This tight junction forms a ring through which the parasite actively penetrates, invaginating the host cell membrane forming the vacuole in which the merozoites resides and develops. The ookinete and sporozoite are motile stages that glide along and often traverse through host cells to reach the host cells they invade. Their apical prominence leads this gliding ensuring it is the primary contact with the host cell to be invaded (Baum *et al.*, 2009).
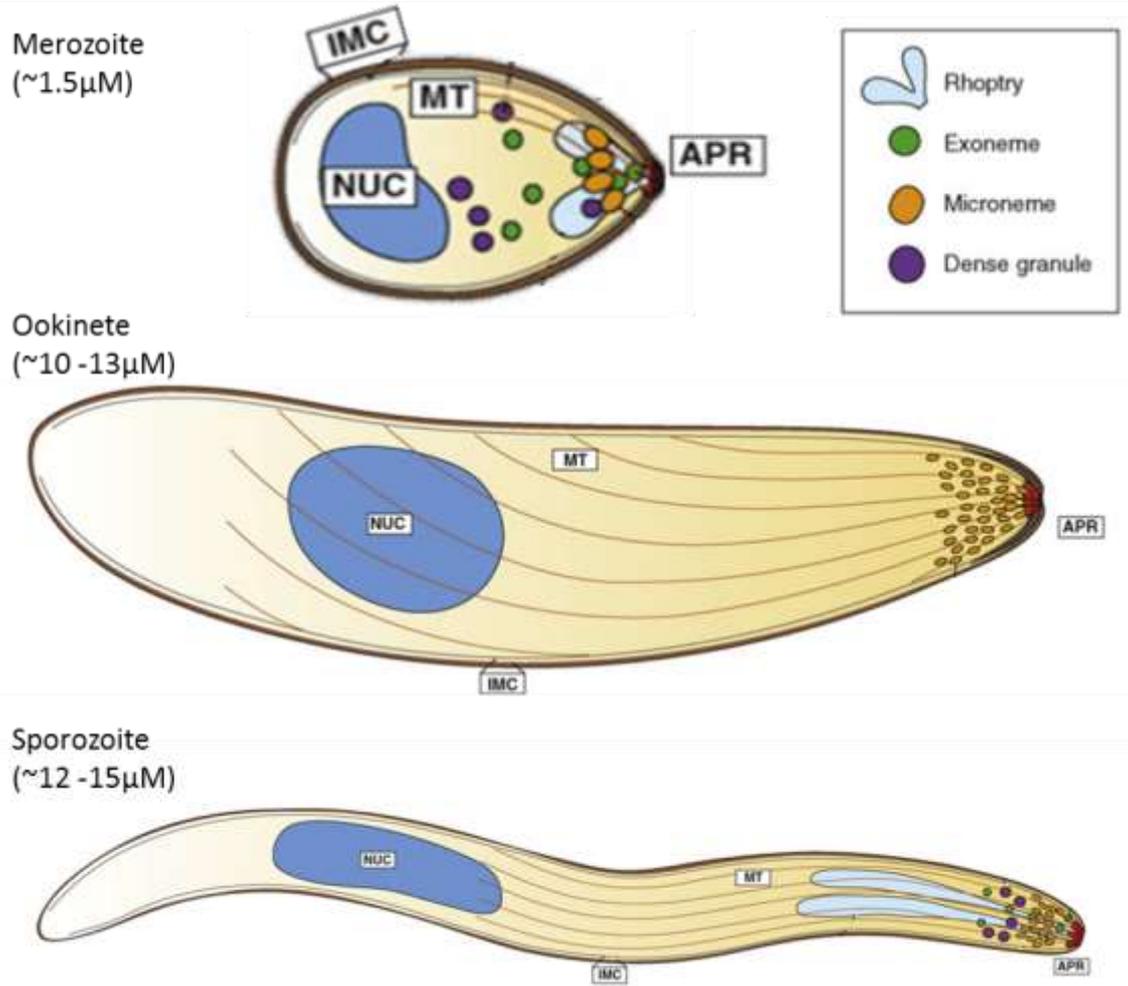
Figure 1.4 Structure of the plasmodium zoite

Invasive zoites of *Plasmodium* specied contain distinct secretory organelles that facilitate motility and the invasion process. The Rhoptries, micronemes and dense granules contain distinct proteins involved in the multiple steps of invasion. The merozoites invade red blood cells in the asexual proliferative cycle and prior to gametocyte commitment. The ookinete traverses and ultimately invades cells of the mosquito midgut. The sporozoite migrated from the oocyst in which it develops in the mosquito midgut to the salivary glands until deposition into the mammalian host. From there it migrates in the skin and goes on to traverse hepatic cells before invading the terminal hepatocyte and replicating to form merozoites (Figure adapted from (Baum *et al.*, 2008))

1.3.2 Morphology of Gametocytes

While *P. falciparum* gametocytes progress through five morphologically distinct phases taking 10 – 12 days to mature (Figure 1.5) *P. berghei* development takes only around 26 hours. In *P. falciparum* the shape of the parasite within the red blood cell alters dramatically resulting in a crescent shaped gametocyte.
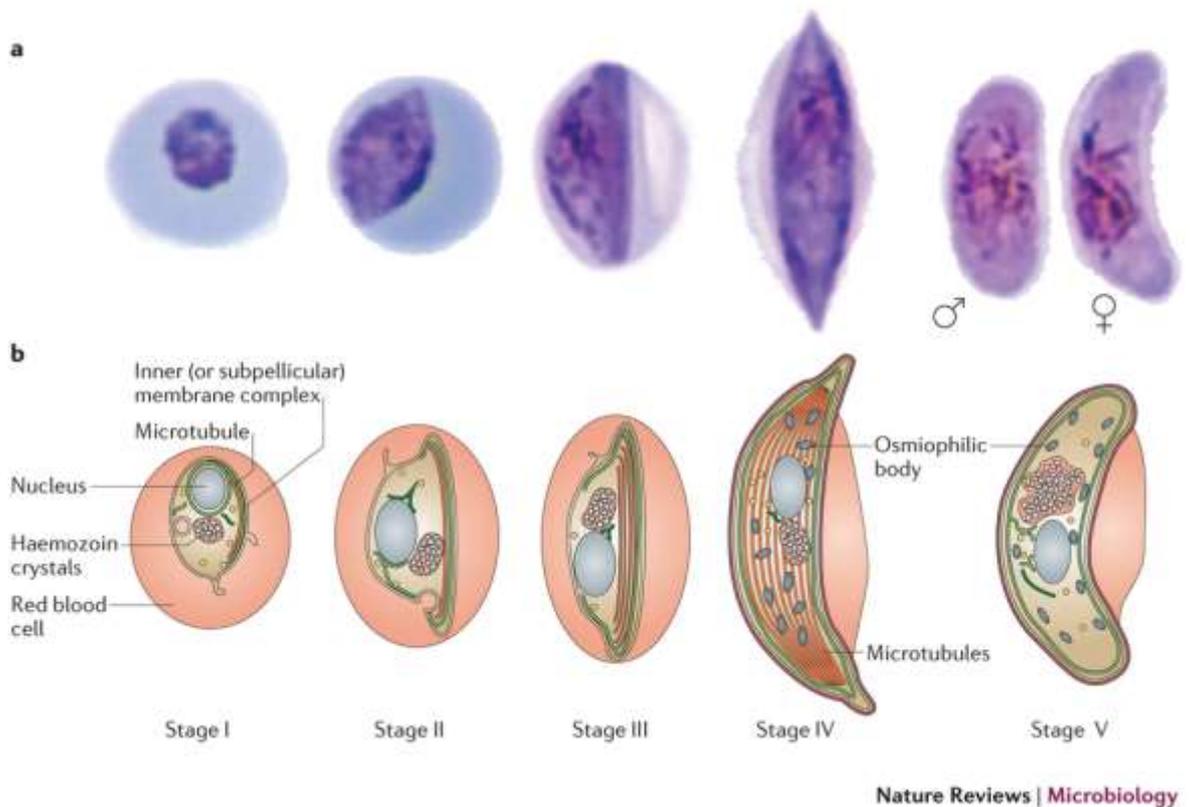
Figure 1.5 Morphological maturation of gametocytes in *P. falciparum*

The morphology of developing *P. falciparum* gametocytes is shown in Giemsa smears (a) and schematic representations (b). From stage I to stage V the parasite elongates until it encompasses the entire cell (stage III) and then induces changes that result in deformation of the host cell and the resulting banana shape of the gametocytes (Figure reprinted with permission, license 3954850712331(Josling & Llinas, 2015)).

P. berghei gametocytes do not alter in overall shape in the same manner as *P. falciparum* gametocytes. Instead they grow in size until they fill the entire host red blood cell. They are granular in appearance than asexual parasites in giemsa smears due to pigmentation within the staining. Female gametocytes have a blue stained cytoplasm attributed to a more basic pH, while male gametocytes have a pink colour which has been attributed to a less basic pH cytoplasm due to a lower number of ribosymes. In mature gametocytes (20 – 26hpi) this pigmentation difference can be used to differentiate gametocyte gender (Mons, 1986a; Mons *et al.*, 1985).

Female gametocyte

Male gametocyte

N = nucleus

→ = haematin

➤ = osmiophilic bodies
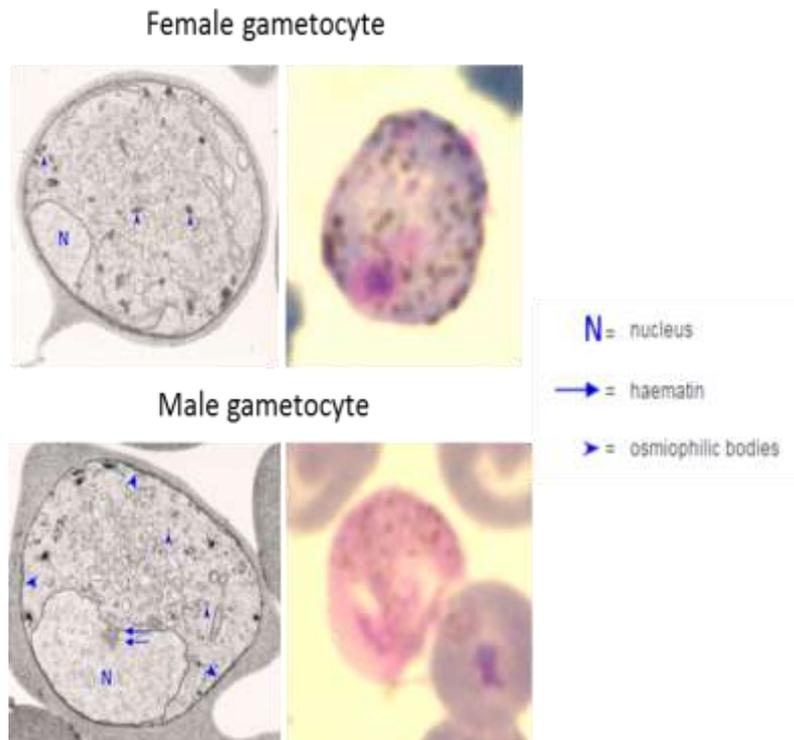
Figure 1.6 Morphology of *P. berghei* gametocytes

P. berghei gametocytes fill the entire red blood cell once mature. They commonly display dense nuclei and a pigmented appearance. They contain osmophilic bodies throughout the cytoplasm. Mature (26hpi) male and female gametocytes are distinguishable based on their cytoplasmic colouration. Female gametocytes contain a blue stained cytoplasm while male gametocytes stain with a pink cytoplasm attributed to differences in cytoplasmic pH (Mons, 1986b; Mons, 1986a).

1.4 *Plasmodium berghei:* **A rodent model**

Due to the complexity of malaria infections observational research is insufficient for determining causal effects during an infection. While tissue samples and autopsy samples can provide insight into disease progression manipulation of the parasite is necessary to support these observations. *In vitro* study of the human malaria parasite *P. falciparum* allows direct study of the human infective species but cannot be utilised to examine the entire life cycle or analyse interactions between the parasite and the host immune system, furthermore adaptations to culture will have accumulated within laboratory parasite strains (Trager & Jensen, 1976; Wu *et al.*, 1995). An alternative to *In vitro* study of the human infective species is *In vivo* study of a representative model. There is a high degree of conservation between *P. falciparum* and the rodent model *P. berghei* with approximately 80% of genes containing orthologs. Some aspects of malaria infection in the *P. berghei* rodent model are not well conserved, for example the progression of cerebral malaria and parasite sequestration in the host brain; other aspects have been shown to be well conserved, for example the interleukin-10 (Il-10) response to malaria in humans and rodents (Langhorne *et al.*, 2011; Li *et al.*, 2003; Peyron *et al.*, 1994; White *et al.*, 2010). Additionally, the preference of *P. berghei* to the invasion of reticulocytes makes it a representative model for *P. vivax* infections. Though *P. berghei* and *P. yoelii* are the most commonly used rodent models due to their tractability for genetic manipulation *P. chabaudi* is considered the closest model for immune evasion modelling (Janse *et al.*, 2011).

Other benefits to use of the rodent model include; its increased transfection efficiency, calculated as in Janse *et al* 2006 is in the range of 10^2 – 10^3 (Janse *et al.*, 2006) compared to 10^6 for *P. falciparum* (Janse *et al.*, 2006; O'Donnell *et al.*, 2002); the availability of direct knockout vectors with high integration efficiency (see section 1.6.4) in the form of recombineering vectors; its updated annotation (2001; Aurrecoechea *et al.*, 2009; Bahl *et al.*, 2002) and the rodent malaria genetically modified parasite database (Janse *et al.*, 2011; Khan *et al.*, 2013), which details genetic modifications that have been achieved or attempted in rodent malaria parasites by laboratories worldwide.

1.5 **Transcriptional regulation in *P. berghei***

Initial studies to identify transcription factors in the further branched eukaryotes, such as apicomplexans, showed a distinct lack of conserved DNA binding domains which are found in other eukaryotic transcription factors (Gardner *et al.*, 2002; Templeton *et al.*, 2004). In *P. falciparum,* for each proposed DNA-binding transcription factor, there are estimated to be 800 other gene targets. This is in contrast to yeast whose ratio of transcription factor to target genes is 1:29. Analysis of DNA binding motifs in *P. falciparum* corroborated the lack of conserved transcription factors but, in the search for genuine transcription factors, a protein (PF14_0633) containing an AT-hook DNA binding domain was studied further. This protein was structurally analysed and the secondary structure was shown to be consistent with the apetela2-integrase (AP2) DNA binding domain in *Arabidopsis*. Further analysis of apicomplexan genomes identified the ApiAP2 family consisting of 27 members (Balaji *et al.*, 2005). This remains the only family of transcription factors identified in *Plasmodium*.

1.5.1 The Api-Ap2 family of TF

After the identification of the ApiAP2 transcription factor family their expression patterns were analysed in datasets collected for transcriptional profiling in asexual blood stages (Bozdech *et al.*, 2003; Otto *et al.*, 2010), gametocytes (Silvestrini *et al.*, 2005; Yuda *et al.*, 2015), mosquito stages (Yuda *et al.*, 2009) and liver stages (Tarun *et al.*, 2008; Yuda *et al.*, 2010b) as well as datasets collected throughout the entire life cycle (Le Roch *et al.*, 2003) in *P. falciparum* and *P. berghei*. In different *Plasmodium* species, ApiAP2 members are expressed at all stages of the lifecycle and have been implicated in a number of different processes (Painter *et al.*, 2011). In *P. berghei* 26 ApiAP2 family members have been identified. All contain between 1 and 3 AP2 domains and seem to show stage specific expression which may be linked to their roles (Figure 1.7).
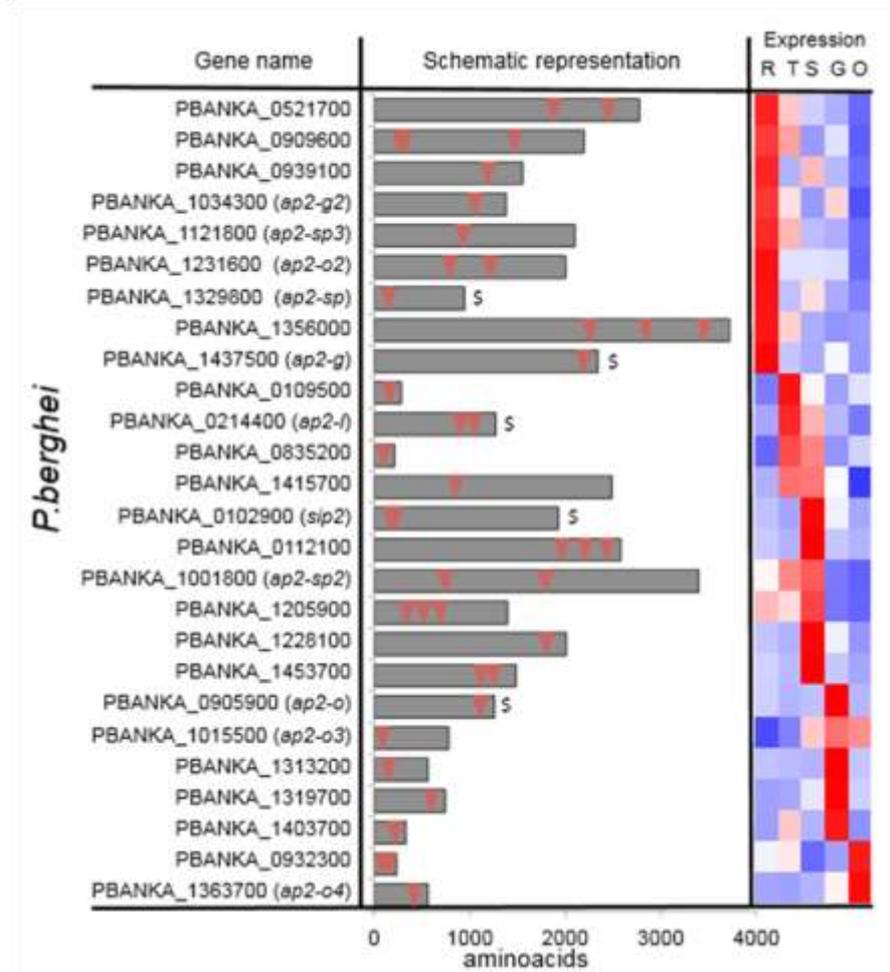
Figure 1.7 Structure and expression analysis of AP2 family members in *P. berghei.*

For each of the 26 AP2 family members in P. berghei their structure was analysed to determine the presence of AP2 domains (pink arrowheads). All of the family members contain between 1 and 3 of these domains. Expression of each AP2 was determined by RNAseq data at 5 life cycle stages; Rings (R), Trophozoites (T), Schizonts (S), Gametocytes (G) and Ookinetes (O) to determine when they were transcribed (Otto paper REF, Kasia screen paper, Figure adapted with author's permission). All genes indicated $ have been characterised further.

In *P. falciparum* the binding sequences and stage specific expression of several ApiAP2 family members were analysed showing similar specificity of stage specific expression based on mRNA abundance (Campbell *et al.*, 2010; De Silva *et al.*, 2008).

Though this study is focussed on sexual commitment and will therefore focus on the transcription factors identified as playing a role in sexual commitment, several other AP2 transcription factors have been characterised with roles at different stages (Figure 1.7). These AP2 factors, AP2-O, AP2-Sp, AP2-L and SIP2 will be discussed briefly.

The first studied AP2 family member was designated AP2-O. To investigate AP2-O expression, a GFP-tagged protein was generated (AP2-O::GFP) and the replacement with this recombinant tagged protein did not appear to affect development and life cycle progression. Expression of the protein was not detected in the blood stages and was only identifiable 8 hours post fertilisation in ookinetes. As expected the signal was localised to the nucleus of the ookinetes and intensity of the signal increased as development and maturation continued. While the AP2-O::GFP protein was not detected in any blood stages, qPCR was performed to quantify the abundance of transcripts in ookinetes and gametocytes (male and female). This analysis showed that though protein was not translated in gametocytes there was a high abundance of mRNA for *ap2-o*. This led to the suggestion that *ap2-o* mRNA was translationally repressed (Mair, 2006), which was confirmed with an immunoprecipitation (IP) of the repression complex found in female gametocytes. Disruption of *ap2-o* had no effect on asexual blood stages and gametocyte development seemed unaffected, however successful infection of mosquitoes could not be observed. Subsequent crosses of *ap2-o* deficient males and females with their wild type counterparts indicated that the inability to generate normal, invasive ookinetes was inherited from the female gametocytes. Direct regulation of known and hypothesised ookinete specific transcripts was also shown for AP2-O through chromatin immunoprecipitation (ChIP) with the AP2-O::GFP line (Kaneko *et al.*, 2015; Yuda *et al.*, 2010b).

To elucidate its function, AP2-Sp was initially tagged with GFP to determine its protein expression (AP2-Sp::GFP). The GFP signal was detected from day 6 post blood feeding in the midgut of mosquitoes and continued in the oocysts until individual sporozoites with high GFP signal were released. Midgut and salivary gland sporozoites exhibited a GFP signal indicating AP2-Sp played a role at the sporozoite development stage. Ap2-Sp knockouts displayed no growth impairment in the blood stages but failed to produce sporozoites in the midgut of mosquitoes after blood feeding. In the midguts of these mosquitoes the size and number of oocysts remained the same for the AP2-Sp deficient line and wild type. Further characterisation showed nuclear divisions occurred as normal in the *ap2-sp* null oocysts but invagination of the plasma membrane and proper segregation of the sporozoites did not occur. Interestingly, expression of an AP2-

Sp chimera with AP2-O, which contains the AP2-Sp AP2 domain in the AP2-O transcription factor, retaining AP2-O's expression profile (native activity in ookinetes) resulted in expression of several known and hypothesised sporozoite specific genes in ookinetes which corroborated its role in sporozoite development (Yuda *et al.*, 2010a).

To analyse the point of action of AP2-L a knockout was produced. In both the blood and mosquito stages proliferation occurred normally and life cycle progression was maintained. However in the liver stages development was dramatically impeded. After invasion of hepatocytes, replication occurred and thousands of merozoites per hepatic schizont were generated. The AP2-L was also tagged with GFP (AP2-L::GFP) and when examining the AP2-L::GFP parasites, the fluorescence intensity in the developing hepatic schizont rapidly increased then depleted as the schizont became mature. This indicated a role for the transcription factor in merozoite development in these hepatic schizonts. In the knockout line, invasion of hepatocytes seemed unaffected with the burden of parasites in each infected animal being the same at 24 hours post infection with salivary gland sporozoites. However development ceased at approximately 36h after infection. This correlates to hepatic schizogony, when nuclei have divided but merozoite segregation has not occurred. These data indicate AP2-L plays a role in the development of hepatic merozoites (Iwanaga *et al.*, 2012).

The AP2 member SIP2 has a role in transcriptional regulation of sub-telomeric tandem arrays found upstream of *var* genes. This transcription factor interacts with the conserved sub-telomeric var promoter element, SPE2 (Voss *et al.*, 2003). In this work SIP2 binding to chromosome end clusters was demonstrated by colocalisation with the known *var* gene interactor heterochromatin protein 1 (HP1 (Flueck *et al.*, 2010)). It was also shown that the entire SIP2 protein does not bind to these SPE2 regions but an N-terminal portion of the transcription factor. High affinity binding to the sub-telomeric regions was demonstrated with this N-terminal portion and the SPE2 regions *in vivo* however binding to single chromosome integral sites where SIP2 was predicted to bind could not be demonstrated. While this binding indicated a role in regulation of expression of these sub-telomeric genes a direct activation effect is unlikely. In support of this hypothesis, overexpression of SIP2 resulted in limited dysregulation of genes and

caused no defects in growth and development. The role of this transcription factor is considered essential as attempts to disrupt the gene were unsuccessful (Flueck *et al.*, 2010).

1.5.2 *ap2-g* and sexual commitment

In independent studies the transcription factor responsible for initiation of sexual commitment was elucidated in *P. falciparum* (Kafsack *et al.*, 2014) and *P. berghei* (Sinha *et al.*, 2014). Both studies utilised genome wide sequencing (GWS) on gametocyte non-producing (GNP) lines to identify genes consistently mutated. Long term culture and serial passage *in vivo* without transmission through mosquitoes has previously been shown to result in a loss of certain traits, including cytoadherence and gametocyte commitment (Day *et al.*, 1993). This has led to the identification of multiple historic GNP lines in *P. falciparum* and *P. berghei*. In the *P. falciparum* study to identify the transcription factor involved in gametocyte commitment, two of these historic GNP lines were sequenced and both contained stop-codons before the DNA-binding domain in *ap2-g*. Further to this, transcript levels of *ap2-g* were linked to gametocyte production. Lines with higher abundance of *ap2-g* showed higher commitment to gametocytogenesis (Kafsack *et al.*, 2014). In *P. berghei* ten lines were serially passaged in mice for 52 weeks and their gametocyte conversion rates monitored over time. Three of these serially passaged lines became GNPs within this time. Several others appeared to reduce their commitment rates. These, along with a historic GNP (line 2.33) were sequenced and the only gene that consistently contained independent mutations disrupting the coding sequence was *ap2-g* (Sinha *et al.*, 2014). In both of these studies *ap2-g* null mutants were generated to prove specificity and in both cases the lines were GNPs. Furthermore, competitive growth assays revealed an advantage in *P. falciparum* and *P. berghei* for the GNP (*ap2-g* null) lines who consistently outgrew their gametocyte producing parental counterparts (Kafsack *et al.*, 2014; Sinha *et al.*, 2014).

In *P. falciparum* several early gametocyte specific genes have been identified through comparisons between a GNP line and a line which consistently produces 10 – 20% gametocytes per cycle (Silvestrini *et al.*, 2005). To further validate the *ap2-g* null mutant and to identify regulatory targets, several of these known

gametocyte transcripts were compared over time in synchronous cultures. Several known gametocyte transcripts were downregulated from as little as 3 hours post invasion in the *ap2-g* null line. Analysis of the upstream regions of several of these rapidly downregulated genes revealed the presence of the identified AP2-G binding motif (Kafsack *et al.*, 2014). In *P. berghei* early gametocyte markers are not known. In an attempt to elucidate downstream proteins in the commitment cascade, the promoters of several consistently downregulated genes in the GNP lines that also contained an AP2-G binding motif were analysed for gametocyte specific expression. In this assay approximately 2kb of the promoter regions was inserted upstream of a *cfp* expression cassette in a silent locus (P230p (van Dijk *et al.*, 2010)) in a parental line (820) expressing GFP under the control of a male specific promoter (PBANKA_041610) and RFP under the control of a female specific promoter (PBANKA_131950). Analysis of CFP expression indicated that 16 of the 18 promoters tested showed specific expression in male (2), female (6) or both gender (8) gametocytes. Together these data indicated that the *ap2-g* transcription factor acts as the trigger to begin a cascade of expression resulting in gametocyte commitment in both *P. falciparum* and *P. berghei* and that key gametocyte specific transcripts can be identified through binding with this transcription factor.

Further to this in *P. berghei* it was noted that, *ap2-g* itself contained several AP2-G binding motifs in its promoter region. Two motifs were identified approximately 1.5kb upstream of the transcription start site and a third was identified 7kb upstream of the start site. To identify if these binding motifs played a role as a positive feedback loop for expression of AP2-G disruptions to the promoter region and mutations to the binding domains were made. If the promoter was disrupted 2kb or 3kb upstream of the transcription start site a reduction in gametocytes was observed when compared to the 820 parental control (Figure 1.8). If the AP2-G binding domains at -1.5kb were also mutated the reduction in gametocytes intensified. This combined data suggests that expression of AP2-G results in its self-promotion through a feedback loop allowing for in-cell levels of AP2-G to increase rapidly and that this loop is necessary for maximal gametocyte commitment.
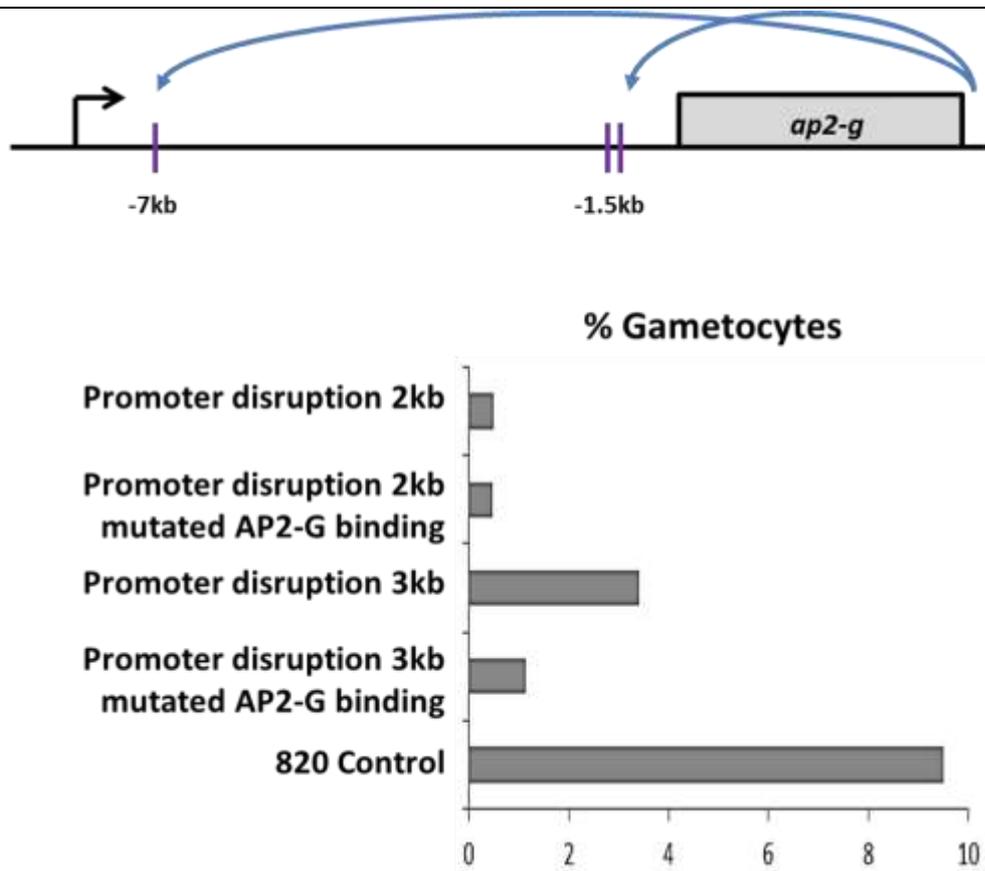
Figure 1.8 *ap2-g* self-regulation

To determine if the AP2-G binding motifs upstream of *ap2-g* were involved in a positive feedback loop the promoter region was disrupted by inserting a selectable marker cassette. This was completed in combination with DNA binding domain mutations. Any interruption of the promoter region (up to -3kb) resulted in a reduction in gametocytes. This reduction was further exacerbated by mutations of the DNA binding domain.

The identified transcription factor in *P. berghei* encodes a protein of 2,339 amino acids. It contains a 60 amino acid AP2 domain at its N-terminus. This domain has been identified in multiple transcription factor proteins.
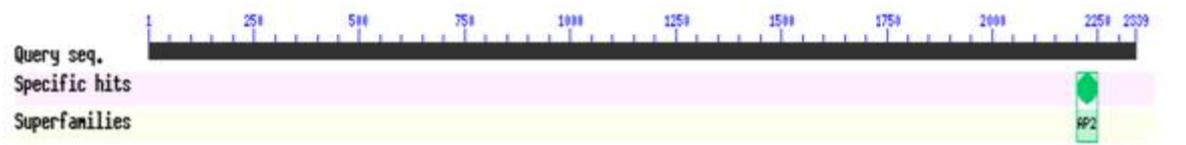
Figure 1.9 AP2-G domain architecture

The identified transcription factor *ap2-g* contains a 60 amino acid DNA binding domain of the AP2 class at its N-terminus (Marchler-Bauer *et al.*, 2014; Marchler-Bauer *et al.*, 2011).

1.5.3 *ap2-g2* and sexual development

A second gametocyte specific transcription factor was identified as important in commitment in *P. berghei* gametocytes, *ap2-g2*. While disruption of *ap2-g* resulted in complete loss of gametocytes disruption of *ap2-g2* still allowed the generation of some morphologically identifiable mature female gametocytes (<5% of WT). However, these female gametocytes were not transmissible through mosquitoes. In contrast to the *ap2-g* null lines the *ap2-g2* mutant displayed WT growth rates indicating that although mature gametocytes were not detectable commitment had occurred and non-viable gametocytes developed. In addition, known gender specific promoters became active (determined by the onset of GFP and RFP fluorescence under the control of known male and female gametocyte specific promoters, PBANKA_041610 and PBANKA_131950 in the 820 line respectively). Combined these data indicate the role of AP2-G2 is post gametocyte commitment and once committed these parasites cannot return to the asexual development pathway (Sinha *et al.*, 2014; Yuda *et al.*, 2015). Further work has identified *ap2-g2* as a transcriptional repressor. In *ap2-g2* null parasites initial sexual differentiation from asexual parasites occurs but transcripts abundant in both gametocytes and asexual stages are prevalent. This disarrayed expression of genes seems to prevent successful commitment and maturation forming identifiable gametocytes. ChIP-seq analysis of AP2-G2 binding identified a large amount of genes downregulated in gametocytes, thereby indicating its role as a repressor. Several of the genes identified are involved in sexual proliferation which would need to be halted to allow for gametocyte maturation. It has been hypothesised that the repressive role of *ap2-g2* is required for correct differentiation of committed gametocytes into mature male or female gametocytes. They hypothesise that the forms of gametocytes generated in the *ap2-g2* null mutant line are aberrant and while they differentiate into males or females they do not express gender specific proxies (GFP for male gametocytes and RFP for female gametocytes) to the same degree as (the 820) parental line gametocytes potentially indicating a reduction in expression of all gender specific transcripts (Sinha *et al.*, 2014; Yuda *et al.*, 2015).

1.6 **Genetic manipulation of *P. berghei***

The ability to manipulate *P. berghei* with high efficiency has allowed the functional knockout of a plethora of genes of interest both on a gene by gene basis and in a high throughput manner (Schwach *et al.*, 2015). The existence of databases that document genetic manipulations and growth and development phenotypes of the genes targeted by laboratories around the world (Janse *et al.*, 2011; Khan *et al.*, 2013; Schwach *et al.*, 2015) allows vast data sharing in the *P. berghei* research community. One of the limitations to the study of gene function is the limited number of selectable markers available in *P. berghei*. For each mutation to the genome a marker is required to allow selection of the integrants. Currently, only three drug selection techniques exist for the positive selection of integrants whereby drug pressure selects for the integrated population. The first selectable marker developed for use in *P. berghei* was the mutated *Plasmodium* or *Toxoplasma dhfr-ts* (*dihydrofolate reductase-thymidylate synthase*) gene which conferred resistance to the drug pyrimethamine which can be readily administered in drinking water. Second, shown to be effective in conjunction with pyrimethamine selection was the introduction of the *hdhfr* (*human dihydrofolate reductase*) gene. This gene confers resistance to the drug WR99210 in background lines with and without resistance to pyrimethamine (via the mutated *dhfr-ts* gene) allowing for additional genetic manipulations (de Koning-Ward *et al.*, 2000). Most recently, the mutated *dhps* (*dihydropteroate synthase*) gene of *P. falciparum* was shown to confer resistance to sulphadiazine administered through drinking water to asexual *P. berghei* parasites (Kobayashi., 2016; Lumb & Sharma, 2011; Triglia *et al.*, 1997; Wang *et al.*, 1997). The limitation for drug selection of *P. berghei* is the efficacy of drugs *in vivo*, their effective dosing and their toxicity to the rodent host. Methods have been developed to allow recycling of the selectable marker or to allow the generation of marker free lines after a mutation.

1.6.1 Selectable marker recycling

To allow multiple manipulations to the parasite genome one of the first techniques developed was that of marker recycling (Braks *et al.*, 2006; Maier *et al.*, 2006). To achieve this, the negative selectable marker yfcu that confers sensitivity to the drug 5-FC was used in conjunction with a positive selectable marker (often *dhfr-ts* which confers resistance to pyrimethamine). In vectors that allowed for this kind of recycling, the positive and negative selection cassettes were flanked by homologous 3'UTR regions. These small homologous regions act not only to regulate the expression of the resistance gene but as a means to recycle the marker (Figure 1.10). Initially the genome manipulation (for example gene knockout by replacement with the selectable marker, as shown in Figure 1.10) was selected for with the positive selectable marker which confers drug resistance through pyrimethamine. Provided these mutant parasites were viable at the blood stage they were then cloned. This cloning ensured that only parasites who had integrated the mutation were present in the population.

The clonal integrated line was then selected with the negative selection drug 5-FC by administering the drug in the drinking water (Orr *et al.*, 2012). As all of the parasites in the clonal line had integrated the mutation and the *dhfr-ts* resistance cassette they had all also inherited the *yfcu* cassette. This cassette conferred sensitivity to the negative selection drug 5-FC. Selection, by administering 5-FC in the drinking water, forced the recombination of the two 3'UTRs that flank the entire selection region. This recombination event removed not only the *yfcu* cassette that was causing the sensitivity to 5-FC but also the *dhfr-ts* cassette that was initially used for positive selection. In the negatively selected line, the alteration initially achieved via integration remained, along with a single 3'UTR used to bring about negative selection (Figure 1.10). Although highly efficient this negative selection left a mixed population of parasites with and without the selection cassettes, and therefore the lines had to be cloned again to ensure a population was made up of only integrants that had lost the selectable markers.
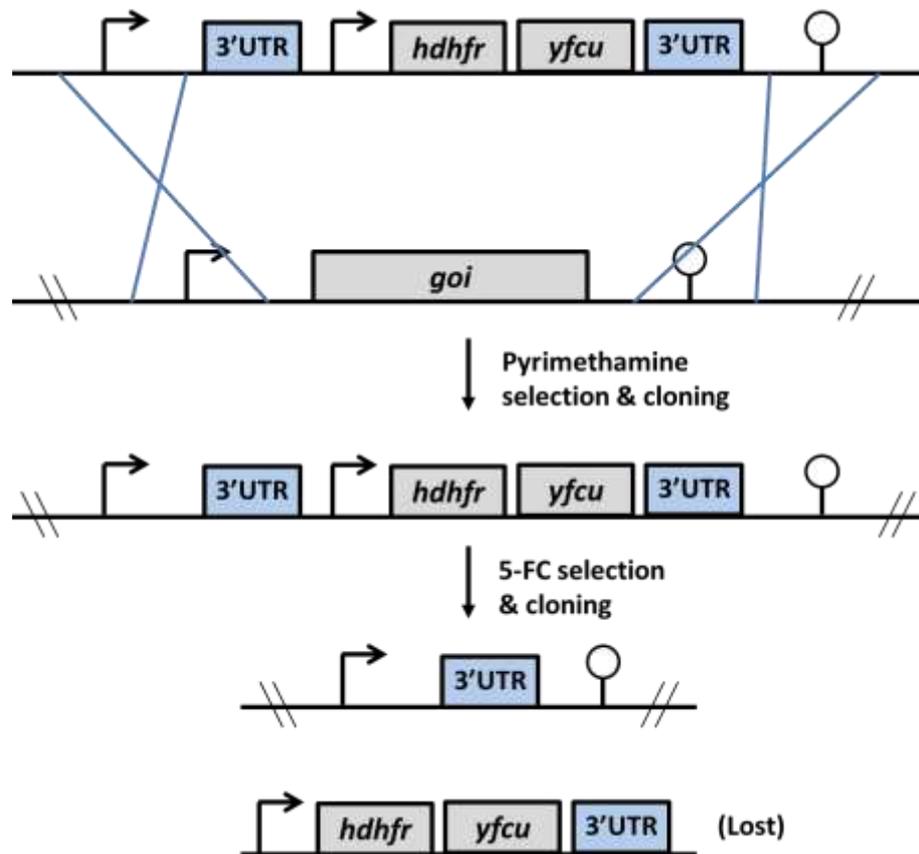
Figure 1.10 Schematic representation of selectable marker recycling

Schematic representation of the process to generate a marker free gene knockout line. In this representation the gene of interest is replaces with the selectable markers *hdhfr-ts* and *yfcu* that confer resistance to pyrimethamine and sensitivity to 5-FC respectively. Selection with pyrimethamine and subsequent cloning resulted in a population exclusively made up of integrants. Selection with 5-FC forces the recombination of the two homologous UTRs (blue) leaving a population of knockout without a selectable marker. Subsequent cloning ensured the whole population had lost the selectable marker.

This selectable marker recycling method allows for subsequent manipulations to be achieved using the same selection drug. While this is of great advantage it does have limitations. Firstly, this process involves two selection steps which can be lengthy and also requires two cloning steps which would increase animal usage. Secondly, the process of negative selection, while removing the selectable marker, leaves behind a footprint (this was the 3'UTR). If subsequent manipulations were required in nearby regions of the genome this footprint could lead to complications and unexpected recombination events.

1.6.2 Marker free manipulations

Another method used to generate mutated lines that do not contain selectable markers is the gene in marker out (GIMO) approach (Figure 1.11). In this method the initial manipulation to the genome integrated a positive/negative selection cassette (often *dhfr-ts* with *yfcu*) conferring resistance to a drug (positive) and sensitivity to a second drug (negative) in the locus of interest. Once this line was selected with the positive selection drug and cloned, generating what is known as a GIMO line, the second mutation was completed. This integrated the DNA that is actually required for the manipulation in the same region (using the same homology arms as integration of the positive/negative cassette, in the example a fluorescent reporter Figure 1.11). After transfection integrants were selected for using the negative selection drug. Only integrants that had replaced the positive/negative selection cassette with the DNA of interest were not sensitive to the negative selection drug. Those that had retained the positive/negative selection cassette were sensitive to the drug. After selection, cloning is completed to ensure a population was only integrants (Lin, 2011).

This method has both advantages over selectable marker recycling and disadvantages. The major advantage to this approach is that a single GIMO line can be used to integrate multiple different DNA's of interest in the same locus. This reduces the time and animal requirements for line generation and is of use for complementation and transgene expression in, for example, overexpression systems. The major disadvantage is that this approach is only really useable for the integration of additional DNA into the genome (transgenes) in regions known to be undisruptive to life cycle progression.
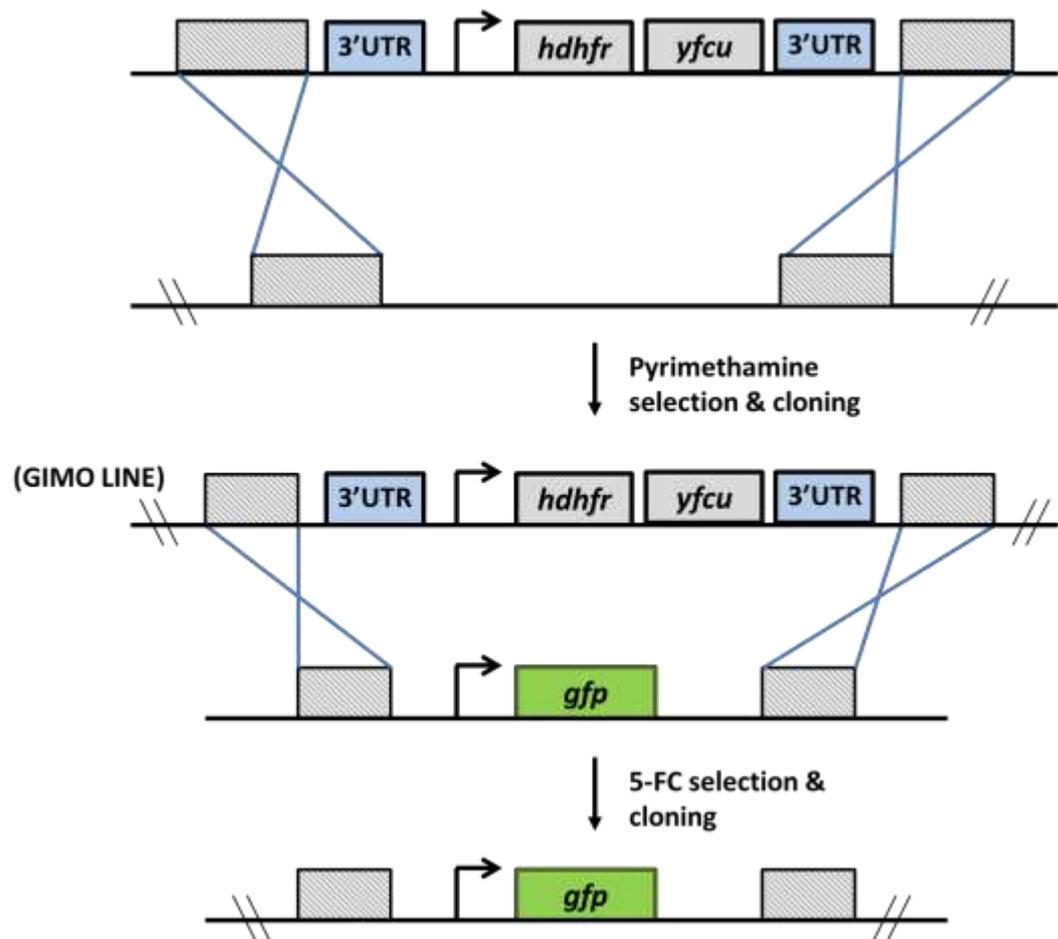
Figure 1.11 Schematic representation of the GIMO (gene in marker out) method

Schematic shows the steps involved in first, generation of a GIMO line and then integration of a reporter into the GIMO line. In the first transfection the region of the genome permissive to manipulation is replaces with the positive/negative selectable cassette. Pyrimethamine selection and subsequent cloning results in a GIMO parent line that is resistant to pyrimethamine and sensitive to 5-FCU. Upon the second transfection, in this case integration of a fluorescent reporter, 5-FC is used to select for parasites that had replaced the selection cassette with the fluorescence reporter. Cloning of this would give rise to a line universally GFP expressing with no trace of the selectable marker.

1.6.3 Functional analysis of non-essential blood-stage genes

Genes that are not essential for the asexual blood stages of the *P. berghei* infection are the most amenable to study. To study the function of genes that cause little to no impediment to asexual proliferation the gene of interest can be directly knocked out. This can be achieved by disruption or replacement of the gene of interest with a selection cassette allowing for the detection and selection of mutants. Wide scale genome disruption has been achieved using a high-throughput knockout technique whereby pools of barcoded knockouts are attempted *in vivo* and the proportion of the barcodes detected post-selection gives insight into the essentiality or dispensability of the gene of interest (for full details see section 1.6.4). In this work, approximately 50% of the *P. berghei* genome was targeted for direct knockout in a screen by the PlasmoGEM project. Relative growth rates for the genes targeted show that only 36% of genes could be knocked out at the asexual blood-stage without impact on the growth rate. 46% of the genes had been identified as likely essential at this stage, and therefore not targetable by direct knockout, and a further 17% could be targeted but the resulting parasites show growth deficiencies in the asexual blood stages (Gomes *et al.*, 2015; Schwach *et al.*, 2015). If we extrapolate this data as representative for the entire genome, this means only 36% of the genome would be amenable to study in other stages as only genes that cause no phenotype or defect in the asexual blood stages can be studied for their role at different life cycle stages.

1.6.4 High throughput vectors for gene analysis

One of the major challenges in *Plasmodium* research is the difficulty in manipulating A-T rich genomes in *E. coli*. Long stretches of A-T rich sequence are unstable in *E. coli* vectors and prevent the use of long homology regions to target genes of interest. In many systems bacterial artificial chromosomes form the basis of large scale genome targeting however these are not permissive to the A-T rich genome segments in *P. berghei* or *P. falciparum*. A method was developed to overcome this (Gomes *et al.*, 2015).

Initially, the *P. berghei* genome was fragmented and combined with the bacteriophage N15 derived vector (pJazz) which can tolerate A-T rich and repetitive stretches of DNA because of its low copy number and hairpin structure for propagation in *E. coli*. This allowed the generation of a library that contained approximately 76% of all genes in the *P. berghei* genome in their entirety. Many genes also contained multiple different library clones. While the library clones can range in length from 4kb to 20kb the average library clone is approximately 8.7kb. If a gene is not entirely covered in a library clone it can still be targeted for knockout. A gene can be targeted if > 50% of its coding sequence can be disrupted by the vector resulting in a gene disruption. If we consider these library clones approximately 91% of the *P. berghei* genome can be targeted for disruption. Selection of library clones was achieved by kanamycin selection of the *E. coli* containing the libraries (Figure 1.12 A). To genetically manipulate these library clones, generation of knockout vectors, in a high throughput way gateway technology[R] (Invitrogen) was combined with the developed recombinase mediated genetic engineering (recombineering (Wang *et al.*, 2006; Zhang *et al.*, 1998)).

The library clones in the linear pJazz vector were first made recombineering competent by electroporation with a recombinase plasmid (pSC101gbdA-tet). Selection of recombinase positive cells was achieved by retaining kanamycin pressure (library selection) and adding tetracycline (recombinase selection). Expression of the recombinase was prevented by culture at 30°C and by withholding l-arabinose (Figure 1.12 B). Induction of the recombinase was stimulated by the addition of l-arabinose to the culture and replication of the plasmid was restricted by increasing culture temperature to 37°C. At this time bacteria were again electroporated to integrate a gene specific zeo/pheS selection cassette. This cassette was provided as a PCR product containing two 50bp homology sequences to a specific target locus surrounding the selection cassette that is flanked by attR1 and attR2 sequences and in some cases a unique barcode (Figure 1.12 C). Now selection of bacterial clones occurred with zeomycin. To generate gene knockout vectors for selection in *P. berghei* the zeo/pheS cassette was replaced with the appropriate selection cassettes. To do this gateway technology was used. A generic gateway selection cassette was generated containing the *hdhfr* and the *yfcu* selection cassettes with a

promoter, all flanked by attL1 and attL2 sequences. *In vitro,* in a gateway reaction the attL1 and attL2 sequences recognise and recombine with the attR1 and attR2 sequences in the specific vectors. This resulted in the exchange of the zeo/pheS resistance cassette with the hdhfr/yfcu cassette (Figure 1.12 D).

After the gateway reaction *E. coli* were electroporated to take up the vector and screened with chloro-phenylalanine (Figure 1.12 E). Screening of clones generated using this method have provided a large number of knockout vectors that simply require release from the pJazz vector for transfection into *P. berghei* lines (Figure 1.12 F). This recombineering strategy has also been adapted and modified for the C-terminal tagging of genes of interest with a variety of different tags (Gomes *et al.,* 2015; Pfander *et al.,* 2011; Schwach *et al.,* 2015).

Figure 1.12 Schemtic representation of the generation of recombineering vectors
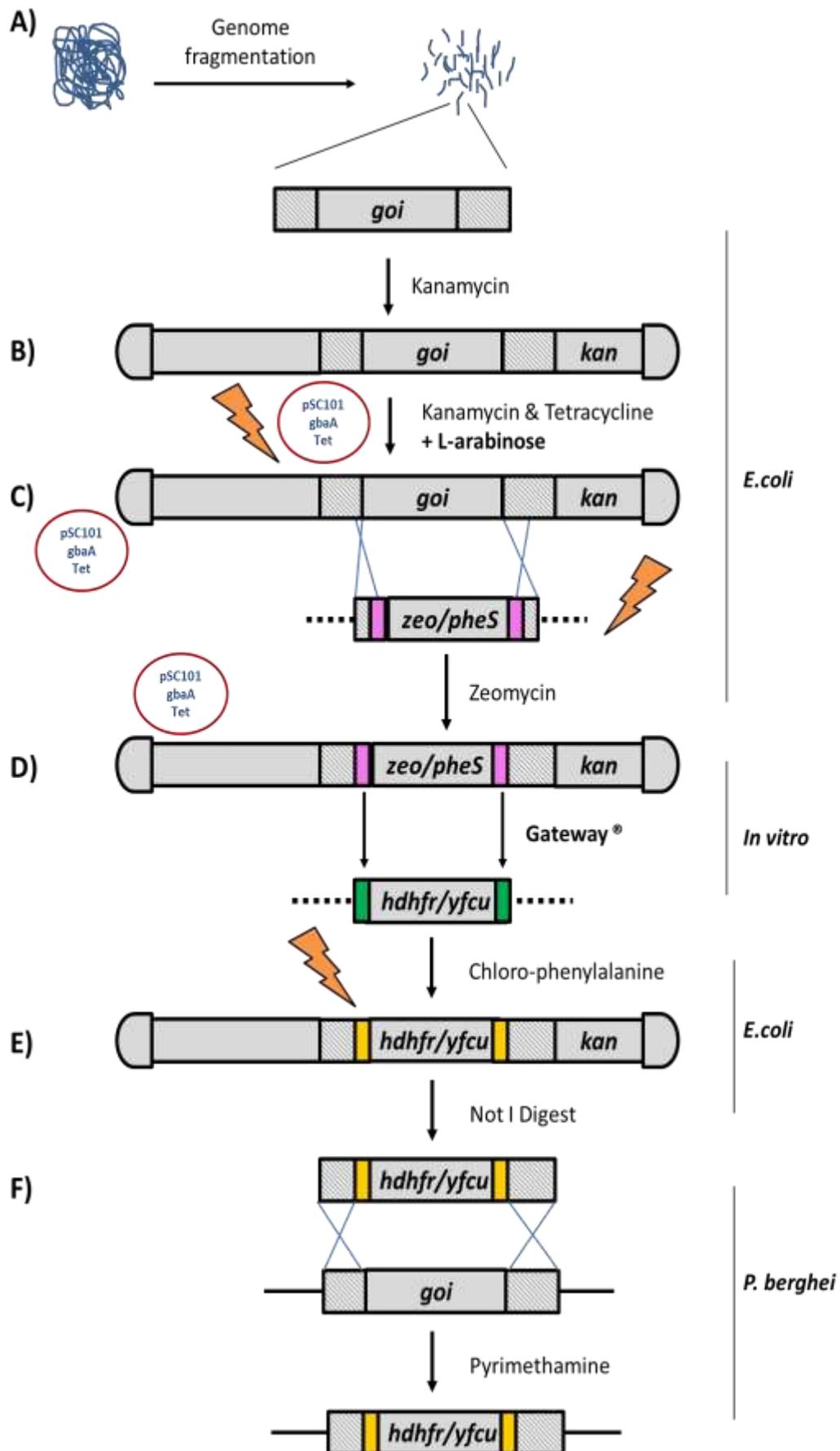
Figure 1.12 Schematic representation of the generation of recombineering vectors

A) Initially the genome was fragmented and made into a library with a N15 derived vector (PJazz). Library clones were selected with kanamycin whose resistance is conferred by the Pjazz vector.

B) The library is made recombineering competent by transfection with the pSC101-gbaA-tet vector. Selection for cells containing both the library clone and the recombineering competent plasmid was achieved with a combination of tetracycline (plasmid) and kanamycin (library) selection pressure.

C) L-arabinose was added to the cultures before electroporation to uptake a gene specific selection product. This specific selection product was made up of two 50bp homology arms (grey striped boxes) flanking attR1 and attR2 sequences (pink boxes) that in turn surround the zeo/pheS selection cassette. Selection with zeomycin resulted in vectors containing the zeo.pheS cassette.

D) The *in vitro* gateway ® reaction was then completed. Vectors were provided with a generic vector in which the *hdhfr/yfcu* selectable marker was flanked by attL1 and attL2 sequences. This replaced the *zeo/pheS* cassette in the vector leaving attB1 and attB2 sequences (from the combination or R and L sequences).

E) To allow selection after the gateway ® reaction *E. coli* were transformed and selected with chloro-phenylalanine.

F) To prepare the specific recombineering vector for transfection in *P. berghei* the vector is digested with Not I, releasing the PJazz arms required for bacterial propagation. After transfection selection occurs with pyrimethamine.

1.7 Inducible technologies in apicomplexa

A technique that has, until recently, been lacking in reverse genetics in *Plasmodium berghei* research is the ability to conditionally regulate gene expression. The asexual blood stage is the only life cycle stage amenable to transfection and manipulation, therefore only genes that are non-essential at this stage can be analysed at other stages when the gene is directly disrupted or knocked out. The high proportion of genes having an essential or important roles in the asexual blood stage mean conditional systems are vital for the study of genes at key points during the asexual blood cycle and for any functional analysis at other stages. One of the key advantages of study with *P. berghei* is that the entire life cycle is accessible for *in vivo* study. By overcoming the limitation derived from only being able to manipulate the asexual blood stages with a conditional system to study function, the entire life cycle becomes accessible for *in vivo* study of all genes.

In recent years many technologies have been adapted for conditional regulation or optimised for apicomplexans. This section compares several different technologies available before the beginning of this study.

1.7.1 Conditional regulation of proteins

Conditional regulation of proteins post-translation is an ideal target for rapid characterisation of function throughout the life cycle and multiple techniques have been developed or adapted for use in apicomplexan. The biggest limitation of protein degradation systems is that to target them for degradation they must be tagged with a targeting sequence or additional proteins. Many proteins cannot tolerate the addition of a tag at all. Its presence either results in mislocalisation of the protein or impedes its function. Some proteins can tolerate a small tag, can still function and are correctly localised, but the presence of larger tags impedes their role. The universal benefit of protein targeting conditional systems is their rapidity. Another general disadvantage is that to maintain a phenotype the protein must continually be dysregulated; however the upside to this is that the processes are generally reversible.

1.7.1.1 Destabilisation domains

The most widely utilised protein degradation method in apicomplexans was established by utilising a mutated domain of the human protein FKBP12 (a destabilisation domain, DD) that confers ligand-dependant stability to proteins that it is fused to. When not stabilised by the appropriate ligand the fusion protein is identified by the proteasome as unstable and subsequently degraded. When the ligand, Shield, is present the protein is stabilised and therefore correctly expressed (Banaszynski *et al.*, 2006). This method has been successfully demonstrated in *Toxoplasma gondii* and *Plasmodium falciparum* (Armstrong & Goldberg, 2007; Herm-Gotz *et al.*, 2007)
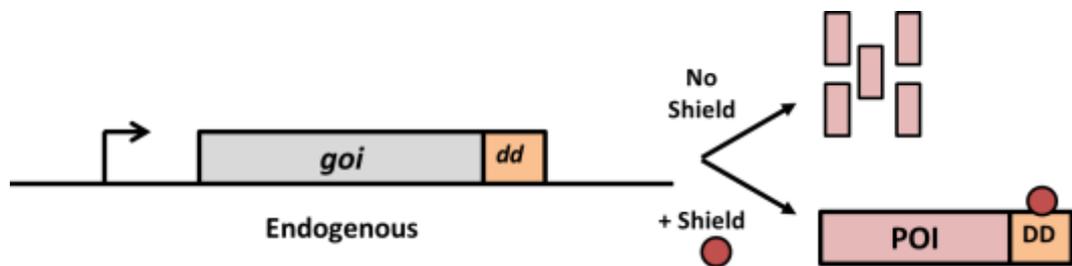
Figure 1.13 Protein knockdown using destabilisation domains

The endogenous gene of interest (*goi*) is tagged with the Destabilisation Domain (DD) and the expressed protein of interest (POI) is stabilised by the ligand Shield during selection and cloning. Removal of the ligand results in polyubiquitination of the protein of interest and subsequent degradation by the proteasome resulting in a protein knockdown

In *T. gondii*, it was shown that endogenous proteins can be replaced with a DD-tagged version and selected and cloned in the presence of Shield. Upon removal of Shield, parasites are knockdown for the protein of interest (Figure 1.13).

This technology has been widely utilised in apicomplexan research due to its many benefits. Firstly, as a protein degradation technique it rapidly removes the protein and allows functional analysis. Secondly, the technique is reversible, after removal of Shield the protein of interest is degraded but re-application of Shield restores protein stability and prevents degradation. However, it does have some limitations. Firstly, as this is a drug-on approach, meaning that Shield must continually be used in culture when selecting and cloning lines of interest, it is not only costly but continued drug application can have implications on parasite biology which may complicate phenotypic analysis. Secondly, though rapid, the technique generates only a knockdown of proteins, meaning the essential function of genes cannot always be truly identified as there is the likelihood that

there will be some residual expression. Third, to determine the extent of protein knockdown an antibody to the protein of interest must be available or the protein must be further tagged which may influence function or expression prior to Shield-induced knockdown. Detection of the protein of interest and quantification of its knockdown requires its expression to be high enough to be detected by Western blot or similar methods. Finally, a fundamental limitation to this technology is the expense and accessibility of the ligand, Shield. It shows poor pharmacokinetics *in vivo* and is not accessible to many cell types and organelles. This has widespread implications to its use as many proteins will not be targetable (Banaszynski *et al.*, 2006).

In an extension of the DD-system, additional copies of the DD-tagged protein can be inserted in the genome or expressed from an episome. Upon addition of Shield protein overexpression occurs. This overexpression alone can give key insights into protein function. With this method the phenomenon of dominant negative phenotypes may also be exploited (Herm-Gotz *et al.*, 2007; Herskowitz, 1987). This phenomenon has been predominantly utilised in *T. gondii* to uncover protein function. There are two ways in which a dominant negative phenotype can be observed. The classical way describes the overexpression of a mutated version of the gene of interest (*goi)* known to inhibit the WT-Protein of interest (POI) (Herskowitz, 1987). The second method describes the overexpression of the additional WT-Protein which gives rise to a dominant-negative phenotype (Figure 1.14). In both cases, before stabilisation the endogenous protein is fully functional however upon shield stabilisation the mutated DD-fusion protein or the overexpressed *wt-goi* inhibits the natively expressed protein resulting in a knockdown.
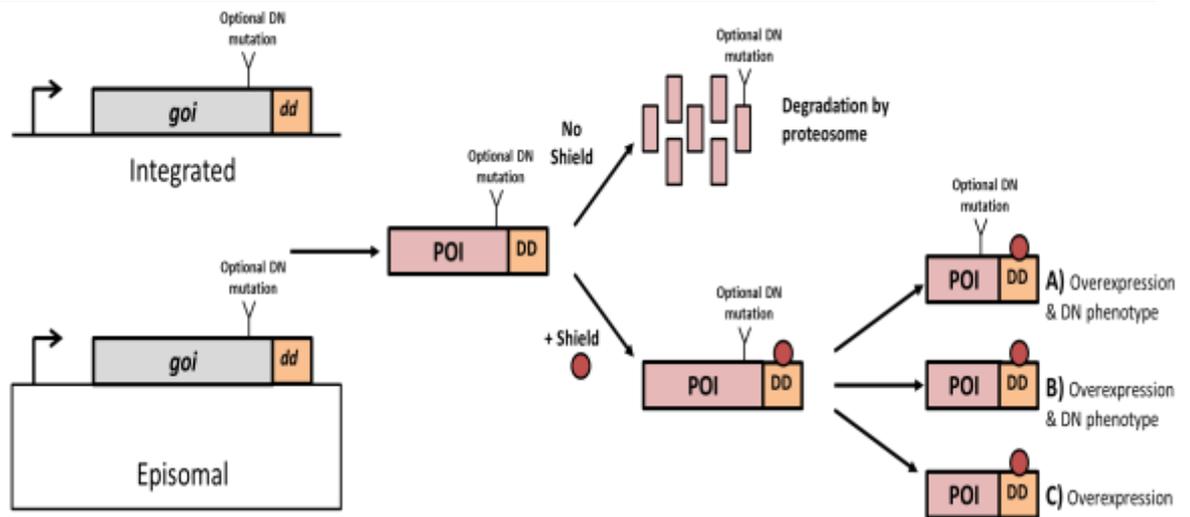
Figure 1.14 Overexpression and dominant negative phenotypes using the DD-system

An additional WT or mutated *goi* can be randomly integrated into the genome or episomally maintained and expressed by the parasite. The overexpressed protein of interest (POI) is stabilised in the presence of the ligand shield and this overexpression can have multiple outcomes.

A) Overexpression of a DN mutant-POI (protein of interest) can impede the WT-POI function giving a dominant negative phenotype.

B) Overexpression of the WT-POI can give rise to a dominant negative phenotype.

C) Overexpression of the WT-POI can have no dominant negative effect.

The benefits and limitations of this technique are similar to the endogenous degradation method described above. However, overexpression and dominant negative protein disruption being drug-off approaches that are reversible is a unique benefit. However, phenotypes identified using a dominant negative strategy often requires direct conditional manipulation to confirm the specificity of the phenotype.

Additional destabilisation domains have been developed in attempts to optimise the technique and reduce the cost associated with the shield stabilisation ligand. The *Escherichia coli* DHFR protein was engineered to be degraded along with proteins it is fused to, a technique called DDD. This system shows more rapid degradation kinetics than the FKBP12 based destabilisation domain, and its ligand trimethoprim is commercially available and much less costly than Shield (Iwamoto *et al.*, 2010). This method has been adapted and shown to be effective in *P. falciparum* with good protein knockdown (<10% maximum remaining after degradation) exhibited within ~7 hours for *pf*Rpn6 (Muralidharan *et al.*, 2011). While this overcomes the high cost of the originally reported destabilisation technique its widespread use has not been reported as yet.

1.7.1.2 Auxin inducible degradation

In apicomplexans the sole role of the Skp, Cullin, F-Box containing complex (SCF complex) is the catalysis of ubiquitination of proteins to be targeted for proteasomal degradation. Plants have evolved an additional conditional role for this complex whereby the plant hormone auxin induces the degradation of a target protein containing a degron sequence This is mediated by an interaction of the SCF complex with TIR1 (transport inhibitor response 1) which is the F-box component of the complex. In plants, the naturally occurring, highly conserved degron sequences (Aux/IAAs) involved in auxin mediated degradation are recognised by the SCF-TIR1 complex. It is the TIR1 F-box component that recognises the degron and the binding between the two is facilitated by auxin (Tan *et al.*, 2007). To utilise this system in other organisms the natural degron sequence was adapted for insertion after a *goi* generating the auxin inducible degron (AID) used as a tag. In the presence of auxin, this degron will bind the whole SCF-TIR1 complex resulting in polyubiquitination of the protein of interest and proteasomal degradation (Nishimura *et al.*, 2009).

To use this system in *P. berghei* it was necessary to generate parental lines which express the plant specific F-box protein TIR1 which directly interacts with the SCF complex and the degron as this protein is not naturally present in the genome. In these parasites, a *goi* is then fused to the AID and HA-tags. In successful conditional regulation, a functional protein of interest tagged with the AID and multiple HA-tags (to allow for immunodetection) replaces the endogenous protein in parasite lines expressing the *oryza sativa* TIR1. Upon incubation with auxin, the protein is engaged by a complex of proteins, comprising *pb*CUL1, *pb*SKP1, *pb*RBX, *pb*E2 and the *os*TIR1 that recognises the AID tag and results in ubiquitination (Figure 1.15). This modification of the POI causes recognition and rapid (within 15 minutes) degradation by the proteasome resulting in a rapid knockdown (Philip & Waters, 2015).

As with other proteasome degradation methods one of the major advantages of this system is the rapid depletion of the protein of interest. A protein knockdown using the auxin system holds advantages over the more established DD-method in that it is a drug-off method so selection and parasite growth takes place under no drug pressure. However, it also has its own disadvantages. As yet the hormone auxin has not been tested for *in vivo* knockdowns. Like with other protein degradation systems that the protein of interest must be tagged with the degron

and HA-tags, for many proteins this could alter and impede function before the knockdown has even been attempted. Furthermore, degradation is reliant on the ubiquitination complex which marks the protein for degradation and the proteasome for the actual degradation. Many proteins may be expressed in parts of the cell where one of both of these components is not present and therefore degradation would not occur. A limited number of proteins have been successfully degraded with this system and therefore the proportion accessible for this type of degradation is not known. Finally, the analysis of a knockdown requires the protein levels to be detectable by western blot at endogenous levels, for some proteins this may not be possible.
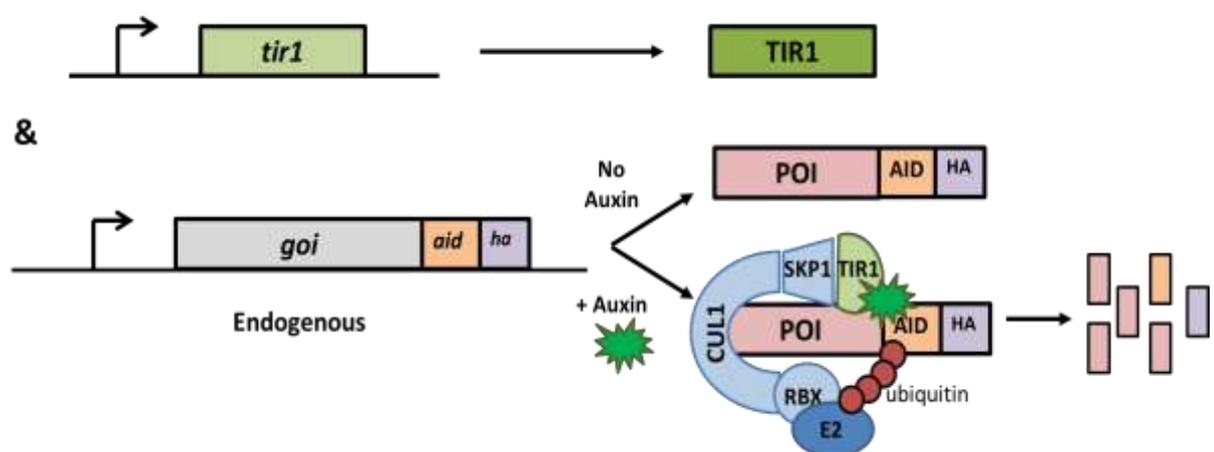
Figure 1.15 Protein knockdown using the auxin degron system

In a parental line expressing a plant F-box protein TIR1, proteins of interest are targeted for ubiquitination and degradation by an AID tag in the presence of the hormone auxin. The protein of interest (POI) is tagged with an auxin inducible degron (AID) and an HA tag. In the presence of the plant hormone auxin the protein is targeted for ubiquitination by the plant TIR1 and the other endogenous components of the SCF complex (Cul1, Skp1, RBX and E2 ligase). This polyubiquitination targets the protein for degradation by the proteasome resulting in a knockdown.

1.7.1.3 Knocksideways

A final method of conditional protein regulation, termed knocksideways (KS) renders already expressed proteins non-functional by rerouting and sequestering them to a membrane or structures, often the mitochondria, within the organism removing them from their site of function (Belshaw *et al.*, 1996; Robinson *et al.*, 2010). Initially, this technique requires a parental parasite line with a specific protein (anchor fragment) highly expressed only in a structure or membrane labelled with a fluorescent protein and the rapamycin binding protein domain FRB. In this line, endogenous proteins are C-terminally tagged with a different fluorescent protein or tag and the rapamycin binding protein domain FKBP12.

When rapamycin is administered *in vivo* or added to culture, the FKBP12::POI fusion is rerouted and heterodimerises with the anchor. This removes the protein from its active locale and putatively inhibits or prevents it functioning (Figure 1.16).
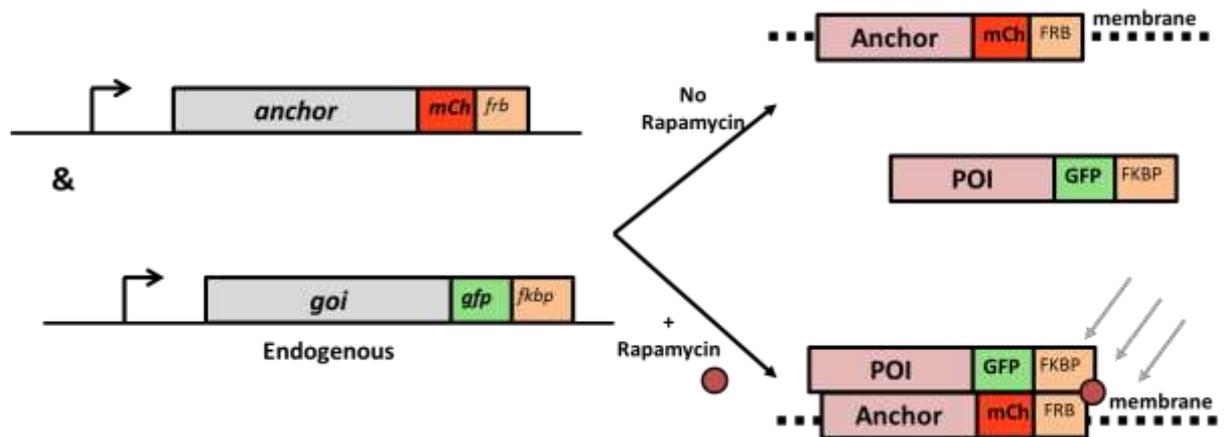
Figure 1.16 Protein sequestration and phenotypic knockdown with knocksideways

In a parental parasite line highly expressing an anchored fluorescent protein with an FRB domain, the endogenous gene of interest (*goi)* is tagged with fkbp::gfp. In the absence of rapamycin, proteins are expressed and carry out their function. Upon the addition of rapamycin, proteins are rerouted and incapacitated. This results from the dimerization of FKBP with the membrane bound FRB. This relocalisation of the protein of interest (POI) can impede protein function resulting from a knockdown in targeted proteins.

One of the key advantages to the knocksideways approach is that the translocation of proteins from their site of activity to a membrane bound structure in the cell, where they cannot perform their function is a very rapid process (< 10 s), even more so than degradation where knock down is typically of the order of 10's of minutes to hours), and has shown great promise with highly stable proteins (Robinson *et al.*, 2010). A second advantage is rapamycin has been shown to be efficacious both *in vivo* and *in vitro* therefore this technique would be applicable in a rodent model as well as in culture. However, the disadvantages are that only proteins known to be unaffected by a tag are targetable by this method, the localisation of the protein must be known to effectively reroute it and disrupt function and ideally the protein should be expressed to levels that are detectable by immunofluorescence to determine that relocalisation has occurred. This system has been adapted for application to *P. berghei* in my host laboratory (K. Hughes, unpublished data).

1.7.2 Translation regulation systems

Regulation of gene expression by controlling when expression occurs has been successfully utilised in eukaryotes and was first adapted for *T. gondii*. Since its first adaptation to apicomplexa it has been optimised for use in multiple organisms, in several ways.

1.7.2.1 Tet-repressor system

The first system to be adapted to *T. gondii* regulated gene expression by introducing Tet operator (TetO) sequences upstream of the *goi* transcriptional start site in a parasite line which constitutively expressed a Tet repressor (TetR). TetR binds the TetO operator sequences adjacent to the *goi* resulting in reduced expression of the gene. In the presence of the tetracycline derivative anhydrotetracycline (ATc) TetR interacts preferentially with ATc resulting in a conformational change that inhibits its binding with the TetO sequences and therefore maintains gene expression (Meissner *et al.*, 2001), Figure 1.17).

There are several limitations with this technique. One of the key issues is that insertion of the TetO sequences can alter endogenous gene expression levels. Another is that the system will never completely silence expression, residual protein expression occurs even in the absence of ATc. Finally, the technique is a drug-on method meaning selection has to occur in the presence of ATc which can impact on parasites long term, potentially leading to the variability in expression control seen in populations and with long term culture.
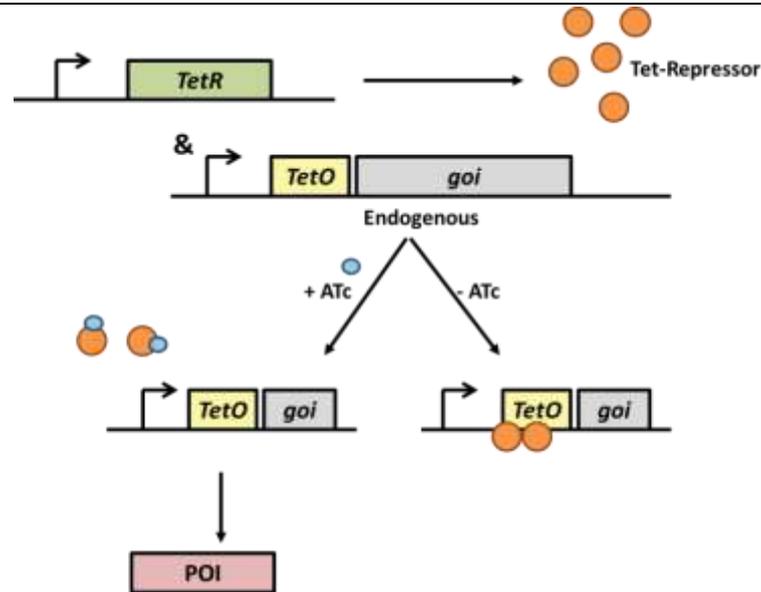
Figure 1.17 Expression regulation using an inducible tetracycline repressor

In a line which expresses the tet repressor (tetR) protein constitutively, the promoter of the gene of interest (*goi*) is altered to contain the tet operator (tetO) sequences upstream of the *goi*. In the presence of ATc the TetR protein is quenched and the protein of interest is expressed. All selection and cloning must take place in the presence of ATc to maintain this expression. In the absence of ATc, tetR binds to the TetO sequences and the endogenous promoter is silenced resulting in protein of Interest (POI) knockdown.

1.7.2.2 Tet-transactivator systems

This system is based on using a Tet-responsive minimal promoter upstream of a *goi* with a Tet-transactivator. In most eukaryotes the Tet-transactivator was generated by fusing TetR with the C-terminal domain of *Herpes simplex* protein HSV-VP16 known to be essential for transactivator activity. The Tet-responsive promoter was constructed by inserting TetO sequences upstream of a minimal promoter and replacing the endogenous promoter. In this circumstance the *goi* is expressed by the minimal promoter until the addition of ATc, whereby the *goi* is knocked down (Gossen & Bujard, 1992).

In *T. gondii,* direct adaptation of this extension of the Tet-system was not successful (Meissner *et al.*, 2001). To generate a functional system, a sequence from *T. gondii* that when fused with TetR has transactivator activity was identified. This is the transactivator TATi-1 (trans-activator trap identified) and is used instead of the non-functional HSV-VP16-based eukaryotic Tet-transactivator (Meissner *et al.*, 2002). To use this system, the endogenous promoter is replaced with a minimal promoter with TetO sequence upstream in a

line expressing TATi-1. The *goi* is then expressed by the minimal promoter. Upon addition of ATc, TATi-1 interacts with ATc preferentially and expression of the *goi* is reduced (Figure 1.18).
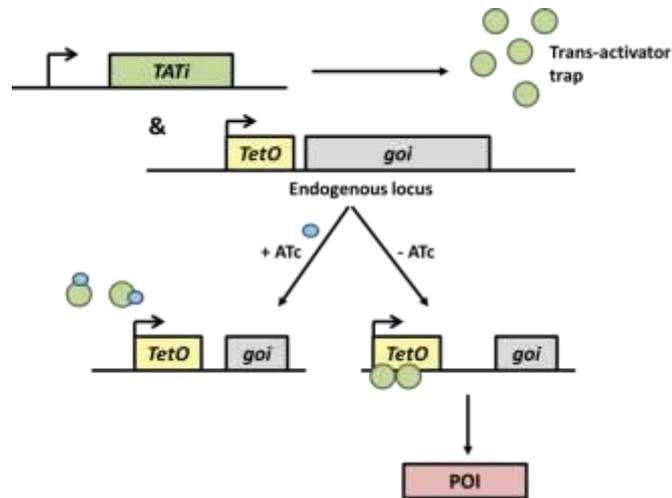
Figure 1.18 Expression regulation using an inducible tetracycline trans-activator

In a line which expresses the transactivator (TATi) protein constitutively the promoter of the gene of interest (*goi*) is replaced with a minimal promoter with tet operator (tetO) sequences. In the absence of ATc the TATi protein binds the minimal promoter resulting in *goi* expression. In the presence of ATc, TATi binding is reduced as a proportion binds to the ATc in place of the TetO sequences and gene expression is reduced. This results in knockdown of the protein of interest (POI).

Direct adaptation of this trans-activator system from *Toxoplasma gondii* to *Plasmodium spp.* was successful in regulation of episomal transgenes (Meissner *et al.*, 2005) but not for generating conditional knockdowns (Pino *et al.*, 2012). In this case a more targeted approach to identify transactivators was taken, focussing on ApiAP2 protein fragments to identify ones with transactivator activity. This screen identified four activating domains (ADs) that when fused with the Tet-repressor had transactivator activity, called TRADs. These *Plasmodium* specific transactivators are used in the same way as TATi for *T. gondii* (Figure 1.19). These *Plasmodium* specific TRADs recognise and are used in conjunction with the Tet-responsive promoter Tet07 (Pino *et al.*, 2012).
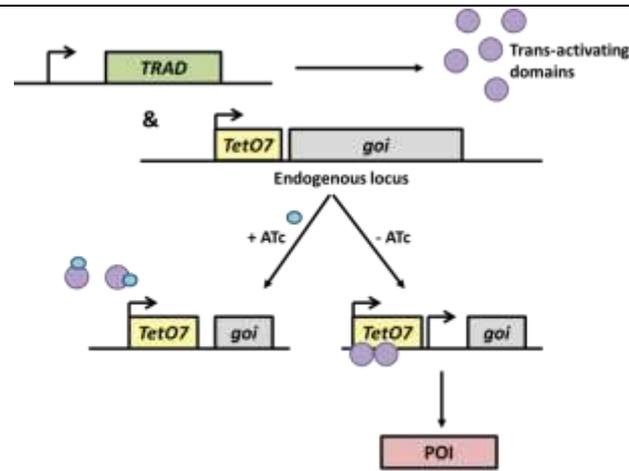
Figure 1.19 Expression regulation using an inducible tetracycline rans-activator

In a line which constitutively expresses one of the *Plasmodium* transactivator activating domain (TRAD) proteins the promoter of the gene of interest (*goi*) is replaced with a minimal promoter with the tet operator sequences recognised by this domain (TetO7). In the absence of ATc the TRAD protein binds the minimal promoter resulting in *goi* expression. In the presence of ATc TRAD binding is reduced as a proportion binds to the ATc in place of the TetO7 sequences and gene expression is reduced. This results in knockdown of the protein of interest (POI).

Although these systems have been extensively used for functional analysis of essential genes, especially in *T. gondii,* they have several limitations. Firstly, as a drug-off technique resulting in gene knockdown the effect of the loss of gene expression requires the natural degradation of the existing protein which can be a slow (many hours) process. Secondly, there will always be residual expression of the *goi* in the presence of ATc therefore conditional gene silencing is not complete with potential phenotypic consequences. Finally, because the endogenous promoter is replaced with a minimal promoter, the *goi*'s temporal expression and level of expression may differ from the WT even before the addition of ATc.

1.7.3 Post-transcriptional systems

One of the most widely utilised systems for the conditional control of protein level is small interfering RNA (siRNA). This system operates when small double-stranded RNA species are expressed complementary to a *goi*. The siRNA binds to a protein complex called Dicer which fragments the siRNA for loading into another complex called the RISC complex (RNA-Induced Silencing Complex). The loaded siRNA then directs the complex to the mRNA produced by the specific *goi* to be targeted and the mRNA is cleaved preventing protein translation (Sen & Blau, 2006). However, this technique does not work in any apicomplexan

parasites due to the divergence of RNAi related proteins like Dicer and Argonaute, key members of the RISC complex, necessary for the pathway (Baum *et al.*, 2009). Therefore, alternative systems have been modified for use in apicomplexa in an attempt to control protein expression post-transcription.

1.7.3.1 TetR-aptamer translation blocking

An adaptation of the TetR system has been established to use the TetR as an RNA-binding protein that will, in an inducible way, inhibit protein translation (Belmont & Niles, 2010; Goldfless *et al.*, 2014). Aptamer sequences integrated immediately upstream of the *goi* are transcribed with the mRNA. The aptamer sequences bind with high affinity to the TetR and blocks ribosome binding and therefore translation (Figure 1.20). To make this system compatible with *P. falciparum* a single integration method was required. In this system the 5' and 3' UTRs of the *goi* were used as homology arms for recombination. The *goi* was replaced with a copy containing the upstream aptamer sequence and a suitable, alternative 3'UTR. Downstream of this an appropriate promoter was used to drive the selectable marker and the *tetR* expression cassette separated by 2a-like peptides to yield individual proteins. This integrated the required aptamer and the *tetR* expression cassettes in a single step (Goldfless *et al.*, 2014). In the absence of ATc the expressed TetR bound the aptamer on the *goi* mRNA preventing ribosome binding and translation. The presence of ATc inhibited binding of TetR to the aptamer and therefore ribosome binding occurs, thereby allowing translation (Figure 1.20).
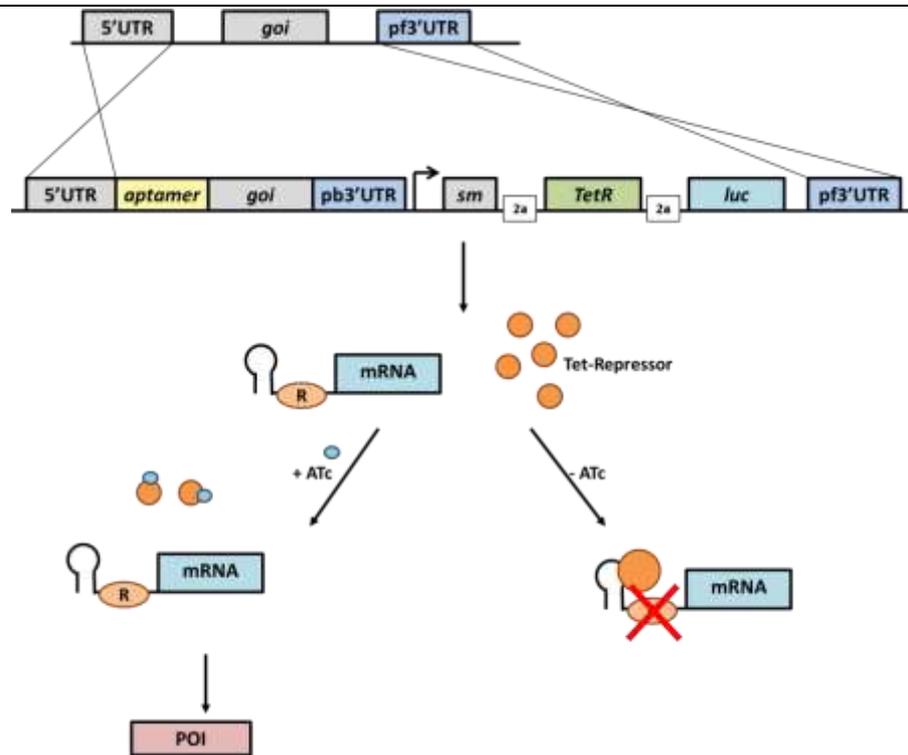
Figure 1.20 Translation blocking by TetR aptamer binding

In *P. falciparum* the *goi* is replaced with a version containing a TetR aptamer sequence immediately upstream. In this same integration a selectable marker, the TetR expression sequence and a luciferase cassette are integrated, separated by 2a-like peptides and driven by a constitutive promoter. The mRNA transcribed will always contain the TetR aptamer and the TetR will be expressed. In the presence of ATc the TetR cannot bind the TetR aptamer and therefore a ribosome can bind the mRNA and translation can occur. In the absence of ATc the TetR binds the aptamer sequence and blocks ribosome binding. This prevents translation of the protein of interest (POI).

This system has several benefits over other systems. The single integration for the TetR expression sequence and insertion of the TetR-aptamer alleviates the need for TetR expressing parental lines. The integration of the TetR-aptamer upstream of the *goi* limits the effect of insertion of additional sequences on the endogenous promoters' activity, though it has been shown to have a potential effect of endogenous gene expression (Goldfless et al., 2014). However, like many other systems this is a potentially leaky knockdown system where the POI may be reduced in expression even in the presence of ATc and expression can never be completely abolished even in the absence of ATc.

1.7.3.2 U1 snRNP gene silencing

Though its primary function is within the spliceosome, an additional and unusual function of the U1 snRNP in eukaryotic systems is the regulation of polyadenylation (Matera & Wang, 2014). Previous work showed that the recruitment of the U1 snRNP to the 3' end of mRNA prevents polyadenylation.

This system is hypothesised to have developed as a protection mechanism to prevent the premature polyadenylation of immature mRNAs. It has however been manipulated to specifically block polyadenylation of transcripts resulting in their subsequent degradation (Matera & Wang, 2014). In the adaptation of this system to T. gondii multiple U1 recognition sequences were positioned directly downstream of a reporter (2 tandem repeats) or a *goi* (4 tandem repeats). To make the system conditional it was combined with the dimerisable Cre system (DiCre), described fully in section 1.7.4.3.

The U1 recognition sequences are inserted immediately downstream of the *goi* but are separated by a cassette containing either a 3'UTR and selectable marker (in a line expressing DiCre) or a 3'UTR, selectable marker and the cassettes expressing DiCre. The U1 recognition sequences are dormant in the absence of rapamycin. Upon addition of rapamycin the DNA between the *goi* and the U1 recognition sequences is removed by the recombinase resulting in the U1 recognition sequences being adjacent to the *goi*. This prevents the polyadenylation of the mRNA and results in its degradation (Figure 1.21). This method has some advantages; significantly it is possible to knockdown a *goi* with a single manipulation of a haploid genome when integrating the DiCre cassette along with the U1 recognition sequences. However, the method does have some limitations. Firstly, attempts to adapt this method to P. falciparum have been unsuccessful even though the U1 recognition sequences are highly conserved between the species (Pieperhoff et al., 2015). Secondly, it appears that even before addition of rapamycin and the repositioning of the U1 recognition sequences some knockdown of the protein of interest (POI) occurs and this may be problematic for recovery of knocked down *goi*. Finally, while the single step integration is of great benefit, the expression levels of DiCre are variable because of locus variability therefore the excision efficiency will also be variable and as will the knockdown levels (Pieperhoff et al., 2015).
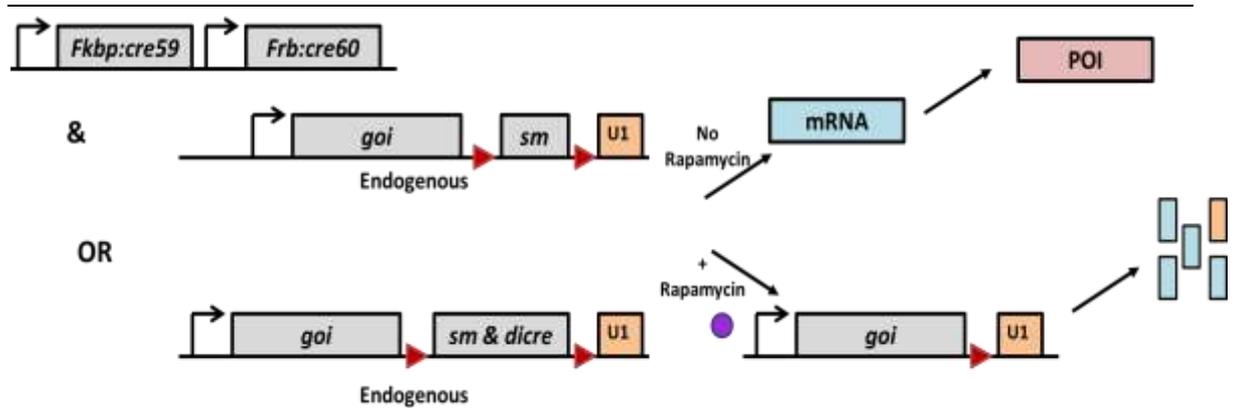
Figure 1.21 Expression regulation using the snRNP, U1

In a line that expresses the two DiCre moieties Cre59 and Cre60 fused to the rapamycin binding proteins Fkbp and FRB respectively, the 3'UTR of the gene of interest (*goi*) is interrupted with a LoxP flanked (floxed) selectable marker cassette and multiple U1 recognition sequences. Protein expression is maintained provided 3'UTR disruption of this gene is not detrimental to the parasite. Upon the addition of rapamycin the selectable marker is excised and the U1 recognition sequences are moved immediately downstream of the *goi*. These sequences prevent polyadenylation and target the mRNA to the exosome resulting in degradation and in a knockdown of protein translation. This in turn results in a knockdown of the protein of interest (POI). An addition to this system makes it a single step conditional alteration whereby Cre50:fkbp and Cre60:frb are expressed with the selectable marker downstream of the *goi* within the floxed cassette. This allows for a single integration to insert the conditional recombinase and the U1 sequences.

1.7.4 Conditional site-specific recombinase systems

With all of the previously discussed conditional regulation systems there is always one fundamental limitation. All of these systems are knockdown systems not complete knockouts. There is no way, currently, to completely abolish protein expression with degradation of the protein or the mRNA. To this end site-specific recombinase systems were explored to completely silence expression in a conditional way.

Site specific recombination (SSR) occurs when a recombinase recognises and binds specific segments of DNA with sequence homology. The recombinase protein brings together the homologous segments of DNA and, depending on their position and orientation in the genome, can insert a sequence, excise the encased DNA, invert it or translocate it (Kolb, 2002). Two systems that have been adapted for multiple organisms are the flp/FRT system and the Cre/LoxP system.

The flp/FRT system uses the flp recombinase which recognises the short flippase recognition target (FRT) sequence (GAAGTTCCTATTC*tctagaaa*G*t*ATAGGAACTTC).

Insertions of DNA can be achieved by positioning an FRT sequence in the desired loci and by providing a vector with the insertion flanked by FRT sequences. Upon flp activity, the locus takes up the FRT flanked insert leaving the vector with its backbone and a single FRT sequence (**Figure 1.22** A). Excision of endogenous DNA is achieved by flanking the region of interest with FRT sites in the same orientation. Upon flp activity, the encased DNA region is removed and only one FRT remains (**Figure 1.22** B). An inversion of the DNA occurs when the FRT sites are in reversed and complementary orientations (head-to-head) surrounding a region of interest. Upon flp activity, the FRT sites recombine and invert the encased DNA region and both sites remain (**Figure 1.22** C). A translocation occurs when two nearby regions of DNA both contain one FRT site. The flp activity then brings the sites together and translocates the regions onto the same loci (Figure 1.22 D).
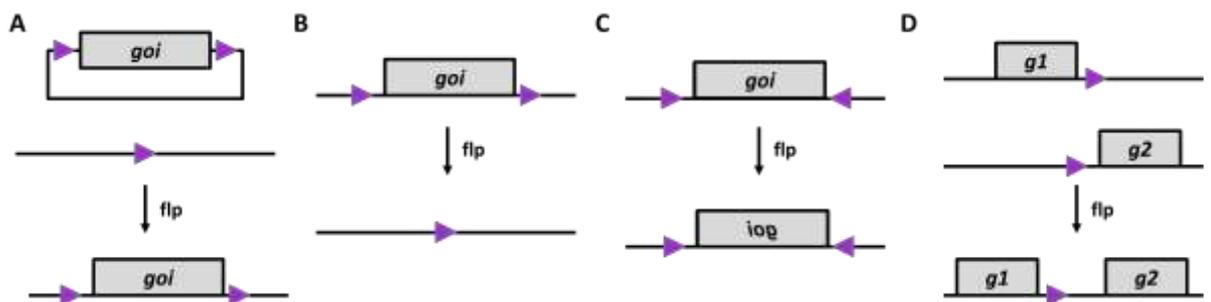
Figure 1.22 General uses for the FRT/flp recombinase system.

The flp recombinase recognises the FRT sequence (purple arrowheads) within the genome and can be used to insert, excise, invert or translocate regions within the genome.

A) Insertions occur where one FRT site is within the genome and a vector is provided with a DNA sequence that is flanked by FRT sites. Upon flp activity, the encased DNA is integrated between the FRT sites at the original FRT location.

B) To excise a gene of interest (*goi)* or a region of DNA it is flanked by FRT sites. Upon flp activity the region between the FRT sites is lost due to recombination of the two sites.

C) Inversion of enclosed DNA can be achieved when two FRT sites, arranged flanking a region of DNA or a gene in a head-to-head orientation, invert in the presence of flp.

D) Translocations occur when FRT sites on nearby loci recombine in the presence of flp translocating entire regions of the genome.

One of the disadvantages of the flp/FRT system was its initial incompatibility in mammalian systems as it denatures at elevated temperatures. This issue has been solved with the identification of a mutated version of the FLP recombinase, FLPe (named as it was the eighth generation mutation) which was stable at the higher temperatures in mammalian systems (Buchholz *et al.*, 1998).

Another recombinase system that needed no adaptation for use in mammalian systems was the Cre/LoxP system. This system operates in the same way as the flp/FRT system. The Cre recombinase recognises the palindromic LoxP sites in the genome acting the same ways as described above for flp/FRT (Figure 1.22).

Additional adaptations to the Cre recombinase system have allowed it to be used conditionally and to fix inversions within the genome. The Cre recombinase has been successfully split into two fragments that alone are inactive, but when they reconstitute regain their recombinase activity (Jullien *et al.*, 2003). These two fragments Cre59 and Cre60, named due to the position of the split at amino acid position 59/60, can be fused to the rapamycin binding protein domains FRB and FKBP whose reconstitution is controlled by the addition of rapamycin (Jullien *et al.*, 2003). The LoxP recognition sequences have also been adapted to contain mutations that allow an inversion to be fixed in the genome. This occurs when two compatible LoxP sites recombine resulting in an inversion and leave behind two incompatible LoxP sites flanking the genome region (Albert *et al.*, 1995).

1.7.4.1 Adaptation of FLP recombinase systems for use in apicomplexan parasites

The first apicomplexan the flp/FRT system was adapted for was *P. berghei*. SSR in this organism has been focussed on genome editing by excision of regions between two FRT sites upon expression of the flp recombinase. Three methods of excision have been developed and shown efficacy. Firstly, the entire *goi* can be flanked by FRT sites and upon recognition and activity of the flp recombinase is entirely excised (Figure 1.23 A). Secondly, an intron of the *goi* can be flanked with the FRT sites. Upon recombinase activity, the intron will be excised leaving behind a truncated gene and potentially resulting in no protein expression or the expression of a truncated and non-functional protein (Figure 1.23 B). Finally the 3'UTR of a *goi* can be flanked by FRT sites, in many cases the loss of the 3'UTR after excision by the recombinase results in loss of protein expression (Figure 1.23 C).
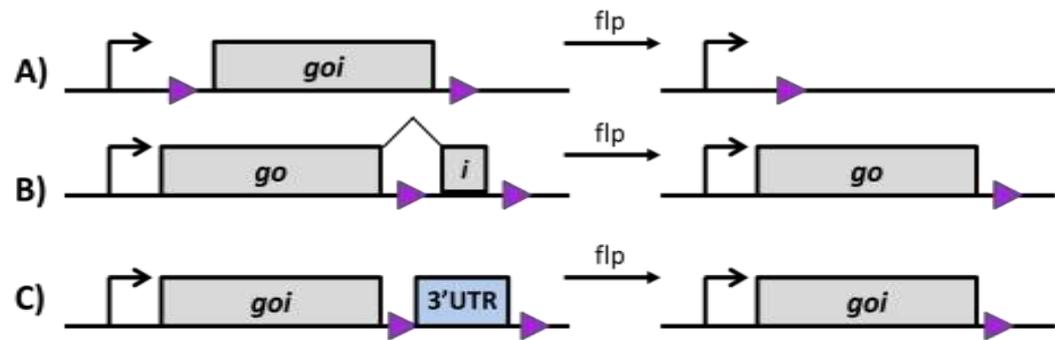
Figure 1.23 Different gene regulation models using the flp/FRT recombinase

A) The gene of interest (*goi*) can be flanked by FRT sequences and the entire gene excised upon flp recombinase expression.

B) An intron of the *goi* can be flanked by FRT sites and this portion of the gene excised upon flp recombinase expression.

C) The 3'UTR of the *goi* can be flanked by FRT sites and removed upon flp expression.

For simplicity, all methods of using the flp/FRT system will be described only with a gene flanked by FRT sites, though in theory it is possible for all described approaches to be used with any of the gene modifying methods described above (Figure 1.23).

The system can be used to investigate protein function after sexual development by exploiting genetic crosses (Figure 1.24). In this method, 2 parental lines, a parental line expressing the flp recombinase with a fluorescent protein and a parental line expressing the FRT flanked *goi* with a different fluorescent protein, are crossed in mosquitoes. After the cross, the progeny are a mix of homozygous (either parent 1 expressing the recombinase plus fluorescent protein, or parent 2 expressing the mutated *goi* with a single colour fluorescence) and heterozygous (both non-fluorescent and negative for the recombinase and the mutated *goi*, or expressing both the recombinase and the mutated *goi* and both fluorescent proteins). Only the progeny containing both the recombinase and the mutated *goi* will have the differentially expressed *goi* in the mosquito stages and sporozoites, when the cells are haploid again, and will also be bi-fluorescent (Figure 1.24 B). Genes can also be targeted for depletion by flanking an intron or the 3'UTR with FRT sites (Figure 1.23).
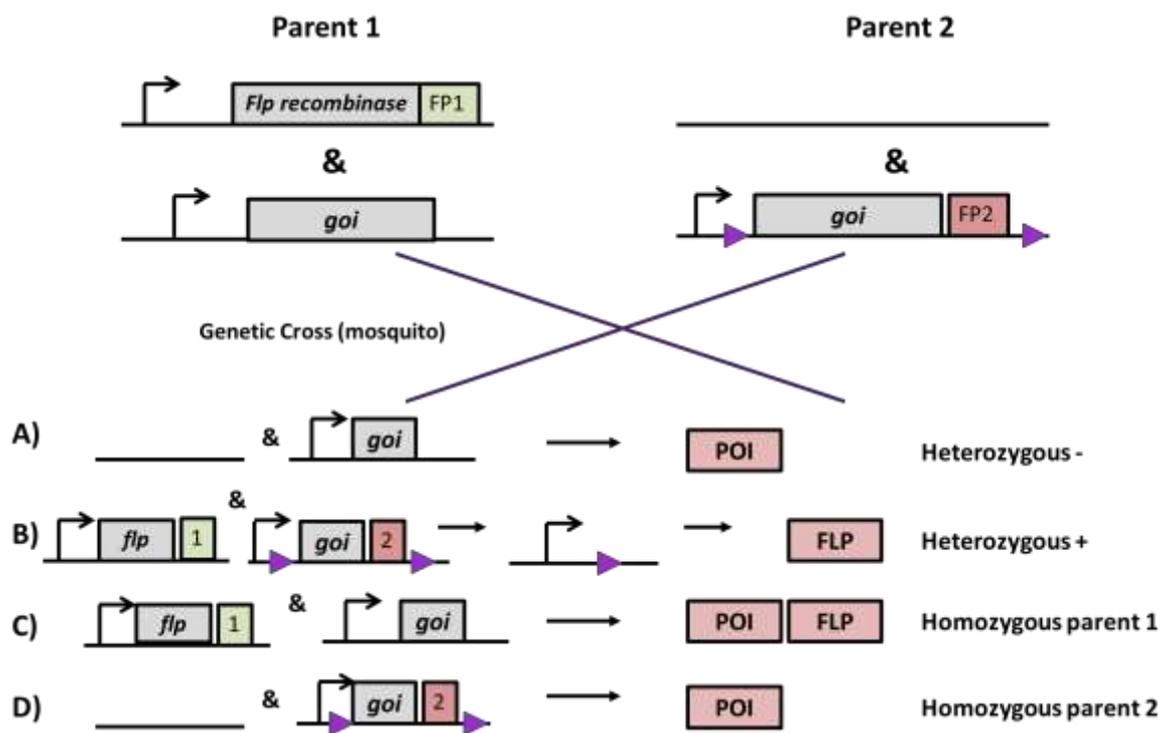
Figure 1.24 Conditional gene expression using genetic crosses and the flp/FRT recombinase system.

As an example of the genetic cross system, the gene of interest (*goi*) has been flanked by FRT sites. One parental line in the cross expresses the recombinase constitutively; the other an FRT flanked *goi*. After sexual reproduction in the mosquito the progeny can be either homozygous replicating one of the parents (C or D) or heterozygous, expressing a wild type *goi* and no recombinase (A) or expressing the recombinase and the FRT flanked *goi* (B). The combination of the FRT flanked *goi* and the constitutive recombinase expression will result in excision of the *goi*. The protein of interest (POI) is only not present in the heterozygous progeny (B).

This system allows effective dissection of gene function in the ookinete, oocyst and sporozoite stages. However it offers no further analysis of function for blood stage essential genes immediately after the asexual cycle (Carvalho *et al.*, 2004). Another issue is that crosses can be variable in their efficiency at producing the different progeny and only a percentage of the offspring will be lacking the POI.

Another way to use the flp/FRT system requires the *goi* to be targeted in one of the three ways described above (Figure 1.23) in a parasite line expressing FLP driven by a stage specific promoter (Figure 1.6.13).

Until the stage specific promoter driving the FLP expression becomes active the *goi* is correctly expressed from its native promoter. However, upon stage specific expression of FLP, the gene is excised and the protein is no longer translated (Lacroix *et al.*, 2011). The limitation with this technology is that each

stage investigated requires a different parental line expressing FLP at the specific stage of interest, which can be labour and animal intensive to generate. Furthermore, the identification of promoters which can drive stage specific expression to a high enough level for activity and unique stage specificity is limited.
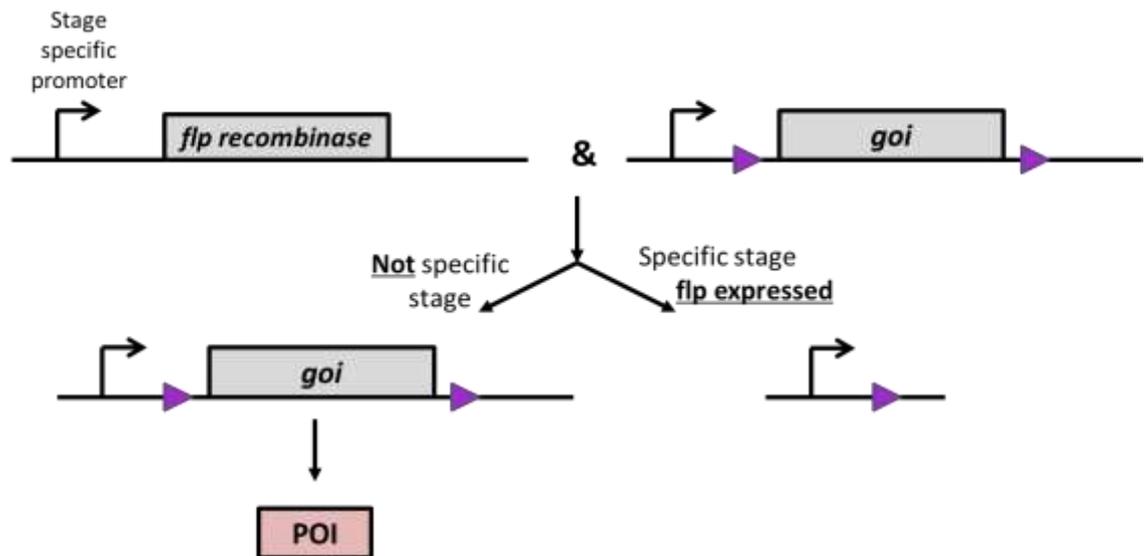
Figure 1.25 Conditional gene expression using stage specific flp expression

A parental line expressing the flp recombinase under a stage specific promoter is transfected with a gene of interest (*goi*) flanked by FRT sites. Until the promoter driving flp expression becomes active the recombinase is not expressed and therefore the protein of interest (POI) is expressed. Upon expression of the recombinase the entire *goi* is excised and expression is depleted allowing analysis of the knockout. Genes can also be targeted for depletion by flanking an intron or the 3'UTR instead of the entire gene (Figure 1.23).

This system has also been adapted and shown to be functional in *P. falciparum* whereby it utilises conditional FLP expression from an episome (Figure 1.26). Parasite lines containing an FRT flanked copy of the *goi* are transfected to uptake an episome expressing FLP from an ATc-regulatable promoter (Meissner *et al.*, 2005). When cultured in the presence of ATc, FLP expression is repressed and the *goi* remains intact. However, when ATc is removed, FLP is rapidly expressed and the *goi* is excised (O'Neill *et al.*, 2011). While this technique is attractive for adaptation, tetracycline inducible systems display some fundamental limitations. Firstly, there is some leaky activity of expression under the ATc repressed promoter resulting in gene excision and therefore gene knockout to occur in some parasites before the removal of ATc. Secondly, protein loss after gene excision has been shown in some cases to take multiple cycles.
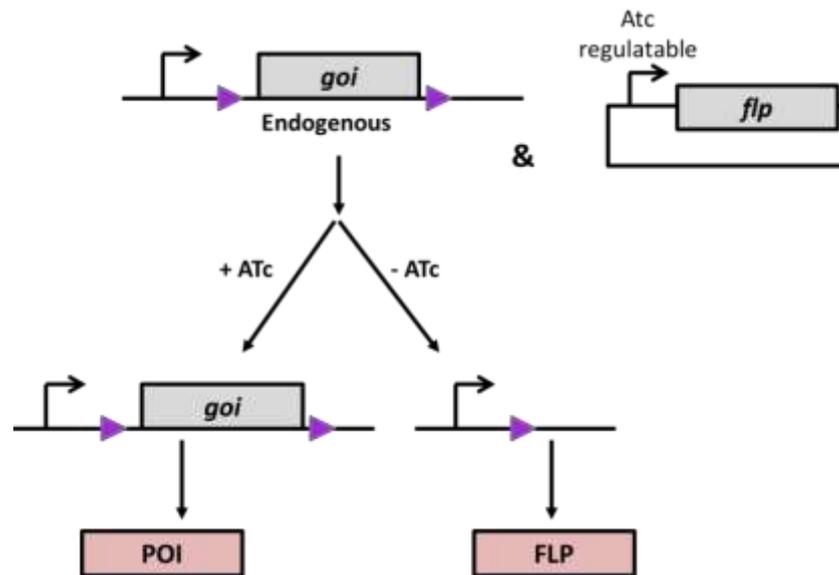
Figure 1.26 Conditional genome editing with the flp recombinase from an episome

The gene of interest (*goi*) is flanked by FRT sites which do not affect expression. A second transfection introduces the flp recombinase under an ATc regulatable promoter. In the presence of ATc the FRT sites remain intact and gene expression is maintained. In the absence of ATc the recombinase is expressed and the FRT sites recombine resulting in *goi* excision and loss of expression of the protein of interest (POI).

1.7.4.2 Cre recombinase

Another SSR method which has been utilised in mammalian cells, *T. gondii* and *P. falciparum* utilises the Cre recombinase. This recombinase recognises LoxP sites made up of two 13pb palindromic sequences at both ends and an 8bp segment in the middle, giving directionality. LoxP sites flanking (floxing) a segment of DNA in the same orientation will result in excision while sites in opposing orientations will result in inversions. Constitutive expression of the recombinase will not allow temporal control of gene expression but multiple methods have been adapted to provide this control, utilising conditional CRE expression or constitutive expression of inactive fragments of the recombinase that can be conditionally heterodimerised restoring activity (DiCre, see section 1.7.4.3).

In *P. falciparum* temporal control of CRE expression has been shown in the same way as the FLP recombinase (Figure 1.26). After flanking the *goi* with LoxP sites, an episome expressing the Cre recombinase under control of an ATc regulatable promoter was transfected and maintained by the parasite. In the presence of ATc the recombinase was supressed and the *goi* was expressed. Upon expression

of Cre (by removal of ATc) the *goi* was excised and protein expression ceases (Figure 1.27).

The limitations of this system mirror those of the FLP recombinase described in 1.7.4.1 (O'Neill *et al.*, 2011).
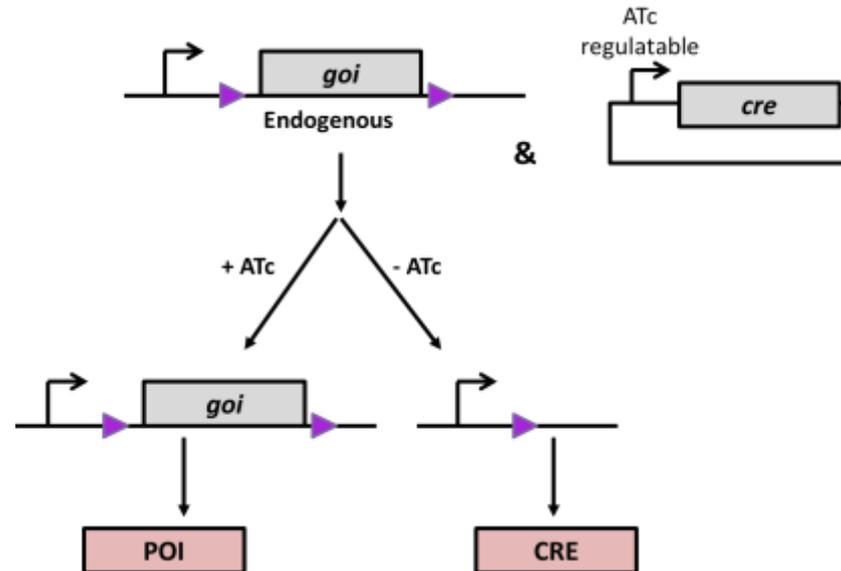
Figure 1.27 Conditional genome editing with the cre recombinase from an episome

The gene of interest (*goi)* is flanked by LoxP sites (floxed) which does not affect expression. A second transfection introduces the Cre recombinase under an ATc regulatable promoter. In the presence of ATc the LoxP sites remain intact and gene expression is maintained. In the absence of ATc the CRE recombinase is expressed and the LoxP sites recombine resulting in *goi* excision and the protein of interest (POI) is no longer expression.

1.7.4.3 DiCre recombinase

To make the Cre recombinase system conditional, adaptations were made to utilise rapamycin binding proteins to bring together two inactive fragments of the recombinase restoring its function (Jullien *et al.*, 2003). The system, successfully adapted to *T. gondii* and *P. falciparum* (Andenmatten *et al.*, 2013; Collins *et al.*, 2013), utilised the FKBP12 – FRB - Rapamycin complex and the split of Cre into two inactive fragments, dimerisable Cre (DiCre), between amino acids 59 and 60. The two moieties, Cre59 and Cre60, are fused to the rapamycin binding protein domains FKBP12 and FRB respectively. Upon addition of rapamycin the Cre recombinase heterodimerises and its activity is restored (Andenmatten *et al.*, 2013; Jullien *et al.*, 2003).

In *T. gondii,* temporal control of gene expression can be achieved when the endogenous *goi* is replaced with its *LoxP* flanked cDNA, followed by a fluorescent protein (YFP) and a selectable marker (Figure 1.28) in a parental line expressing the DiCre fragments. In the absence of rapamycin, parasites are able to express the endogenous *goi* from its cDNA. Recombinase-mediated excision is achieved after the addition of rapamycin and results in the loss of the gene and expression of YFP in its place (Andenmatten *et al.*, 2013). The conditional deletion of many *goi's* can be achieved from a single parental line expressing DiCre. This system has several benefits, arising from the fact that the gene is normally expressed from its own promoter before addition of rapamycin and the fact that gene excision is very efficient and rapid. Another benefit of this system is that knockouts are readily distinguishable from parasites still expressing the protein due to their expression of the fluorescent protein. An advantage of this method is that even if a knockout line is not cloneable (indicating the *goi* is essential for a proliferative process) the phenotype of a knockout can be studied in more depth as the parasites are readily distinguishable (Andenmatten *et al.*, 2013).
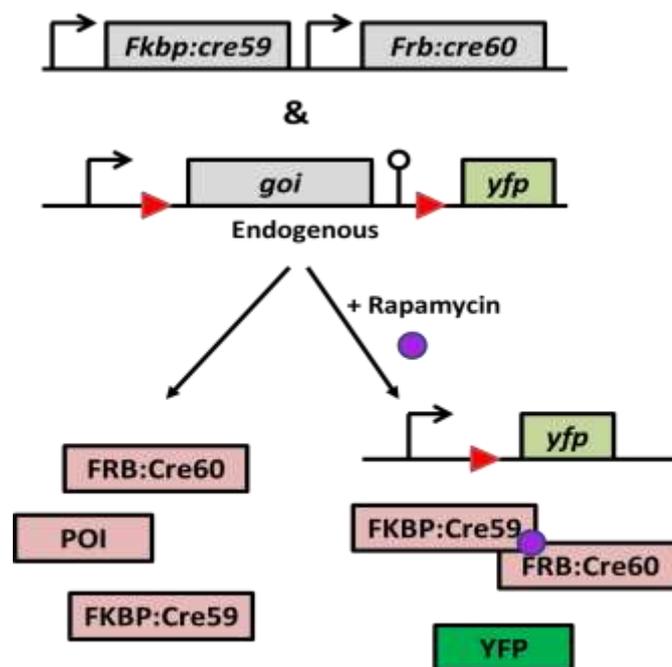
Figure 1.28 Conditional gene deletion in *T. gondii* using DiCre

Parental lines expressing the two DiCre fragments Cre59 and Cre60 fused to the rapamycin binding domains fkbp and frb respectively are subsequently transfected to replace the endogenous gene of interest (*goi*) with a LoxP site flanked (floxed) version with *yfp* immediately downstream. In the absence of rapamycin the DiCre fragments are expressed but inactive and the protein of interest is expressed as normal. In the presence of rapamycin the DiCre fragments heterodimerise and recombinase activity is restored. The recombinase recognises the LoxP sites and excises the *goi* resulting in a loss of protein of interest (POI) expression. Under the endogenous promoter YFP is expressed thus making the gene knockout parasites identifiable by fluorescence.

The DiCre system has been utilised in *P. falciparum* in a multitude of ways. Firstly, both DiCre proteins bound to their respective rapamycin binding domains with nuclear localisation signals (NLS) are constitutively expressed from an episome which is transfected into a line containing a floxed *goi*. After floxing, the gene is still under the control of the endogenous promoter therefore normal function should be maintained. The second transfection provides the episome expressing the NLS:FKBP:Cre59 and NLS:FRB:Cre60 proteins. Addition of rapamycin results in dimerisation of these proteins, formation of the active recombinase and excision of the floxed gene (Figure 1.29). This method has been successfully used to conditionally knockdown proteins within one erythrocytic cycle, however its utility at other life stages has not been investigated and may be limited by episome maintenance (Yap et al., 2014).
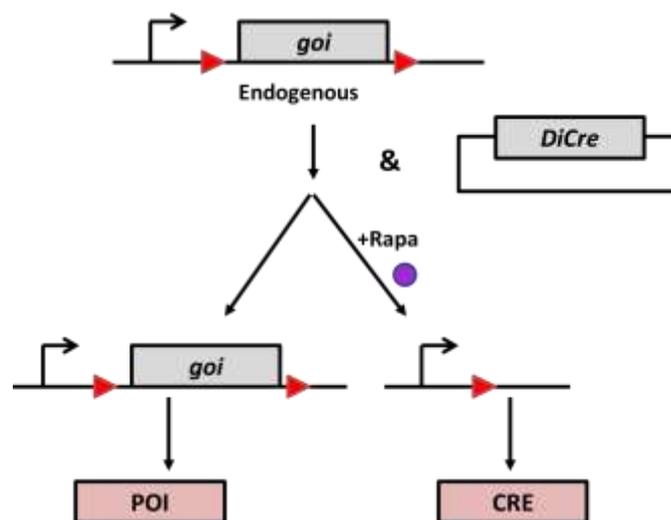
Figure 1.29 Conditional genome editing with the DiCre recombinase expressed from an episome

The gene of interest (*goi)* is flanked by LoxP sites (floxed) which does not affect expression. A second transfection introduces a single episome expressing the two DiCre recombinase fragments under constitutive promoters. In the absence of rapamycin the fragments are inactive, the LoxP sites remain intact and gene expression is maintained. In the presence of rapamycin the recombinase heterodimerises restoring activity and the LoxP sites recombine resulting in *goi* excision and loss of protein of interest (POI) expression.

 A final method has been used in *P. falciparum* that stably integrates the DiCre expression cassette in a single step while floxing the 3'UTR of the *goi* to be excised along with the selectable marker (Figure 1.30). This provides two benefits, the first being floxing of the genome region to target for knockdown or knockout and the DiCre expression system are provided in a single transfection, the second being the excision of the selectable marker along with the *goi* enabling further manipulations to occur when resistance markers are scarce (Collins *et al.*, 2013).

In this method the 3'UTR is targeted as it can theoretically be replaced and floxed without altering gene function. Before addition of rapamycin, the *goi* is expressed normally with the replacement 3'UTR contained within the LoxP sites. Upon rapamycin addition the 3'UTR is removed along with the selectable marker. In this case the excision occurred rapidly, however gene expression was not silenced due to the presence of an alternative polyadenylation site allowing transcription to continue in the absence of a 3'UTR (Collins et al., 2013).
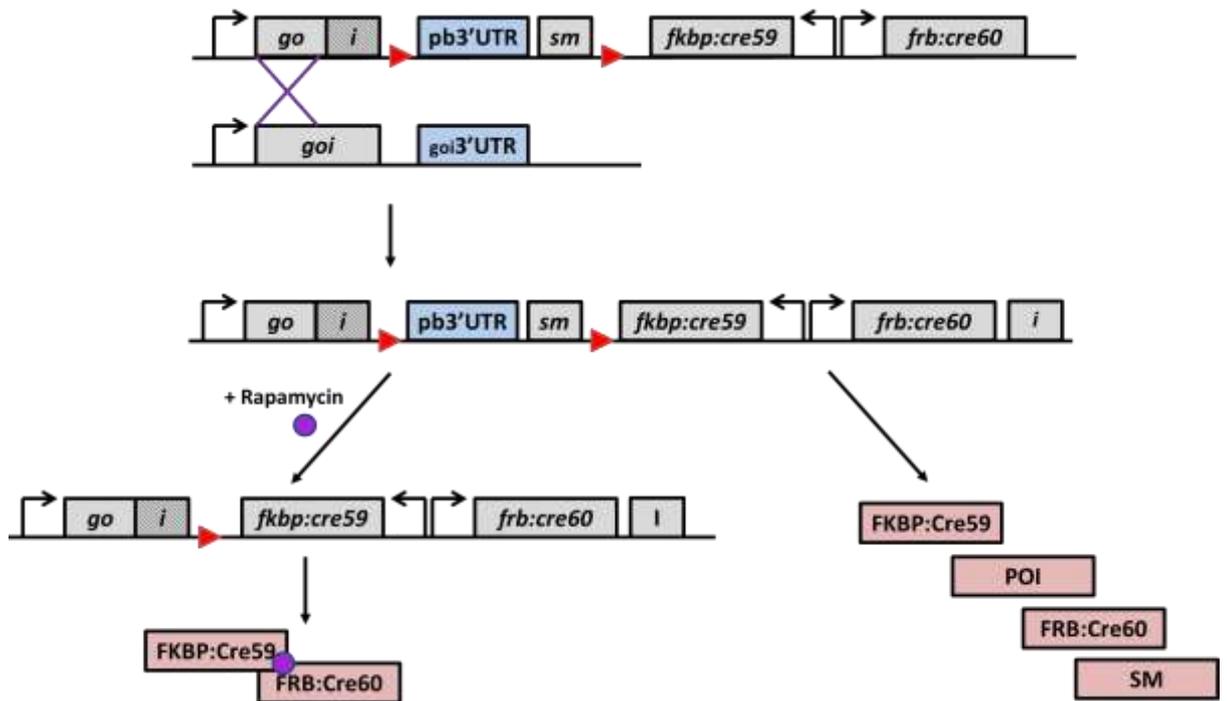
Figure 1.30 Conditional genome editing with the DiCre recombinase integrated at the same time as floxing the *goi*

The gene of interest (*goi)* is targeted by disruption and loss of its 3'UTR which is often sufficient for dysregulation. In a single transfection the 3'UTR is replaced with a standard 3'UTR and selectable marker (SM) flanked by LoxP sites (floxed). Downstream of this are the two DiCre cassettes whose expression is driven by constitutive promoters. Pre-rapamycin the *goi* is expressed as normal from its endogenous promoter and the two DiCre fragments are expressed but inactive. Upon addition of rapamycin the two DiCre fragments heterodimerise and regain recombinase activity excising the 3'UTR and SM. This UTR disruption can disrupt gene regulation and expression. This disruption of the 3'UTR is often sufficient to reduce or ablate expression of the protein of interest (POI) or to result in non-functional POI expression.

In light of the limitations and advantages of all these methods, shown to be effective in apicomplexan, we elected to use an approach similar to that used in *T. gondii.* We will generate a plethora of parental DiCre lines to allow phenotyping of multiple stages throughout the life cycle and, if the system works as efficiently as demonstrated in *P. falciparum,* effectively excise genes of interest within a cycle.

1.8 **Transcriptomics**

1.8.1 Introduction & background

In this work we want to utilise an untargeted method to identify the differences in asexual proliferative parasites and sexually differentiated parasites. This approach will allow us to understand not only how they differ when they are mature but how the process begins and the key factors involved at each stage of commitment. Genome wide sequencing gave the first insight into the molecular mechanisms beginning the cascade of events that result in sexual differentiation. Here we wanted to look at the overall changes in cells that bring about the massive differences in parasite morphology and lifestyle.

1.8.2 Why use transcriptome & RNAseq

Several "Omic" technologies exist for the collective identification of genes (genomics), mRNA transcripts (transcriptomics), proteins (proteomics) and metabolites (metabolomics) each able to answer different questions. Ideally untargeted experiments would utilise a multitude of unbiasedly look at the differences in states. However, as the processes being investigated in this work are dynamic and we wanted to examine them over an extended time a single "omic" was selected for the analysis.

The foundation for this work was to use the recently acquired knowledge about *ap2-g* as the trigger that begins the cascade of sexual commitment to control the onset of gametogenesis and compare populations with the same genetic background that have no gametocyte commitment and those that have increased gametocyte commitment. We hypothesised that changes would be both rapid and dynamic in the processes that led to differentiation and therefore wanted a technology with high sensitivity, good depth coverage and good dynamic range for differential expression between samples.

While both "shotgun" proteomics and RNAseq will allow overall abundance analysis for peptides and transcripts respectively their sensitivity differs. To detect peptides with confidence their expression must be higher than that of transcripts. This is owing to the increased sensitivity of RNAseq. As we want to identify the novel changes in gene expression as soon as they happen within the

differentiating cells RNAseq gives us the more sensitive detection (Conesa *et al.*, 2016; Hebenstreit *et al.*, 2011).

Advances in RNAseq and mRNA enrichment techniques has made it a technology that can sample the entire transcriptome without the need for amplification of material, can quantify accurately across a large range of expressions with little background and can detect transcript abundances with high sensitivity. The selection of only mRNA (which makes up only 2% of all RNA in the cell) using polyA selection has allowed an increase in sensitivity and a reduction in background noise (Conesa *et al.*, 2016; Mortazavi *et al.*, 2008).

While the additional sensitivity of RNAseq is considered a benefit in many ways the non-correlative relationship between transcript detection and peptide detection can be considered the fault of over-sensitivity in transcript detection. Whole transcriptome experiments have identified that in metazoan cells there are two distinct populations of mRNA transcripts (Hebenstreit *et al.*, 2011). These are considered the high expressers (HE) and the low expressers (LE). If sensitivity is lost the many LE transcripts drop below the threshold of detection and the HE are the only transcripts retained in an analysis. This gives a bias to the detection and quantification of highly expressed genes generating more transcripts than lowly expressed genes with fewer transcripts. However, there is some evidence to suggest that a higher proportion of the LE genes constitute genes are transcribed but not actively translated into protein. This in turn led to the hypothesis that only the HE genes constitute the active transcriptome, that is the portion that goes on to generate proteins involved in cellular processes. GO-term analysis revealed that many HE genes code for housekeeping genes essential for normal cell processes. Of particular interest in this work was the identification of many of the LE genes (the genes that could potentially be missed using technologies such as proteomics and microarray) had GO terms associated with cell type differentiation. While it appears analysing transcripts in the LE class might lead to the identification of more transcripts not making functional proteins it might also allow the detection of low abundance transcripts and proteins key in the process we are looking at (Hebenstreit *et al.*, 2011).

1.9 **Aims & objectives**

This work aims to develop and use a novel inducible technology to control the onset of gametocyte commitment and take an untargeted approach to examine the early stages of commitment and gametocyte development with the aim of elucidating novel factors in the processes of commitment and gender determination. It further aims to validate the identification of these novel factors by examining several candidates for their role(s) in gametocytes.

Chapter 2   Materials and Methods

2.1 **Materials**

For this study a variety of equipment and reagents were used. For ease of description compositions, sources and descriptions are listed below based on their use.

2.1.1 Equipment

All equipment used in this work is listed below, in Table 2.1.

Amaxa	Amaxa Nucleofector® transfection machine
Applied Biosystems	StepOne qPCR , 48-well qPCR plates, adhesive qPCR lids (48well)
BD Biosciences	Syringes, Needles (23 gauge), FACS tubes, Insulin syringes, LSRII flow cytometer, aria II sorter, AriaIII sorter.
Beckman	Allegra X-22 centrifuge, CyanADB flow cytometer
BioRad	Agarose gel electrophoreses equipment, UV Transilluminator, Gel documentation imaging system, Gene Pulser Xcell, Micropulser, SDS-PAGE system, Blotting apparatus (Transblot SD), mini protrean tetra system, mini protrean gradient gels,
Carestream	X-ray film
Corning	tissue culture flasks
Eppendorf	5424 centrifuge, 5417R centrifuge PCR thermocycler (Mastercycler Epigradient), Thermomixer compact
Europroxima	Plasmodipur filters
Fisher Scientific	Ultrasound water bath FB15047
GE	Nitrocellulose
Grant	Water bath
Genlab	Incubator
IKA	vortex genius 3
Illumina	Illumina HiSeq2500 system
Kuhner	Shaker X Incubator
Leica	Leica M205 FA Fluorescence Stereomicroscope
MACS	Magnetic columns
Millipore	MilliQ water deionising facility, millipore filter units, stericup
Paxcam	Paxcam 5 camera
Sartorius	Analytical balances
StarLab	ErgoOne® Pipettes, StarPet Pro Pipette Controller

Stuart	Heat block US152, Roller mixer SRT6D, Orbital Shaker
Thermo Scientific	Nanodrop spectrophotometer, slides
Zeiss	Axioskop 2 (mot plus) fluorescence microscope with Axiocam MRm CCD camera, Primo Vert (light microscope), PrimoStar light microscope

Table 2.1 Equipment utilised in this work

2.1.2 Computer Software

Computer software used in this work is listed in Table 2.2.

Adobe Systems Inc.	Photoshop CS4, Illustrator CS4 and Acrobat Reader DC
Babraham	TrimGalore!
CCB Johns Hopkins	HiSat2
CLC Bio	CLC Genomics Workbench 6
Microsoft Corporation	Windows 7, Microsoft Office 2007, 2010
National Center for Biotechnology Information (NCBI)	Basic Local Alignment Search Tool (BLAST)
National Institute of Allergy and Infectious Diseases (*NIAID*)	PlasmoDB, ToxoDB and EuPathDB
OligoCalc	Oligo Analysis tool http://www.basic.northwestern.edu/biotools/oligocalc.html
PuTTY	Windows PuTTY (release 0.62)
Cran	RStudio64 (version 3.1.2), ggplot2, factomineR
Thomson Scientific	Endnote X6
Trapnell lab	Cuffquant package

Table 2.2 Computer software used for analysis in this work

2.1.3 Biological and chemical reagents

Basic reagents used in this study and for the generation of solutions are listed in Table 2.3.

Alpha laboratories	Heparin
Ambion	Trizol$^®$ reagent
Biolabs	DNA ladder
Fermentas	Protein ladder
Fluka	Triton
Formedium	Tryptone, yeast extract

Life Technologies	DNaseI, DNA ladder (1 kb plus),FBS, Hepes, HT supplement, NuPage SDS loading buffer and reducing agent, RPMI plus l-glutamine, Dyecycle Ruby, Hoechst
Marvel	milk powder (skimmed)
Melford	Agar
New England Biolabs	All restriction endonucleases and associated buffers, *Taq* polymerase, USER enzyme
Roche	10x PBS, Agarose
Sigma	Ponceau S, isopropanol, sodium dodecyl sulfate (SDS), dimethyl sulfoxide (DMSO), triton X-100, rapamycin, ß-mercaptoethanol, Tween20, Giemsa stain, Pen-strep, immersion oil, ficoli, bromophenol blue, buffered phenol, phenol:chloroform:isoamylalcohol, chlorophorm:isoamlyalcohol
Thermo Fisher Scientific	Bovine serum albumin, ethylene diamine tetraacetic acid, glycerol, glycine, methanol, Tris, Sodium Chloride, PageRuler Prestained Protein Ladder, Platinum *Taq* DNA Polymerase High Fidelity, Tween
VWR	$CaCl_2*2H_2O$, glacial acetic acid, ethanol, methanol,
Zeiss	immersion oil

Table 2.3 Biological and chemical reagents used

2.1.4 Drugs and Antibiotics

Selection drugs and antibiotics used for bacterial and parasite selection are listed in Table 2.4

Sigma	Ampicillin sodium salt, Kanamycin, Pyrimethamine, Rapamycin, 5-Fluorocytosine
Merck	Phenylhydrazine

Table 2.4 Sources of drugs and antibiotics used for this work

2.1.5 Kits

A selection of kits were used for the preparation of bacterial material in this study. Their source can be found in Table 2.5

Agilent	High sensitivity DNA chip kit
Beckman coulter	Agencourt AMPure XP purification beads
Kappa Biosystems	Kappa HotStart PCR kit
NEB	Magnetic mRNA isolation kit, NEBNext DNA Library Prep Master mix set for Illumina
Qiagen	Spin Mini-Prep, Plasmid Midi-Prep, QIAquick gel extraction kit, Sybr qPCR kit

Table 2.5 Sources of preparatory kits used for this work

2.1.6 Buffers, Solutions and Media

The culture of different bacterial strains required multiple medias and selection antibiotics (Table 2.6). Many buffers and reagents were required for the isolation of DNA, RNA and Protein from parasite lines. They chemical composition of these can be found in Table 2.7, Table 2.8, Table 2.9.

LB medium	10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl
LB agar	1.5 % (w/v) agar in LB medium
Ampicillin (1000x)	100 mg/ml in ddH$_2$O
Kanamycin (1000x)	50mg/ml
Terrific Broth	12g Tryptone, 24g Yeast extract, 4ml glycerol, 0.17M KH$_2$PO$_4$ and .72M K$_2$HPO$_4$ in ddH$_2$O

Table 2.6 Bacterial culture medias and selection used in this work

phenol chloroform isamylalcohol	Sigma
e lysis buffer	1.5mM NH$_4$Cl, 0.1< KHCO$_3$, 0.01M EDTA
50X TAE buffer	2 M Tris, 0.5 M Na$_2$EDTA, 5.71 % glacial acetic acid (v/v)
TNE buffer	10mM Tris (pH8.0), 5mM EDTA (pH8.0), 100mM NaCl
5X Loading dye	15 % Ficoll (v/v), 20 mM EDTA, 0.25 % Bromophenol Blue (w/v) in H$_2$O
Thermo 1 kb+ DNA ladder	150 µl 1kb+ ladder (1 µg/µl), 300 µl 5X DNA loading buffer, 1050 µl H$_2$O

Table 2.7 Buffers used for the preparation and analysis of DNA

chloroform, chloroform:isoamylalcohol	Sigma
isopropanol	Sigma
buffered phenol	Sigma
RNAse free H_2O	Sigma

Table 2.8 Buffers used for the preparation and analysis of RNA

RIPA buffer	50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 0.5 % sodium deoxycholate, 0.1 % SDS (w/v), 1 % triton X-100 (v/v)
SDS PAGE running buffer	25 mM Tris, 192 mM glycine, 0.1 % SDS (w/v)
Transfer buffer for wet blot	48 mM Tris, 39 mM glycine, 20 % methanol (v/v)
Washing solution or TBS-T	0.2 % tween (v/v) in TBS
PageRuler Prestained Protein Ladder	62.5 mM Tris-H_3PO_4 (pH 7.5 at 25°C), 1 mM EDTA, 2 % SDS, 10 mM DTT, 1 mM NaN_3 and 33 % glycerol.

Table 2.9 Buffers used for the preparation and analysis of Proteins

Short term culture and storage of P. berghei lines required multiple medias and reagents for successful growth. Purification and isolation of specific stages of the life cycle required a plethora of reagents (Table 2.10).

10 x PBS	137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 1.8 mM KH_2PO_4 (pH 7.4)
2x Freezing solution	Heparin, glycerol, rich PBS
5-FC	1.5mg/ml 5-fluorocytosine, tap water
Complete media	RPMI1640, Sodium Bicarb, Hepes, HT suppl. FBS
erythrocyte lysis buffer	15mM NH_4Cl, 1mM $KHCO_3$, 0.1mM EDTA
Gas mix	containing 5% CO_2, 5% O_2 and 90% N_2
Giemsa staining solution	12 % Giemsa stain (v/v) in H_2O
Heparin	200units/ml
Nycodenz	Lucron bioproduct (138g in buffered medium 500ml)
Ookinete culture media	RPMI1640, 10% FCS, xanthurenic acid
Phenylhydrazine	12.5mg/ml phenylhydrazine, ddH_2O water
Pyrimethamine	70µg/ml pyrimethamine, tap water
Rapamycin	4mg/mL in DMSO
rich PBS	20mM Hepes, 20mM Glucose, 4mM $NaHCO_3$, 0.1% BSA
Sulphadiazine	30mg/ml
FACS buffer	10% rPBS, 1 mM EDTA in PBS

Table 2.10 Media and buffers used for the culture and purification of P. berghei in vitro

2.1.7 Antibodies

In this work several antibodies were used for the detection and identification of proteins by Western blot. Table 2.11 details their source and the dilutions successful probing was achieved with.

Name	Species	Dilution	Source
FKBP-12	Mouse	1:500	Abcam Pierce
Cre	Mouse	1:500	Abcam
Enolase	Rabbit	1:1000	Abcam
α-hrp secondary	mouse	1:3000	Dako
α-hrp secondary	rabbit	1:3000	Dako

Table 2.11 Antibodies used for Western blot analysis in this work

2.2 **Methods**

2.2.1 *P. berghei* culture and purification methods

2.2.1.1 Infection of rodents

All rodent infections were carried out in female, outbred animals (Envigo). Either, Wistar rats (150 – 174g), NIH Swiss mice (26 – 30g) or TO mice (26 – 30g). Standard infections were carried out by intraperitoneal injection with *P. berghei* infected blood either cryopreserved with a freezing solution (see section 2.2.1.3) or directly from cardiac puncture or tail drops diluted in rich PBS. Synchronous infections were established with intravenous infection of purified schizonts (for details see section 2.2.1.7). If a mouse or rat was infected with a parasite line that was non-clonal immediately after infection the drinking water of the mice was supplemented with pyrimethamine or 5-fluorocytosine dependant on the selection pressure required (see section 2.2.2.2).

As *P. berghei* invades and infects reticulocytes during an infection maximising the amount of these cells within the blood stream is beneficial. Unfortunately, phenylhydrazine which induces reticulocyte production results in high autofluorescence for all blood cells and is not compatible with flow cytometry or microscopic analysis. Therefore when these techniques were not used, animals were treated with 0.1ml phenylhydrazine to increase the proportion of reticulocytes two to three days prior to infection.

2.2.1.2 Monitoring *in vivo P. berghei* infections

Parasite burden was determined by counting of Giemsa stained blood smears made from a drop (~5µl) of tail blood. Thin blood smears were made by smearing a tail drop on a standard microscope slide and fixing by submerging dried blood smears in methanol for 10 – 30 seconds. Fixed slides were dried before submersion in Giemsa solution for 15 – 20 minutes. Slides were then rinsed in H_2O and air dried before counting with a standard light microscope with a 100x objective and immersion oil.

2.2.1.3 Cryopreserving *P. berghei* parasite lines

To store parasite lines for future analysis *in vivo* blood was collected via cardiac puncture when parasitaemia reached 3 – 8%. Blood was collected in a 2ml syringe with a 23G needle, pre-loaded with 0.05 – 0.1ml heparin. After collection blood was mixed (50:50) with freezing solution and 500µl aliquots stores in cryovials at -80°C (short term storage, < 6 months) or in liquid N_2 (long term storage > 6 months).

2.2.1.4 Transmission through mosquitoes

For all mosquito feeds naïve mice were infected with *P. berghei* without any prior drug treatments. Cages of 200 – 250 mosquitoes were starved for 24 hours before the feed. Mice were anaesthetised with Ketaset$^®$ before feeding for 10 – 15 minutes on a cage. Mature oocysts were identified in the mosquito midguts between days 10 and 13 using a Leica M205 FA Fluorescence Stereomicroscope. Salivary gland sporozoites were extracted between days 18 and 21 and identified by rupturing salivary glands and Giemsa staining methanol fixed salivary gland extract. Infected mosquitoes were then allowed to feed on anaesthetised naïve mice for 10 minutes between days 20 and 23. These mice were monitored for parasitaemia (for details see section 2.2.1.2) between days 4 and 14 to ascertain if transmission was successful.

2.2.1.5 *In vitro* culture of blood stage parasites

Short term culture of *P. berghei* blood stage parasites can be carried out to obtain mature schizonts or gametocytes. Blood was collected from mice (0.8 – 1.5 ml) or rats (6 – 7 ml) in pre-heparinised syringes via cardiac puncture. For every 1ml of infected blood a large flask (150cm, unvented) of culture is set up with 105 ml complete culture media and gassed for 30 seconds with a gas mix. Flasks are incubated overnight at 37°c on a shaker at 35 - 40rpm. To check for maturation, 1ml of cultures parasites was taken and Giemsa-smears made to count parasites present.

2.2.1.6 *In vitro* culture of ookinetes

Culture of ookinetes was completed with 200 – 250µl infected blood collected via cardiac puncture. Blood was incubated overnight at 21°C with 6 – 8ml ookinete media pre-warmed/chilled to 21°C in a small flask (25cm, unvented). Ookinetes were identified in Giemsa smears from 1ml of pelleted culture (spun for 1 minute at 10,000g) made after overnight culture and again 24 hours later.

2.2.1.7 Purification of schizonts or gametocytes

When purifying schizonts a maximum of 2ml of infected blood was used for a single purification. Overnight cultures, checked for the presence of schizonts or gametocytes, was condensed and combined from culture by centrifugation for 8 minutes at 450g. This resulted in a final volume of 35ml per purification. A 55% (schizonts) or 53% (gametocytes) Nycodenz diluted in rich PBS solution was then gently laid under the blood culture carefully to ensure a clear delineation between the two solutions. The suspension was centrifuged for 20 minutes at 450g without break. After centrifugation a dark brown layer was visible between the two solutions. This layer contained predominantly schizonts (55%) or gametocytes (53%) with some contamination of mature schizonts, gametocytes and late stage trophozoites. The layer was carefully syphoned off to collect the mature parasites. The purified cells were then centrifuged for 8 minutes at 450g to collect the schizonts. Dependant of the downstream experiments parasites were resuspended in 250µl – 500µl fresh complete culture media (detailed in (Janse *et al.*, 2006)).

2.2.1.8 Life cycle stage identification

It was often necessary to discriminate between life cycle stages in a mixed population in *P. berghei* infections. Several methods have been exploited in this work to allow identification of key stages.

Identification by Giemsa smears

It was possible to classify the stage in the life cycle that parasites are in with a Giemsa stained blood smear. From as little as 15 minutes after merozoite invasion of a reticulocyte ring stage parasites were identifiable. Throughout the life cycle *in vitro* all stages were identifiable. From 18 – 24 hours post merozoites invasion schizonts sequester *in vivo* and were therefore not detectable in the peripheral blood used for blood smears. From around 18 hours post merozoites invasion gametocytes became identifiable in the circulating blood however gender discrimination is not often determinable until 28 – 30 hours post invasion when they were fully mature (Janse *et al.*, 2011).

Gametocyte identification/conversion by fluorescence

A parental line widely used in the laboratory, and in this work, contains a male specific promoter driving GFP expression and a female specific promoter driving RFP expression from a silent locus (P230p). The male specific promoter, PBANKA_041610, corresponds to a dynein heavy chain and the female specific promoter, PBANKA_131950, corresponds to a LCCL protein, CCP2. The expression of GFP and RFP allowed the identification of male and female gametocytes respectively by microscopy of flow cytometry. All fluorescence analysis was completed on non-phenylhydrazine treated mice to ensure no background auto-fluorescence was exhibited.

A second line has been developed in the lab that identified gametocytes of both genders. This line contains the PBANKA_101870 promoter driving RFP expression in the P230p silent locus. As this promoter was active in gametocytes from an earlier time point than morphological discrimination it was termed the early gametocyte line (EG). By microscopy and flow cytometry this fluorescence allowed identification of gametocytes but not gender discrimination.

Analysis by flow cytometry in these lines was achieved by resuspending blood in 500µl rich PBS with Hoechst and incubating at 37°C for 30 minutes to stain the nucleus. Blood was the pelleted by centrifugation for 30 seconds at 10,000rpm and the supernatant removed. The pellet was then resuspended in 1mL FACS buffer (Table 2.3) and filtered through a Nitex membrane (0.45 aperture). For full details on flow cytometry gating strategies see section 2.2.5. Briefly, in both lines red blood cells were first discriminated based on their size and granularity (Forward scatter (FSC) and Side scatter (SSC)) before small aggregates were excluded by doublet discrimination (FSC-Area and FSC-Height) which compared the cell volume to its diameter to gate for single cells only. Infected red blood cells were then identified based on their Hoechst signal (excited with a violet laser, V450). Once infected cells were identified gametocytes were identified and quantified in the 820 line if they were GFP (male) or RFP (female) positive. In the early gametocyte line gametocytes of either gender were identified and quantified based on their RFP signal.

2.2.2 Genetic modification of P. berghei

2.2.2.1 Transfection of *P. berghei*

An efficient method of *P .berghei* transfection has been previously described (Janse *et al.*, 2006). Briefly, *in vitro* cultures (see section 2.2.1.5) were set-up from infected blood collected by cardiac puncture from a phenylhydrazine pre-treated mouse and cultured overnight. After Giemsa smears were analysed to ensure maturation of schizonts, cultures were purified as described using a 55% nycodenz gradient (see section 2.2.1.7). Transfection DNA was prepared (see section 2.2.4.3) by initially mixing 10µl DNA (3µg - 10µg linearised DNA) with 100µl Nucleofector ® solution (Table 2.3). Schizonts collected in 500µl complete culture media were centrifuged to pellet and all remaining media removed, schizonts were then resuspended in the Nucleofector ® / DNA suspension and immediately transferred to an electroporation cuvette. The electroporation cuvette was then placed in an Amaxa Nucleofector ® transcfection machine (Table 2.1) and program U-33 used. Immediately after electroporation 200µl fresh complete culture media (37°) was added to the suspension and injected in to the tail vein of a naïve mouse pre-warmed in a 37°C hot-box (Table 2.1) so the veins were dilated for injection.

2.2.2.2 Selection of transfected *P. berghei* parasites

In many cases the drug resistant selectable marker used in genetic manipulations was DHFR/TS. This gene confers resistance to pyrimethamine. In these cases, 24 – 30 hours after transfected parasites were intravenously injected in to the mice their drinking water was replaced with water containing pyrimethamine. This drug supplemented water was maintained for 5 – 14 days until parasites were detectable in the blood stream of mice.

In other cases transfections were completed on parasites that already contain a drug resistance cassette. These lines, known as GIMO (gene in marker out, see section 1.6.2) lines, contain a dual cassette conferring resistance to pyrimethamine (DHFR/TS) and susceptibility to 5-fluorocytosine (yeast cytosine deaminase and uridyl phosphoribosyl transferase (yFCU)). The original integration of this cassette was achieved with pyrimethamine selection, subsequent manipulations to the GIMO locus exploits the 5-fluorocytosine (5-FC) sensitivity from yFCU. One day after transfection into a GIMO line, the drinking water of mice was substituted with water containing 5-fluorocytosine. Water supplemented with 5-fluorocytosine was maintained for 5 – 14 days post transfection, until parasites are detectable in the blood stream of infected mice.

2.2.2.3 Cloning selected parasite populations

After transfection and selection by either pyrimethamine or 5-fluorocytosine the majority of the surviving population are transfectants but a subset may be wild-type parasites that have survived the selection procedure. In some cases transfectants can be identified by their fluorescence and therefore a pure transfectant population can be sorted based on fluorescence. In other cases transfectants and wild-type parasites are not fluorescently discernible and must be cloned by limiting dilution.

Isogenic sorting

If a transfected line was fluorescently identifiable from the wild-type background line then it was sorted to a non-clonal population of fluorescent transfectants. Due to the damage induced by many DNA stains parasites are identified only on their fluorescence signal. Singlet red blood cells were identified based on their size (forward scatter) and granularity (side scatter) and

gated. From this population infected red blood cells were determined by their fluorescence intensity (CFP or GFP). Sorting was carried out on a BD FACSAria II, a BD FACSAria III or BioRad S3e cell sorter with samples resuspended in FACS buffer and maintained at 37°C before sorting into rich PBS. Each mouse infected for isogenic cloning was intravenously injected with 50 sorted parasites. In a standard isogenic cloning 2 naïve mice were each infected with 50 parasites.

Cloning by limiting dilution

If a line was not discernible from the wild-type population cloning was achieved by intravenous injection of 0.8 parasites per mouse. In each cloning 10 naïve mice were injected. Briefly, the donor for the line to be transfected was intraperitoneally injected with 100µl phenylhydrazine three days before infection with 500µl of the cryopreserved (see section 2.2.1.3 for details on parasite cryopreservation) line to be cloned. One day after infection parasitaemia was accurately calculated (a minimum of 25 fields counted) from Giemsa smears and cloning by limiting dilution completed if parasitaemia was between 0.1% and 1%. If the parasitaemia was not within these parameters two days after infection the parasitaemia was again analysed but only if these parameters were met was cloning be completed. From this donor mouse approximately 5µl of tail blood was collected (via a tail drop) in 1ml rich PBS. A homogeneous sample of this dilution (10µl) was used to count red blood cells on a haemocytometer. The concentration of red blood cells in the dilution was calculated from this and from the accurate parasitaemia calculation the quantity of infected red blood cells determined. Two dilution of 1:10 were completed before a final dilution to obtain 5ml of a solution that contained 0.8 infected red blood cells per 200µl (20 infected red blood cells/5ml). For the cloning 200µl of this dilution was intravenously injected into 10 naïve mice pre-warmed in a hot box at 37°C to dilate their veins. From day 10 after infection mice were checked for parasitaemia.

2.2.2.4 Negative selection of *P. berghei* parasites

In several of the lines used in this study it was necessary to recycle the selectable marker used to integrate the original genome modification. Vectors with the ability to recycle the selectable markers contain both the positive selection cassette *dhfr/ts* and the negative selection cassette *yfcu* flanked by identical 3'UTR sequences (Figure 2.1). Initial integration of the vector into the

genome was achieved by pyrimethamine selection (see section 2.2.2.2) and then the line was cloned (see section 2.2.2.3) so all parasites were integrants. To recycle the marker 5-FC selection pressure was applied to the mice through their drinking water. The addition of 5-FC caused homologous recombination between the matching 3'UTRs that flank the entire selection cassette (Figure 2.1). Parasites that lost the entire selection cassette no longer carried the *yfcu* gene that confers susceptibility to 5-FC. In the process they also lose the *dhfr/ts* gene that confers resistance to pyrimethamine.
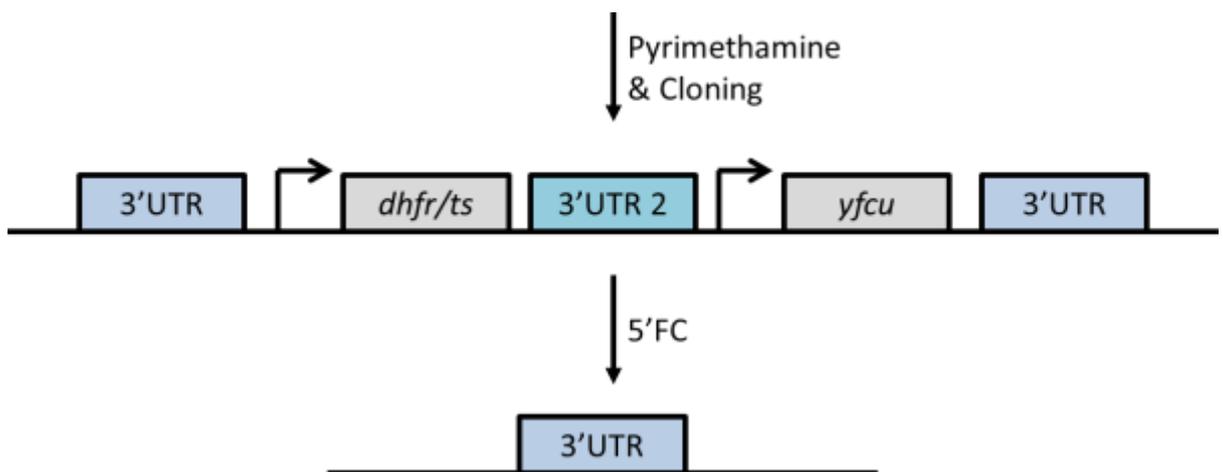
Figure 2.1 Schematic representation of negative selection in P. berghei

To recycle the selectable marker, after integration, selection and cloning, clonal integrated lines are selected with the drug 5-fluorecytosine (5-FC). Only parasites containing the *yfcu* gene are susceptible to this drug. Negative selection is achieved when the parasites use the two homologour 3'UTRs surrounding the selection genes (*dhfr/ts* and *yfcu*) to loop-out both genes. These parasites are able to survive drug treatment with 5'FC and become susceptible to pyrimethamine again. The only remaining footprint of the selection is the 3'UTR.

Once these parasites had been selected they were again cloned (see section 2.2.2.3) to ensure the entire population had lost both the susceptibility to 5-FC (*yfcu* gene) and the resistance to pyrimethamine (*dhfr/ts* gene).

2.2.3 Extraction of *P. berghei* parasite material

Further characterisation of the parasite lines required purification of the parasites from whole mouse blood collected by cardiac puncture with 0.05 – 0.1ml heparin in a 2ml syringe. The subsequent methodologies involved in the processing varied dependant on the downstream analysis.

2.2.3.1 Parasite extraction for genomic DNA

For isolation of DNA, parasites were isolated from heparinised whole blood after cardiac puncture. DNA was isolated from 300µl - 500µl blood lysed, on ice, in 50ml pre-chilled (4^oC) erythrocyte lysis buffer for 3 – 5 minutes. Parasites were pelleted by centrifugation for 8 minutes at 450g. Lysed parasites were washed with 1ml PBS before complete removal of the supernatant and storage at -20^oC.

2.2.3.2 Parasite extraction for Protein lysate

For isolation of protein lysate, parasites were isolated from 0.5 – 1ml heparinised whole blood obtained from cardiac puncture. Blood was first leucocyte depleted by filtration through a plasmodipur® filter. For depletion, whole blood was diluted in 5ml rich PBS and passed through a filter pre-wetted with 5ml rich PBS using a 20ml syringe. The filter was then washed with a further 10ml rich PBS. Leucocyte depleted blood was pelleted by centrifugation for 8 minutes at 450g. Blood was resuspended in 50ml pre-chilled (4^oC) erythrocyte lysis buffer and incubated on ice for 3 – 5 minutes before parasites were pelleted by centrifugation for 8 minutes at 450g. Parasites were washed with 1ml PBS before complete removal of the supernatant and storage at -80^oC.

2.2.3.3 Parasite extraction for RNA

For isolation of RNA parasites were extracted from 1 – 1.5ml heparinised whole blood obtained from cardiac puncture. Blood was immediately diluted in 10ml ice cold filter sterilised rich PBS and kept on ice. Blood was leucocyte depleted by filtration through two plasmodipur® filters sequentially (as described in 2.2.3.2, with sterile PBS). Double filtered, leucocyte depleted blood was centrifuged at 4^oC for 8 minutes at 450g to pellet blood. After removal of the rich PBS the blood pellet was resuspended in 50ml filter sterilised erythrocyte lysis buffer and incubated on ice for 5 minutes. Lysed blood was pelleted at 4^oC for 8 minutes at 450g before resuspension in 50ml filter sterilised erythrocyte lysis buffer and incubation on ice for a further 5 minutes. Double lysed blood was pelleted as before. After two incubations with erythrocyte lysis buffer parasite pellets should have been completely lysed and free from red blood cell material. However, if the dark brown/red parasite pellet retained a lighter red halo the supernatant was removed and a third erythrocyte lysis step was

repeated as above. Once the parasite pellet was completely lysed the supernatant was removed and the pellet washed in 1ml PBS and transferred to a 2ml Eppendorf® tube. Parasites were then pelleted (10,000 rpm) and the supernatant completed removed before resuspension in 1ml TRIzol® reagent and stored at -80°C.

2.2.4 Molecular biology methods

2.2.4.1 Genomic DNA extraction

For genomic DNA extraction, the parasite pellet was resuspended in 700µl TNE buffer (Table 2.7) and transferred to a 2ml Eppendorf® then 200µg RNAse (Table 2.3) and 1% SDS (100µl of a 10% solution) was added. This mixture was incubated for 10 – 15 minutes at 37°C. 200µg Proteinase K (Table 2.3) was added and the solution incubated for a further 1 hour at 37°C. After incubation 500µl buffered phenol (Table 2.3) was added and mixed by inverting the tube several times. The mixture was centrifuged for 3 minutes at 14,000rpm. The aqueous upper phase was carefully removed and transferred to a new 2ml Eppendorf® where 500µl phenol:chloroform:isoamylalcohol (25: 24: 1, Table 2.7) was added before mixing by inversions of the tube. To separate the layers the mixture was again centrifuged for 3 minutes at 14,000rpm. The aqueous phase was transferred to a new 2ml Eppendorf® and 500µl chloroform:isoamylalcohol (24: 1) was added and mixed by inversions. A final centrifuge of 3 minutes at 14,000rpm was completed before the aqueous phase was transferred to a 1.5ml Eppendorf®. DNA was ethanol precipitated with 2.5 volumes 96% ethanol and 5µl 3M Sodium acetate (pH 5.2) at -20°C overnight. The precipitated DNA was spun at maximum speed for 20 minutes to precipitate the DNA. The supernatant was removed before two washes with 500µl 70% ethanol were completed to remove residual ethanol. The DNA pellet was left to dry for 5 – 30 minutes at room temperature. When completely dry the DNA was resuspended in 100µl ddH$_2$O.

2.2.4.2 RNA extraction

The RNA extraction protocol was completed at the Sanger institute by Katarzyna Modrzynska. Extraction of RNA was completed following the TRI Reagent protocol (Chomczynski, 1993). Once thawed parasites were homogenised in the TRIzol® reagent and incubated at room temperature for 5 minutes. 200µl

chloroform was added to the mixture and vigorously shaken for 15 seconds before incubation at room temperature for 5 – 15 minutes. Phase separation was achieved by centrifugation at 4°C, 12,000g for 15 minutes. RNA was found only in the aqueous phase (three layers were discernible, the phenol-chloroform phase (red), the interphase and the aqueous phase) and was transferred to a new Eppendorf$^{®}$. RNA was precipitated with ½ the volume of the original homogenised volume of isopropanol. To precipitate the mixture was incubated at room temperature for 10 minutes before centrifugation at 12,000g for 10 minutes at 4°C. RNA was washed twice by resuspending the pellet in 1ml in 75% ethanol and vortexing. The pellet was obtained by centrifuging for 10 minutes at 12,000g at 4°C. The washed pellet was then air dried for 3 – 5 minutes and resuspended in RNAse free H$_2$O.

Between 1μg and 2μg of the total RNA extracted (above) was then used for mRNA isolation (Magnetic mRNA isolation kit, Table 2.5). First strand cDNA synthesis was performed using the SuperScript III first-strand system with Oligo(dT) and random primers (1:1, Table 2.3). The DNA/RNA hybrids were purified using Agencourt RNAClean XP beads. Briefly, 1.8 volumes of Agencourt RNAClean XP was added to the sample and mixed by vortexing for 30 seconds. Beads were separated from the solution on a super magnet plate for 5 – 10 minutes. The supernatant was gently removed (still on the magnet) and the beads were washed three times with 70% ethanol. Beads were then air dried for 10 minutes to remove residual ethanol. Elution was achieved by adding 40μl RNase free H$_2$O and vortexing thoroughly to dissociate from the beads. Second strand cDNA was synthesised using a mix of 10mM dUTPs, DNA polymerase I and RNAseH (Table 2.3) incubated at 16°C for 2.5 hours. Second strand cDNA was then purified using Agencourt RNAclean XP (as above). The long cDNA fragments generated were then fragmented using a Covaris S220 system (duty cycle = 20, intensity = 5, cycles/burst = 200, time = 30 seconds). The yielded ~200bp fragments were then end-repaired, poly-A tailed, indexed using NEBNext DNA library prep master mix set for Illumina, following manufacturer instructions, and ligated for sequencing using the no-PCR method (Kozarewa *et al.*, 2009). Excess adapters were then removed by two rounds of clean-up using 1x volume of Agencourt AMPure beads with other steps as described above. The final libraries were eluted in 30μl RNAse free H$_2$O. Quality control was performed on

all samples with High sensitivity DNA chip kit (Table 2.5). For successful indexing samples were digested with the USER enzyme (Table 2.3) and quantified by qPCR. If any sample in a pair yielded insufficient quantity for sequencing 5 cycles of PCR amplification with Kapa HiFi Hotstart PCR mix (Table 2.5) and Illumina tag-specific primers were performed on both the first and second strand and its partner in the time course. Pools of 12 Indexed libraries were sequenced using an Illumina HiSeq2500 system resulting in 100bp, paired end reads.

2.2.4.3 Protein extraction for Western blot

Protein was only isolated for Western blot analysis, therefore, protein pellets were resuspended in 5x pellet volume RIPA lysis buffer (Table 2.9) and incubated on ice for 30 minutes – 1 hour. As only soluble proteins were analysed in this work the proteins were spun for 10 minutes at 4°C, 14,000rpm and the supernatant transferred to a new Eppendorf®. 4x SDS gel-loading buffer (Table 2.9) with fresh fresh (15%) ß-mercaptoethanol was then added and the samples boiled at 100°C for 5 minutes in a hot block (Table 2.1).

2.2.4.4 Generation of recombineering vectors (PlasmoGem, Sanger)

All knockout vectors were obtained through the PlasmoGem resource operated by the Sanger institute. Though each gene's knockout vector differs in its specific makeup the general structure of vectors remains consistent (Figure 2.2). pJazz arms were always 2kb and 10kb and are released from the specific gene knockout vector by Not I digest. Their orientation on the vector can be 5' short arm and 3' long arm (as shown) or 5' long arm and 3' short arm. The homology arms were variable in size, though all vectors contained a minimum homology arm of 2kb on both sides. The homology arms included (in some cases) multiple genes dependant on the region and some also included part of the gene being targeted. After double homologous recombination part or all of the *goi* was replaced with the positive/negative selectable marker cassette driven by an ef1α constitutive promoter. Each of the individual vectors was selected for using pyrimethamine and integration was checked with a primer specific to the vector, in the *yfcu* or *dhfr* selection cassettes or in the *ef1a* promoter, and one specific to the region integrated in to (Table 2.12, Supplementary 1 and Supplementary Figure 4). This specific primer could have been in an intergenic region, within the *goi* itself or within one of the genes surrounding the *goi*.
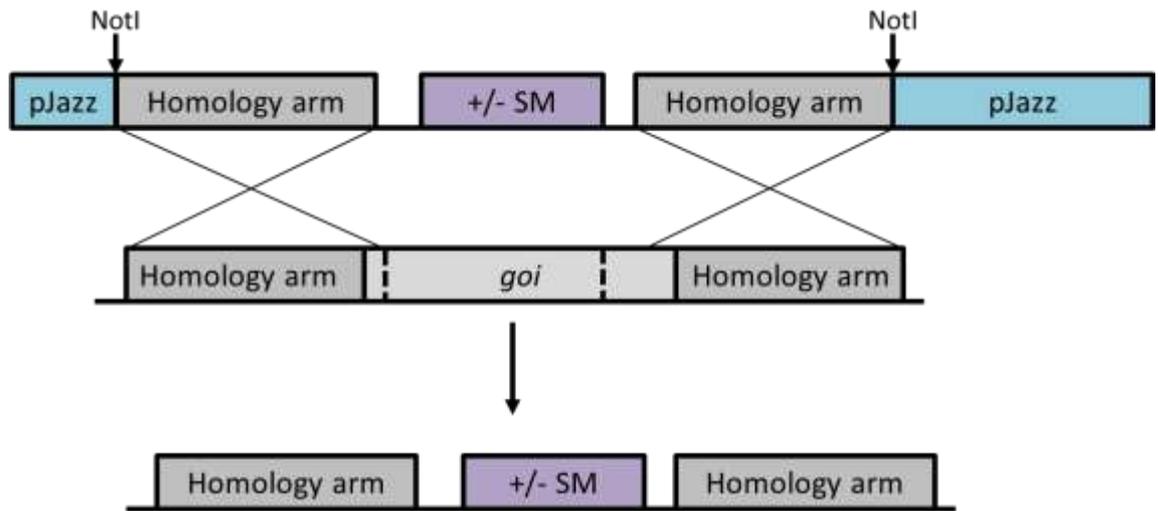
Figure 2.2 Schematic representation of the PlasmoGem knockout vectors

For bacterial growth all vectors contain PJazz hairpin telomers allowing for replication. These are released from the vector by a NotI digest that releases the 2kb and 10kb pJazz arms. These can be oriented as shown (5' short, 3'long) or in the opposite orientation (5'long, 3'short). The homology arms for each vector differ based on the gene targeted. The homology arms can encompass several neighbouring genes to the gene of interest *(goi)* or can be made up of part of the *goi*. The selectable marker, made up of a positive DHFR selection cassette and the negative selection cassette YFCU replaces part or all of the *goi* after successful transfection.

2.2.4.3 Preparation of DNA constructs

All DNA constructs used in this study were generated prior to this work being completed by lab colleagues (RC, KH, RL & KM) or obtained from the PlasmoGem resource. All of the vectors were grown and prepared for transfection. PlasmoGem obtained vectors (Table 2.12, Supplementary file 1, Recombineering vector maps) were all grown in terrific broth supplemented with 0.4% glycerol and with kanamycin selection. As PlasmoGem vectors need to be maintained as low copy number vectors and extended growth has been shown to increase mutation rate they were cultures overnight for a maximum of 16 hours at 37°C in a shaking incubator (Table 2.1). For transfection of the PlasmoGem vectors, 10mL of overnight culture was used for a mini-prep kit (Quiagen, Table 2.5) following the manufacturer's instructions. Usually this yielded between 5μg and 10μg DNA; all of which was used for the transfection. All recombineering vectors were released from their bacterial growth arms (pJazz) by NotI digest and due to their size were not purified before transfection. The pJazz arms show no sequence homology to the *P. berghei* genome and therefore cannot integrate.

Gene	PbGem reference	Map location	Integration (F)		Integration (R)		Integration product	
			Description	No.	Description	no.	Size	End
PBANKA_090850	PbGEM-237320	S1-A	YFCU	G3614	PBANKA_090850	G3611	2.1kb	3'
PBANKA_101830	PbGEM-263324	S1- B	PBANKA_101810	G1319	YFCU	G3614	3.2kb	5'
PBANKA_112040	PbGEM-099323	S1-C	PBANKA_112030	G3620	hDHFR	G3612	3kb	5'
PBANKA_121830	PbGEM-248816	S1-D	Ef1α promoter	G3612	PBANKA_121840	G3621	4kb	3'
PBANKA_070470	PbGEM-330627	S1-E	Ef1α promoter	G3612	PBANKA_070460	G3622	2.2kb	3'
PBANKA_060120	PbGEM-329947	S1-F	Ef1α promoter	G3612	PBANKA_060120	G3609	2.2kb	3'
PBANKA_142920	PbGEM-111063	S1-G	Ef1α promoter	G3612	PBANKA_142930	G3624	3.2kb	3'
PBANKA_081180	PbGEM-237576	S1-H	PBANKA_081170	G3610	YFCU	G3614	3.4kb	5'
PBANKA_093360	PbGEM-291200	S1-I	PBANKA_093350	G3613	Ef1α promoter	G3612	3.3kb	5'
PBANKA_021390	PbGEM-227253	S1-J	PBANKA_021380	G3615	YFCU	G3614	2.4kb	5'
PBANKA_111500	PbGEM-299538	S1-K	Ef1α promoter (2)	G3317	PBANKA_111510	G3618	2.1kb	3'
PBANKA_050830	PbGEM-014629	S1-L	PBANKA_050820	G3634	YFCU	G3614	2.5kb	5'
PBANKA_133500	PbGEM-057898	S1-M	YFCU	G3614	PBANKA_133510	G3625	5.3kb	3'
PBANKA_133810	PbGEM-058356	S1-N	YFCU	G3614	PBANKA_133830	G3626	4.6kb	3'
PBANKA_113860	PbGEM-247164	S1-O	Ef1α promoter	G3612	PBANKA_113860	G3627	2.9kb	3'
PBANKA_010510	PbGEM-005773	S1-P	YFCU	G3614	PBANKA_010490	G3629	4.3kb	5'
PBANKA_122550	PbGEM-050079	S1-Q	PBANKA_122540	G3631	YFCU	G3614	3.2KB	5'
PBANKA_122990	PbGEM-050692	S1-R	YFCU	G3614	PBANKA_123000	G3632	3.6kb	3'
PBANKA_136040	PbGEM-254107	S1-S	YFCU	G3614	PBANKA_136040	G3633	2.1kb	5'
PBANKA_121260	PbGEM-102303	S1-T	YFCU	G3614	PBANKA_121260	G3732	2.4kb	3'
PBANKA_082800	PbGEM-091835	S1-U	YFCU	G3614	PBANKA_082800	G3735	2.1kb	3'
PBANKA_090240	PbGEM-287210	S1-V	PBANKA_090240	G3738	YFCU	G3614	2.2kb	5'
PBANKA_112510	PbGEM-043599	S1-W	Ef1α promoter	G3612	PBANKA_112510	G3740	2.8kb	3'
PBANKA_142930	PbGEM-340051	S1-X	PBANKA_142930	G3741	YFCU	G3614	2.8kb	5'
PBANKA_140430	PbGEM-339443	S1-Y	Ef1α promoter	G3612	PBANKA_140440	G3749	2.9kb	3'
PBANKA_050500	PbGEM-085541	S1-Z	PBANKA_050490	G3750	Ef1α promoter	G3612	2.3kb	5'
PBANKA_101830	PbGEM-263324	S1-AA	PBANKA_101810	G3619	YFCU	G3614	3.2kb	5'
PBANKA_051060	PbGEM-329547	S1-BB	PBANKA_051060	G3753	Ef1α promoter	G3612	2.6kb	5'
PBANKA_050440	PbGEM-014061	S1-CC	Ef1α promoter	G3612	PBANKA_050450	G3628	2.4kb	3'
PBANKA_080720	PbGEM-236591	S1-DD	Ef1α promoter	G3612	PBANKA_080730	G3630	3.6kb	3'
PBANKA_031270	PbGEM-229846	S1-EE	PBANKA_031260	G3316	YFCU	G3614	5.3kb	5'
PBANKA_130810	PbGEM-072322	S1-FF	PBANKA_130800	G3751	Ef1α promoter	G3612	4.9kb	5'
PBANKA_083040	PbGEM-027375	S1-GG	PBANKA_083020	G3742	YFCU	G3614	3.4kb	5'
PBANKA_041340	PbGEM-329075	S1-HH	PBANKA_041340	G3739	YFCU	G3614	2.8kb	5'
PBANKA_071650	PbGEM-282066	S1-II	YFCU	G3614	PBANKA_071660	G3733	3kb	3'
PBANKA_135250	PbGEM-060427	S1-JJ	PBANKA_135250	G4103	YFCU	G3614	2.3kb	5'
PBANKA_141810	PbGEM-064794	S1-KK	YFCU	G3614	PBANKA_141820	G3736	2.7kb	3'
PBANKA_143520	PbGEM-067210	S1-LL	Ef1α promoter	G3612	PBANKA_143520	G3737	2.6kb	3'
PBANKA_041720	PbGEM-231150	S1-MM	PBANKA_041710	G3752	Ef1α promoter	G3612	3.8kb	5'
PBANKA_010240	PbGEM-327555	S1-NN	PBANKA_010230	G3734	Ef1α promoter	G3612	2.8kb	5'

Table 2.12 Plasmogem vectors and their integration strategies

Each of the genes of interest (*goi*) targeted for knockout with a recombineering vector has an individual vector map (supplementary 1) and integration strategy specific to the *goi.* In all cases due to the size of the homologous regions only 5' or 3' integration is possible. For primer sequences see Supplementary Figure 4.

Laboratory prepared constructs (Table 2.13, Supplementary 2, Lab generated vector maps) were all grown at 37°C overnight in a shaking incubator (16 -20 hours) in LB media with ampicillin selection as all contain the ampicillin resistance cassette. For transfection of the lab prepared vectors, a 50mL overnight culture was prepared using a midi-prep kit (Quiagen, Table 2.5) following manufacturer's instructions. This usually yielded between 50µg and 150µg plasmid DNA which was sufficient for multiple transfections. The different vectors were linearised with one or two enzymes to release the region to be integrated from the bacterial resistance cassette (ampicillin) before purification of the fragment of interest (Table 2.12, Supplementary 2, Lab generated vector maps). Purification was achieved by gel purification of the required fragment (QIAquick gel extraction kit, Table 2.5), following manufacturer's instructions. As the purification step results in a loss of DNA initial digests are set up with 15 - 20µg DNA to ensure retention of 5µg - 10µg of purified DNA.

Vector Name	Description	Linearisation	Integration product size	Integration 5' Primers	Size	3' Primers	Size	WT Primers	Size (Int)	Size (WT)	Map location
P1-GIMO	p230P P1 +/-ve selection cassettes.	SacII	4.2kb	P1 - P5	1.3kb	P6 - P2	2kb	P1 - P2	5.5kb	3.8kb	S2-A
P2-GIMO	p230P P2 +/-ve selection cassettes.	SacII	4.2kb	P3 - P5	1.6kb	P6 - P4	0.8kb	P3 - P4	4.6kb	2.9kb	S2-B
P1 - DiCre	p230P P1 DiCre (in to GIMO)	PvuI & SapI	5.5kb	P1 - P7	1.2kb	P8 - P2	2.3kb	P1 - P2	6.8kb	5.5kb	S2-C
P2 - DiCre	p230P P2 DiCre (in to GIMO)	PvuI & SapI	5.5kb	P3 - P7	1.5kb	P8 - P4	1.1kb	P3 - P4	5.9kb	4.6kb	S2-D
P2 - GFP	p230P P2 constitutive GFP	SacII	7.5kb	P3 - P9	2.5kb	P10 - P4	1.4kb	P3 - P4	7.8kb	2.9kb	S2-E
P2 - GFP (-ve sel)	p230P P2 constitutive GFP negative selected.	SacII	n/a	n/a	n/a	P10 - P4	1.4kb	P3 - P4	5kb	7.8kb	S2 -F
P2 - CFP	p230P P2 constitutive CFP	SacII	7.5kb	P3 - P9	2.5kb	P10 - P4	1.4kb	P3 - P4	7.8kb	2.9kb	S2-G
P2 - CFP (-ve sel)	p230P P2 constitutive CFP negative selected.	SacII	n/a	n/a	n/a	P10 - P4	1.4kb	P3 - P4	5kb	7.8kb	S2 -H
P2 - early gam RFP	p230P P2 early gam RFP	SacII	8.1kb	P3 - P9	2.5kb	P10 - P4	1.4kb	P3 - P4	8.4kb	2.9kn	S2- I
P2 - early gam RFP (-ve sel)	p230P P2 early gam RFP negative selected.	SacII	n/a	n/a	n/a	P10 - P4	1.4kb	P3 - P4	6kb	8.4kb	S2 -H

Table 2.13 Lab generated vectors and their integration strategies

Each of the vectors of interest is linearised before transfection releasing the portion of the vector required to be integrated. As these are all less than 10kb they can be purified before tranfection (Quiagen gel purification kit).

Successful integration of each of the vectors was shown through 5' & 3' Integration as well as checking for the WT or Integrated size amplicon surrounding the homology regions. If a line was to be negatively selected the reduction is amplicon size was noted after loop out of the selection cassettes. All vector maps (integrated in to the genome) can be found in supplementary 2). For primer sequences see Supplementary Figure 4.

2.2.4.4 Checking for integration of transfectants by PCR.

From day 7 post-transfection parasitaemia of mice was monitored (by Giemsa smears, see section 2.2.1.2). When parasitaemia reached 3 - 8% blood was collected and approximately 500µl cryopreserved (see section 2.2.1.3). The remainder was lysed and used for genomic DNA preparation (see sections 2.2.3.1 and 2.2.4.1). To ascertain if the population obtained after transfection contained integrants and to estimate the proportion of the population that was wild type (WT) PCR using specific primers was used. In the case of laboratory generated integrants both 5' and 3' Integration could be determined as well as WT-region vs. Integrated-region PCR. The proportion of WT compared to integrated locus was determined using upstream and downstream primers, outside of the homology arms used to integrate the DNA of interest, and the size differences identified a WT or Integrated region. To determine 5' and 3' integration these 5' and 3' homology external primers were used in conjunction with a DNA of interest specific primer at the 5' and at the 3' end. These amplified 5' and 3' Integration products only if the DNA of interest was present in the population (Figure 2.3 A). In the case of the recombineering vectors the regions of homology used are too large to amplify the entire region therefore it was not possible to compare the proportion of WT and integrated in the population. Furthermore, it is common for these vectors to have one very large homology arm (>5kb) which is too large to amplify across. Therefore, all recombineering vectors were considered integrated if the 5' or 3' integration product that amplifies across the smaller homology arm was successfully amplified (Figure 2.3 B). In lab generated lines that were subsequently cloned the WT region amplified (red primers) disappeared from the population leaving only the Integrated region.
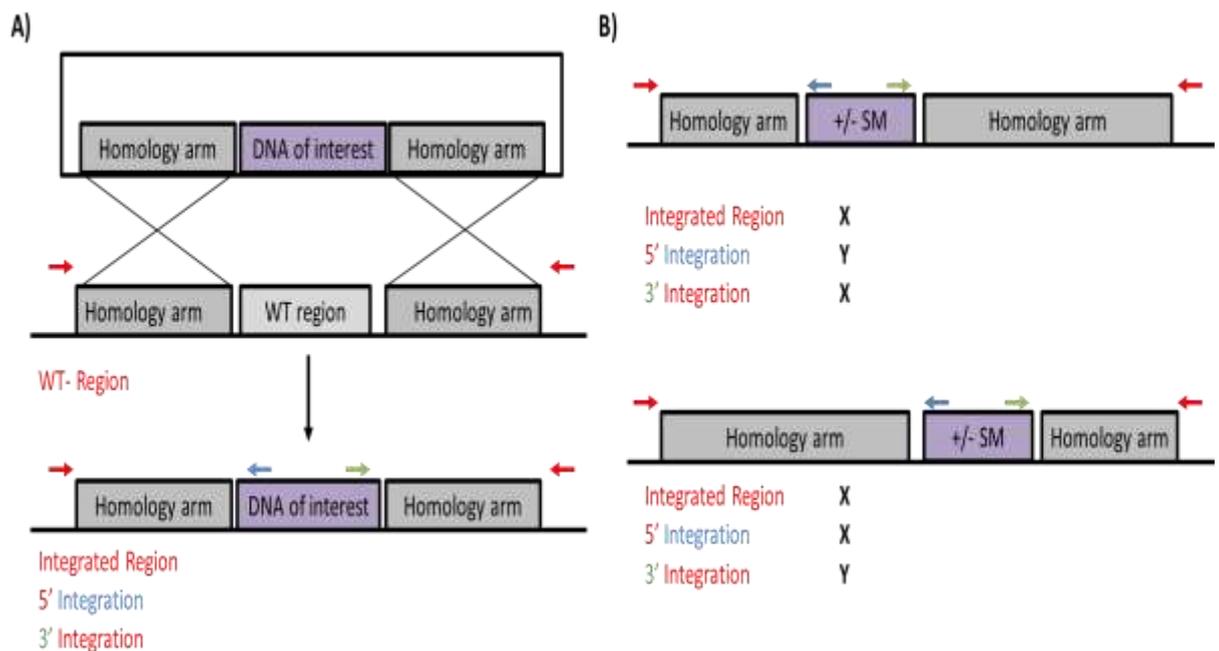
Figure 2.3 Schematic representation of Integration of recombineering and lab made vectors

A) For the laboratory generated vectors the DNA of interest was considered integrated if the Integrated region (red primers) was amplified along with the 5' product (red/blue primers) and the 3' product (green/red primers). Wild type remaining in the population was identified by the amplification of a different (usually smaller) sized amplicon with the primers external to the homology arms (red).

B) For the recombineering vectors the homology arms are always too large to allow for WT or integrated region amplification to occur (red primers). One homology arm is always small enough to allow for amplification with an internal primer and one outside of the homology arm. However, this could be 5' (red/blue) or 3' (green/red) dependant on the vector and gene orientation. Therefore, a recombineering vector was considered integrated if the 5' or 3' integration amplicon was amplified.

2.2.4.5 Confirmation of negative selection by PCR

To check for successful negative selection PCR was also used. In this case the loss of the selectable marker was confirmed in two ways. Firstly the primers used to compare the wild type region and the integrated regions were used to amplify. In this case parasites in which negative selection has not occurred amplified a larger region than those in which negative selection had been successful. Secondly, primers that recognise the matching 3'UTRs used to drive the homologous recombination resulting in negative selection and the selectable marker were used to amplify the selectable marker. Only in parasites where the selection cassette remains, no successful negative selection, was a product be amplified (Figure 2.4).
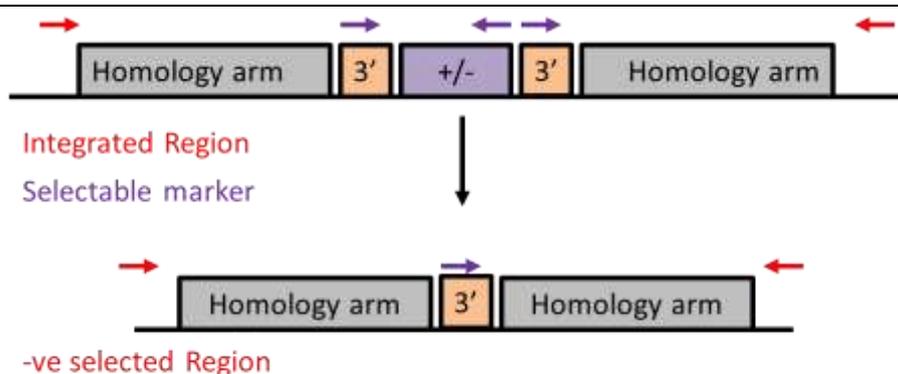
Figure 2.4 Schematic representation of negative selection confirmation by PCR

In negatively selected lines successful recycling of the selectable marker is confirmed by a size reduction (corresponding to a loss of the selectable markers and one 3'UTR) when amplifying with primers external to the homology arms (red) and by the loss of the amplicon that is amplified by primers specific to the 3'UTR and the selection cassette (purple).

2.2.4.6 General PCR conditions

Parasite DNA was isolated (see section 2.2.4.1) as previously described and the strategy of amplification designed based on the genomic locus in question. The general protocol for PCR amplification used a master mix of the following reagents:

Component	Volume (µl)	Final concentration
10x PCR buffer	2.5	1x
25mM dNTPs	0.5	0.5mM
10pmol F primer	1	1µM
10pmol R primer	1	1µM
50mM MgCl2	1	2mM
DNA template	1	(50 – 100ng/reaction)
Taq polymerase	0.2	1 Unit
ddH$_2$O	to 50µl	

The following thermocycler program was used for all reactions:

Denaturation		Annealing	Extension		Hold
94°C	94°C	Tm ^{o}C	72°C	72°C	4°C
30 sec	15 sec	30 sec	1 min/kb	10 min	∞
	22 – 30 cycles				

Where the Tm was determined by the primer pair being used (Table 2.2 for Tm calculation software) to amplify the fragment and the elongation was determined by the length of the amplicon (30 seconds per kb of amplification).

Analysis of all PCR amplifications was carried out by running 1% agarose gels with SybrSafe ® DNA stain for 30 – 45 minutes at 100V. Visualisation of gels was achieved by UV illumination in a Biorad gel doc.

2.2.4.7 qPCR procedure

Parasite DNA was isolated (see section 2.2.4.1) as previously described. The general protocol for qPCR amplification used a master mix of the following reagents:

Component	Volume (µl)	Final concentration
2x Master mix	12.5	1x
10pmol F primer	0.4	0.5µM
10pmol R primer	0.4	0.5µM
DNA template	5	(2ng/µl)
ddH$_2$O	to 25µl	

The protocol for analysis is a standard program in the StepOne™ Software:

Denaturation		Annealing	Melt curve		
95°C	95°C	60°C	95°C	60°C	95°C
10 min	15 sec	1 min	15 sec	1 min	15 sec
	40 cycles			Gradient + 0.3^{o}C	

All reactions were carried out according to the manufacturers instructions (Quiagen, Table 2.5). Subsequent quantifications and calculations were based on the ΔΔCt method.

2.2.4.8 Western blot analysis

Western blot analysis was completed on parasite pellets isolated for protein analysis (see section 2.2.3.2) and proteins were lysed and the soluble fraction prepared for Western blot analysis.

Protein samples were run on an SDS-PAGE gradient gel in running buffer (Table 2.9) for approximately 2 hours at 120V. Transfer to a Whatman® Protran® nitorocellulose membrane was completed in a Trans-blot® electrophoretic transfer system (Table 2.1). Transfer buffer (Table 2.9) was pre-chilled and all transfer elements were pre-soaked in the buffer. The transfer cassette was

assembled (Figure 2.5) to ensure transfer from the gel to the nitrocellulose membrane. All transfers were run at 4°C with ice blocks in the transfer buffer for 1 hour at 100V. To ensure successful protein transfer and to ascertain if the samples had similar loading all membranes were stained with Ponceau-S (Table 2.9) by immersing the membrane in the solution for 10 – 30 seconds and washing 5 times with PBS (Table 2.3). Once successful transfer was confirmed, by the presence of multiple bands, the membrane was blocked with 5% milk (Table 2.3) made up in PBST (Table 2.9) for 30 minutes at room temperature. This removal of residual Ponseau-S turned the milk pink. All membranes were then probed with the primary antibody of choice. Primary antibodies were diluted to their appropriate dilution (Table 2.4) in 5% milk made up in PBST (Table 2.11) and incubated overnight at 4°C on a rotating platform.

After probing with the primary antibody and washing excess antibody away with three 15 minute PBS washes at room temperature the membrane was probed with the compatible HRP-labelled secondary antibody. All secondary antibodies were diluted 1:3000 in 5% milk made up in PBST and incubated for approximately 1 hour. Excess secondary antibody was washed off as described above before protein detection by enhanced chemiluminescence (ECL) detection system.
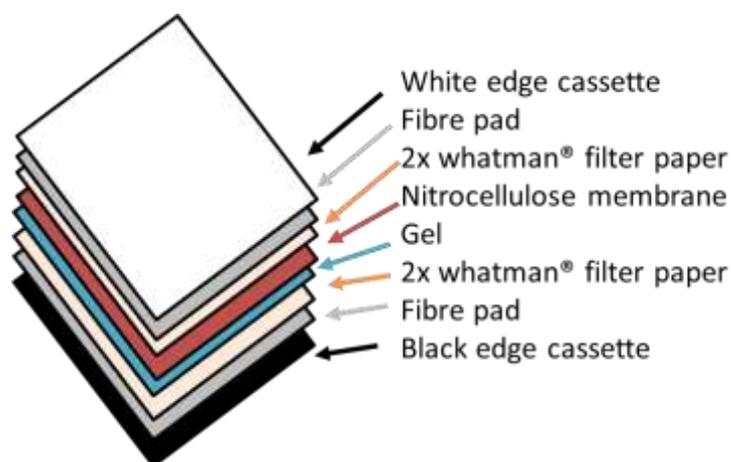
Figure 2.5 Schematic of membrane transfer for Western blot

The washed membrane was gently dried to remove excess PBS. The ECL mix was always freshly prepared at a ratio of 50:50 and poured on to clingfilm. The membrane was immediately placed face down on the ECL mixture and incubated for 3 – 5 minutes. Visualisation was achieved with X-ray film and photographic development.

2.2.5 Flow cytometry

2.2.5.1 Standard Flow cytometry methods

To analyse many of the lines generated in this study flow cytometry was used. For all analysis samples were isolated from tail drops, cardiac puncture or overnight culture. After pelleting the blood sample by centrifugation at 10,000rpm for 1 minute and removing the supernatant the blood was resuspended in pre-warmed (37°C) rich PBS (Table 2.10) with Hoechst (Table 2.3) or dyecycle ruby$^{®}$ (dependant on the complementary fluorescence being analysed (see below). Both of these dyes were used to stain the nucleus of the parasites. This mixture was incubated for 30 minutes at 37°C in a shaking incubator, 35 – 40rpm. After staining was completed the blood was pelleted as above and resuspended in 500μl - 1000μl FACS buffer (Table 2.10). The resuspended blood was passed through a nitex membrane (45 micron aperture) to limit blood cells aggregating. Samples were run at a maximum of 10,000 events per second through one of four flow cytometry machines (Table 2.1). Majority of analysis was carried out on the Cyan ADP or the LSR II. Initially the population was gated not based on fluorescence. Samples were gated on forward scatter (FSA) which determines the size of cells and side scatter (SSC) which determines the granularity of cells. In this gating we excluded anything very large or granular. Secondly, we used doublet discrimination to remove and cells which do not have proportions of cell area (FSA) and diameter (FSH) that correlate to a single cell. This excluded smaller aggregates of cells in our samples from subsequent analysis, hence its name, doublet discrimination. From this stage the method of gating and analysis varied dependant on the line being analysed.

2.2.5.2 Analysis of 820 derived lines (no additional fluorescence)

One of the most commonly used parental lines in this work was the 820 line. This line expressed GFP under the control of a male specific promoter (PBANKA_ 041610, dynein heavy) and RFP under the control of a female specific promoter (PBANKA_131950, CCP2). When analysing this line the nucleus was stained with Hoechst as described above. After gating based on the cells (as above), our third gating strategy identified Hoechst positive cells. As the samples we are now analysing only contain non-nucleated red blood cells any cell that were not

infected with a parasite were Hoechst negative. Those infected with a parasite were Hoechst positive. As so many different 820 based lines were used in this study both the Cyan ADP and the LSR II were routinely used to analyse fluorescence in male and female gametocytes (as determined by their GFP and RFP proxies). On the Cyan ADP the Hoechst signal is excited by the Violet laser and filter set (Figure 2.6 A) while on the LSR II is it excited by the V-450 laser and filter set (Figure 2.6 B). Once the distinct population of Hoechst positive cells had been gated for we analysed populations of GFP and/or RFP positive parasites which were considered male and female gametocytes respectively. On both the Cyan ADP and the LSR II RFP is excited and detected by the PE filter channel and GFP is excited and detected by the FITC channel. In many cases this analysis was used to quantify the populations in each of these gates to determine gametocyte conversion and gender segregation.
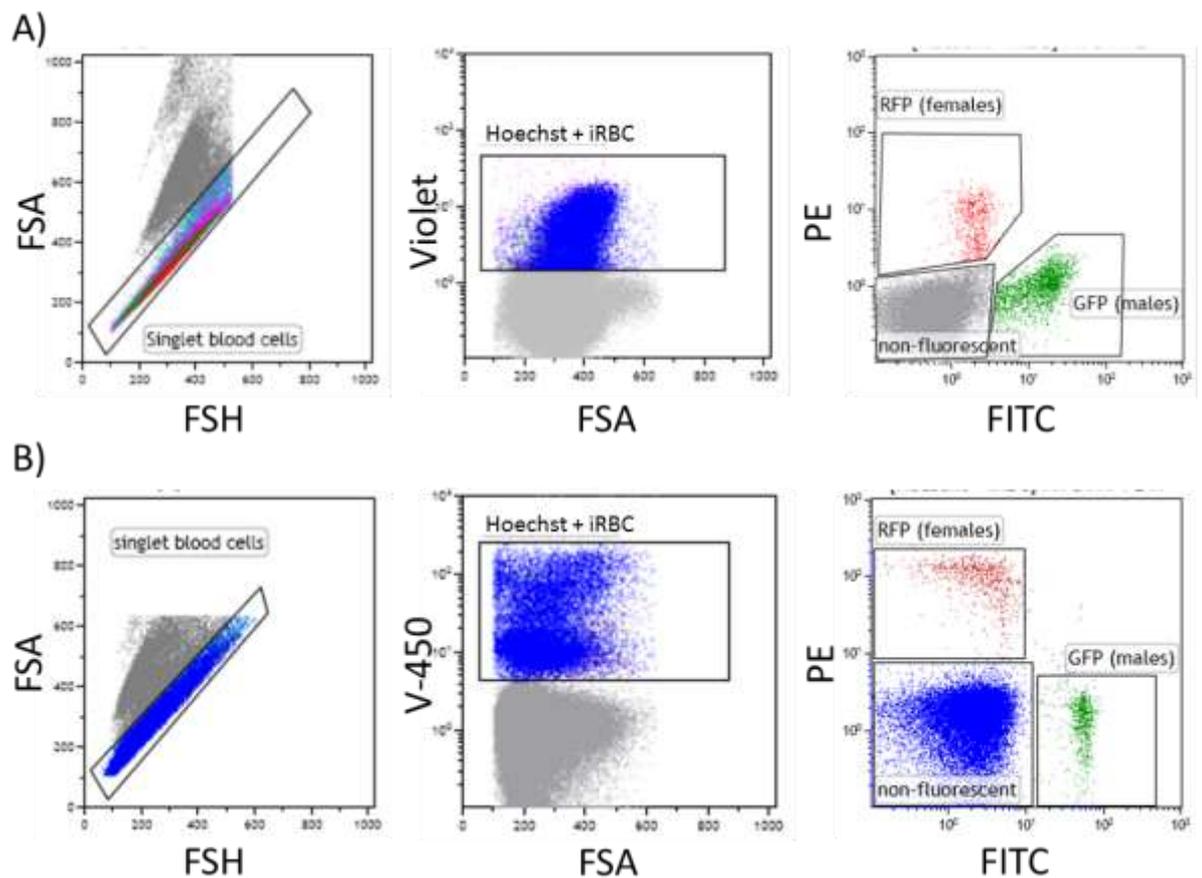
Figure 2.6 Flow cytometry analysis of the 820 parental line

Analysis on the Cyan ADP (A) and the LSRII (B). Cells are identified as singlets based on their ratio of cell diameter (FSH) and cell area (FSA) before identification of infected red blood cells by the Hoechst signal only exhibited in blood cells containing a parasite, whose nucleus was stained by Hoechst. Finally male (GFP) and female (RFP) gametocytes are identified by their fluorescence signal driven by gametocyte specific promoters PBANKA_041610 (male) and PBANKA_131950 (female).

2.2.5.3 Analysis of 820 derived lines (additional fluorescence)

Several lines were generated that not only express GFP under a male specific promoter and RFP under a female specific promoter but express CFP under the control of a promoter with an unknown expression profile. In these lines the Hoechst nuclear stain is not useable as excitation and emission of CFP and Hoechst overlap which did not allow separation of the two signals (Figure 2.7 A). Therefore a nuclear dye that does not share significant excitation with any of the fluorophores in the line was used. This was vibrant dyecycle ruby® whose excitation (Figure 2.7 B) and emission (Figure 2.7 C) did not significantly overlap with the other fluorophores.
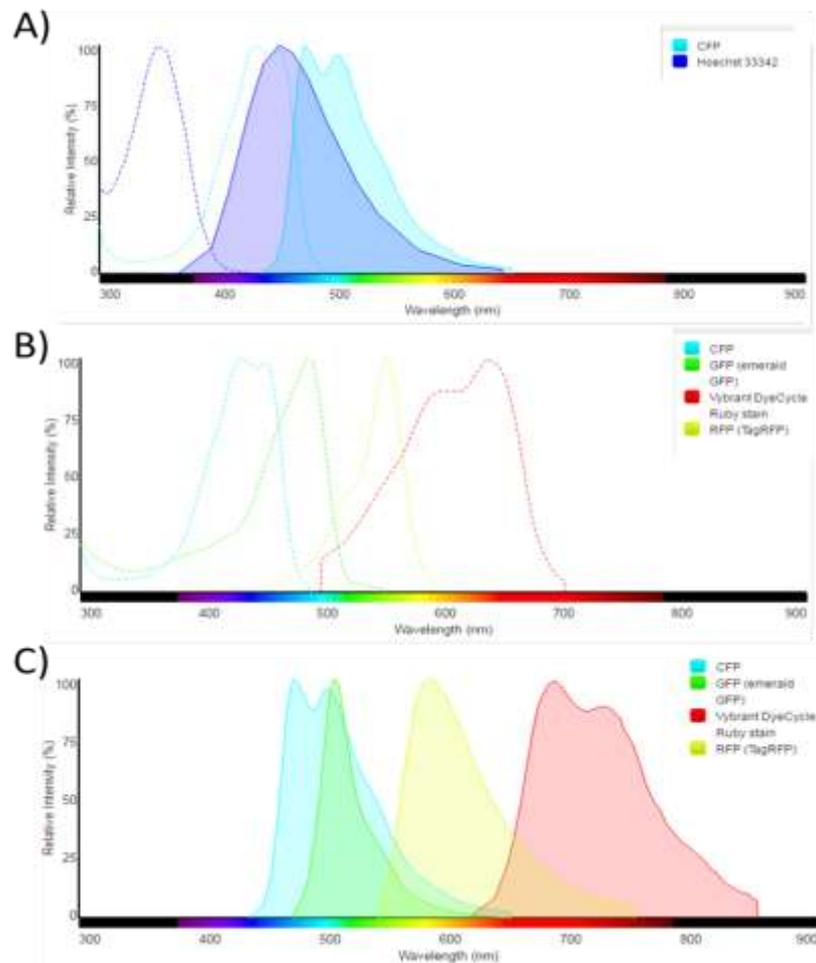
Figure 2.7 Excitation and emission spectra of potential nuclear dyes

A) Hoechst, the standard nuclear dye used in the lab shows too much overlap with the CFP in these lines to allow successful discrimination of the two fluorescence's.

B) Excitation of the three fluorophores (CFP, GFP & RFP) and the nuclear stain dyecycle ruby are very distinct utilising a multitude of laser sources.

C) The emission spectra for the fluorophores and dyecycle ruby are quite distinct, with minor overlap between CFP and GFP whose excitation occurs from different laser sources.

In these lines, samples were therefore incubated with dyecycle ruby® to stain the nucleus of parasites inside red blood cells. Analysis identified the distinct population of ruby positive cells and infected red blood cells, due to the presence of a nucleus (Figure 2.8 A). In this line the GFP (male gametocytes) and RFP (female gametocytes) were clear (Figure 2.8 B), as was a distinct CFP positive population (Figure 2.8 C). To ascertain if the CFP population was specific to gametocytes, one gender or constitutive in all stages the CFP positive population was re-plotted on the male and female gametocyte population (Figure 2.8 D). Here it became evident that a proportion of the asexual parasites (grey) were also CFP positive (blue) and that almost all female (red) and male (green) gametocytes were CFP positive as well as their characteristic green (male) and red (female). This assay was only ever carried out on the Cyan ADP.
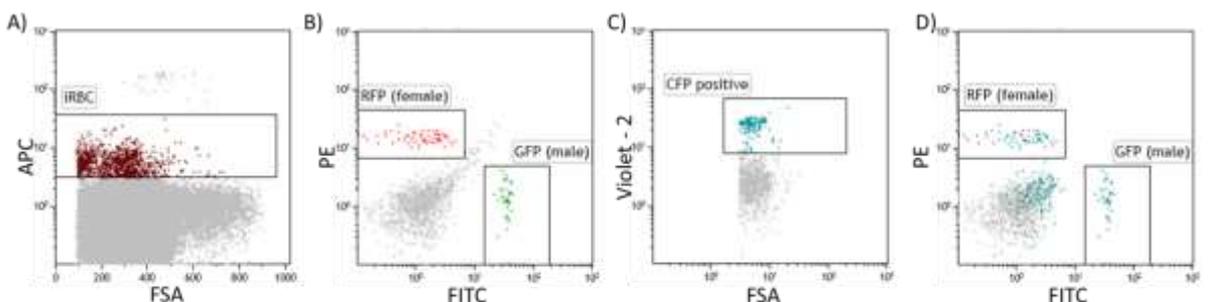
Figure 2.8 Analysis of a line expressing GFP and RFP under male and female gametocyte specific promoters (respectively) and CFP under an unknown promoter

A) To identify infected red blood cells vibrant dyecycle ruby was used to stain the nucleus of parasites. Only infected red blood cells will be positive in the APC channel.

B) In this line male (GFP) and female (RFP) gametocytes are identifiable by their fluorescence.

C) CFP positive parasites are also detectable in the population.

D) To determine is CFP expression is restricted to gametocytes, or even one gender of gametocytes the positive CFP population is back gated on the gametocyte populations based on their GFP (male) and RFP (female) signal. Here we see all gametocytes are also CFP positive and a small proportion of the asexual parasites (grey) are also CFP positive.

2.2.5.4 Analysis of single fluorescence lines

In this work three lines that express only one fluorescent protein were used. Two of these constitutively express a fluorescent protein, GFP or CFP, and one was used where only gametocytes express RFP. In the GFP and RFP lines Hoechst was used as the nuclear stain. In the GFP line, all Hoechst positive cells were also GFP positive (Figure 2.9 A). In the RFP line, all life cycle stages were Hoechst

positive and approximately 5 – 15% of the population was RFP positive, accounting for the gametocytes (Figure 2.9 C). In the CFP line, dyecycle ruby® was used as the nuclear stain. All dyecycle ruby® positive cells were also CFP positive (Figure 2.9 B). In the constitutive CFP expressing line, the variability in CFP expression was more pronounced than in the constitutive GFP expressing line.
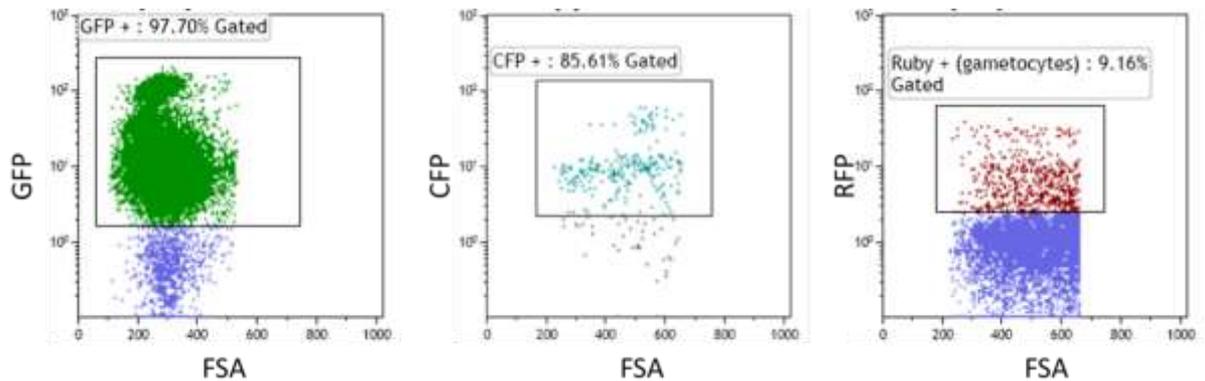
Figure 2.9 Analysis of the three single fluorescent lines used in this work

A) The constitutive GFP expressing line with Hoechst nuclear stain shows only dual positive cells.

B) The constitutive CFP expressing line with dye cycle ruby nuclear stain shows only dual positive cells, however high variability in CFP intensity is evident.

C) The gametocyte specific RFP expressing line with Hoechst nuclear stain shows a mixed population of Hoechst positive cells with no RFP expression (asexual parasites) and dual positive cells (gametocytes).

2.2.5.5 Analysis of fluorescence switching lines

In this work we needed to analyse the rate and efficiency of a switch in fluorescent protein expression. These samples were stained with the Hoechst nuclear stain as described above and gated based on size and granularity, for singlets only and for infected red blood cells based on this Hoechst stain (All described in section 2.2.5.1).   For unbiased analysis of fluorescent protein switching it was necessary to identify three different fluorescent populations within a sample. Firstly, those expressing only the original fluorescent protein, in this case GFP. Secondly, those expressing only the switched fluorescent protein, in this case RFP. Thirdly, the intermediate population whose fluorescence contains both fluorescent proteins (GFP and RFP). As an intermediate population containing both RFP and GFP was only present in the experimental samples it was essential to clearly identify the single fluorescent

populations prior to analysis of the experimental lines. To this end we gated, before the switching experiment, for three different populations, a non-fluorescent line (HP), a constitutive GFP expressing line with similar intensity to the switchable line (507) and a constitutive RFP expressing line with similar intensity to the switched line (733) to clearly identify the populations (Figure 2.10 A). After identifying these distinct, individual fluorescent lines analysis of switching was completed over time. It is clear to see over time that the populations in the switching line move from a purely GFP positive population, to a mixed fluorescence population then in to an almost exclusively RFP population (Figure 2.10 B). The prior and independent gating strategy allowed discrimination of the populations without bias.
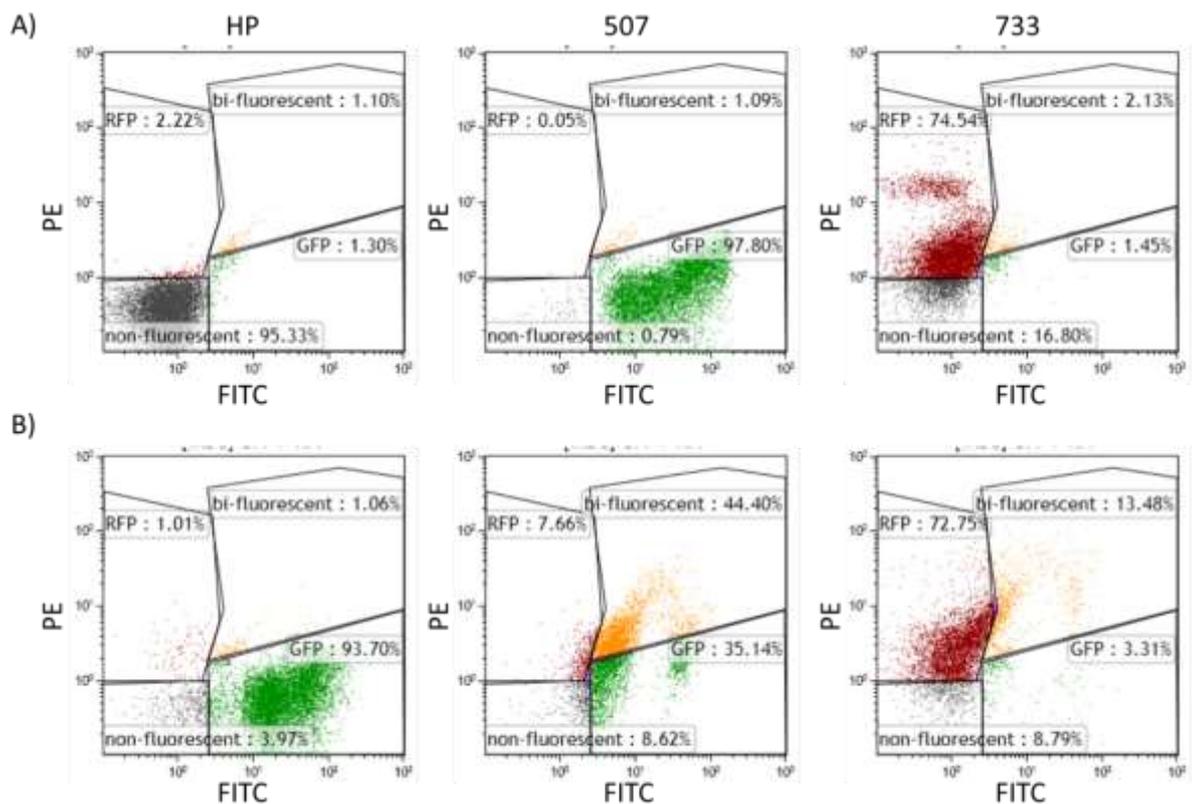
Figure 2.10 Unbiased gating strategy and analysis of the switching line.

All samples were initially gated on their size and granularity, followed by doublet discrimination. Infected red blood cells were identified by their Hoechst nuclear stain.

A) Each of the control lines (HP, 507 and 733) were gated based on their fluorescence profiles. HP is a non-fluorescent line. 507 is a constitutively GFP expressing linr, with expression driven by a promoter of similar intensity as the switching line. 733 is a constitutively RFP expressing line, with expression driven by a promoter of similar intensity to the switched line. Gating strategy was completed independent of the switching line to ensure identification of the three distinct populations was unbiased.

B) Switching reported line analysis was completed with the same instrument settings and the same gating strategy designed from the control lines. This means that all classifications within the populations are unbiased. It is clear that the line begins as almost completely GFP expressing but over time the switch results in the dominant population being bi-fluorescent and then predominantly RFP positive.

2.2.6 Sequencing and bioinformatics methods

The actual sequencing was carried out by Katarzyna Modrzynska (KM) at the Sanger institute. All library prepared RNA samples (see section 2.2.3.3 and 2.2.4.2 for preparation method) were pooled to contain approximately 12 indexed libraries and sequenced using an Illumina HiSeq2500 system resulting in 100bp paired-end reads. Raw data was quality assessed, quality trimmed and aligned to the reference genome. Differential expression between conditions and over time was calculated by quantification of reads and normalisation based on the gene size within the genome.

2.2.6.1   Quality control (FastQC)

The report generated from running FastQC analyses samples in a multitude of ways, only some of these are pertinent for RNAseq samples obtained from *P. berghei*. First, the total number of sequences obtained. This indicated the number of reads that were generated from the sequencing experiment when run on raw samples, or the number of samples retained when run on trimmed samples (see section 2.2.6.2). Next, the per base sequence score which was given as a Phred score. The Phred score gives the base call accuracy, a Phred score of 20 (the cut off for this work) indicated accuracy of 99%, any higher score improves upon this base call accuracy. Secondly, the per tile sequence quality which indicated if there were any problems with the entire cell or tile used for the sequencing. Of particular importance are regions with poor quality reads associated. These may arise from poor cell loading and the introduction of air bubbles. The Phred score was next summarised in the per sequence quality score. This, unlike the per base sequence score which gave call accuracy for each base, gives the mean quality for the entire 100bp sequence obtained. For *P. berghei* sequences the per base sequence content analysis is always considered inconsistent with standards. This is not due to sample preparation or quality, merely the fact that the genome of *P. berghei* is AT rich when compared to other organisms. Likewise with the per sequence GC content which is lower than standard samples for *P. berghei* sequences. Next, the per base N content is checked. This indicated any bases that could not be accurately called by the sequencer. Though at the end of the sequencing reads it is feasible that the sequencer cannot determine a base accurately higher than 5% N content would

indicate a failure of the sequencer to accurately identify bases. To ensure that that the pipeline of sequencing was generating reads of the length expected the sequence length distribution is calculated. For all samples pre-trimming this should be the same length as your library preparation, 100bp. For samples post-trimming the sequence length was not smaller than 20bp as these were discarded at the trimming step (see section 2.2.6.2) and ideal samples will be as close to 100bp as possible, indicating no bases were removed because of low quality (see section 2.2.6.2). In all samples the sequence duplication levels were indicated as problematic. This is because FastQC is designed for genomic sequencing as well as RNAseq. In genomic sequencing high duplication indicated the presence of PCR amplification bias in the library preparation step. In RNAseq analysis this is to be expected as high abundance transcripts make up a large amount of the sequenced reads. When sequencing transcripts you expect a diverse range of fragments aligning to the entire genome. It would be unusual for any single sequence to make up even a low percentage of the entire sequenced population. If any sequence made up more than 0.1% of the total sequenced reads that sequence was identified. If this has been flagged it could have been due to sample amplification bias or could be biologically relevant. Finally, the sequenced reads are analysed for the presence of known adapters (Adapter content) and 7 base pair sequences deemed overrepresented in the reads (Kmer content). While this measure was useful for the identification of adapters the seven base pair repeats were not relevant. In *P. berghei* repetitive regions are quite common and in RNAseq this repetitive nature can be exaggerated if the gene is highly transcribed. Therefore this measure, and the consistent error messages associated are not considered pertinent (Bioinformatics, 2011).

2.2.6.2 Quality trimming

While FastQC allowed us to visualise the quality of the datasets generated (raw files) it didn't remove reads or bases that were not of high enough quality to be used in the analysis. To do this quality trimming was employed in the form of Trim Galore! As our samples were paired end samples each of the two files was trimmed together. All reads were first trimmed to remove any adapter sequences and any bases at the end of the sequence deemed low quality (Phred score below 20 or a 99% confidence in base assignment). Any sequence that has

been trimmed to a length of less than 20 was then discarded from subsequent analysis. Finally, any paired sample whose partner has been discarded due to low quality or short length was also discarded. This analysis was followed by a second FastQC to allow visualisation of the reads after these quality trimming steps had been performed (Krueger, 2015).

2.2.6.3 Read alignment to a reference genome

To identify which genes were represented in the samples and potentially differentially expressed the reads obtained (and retained during quality control) were mapped to the reference genome. The reference genome used in this study was the *P.berghei* ANKA genome sequence and annotation, obtained from PlasmoDB, version 10.0 (Aurrecoechea *et al.*, 2009; Hall *et al.*, 2005). Alignment of the reads was performed using HiSat2 (Kim *et al.*, 2015; Pertea *et al.*, 2016). Two additional options were used in this alignment; -k10, which dictates the maximum number of positions a read can be aligned to with the same confidence in the genome is 10; --max-intronlen 50000, which sets the maximum intron length as 50kb which allows concordant alignments only if the two pairs of a read align within this distance. Aligned reads were then indexed using the SAMtools suite (Li *et al.*, 2009).

2.2.6.4 Quantification of alignments

Once the reads were aligned to the reference genome it was necessary to quantify the reads that align to each gene and normalise these based on the size of the gene in the genome. To do this Cuffquant was used (Trapnell *et al.*, 2012). Several additional/optional arguments were used when quantifying the reads obtained with each condition (Table 2.14) to ensure accurate normalisation. Additionally, standard normalisation to calculate FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values was completed to account for relative gene size within the genome.

Code	What?	Why?
-U	multi-read correction	Weighting of multiply aligned reads. Standard = 2 alignments 0.5/0.5 weighting. This alters weighting based on each of the genes aligned abundances.
-b	frag-bias-correct	Over or Underrepresented transcripts arising from PCR amplification bias in the library steps are identified and normalised.
-library -type	Library type fr-first strand	Paired samples are aligned facing towards each other with orientation of the left hand end of one fragment is read first and the right hand end of the other strand is read first.

Table 2.14 Additional arguments used in Cufflinks quantification of reads aligned

After alignment of the reads they must be quantified for comparison. Normalisation methods above allow for correction based on multi-read alignments, library prep bias and dictate how paired end reads should be aligned.

2.2.6.5 Differential expression analysis

To normalise the entire dataset being analysed the differential expression analysis was completed on all samples and conditions. This reduced the bias that comes from sampling variability. No additional, or non-default, arguments were used when completing differential expression analysis with Cuffdiff (Trapnell *et al.*, 2012).

2.2.6.6 Analysis of gene expression

Initially analysis of gene expression looked at the overall expression profiles of the samples obtained when compared to single stage RNAseq expression data (Hall *et al.*, 2005; Otto *et al.*, 2014). To achieve this principal component analysis (PCA) was completed in R (Team, 2013). The variability in the datasets provided was analysed to identify the predominant 5 vectors that account for the variability between the individual samples provided. This analysis was completed with the FactoMineR PCA tool (Lê *et al.*, 2008). Variables and their contributions to the variation in each of the five dimensions were extracted from this analysis and PCA plots generated using ggplot2 (Kahle & Wickham, 2013). In this case two dimensional plots based on the dimensions that best characterised the variability were plotted.

Differential expression analysis was completed by extracting the FPKM values for each gene in the genome, along with its gene annotation, for each of the paired samples analysed. To identify genes differentially regulated the fold change between the non-induced and induced sample in each pair was calculated for each gene (Induced / non-induced) in excel.

Comparisons were made between genes identified as gametocyte specific from proteomics data (Khan *et al.*, 2005) and with transcript datasets obtained for specific stages (Otto *et al.*, 2014) using excel.

2.3 **Summary of lines used**

Throughout this work lines have been generated to facilitate the identification of novel gametocyte specific genes and to identify the function of several candidates. A summary of the *P. berghei* lines made is shown in Table 2.15.

Line name	Vector	Integration shown?	Clonal	made by?	Vector map location
HP GIMO	P1-GIMO	Y	Y	LUMC	Suppl. 2 A
820 GIMO	P2-GIMO	Y	Y	RC/RK	Suppl. 2 B
507 GIMO	P2-GIMO	Y	Y	RK	Suppl. 2 B
HP::DiCre	P1-DiCre	Y	Y	RC	Suppl. 2 C
820::DiCre	P2-DiCre	Y	Y	RC/RK	Suppl. 2 D
507::DiCre	P2-DiCre	N	N	RK	Suppl. 2 D
CFP::DiCre	P2-CFP	Y	Y	RK	Suppl. 2 G
CFP::DiCre	P2-CFP	Y	Y	RK	Suppl. 2 H
EarlyGam RFP::DiCre	P2-early gam RFP	Y	Y	RK	Suppl. 2 I
EarlyGam RFP::DiCre	P2-early gam RFP	Y	Y	RK	Suppl. 2 J
PBANKA_090850 KO	PBANKA_090850	Y	N	RK	Suppl. 1 A
PBANKA_101830 KO	PBANKA_101830	N	N	RK	Suppl. 1 B
PBANKA_121830 KO	PBANKA_121830	N	N	RK	Suppl. 1D
PBANKA_070470 KO	PBANKA_070470	N	N	RK	Suppl. 1E
PBANKA_142920 KO	PBANKA_142920	N	N	RK	Suppl. 1G
PBANKA_081180 KO	PBANKA_081180	Y	N	RK	Suppl. 1H
PBANKA_093360 KO	PBANKA_093360	N	N	RK	Suppl. 1I
PBANKA_021390 KO	PBANKA_021390	Y	N	RK	Suppl. 1J
PBANKA_111500 KO	PBANKA_111500	N	N	RK	Suppl. 1K
PBANKA_050830 KO	PBANKA_050830	N	N	RK	Suppl. 1L
PBANKA_133500 KO	PBANKA_133500	N	N	RK	Suppl. 1M
PBANKA_133810 KO	PBANKA_133810	Y	N	RK	Suppl. 1N
PBANKA_010510 KO	PBANKA_010510	N	N	RK	Suppl. 1P
PBANKA_122550 KO	PBANKA_122550	Y	N	RK	Suppl. 1Q
PBANKA_122990 KO	PBANKA_122990	N	N	RK	Suppl. 1R
PBANKA_121260 KO	PBANKA_121260	Y	N	RK	Suppl. 1T
PBANKA_090240 KO	PBANKA_090240	N	N	RK	Suppl. 1V
PBANKA_112510 KO	PBANKA_112510	Y	N	RK	Suppl. 1W
PBANKA_140430 KO	PBANKA_140430	N	N	RK	Suppl. 1Y
PBANKA_101830 KO	PBANKA_101830	N	N	RK	Suppl. 1AA
PBANKA_112040 KO	PBANKA_112040	Y	N	RK	Suppl. 1C
PBANKA_060120 KO	PBANKA_060120	Y	N	RK	Suppl. 1F
PBANKA_113860 KO	PBANKA_113860	N	N	RK	Suppl. 1O
PBANKA_051060 KO	PBANKA_051060	N	N	RK	Suppl. 1BB
PBANKA_050500 KO	PBANKA_050500	Y	N	RK	Suppl. 1Z
PBANKA_136040 KO	PBANKA_136040	N	N	RK	Suppl. 1S
PBANKA_082800 KO	PBANKA_082800	N	N	RK	Suppl. 1U
PBANKA_142930 KO	PBANKA_142930	N	N	RK	Suppl. 1X
PBANKA_050440 KO	PBANKA_050440	Y	Y	RK	Suppl. 1CC
PBANKA_080720 KO	PBANKA_080720	Y	Y	RK	Suppl. 1DD
PBANKA_130810 KO	PBANKA_130810	N	N	RK	Suppl. 1FF
PBANKA_031270 KO	PBANKA_031270	N	Y	RK	Suppl. 1EE
PBANKA_083040 KO	PBANKA_083040	Y	N	RK	Suppl. 1GG

Table 2.15 Lines generated and used in this work
RK = Robyn Kent, LUMC = Leiden University Medical center, RC = Rachael Cameron

Chapter 3   Establishment of a DiCre system in *P. berghei*

3.1 Introduction

To determine an appropriate inducible expression system in *P. berghei*, successful methods from other apicomplexa were compared (see section 1.7 for full details) and the dimerisable Cre recombinase system was chosen for adaption (Andenmatten *et al.*, 2013; Collins *et al.*, 2013; Jullien *et al.*, 2003). Briefly, parental parasite lines constitutively express two inactive Cre fragments, CRE59 and CRE60, that are fused to rapamycin binding proteins, FKBP and FRB respectively. Upon the addition of rapamycin, the rapamycin binding proteins heterodimerise and bring together the subunits of Cre reconstituting the enzyme and restoring recombinase activity (Figure 3.1 A). The Cre recombinase recognises LoxP sites and acts to excise or invert the DNA contained between these sites (Figure 3.1 B & C). In multiple systems, including other apicomplexan, LoxP sites arranged in the same orientation have been used to flank a gene of interest. Upon rapamycin addition the recombinase becomes active, recognising the LoxP sites resulting in excision (Figure 3.1 B). In various apicomplexans complete gene deletion has been achieved with this method (Andenmatten *et al.*, 2013; Collins *et al.*, 2013; Jullien *et al.*, 2003). If LoxP sites flank a gene of interest the flanked DNA will invert upon recombinase activity. Unmutated LoxP sites will continually invert in the presence of the Cre recombinase and not fix the inversion within the genome. Optimisation and mutations of the LoxP recognition sequences identified mutations that would allow a single inversion of the region flanked (or floxed) by LoxP sites (Albert *et al.*, 1995; Zhang & Lutz, 2002). This offers great benefit to inversions with unmutated sites that can occur continually. Two mutated LoxP sites, Lox66 and Lox71, each contain one mutation but are able to recombine as a pair. When they recombine they form a non-mutated LoxP site and a double mutated Lox72 site in the regions flanking the region of the genome where they were inserted. These sites (LoxP and Lox72) are incompatible and cannot recombine again, thus preventing further gene inversions. Once this inversion has been achieved, with the addition of rapamycin and activation of the recombinase, the inversion of the genome occurs and remains inverted.
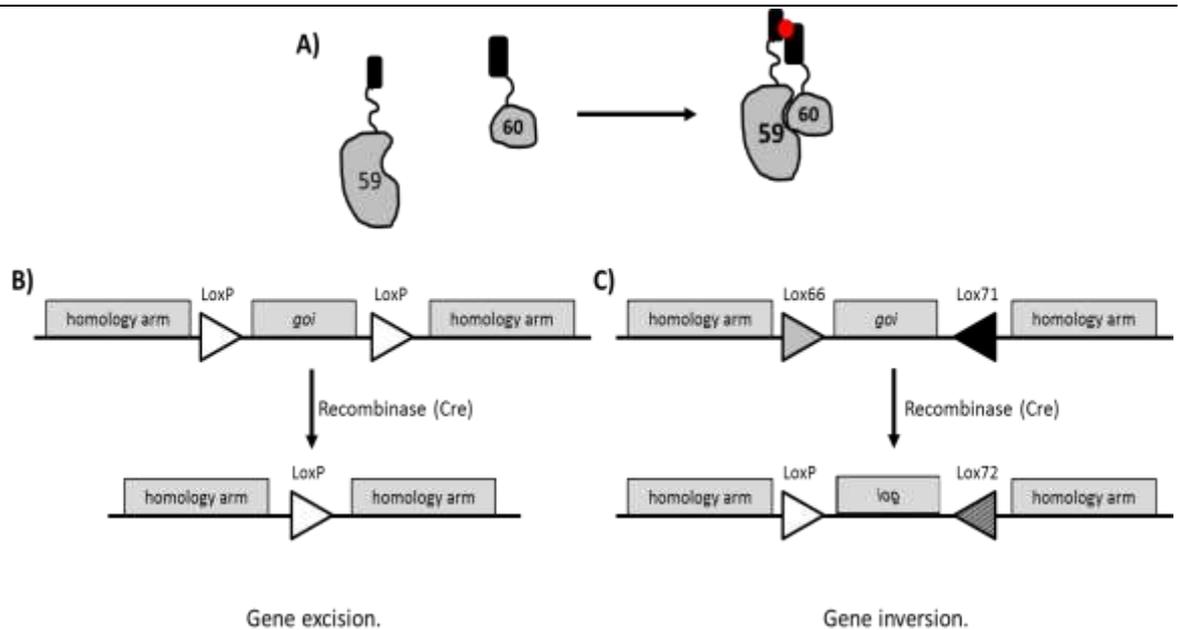
Figure 3.1 schematic representation of DiCre activity

A) Schematic representation of the recombination of the two inactive Cre fragments, CRE59 and CRE60 (59 and 60) that are bound to the Rapamycin binding proteins FKBP and FRB respectively (black boxes). In the absence of rapamycin the fragments remain separate and inactive, upon addition of rapamycin (red circle) the recombinase heterodimerises and regains its activity.

B) The recombinase recognises LoxP sites in the same orientation flanking a gene of interest (goi), resulting in excision of the contained DNA and leaving one LoxP site in situ.

C) If LoxP sites are oriented in opposite orientations surrounding a goi the flanked gene will be reversed upon recombinase activity. If these LoxP sites are the mutated LoxP sites Lox66 and Lox71 their recombination will result in an inversion and alteration in the sites forming a LoxP site and a Lox72 site. These are incompatible and will prevent further inversions occurring.

3.2 **Generation of parental DiCre lines**

To obtain parental lines that would allow for conditional gene regulation throughout the lifecycle and for multiple methods of phenotypic analyses, several lines were generated by a series of genetic modifications to existing *P. berghei* lines already created.

3.2.1 Generation of GIMO (Gene In Marker Out) parasite lines

In these lines the P230p locus is utilised to input additional sequence. This locus has been utilised, without detriment to the parasites, to constitutively express fluorescence throughout the life cycle. Two of the three standard parasite lines used already contain additional genetic modifications in this region (Figure 3.2). The 507 line contains a *gfp* cassette with expression driven by the Ef1α constitutive promoter (PBANKA_113330). The 820 line contains a *gfp* cassette driven by a male specific promoter (PBANKA_0416100) and an *rfp* cassette driven

by a female specific promoter (PBANKA_1319500). Both of these lines contain their alteration in the P1 region of the P230p locus. The final line, HP, contains no modification in the P230p locus and is effectively a wild type parasite.
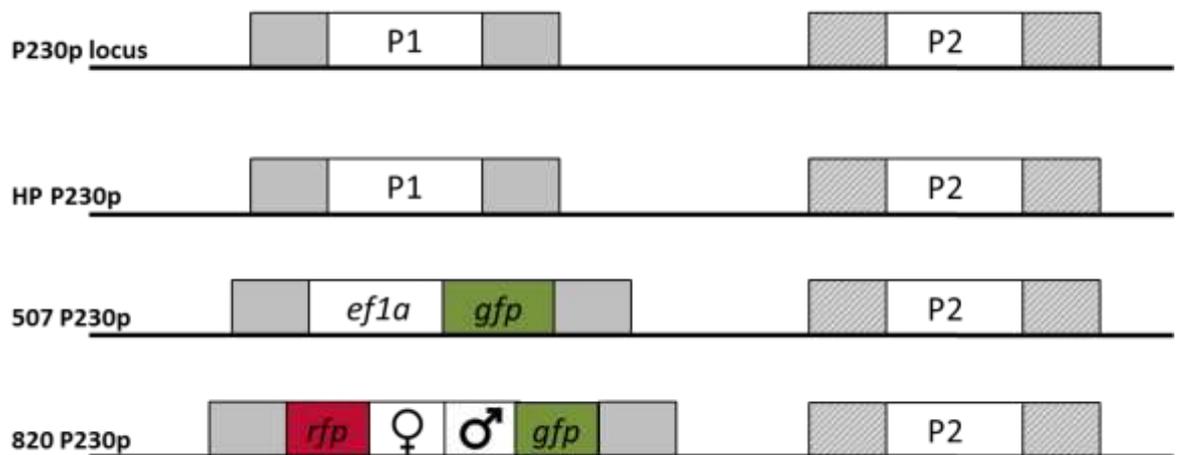
Figure 3.2 Schematic overview of the P230p locus in standard parasite lines used.

The P230p locus contains two regions used for homologous recombination in this work. Existing strains 507 and 820 contain existing modifications in the P1 region. In all lines the P2 region is unaltered.

The previously described Gene In Marker Out (GIMO) strategy of genetic manipulation was used to generate intermediate parasite lines. After integration or the mutation of interest these lines will contain no selectable marker (see section 1.6.2 for full details, (Lin *et al.*, 2011)). Three standard parasite lines were transfected to uptake a GIMO cassette (vector with homology to P230p region P2 generated by Rachael Cameron (RC)), in the previously identified redundant P230p locus (Figure 3.2 A). The transfected cassette contains two homology arms which correspond to either of the P230p locus regions, P1 or P2 flanking an hDHFR::yfcu cassette conferring resistance to pyrimethamine and susceptibility to 5-FC. To generate the parental lines the positive selectable cassette, *hDHFR* conferring resistance to pyrimethamine, is used. To select for subsequent manipulation, insertion of the vector of interest and loss of the GIMO cassette, the negative selectable marker, *yFCU* conferring resistance to 5-FC, was used (Lin *et al.*, 2011).

The three desired parental lines were selected to allow for full analysis and phenotyping throughout the *Plasmodium* life cycle. Firstly, the HP::GIMO line was generated. This essentially WT parasite lab strain had not been previously manipulated and the GIMO cassette was introduced into the P230p locus at a

region we refer to as P1 (Figure 3.3 A, parasite line obtained from Leiden University, A P Waters & C Janse). Secondly, the 507::GIMO line that already contains a constitutively expressed GFP cassette in the P1 region of the P230p locus ((Franke-Fayard *et al.*, 2004) Figure 3.3 A) where the GIMO cassette was integrated into the P2 region of the P230p locus (vector adapted by RC). Finally, the 820::GIMO line which contains a GFP cassette driven by a male gametocyte specific promoter (PBANKA_0416100) and a RFP cassette driven by a female specific promoter (PBANKA_1319500) in the P1 region, integrated the GIMO cassette into the P2 region (line generated with RC, (Khan *et al.*, 2005) Figure 3.3 A). After selection of the parasites with pyrimethamine, the majority of the surviving parasites should contain the GIMO cassette. Integration of the GIMO construct in each of the three lines was readily detectable by PCR. Primers, specific to the native P230p loci (p1 & p2 for the P1 region and p3 & p4 for the P2 region), upstream and downstream of the homology arms used for integration were designed in conjunction with primers specific to the ef1α promoter used to drive hDHFR in the P230p GIMO locus (5' GIMO Integration, primer p5) or the yfcu gene in this same locus (3' GIMO Integration, primer p6). In populations after selection, both the wild type locus and the P230p GIMO locus were detectable. This was demonstrated by amplification of both the WT locus and the GIMO locus as well as amplification of both the 5' and the 3' integration products. Post-cloning only the desired P230p GIMO locus was amplified by PCR. Amplification of the 5' and 3' integration products was also maintained in the clonal line (Figure 3.3 A & B).
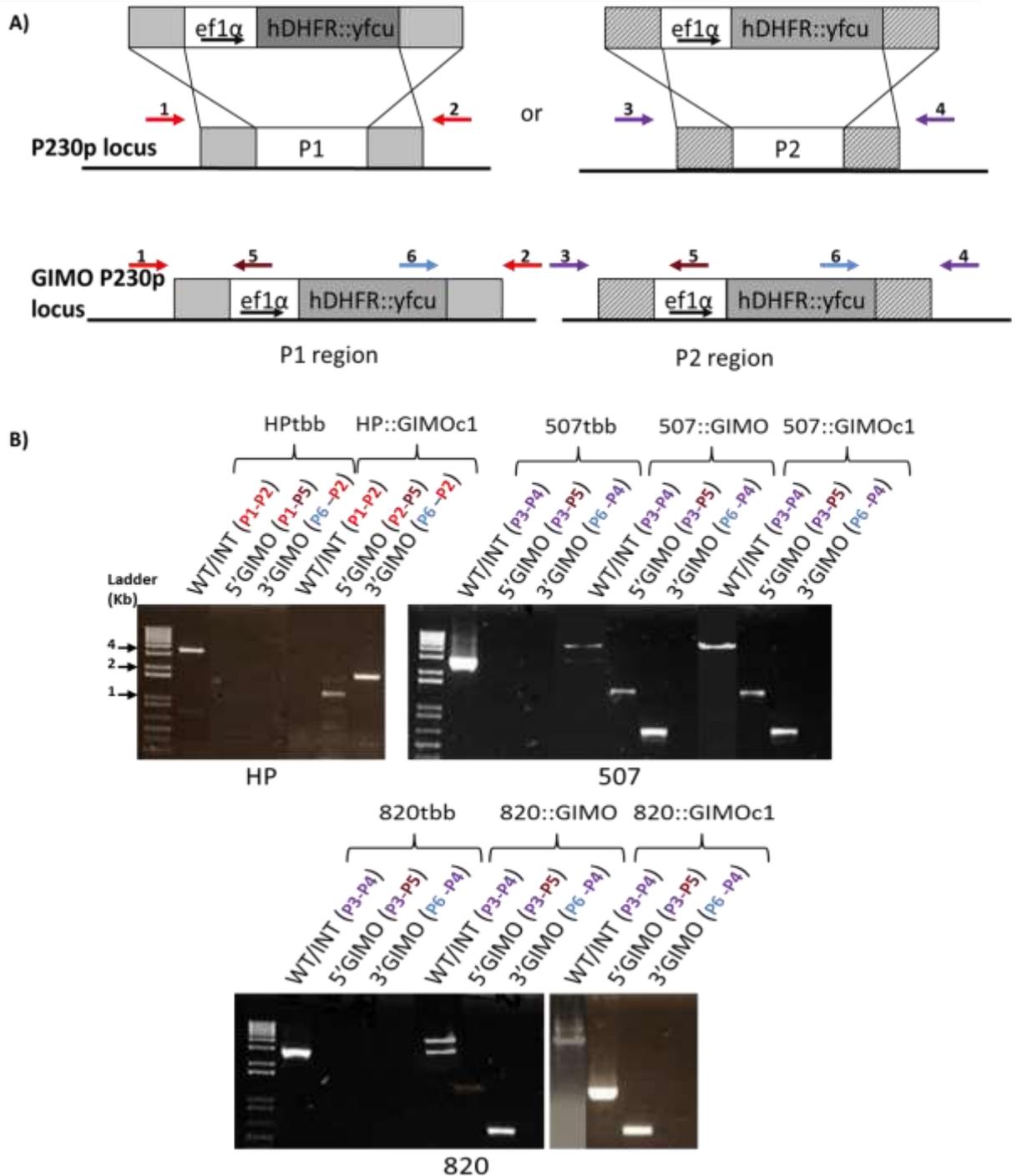
Figure 3.3 GIMO integration into the P230p locus

A) Schematics show integration of the GIMO vector (top panel) into the two regions of the P230p locus (P1 and P2). To determine the proportion of WT locus compared to the desired GIMO locus, amplification with primers external to the vector were used; red p1 & p2 for P1 and purple p3 & p4 for P2. In the WT line a fragment of 3.7kb (P1, HP) or 2.9kb (P2, 507 and 820) is amplified. For the GIMO locus the products would be 5.4kb (P1, HP::GIMO) or 4.5kb (P2, 507::GIMO and 820) respectively. 5' and 3' integration for the P1 region was demonstrated with primers p1 & p5 and p6 & p2 amplifying fragments of 1.3kb and 1.9kb respectively. Likewise, 5' and 3' integration for the P2 region was demonstrated with primers p3 & p5 and p6 & p4 amplifying fragments of 1.5kb and 0.7kb respectively.

B) PCR analysis of the parasite strains (WT HP, 507 and 820) used for transfections shows amplification of only the WT locus, with no 5' or 3' integrations. Amplifications with WT/INT primers on the lines after transfection (507::GIMO and 820::GIMO), show that the population is a mixture of the WT and GIMO loci. Clones used for subsequent experiments or manipulations (HP::GIMOc1, 507::GIMOc1 and 820::GIMOc1) show only the GIMO locus.

3.2.2 Generation of DiCre parasite lines

All three GIMO parental lines generated were then transfected with either the P230p P1 or P2 linearised plasmids containing the two inactive Cre recombinase subunits, *cre59* and *cre60*, fused to rapamycin binding domains, *fkbp* and *frb* respectively (Andenmatten *et al.*, 2013; Collins *et al.*, 2013; Jullien *et al.*, 2003), driven by the bidirectional *ef1a* constitutive promoter (Figure 3.4 A, both vectors generated by RC). Double homologous recombination occurred with the previously used (GIMO) homology regions. After negative selection with 5-fluorocytosine (Orr *et al.*, 2012), which selects for parasites that have replaced the GIMO P230p locus region (which confers sensitivity to 5-fluorocytosine) with the DiCre vector provided during transfection, the DiCre cassette was readily detectable by PCR. Again, primers (p1 and p2 for the P1 region and p3 and p4 for the P2 region) specific to the upstream and downstream regions of the P230p locus outside of the regions of homology were used for detection of wild type, GIMO or DiCre DNA. These primers were also used to detect integration when used in conjunction with new primers that recognise either the *cre59* subunit (5' DiCre Integration, primer p7) or the *cre60* subunit (3' DiCre Integration, primer p8). For all lines (HP::GIMO, 507::GIMO & 820::GIMO) no wild type P230p locus was detected. The populations post-transfection and selection with the negative selection drug 5-flurocytosine, contained both the GIMO P230p locus and the integrated DiCre P230p locus, detectable by PCR (Figure 3.4 A & B). After cloning of the HP::DiCre and 820::DiCre lines, only the integrated DiCre P230p locus was present (Figure 3.4 A & B, 820::DiCre RC, HP::DiCre line generated by RC & RK).

However, after multiple attempts a correctly integrated 507::DiCre clonal line was not obtained (Figure 3.4 A & B). Therefore an alternate strategy was attempted to further amend the HP::DiCre parental line to constitutively express GFP.
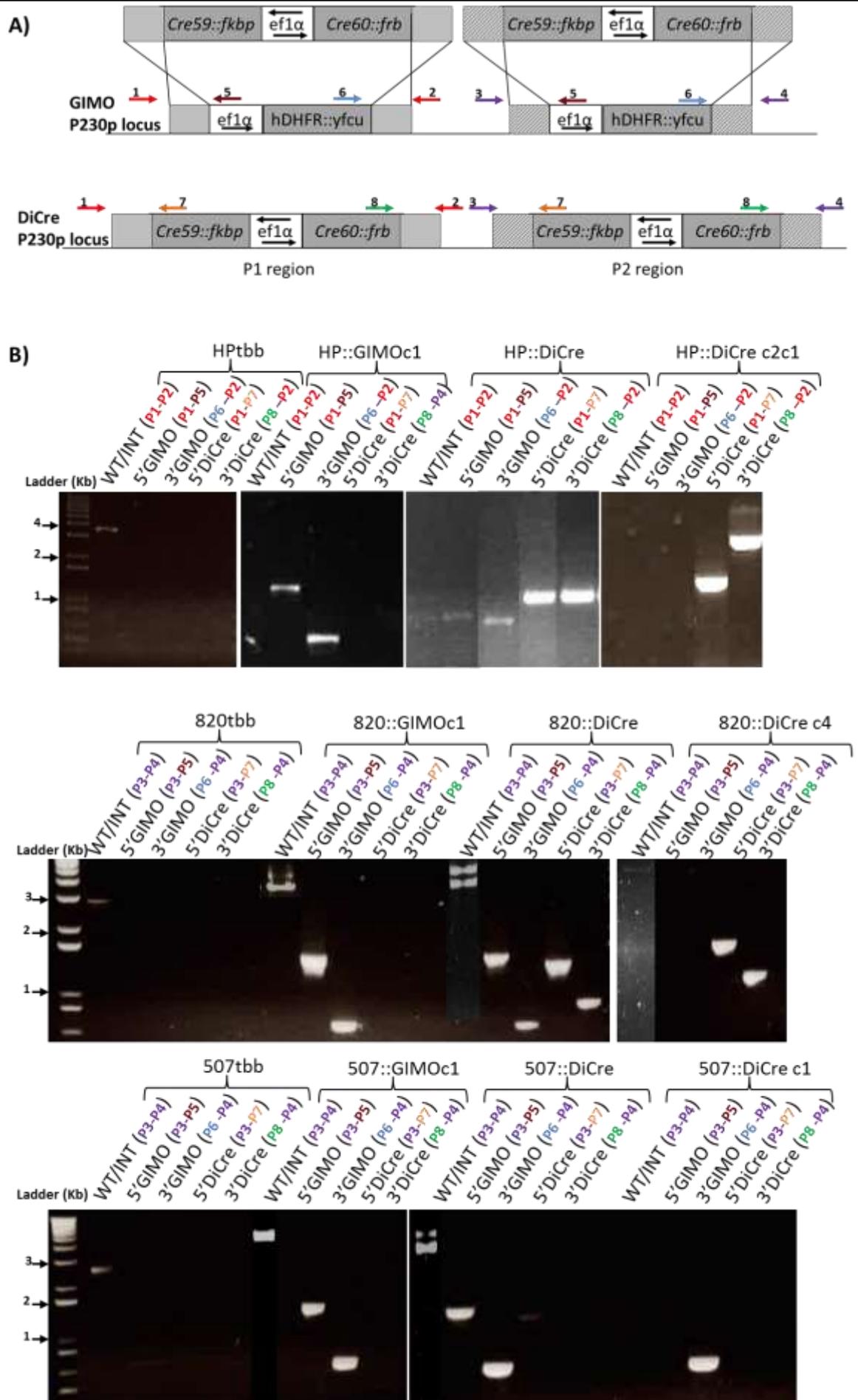
Figure 3.4 DiCre integration into the P230p locus

Figure 3.4 DiCre integration into the P230p locus

A) Schematics show integration of the DiCre vectors (top panel) into the two regions of the GIMO P230p locus (P1 and P2). To determine the proportion of GIMO locus compared to the desired DiCre locus, amplification with primers external to the homology arms were used (p1 and p2 for the P1 region and p3 and p4 for the P2 region) with the primers specific to the 5' end of the GIMO (p5) or DiCre (p7) cassettes and the 3' end of the GIMO (p6) or DiCre (p8) cassettes. 5' and 3' GIMO locus presence in the P1 region was demonstrated with primers p1 & p5 and p6 & p2 amplifying fragments of 1.3kb and 1.9kb respectively. Likewise, 5' and 3' GIMO locus presence in the P2 region was demonstrated with primers p3 & p5 and p6 & p4 amplifying fragments of 1.5kb and 0.7kb respectively. For the desired DiCre cassette integration in the P1 region, primers p1 & p7 amplify a 5' fragment of 4.5kb and primers p8 & p2 amplify a 3' fragment of 2.2kb. In the P2 region primers p3 & p7 amplify a 5' fragment of 3.3kb and primers p8 & p4 amplify a 3' fragment of 1.1kb. To ensure that no wild type locus was present in the parasites, primers 1 & 2, for the P1 region, and p3 & p4, for the P2 region, were used. If a wild type population was present fragments of 3.7kb or 2.9kb respectively would be amplified. In the P1 region the integrated GIMO locus amplified with primer p1 & p2 is 5.5kb and the DiCre locus is 6.8kb. In the P2 region the integrated GIMO locus amplified with primers p3 & p4 is 4.5kb and the DiCre locus is 5.9kb.

B) PCR analysis of the original parental strains used for GIMO transfections (HPtbb, 507tbb & 820tbb) showed amplification of only the WT locus (p1 and p1 for region P1 and p3 and p4 for region P2), with no 5' or 3' Integrations (GIMO P1 with p1 & p5 and p6 & p2; GIMO P2 with p3 & p5 and p6 & p4; DiCre P1 with p1 & p7 and p8 & p2; DiCre P2 with p3 & p7 and p8 & p4). Amplifications of the clonal GIMO parental lines showed only the GIMO locus (evident from amplification with primers p1 & p2 in region P1 and p3 and p4 in region P2). Amplifications on the lines after DiCre transfection showed no WT locus but both the GIMO locus and the DiCre locus (demonstrated by integration PCRs specific to the GIMO cassette or the DiCre cassette). After cloning the HP::DiCre and 820::DiCre lines showed only the presence of the DiCre locus. Successful 5' and 3' integration of the DiCre cassette into the the 507::GIMO line was not obtainable with a clonal line.

3.2.3 Generation of fluorescent DiCre lines

To generate constitutively expressing fluorescent DiCre parental lines, two existing vectors were transfected into the HP::DiCre line. The vectors express either GFP or CFP under the constitutive *hsp70* promoter with a *Pb48/45* 3'UTR. Upstream of the fluorescence cassette is a positive/negative selection cassette (containing both hDHFR and yfcu, conferring resistance to pyrimethamine and sensitivity to 5-FC respectively) surrounded by *Pb dhfr/ts* 3'UTRs (vectors generated by Katie Hughes (KH)). After transfection integrants were selected for with pyrimethamine. Unfortunately in the case of the integration of a constitutive GFP expression cassette no parasites were recovered from three independent transfection experiments. However, the constitutive CFP line was obtained. To confirm successful integration, primers that bind outside of the P230p homology regions (p3 and p4) were used in conjunction with primers binding to the *dhfr/ts* 3'UTR (5' Integration, primer p9) and the *48/45* 3'UTR (3' Integration, primer p10) to demonstrate integration of the fluorescent modifications. The two external primers amplify only the WT P230p P2 locus (Figure 3.5 A). After transfection both the wild type P2 region of the P230p locus and the CFP/SM cassette were detectable by PCR (Figure 3.5 A & B). Alongside this CFP fluorescence was now detected in the line by flow cytometry (Figure 3.5 C). As the WT parasites were non-fluorescent and the integrants were CFP positive isogenic cloning was utilised to obtain a population of only integrants but also reduce animal usage. In this case 50 CFP positive cells were sorted and intravenously injected into naïve mice. Flow cytometry analysis revealed the entire population was CFP positive and by PCR only the CFP/SM loci were detectable (Figure 3.5, A B & C).
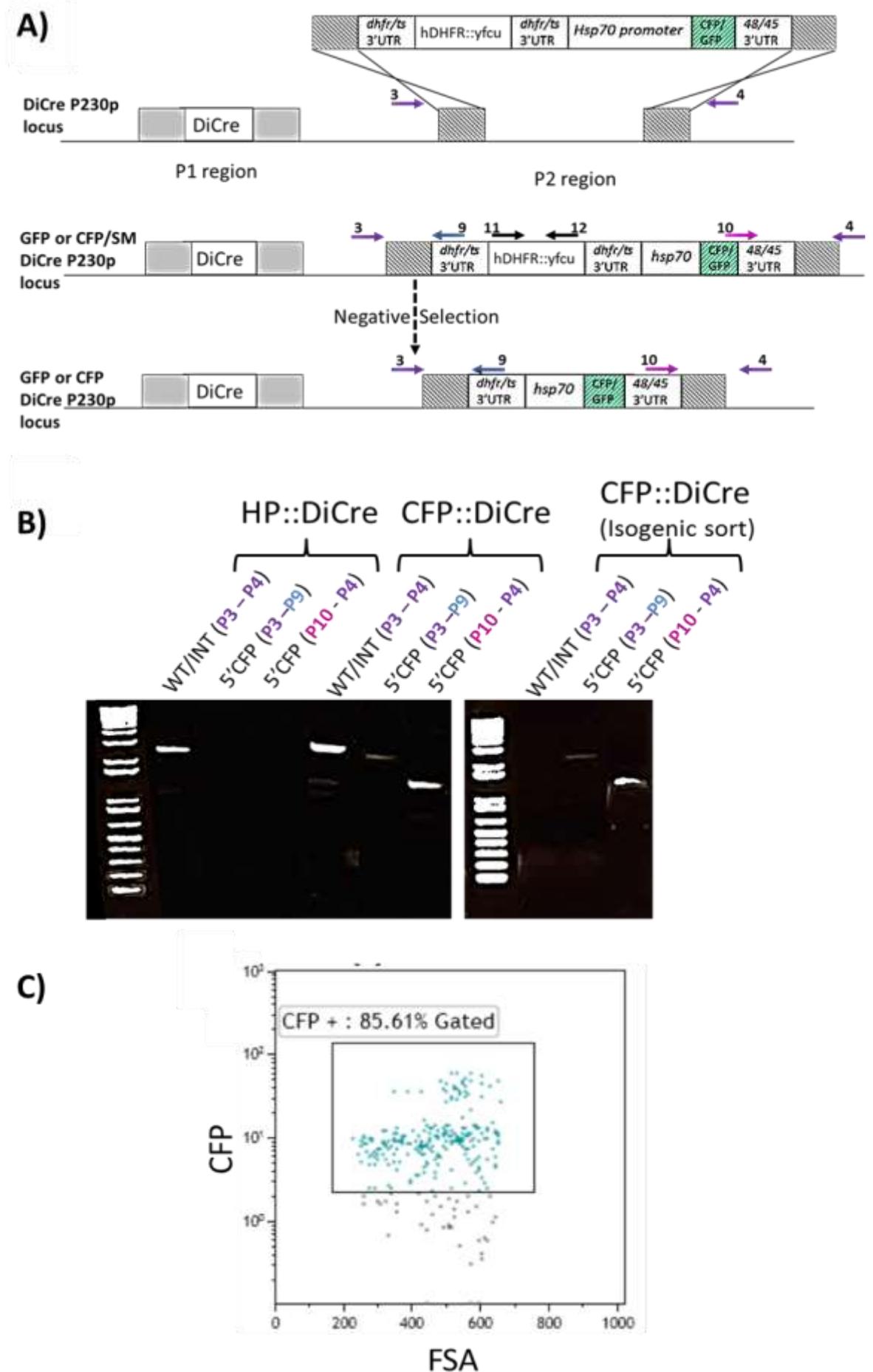
Figure 3.5 Integration of constitutive CFP into the P2 region of the HP::DiCre line.

Figure 3.5 Integration of constitutive CFP into the P2 region of the HP::DiCre line.

A) Schematics show integration of the constitutive fluorescence vectors (top panel) into the P2 region of the P230p locus in the HP::DiCre parental line. Though the GFP line was not obtained its integration would have been in to the same region. To determine the proportion of WT locus in the population, amplification with primers (purple) p3 & p4 external to the vector were used. In the WT line a fragment of 2.9kb (P2) will be amplified, for the integrated locus the products will be 7.8kb. As this product is very large the integrated product cannot be amplified, therefore the proportion of WT to Integrated (INT) could not be determined. 5' and 3' integration was demonstrated with primers p3 & p9 and p10 & p2 amplifying fragments of 2.5kb and 1.3kb respectively.

B) PCR analysis of the HP::DiCre Parental strain used for transfections shows amplification of only the WT P2 locus, with no 5' or 3' integrations. Amplification of the lines after transfection shows the population is a mixture of the WT and the integrated (CFP::DiCre) because of the presence of the WT amplicon (p3 & p4) and the 5' (p3 & p9) and 3' (p10 & p4) integration products. As the proportion of wild type to integrated could not be identified and the line was fluorescent it was sorted by FACS. After isogenic cloning the line contained no WT population demonstrated by the lack of amplification with the P3 and P4 primers.

C) Expression of the fluorescent marker (CFP) was detectable throughout the blood stages by flow cytometry analysis

To allow for subsequent manipulations to this line it will be necessary to recycle the selectable marker used to select for the integration of the CFP/SM cassette. To do this, homologous recombination between the *pb dhfr/ts* 3'UTRs surrounding the positive/negative selection cassette will be utilised (Figure 3.5 A). Applying drug pressure, in the form of 5-FC, selects for parasites that have recombined the UTRs looping out the selectable marker cassette in the process. The only remaining "footprint" of this genetic manipulation will be a single 3'UTR.

3.2.4 Generation of stage specific fluorescent DiCre parasite lines

The HP::DiCre line was subjected to another genetic manipulation to integrate a stage-specific fluorescent marker at the P2 region of the P230p locus. This cassette, (generated by Rebecca Lee (RL)) expresses RFP under an early gametocyte (EG) promoter (PBANKA_101870 (Sinha *et al.*, 2014)) with a *Pb*48/45 3'UTR. It also contains a positive/negative selection cassette (containing both hDHFR and yfcu conferring resistance to pyrimethamine and sensitivity to 5-FC respectively) surrounded by *Pb dhfr/ts* 3'UTRs (Figure 3.6 A). After transfection, integrants were selected for with pyrimethamine. In this population there was a mixture of the P230p DiCre locus with no additional genetic manipulation (amplification with primers p3 and p4) and the P230p DiCre locus with the integrated gametocyte-specific RFP cassette (5' integration with primers p3 & p9 and 3' integration with primers p10 and p4; Figure 3.6 A & B). After cloning, only the P230p locus containing DiCre and the gametocyte-specific RFP was detectable (loss of amplification product with primers p3 & p4 but amplification of 5' and 3' integration with primers p3 & p9 and p10 & p4 respectively). To ensure this line was correctly expressing RFP in gametocytes only it was analysed by flow cytometry (Figure 3.6 C) and the RFP positive population was sorted. Giemsa analysis of this RFP positive population showed that it only contained gametocytes (Figure 3.6 D). Now that we had confirmed the fluorescent signal is associated with the correct parasite stage and expression was correctly restricted the line will be negatively selected to recycle the selectable marker. Once successfully negatively selected the line will again be cloned to ensure the marker was lost in all parasites (Figure 3.6 A).
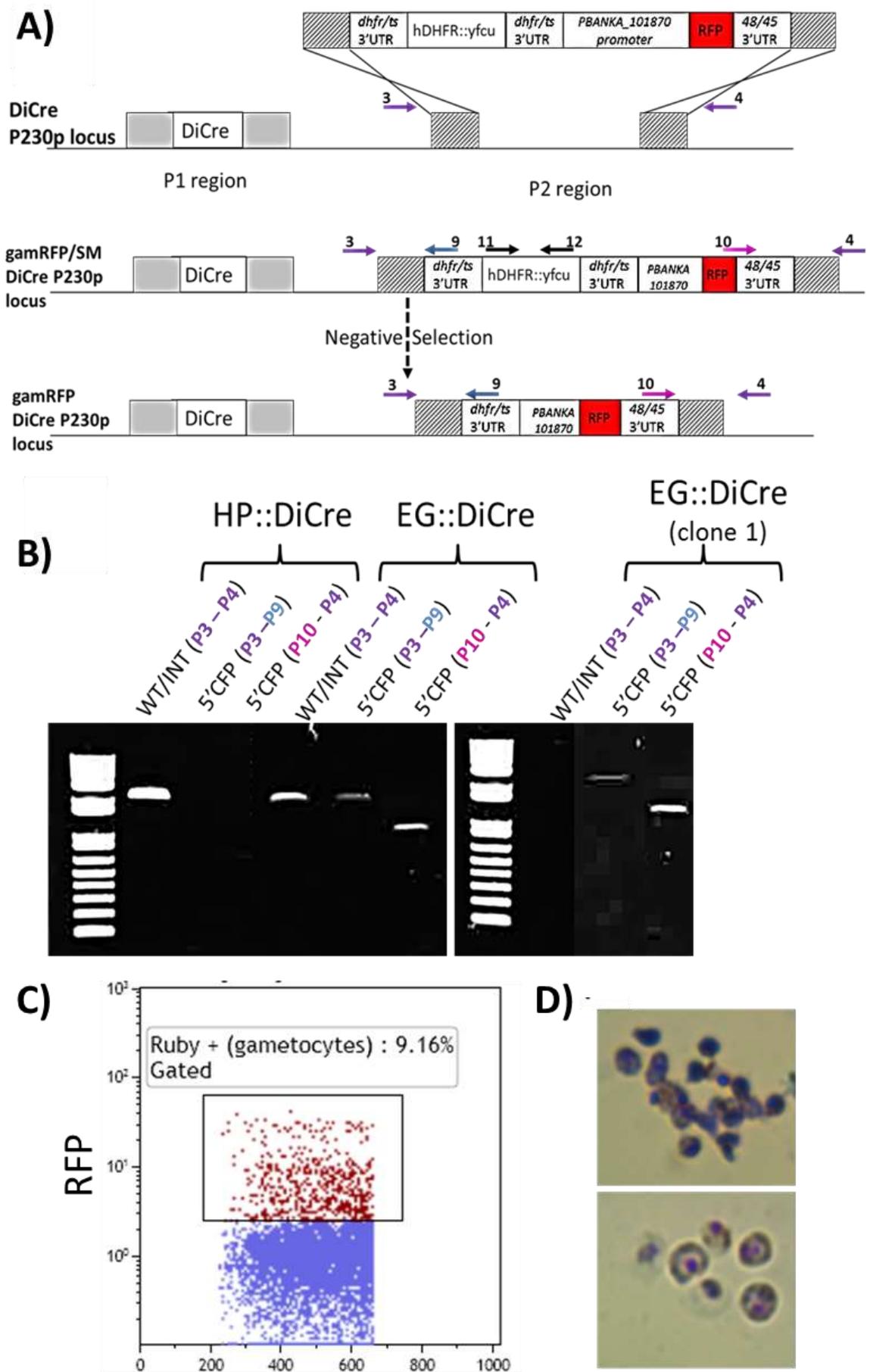
Figure 3.6 Integration of the gametocyte-specific RFP expression vector into the P2 region of the HP::DiCre line.

Figure 3.6 Integration of the gametocyte-specific RFP expression vector into the P2 region of the HP::DiCre line.

A) Schematics show integration of the gametocyte-specific RFP vector (top panel) into the P2 region of the P230p locus. To determine the proportion of WT locus compared to the desired locus amplification with primers (purple) p3 & p4 external to the vector were used. To determine the proportion of WT locus in the population, amplification with primers (purple) p3 & p4 external to the vector were used. In the WT line a fragment of 2.9kb (P2) will be amplified, for the integrated locus the products will be 8.3kb. As this product is very large the integrated product cannot be amplified, therefore the proportion of WT to Integrated (INT) could not be determined.5' and 3' integration was demonstrated with primers p3 & p9 and p10 & p4 amplifying fragments of 2.5kb and 1.3kb respectively.

B) PCR analysis of the HP::DiCre Parental strain used for transfections showed amplification of only the WT locus, with no 5' or 3' integrations. Amplification of the lines after transfection showed a mixture of the WT locus and the EG::DiCre locus. After cloning, the line used for negative selection showed only the EG::DiCre locus.

C) Flow cytometry analysis showed that a proportion of the population expresses RFP, corresponding to the 5 – 15% expected gametocytes.

D) After sorting of the RFP positive population only gametocytes are visible by giemsa smear.

Expression of the two Cre subunits was detectable in the blood stages by Western blot analysis using antibodies against the FKBP rapamycin binding domain for the Cre59 subunit and against the full Cre recombinase for the Cre60 subunit (Figure 3.7).
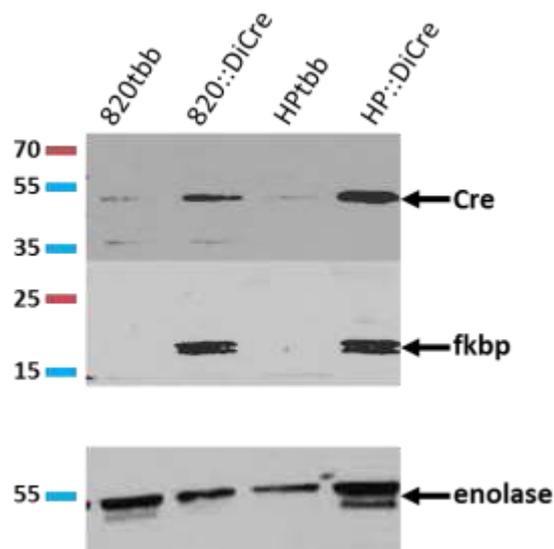
Figure 3.7 Expression of the DiCre subunits in parental parasite lines.

Western blot analysis to show expression of the two Cre recombinase fragments CRE59::FKBP, detected with an FKBP antibody and CRE60::FRB, detected with an anti CRE antibody.

These parental lines will allow for the in depth phenotyping of lines utilising the DiCre technology to conditionally disrupt or overexpress genes of interest.

3.3 **Proof of principle**

In order to establish the efficacy, efficiency and rate of activity of the DiCre system, a system was established that would allow for ready detection of excision with a fluorescent output. The Parental line used for this test was the HP::DiCre line that constitutively expresses the DiCre subunits but is not fluorescent. A vector (Figure 3.8 A, constructed by Daniel Bargieri in the laboratory of Robert Ménard, Institut Pasteur, Paris) that is episomally maintained as a single copy, due to the presence of the *pbCen5* sequence (Iwanaga *et al.*, 2010), was transfected into this line. The vector contains a floxed *gfp* cassette driven by an *hsp70* promoter and a promoterless *rfp* cassette downstream of the floxed *gfp*. To allow selection, the vector also contains an *hdhfr* cassette. Before the addition of rapamycin, parasites constitutively express GFP (under the *hsp70* promoter) which is readily detectable by

microscopy and flow cytometry (Figure 3.8 B) allowing quantification of parasites containing the vector. After gene excision by the Cre recombinase parasites constitutively express RFP under the *hsp70* promoter (Figure 3.8 A).
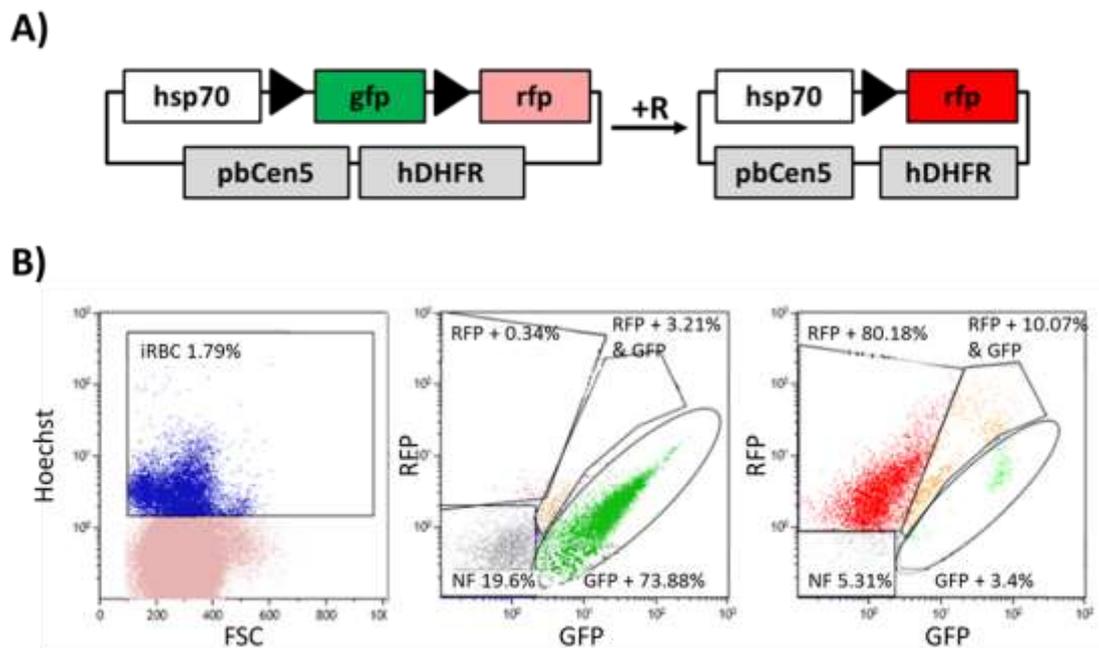
Figure 3.8 Reporter vector and flow cytometry analysis of the line

A) Schematic representation of the reporter vector. Before addition of rapamycin the HSP70 promoter is driving GFP expression. As the *gfp* cassette is flanked by loxP sites after rapamycin addition and Cre excision the *gfp* cassette is excised. This results in RFP expression from the HSP70 promoter. The vector also contains a *pbCen5* sequence which maintains the episome as a centrosome and the *hdhfr* cassette which allows for selection.

B) Analysis of the reporter through flow cytometry shows ready identification of infected red blood cells with the Hoechst nuclear stain and GFP positive parasites without induction (middle panel) and after induction the resulting parasites are predominantly RFP positive (right panel).

To establish the rate at which gene excision occurs *in vitro,* synchronous cultures were set up from ring stage parasites (blood collected 2 hours post invasion) and immediately induced with 200mM rapamycin. The percentage of parasites expressing GFP and/or RFP was quantified (full details of the unbiased gating strategy employed in section 2.2.5.5) every four hours (post induction) by flow cytometry. Over time the percentage of parasites containing GFP diminishes and the proportion with both GFP and RFP present or only RFP present increases (Figure 3.9 A). Approximately 16 hours after induction, synchronous parasites predominantly expressed RFP instead of GFP.
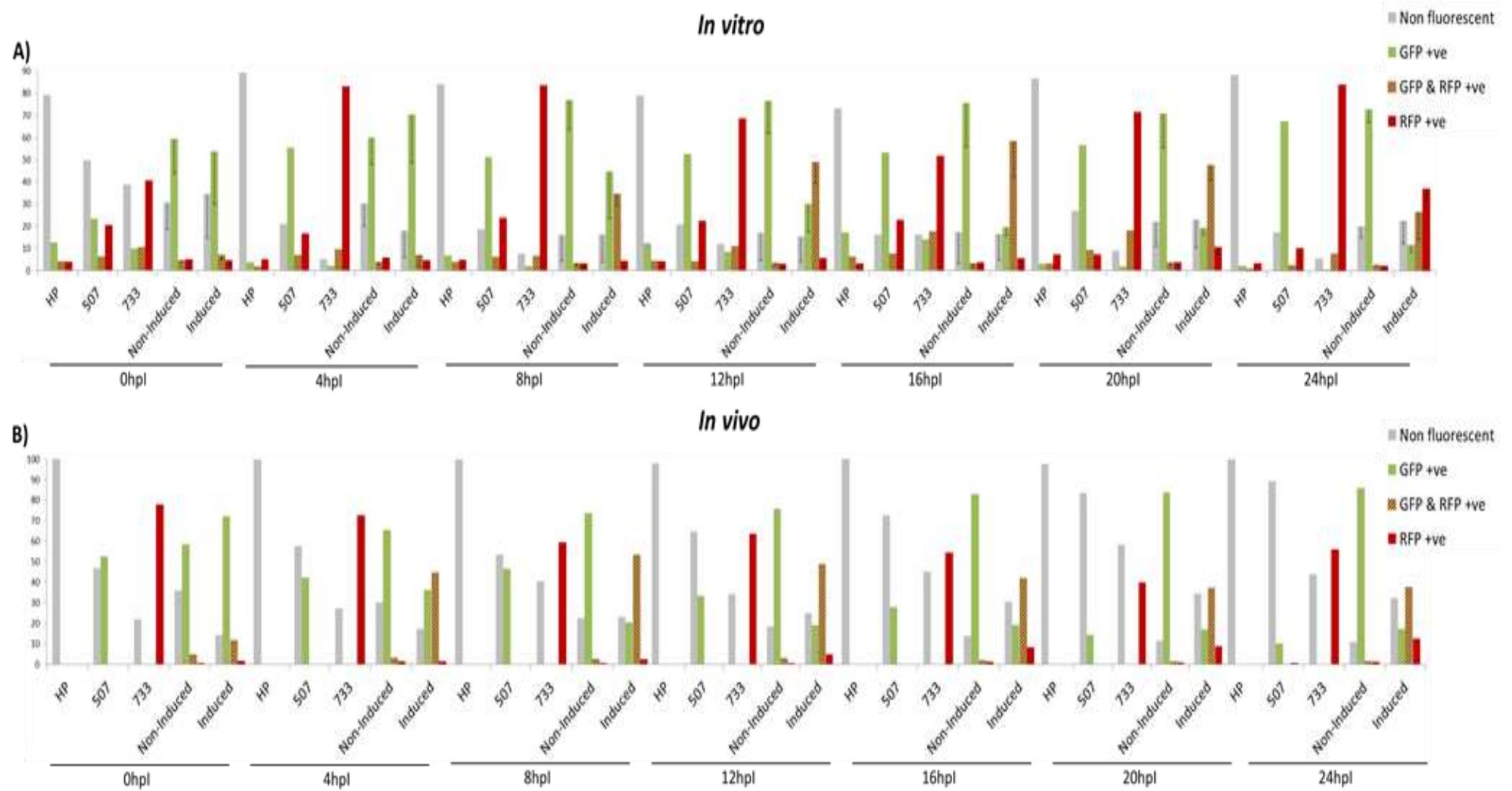
Figure 3.9 Flow cytometry analsysis of fluorescent protein switching post induction.

Figure 3.9 Flow cytometry analsysis of fluorescent protein switching post induction

A) *In vivo* addition of 200nM rapamycin to synchronous cultures (starting from 2h rings) the proportion of parasites positive for GFP, RFP or both GFP&RFP is quantified every four hours for 24 hours. Parasites are classified as non-fluorescent (grey bars), if their fluorescence intensity by flow cytometry is low; GFP positive (green bars), if only their FITC signal is positive; GFP&RFP positive (green and red striped bars), if their FITC and PE signal is positive; RFP positive (red bars), if only their R+PE signal is positive. At each time point the control lines (HP, non-fluorescent; 507, GFP positive and 733, RFP positive) and the reporter line with (I) and without (NI) rapamycin were classified according to these fluorescence criteria. Only upon addition of rapamycin (I) is there a switch from GFP expression to RFP expression in the reporter line.

B) If rapamycin is administered intraperitoneally (4mg/kg) to mice the resulting excision, determined by flow cytometry analysis (as above) of fluorescent protein expression, follows the same kinetics as demonstrated for *in vitro* addition of rapamycin

At multiple time points after induction (0 hours, 1 hour, 4 hours, 12 hours and 24 hours) genomic DNA was extracted and the proportion of parasites that had excised the *gfp* gene was quantified by qPCR using specific primers for the excised and non-excised locus (Figure 3.10 A). At each time point the proportion of the population in the non-excised state is quantified compared to the sample with the least non-excised proportionally (24hpi). Likewise the proportion of the population in the excised state is quantified compared to the sample with the least excised proportionally (NI). To account for sample variability both non-excised and excised are compared and normalised to a control primer pair on the vector that amplifies at the same efficiency in the non-induced and induced samples (Figure 3.10 B). These fold change calculations have to be carried out on these control samples as there is no way to obtain relevant samples that are 100% non-excised or 100% excised. It is clear that over time post induction the population becomes predominantly excised when compared to non-excised and the previous time points. It is also clear that the excision happens very rapidly, with some appearing to have excised at 0hpi. This is likely because the action of rapamycin occurs in the time it takes to process the samples (5 – 10 minutes). Though it seems the maximum excision is reached within 4 – 12 hours post induction (Figure 3.10 C)

It seems that the change in the fluorescence profile of the parasites, from GFP expressing to RFP expressing, has not reached proportion that represents the excision proportion in the genome by 24 hours post induction (Figure 3.9 A), which is the longest time parasites can be cultured in vitro. To assess long term fluorescent protein changes as a read-out for the genetic change we had to carry out the same experiment in vivo.
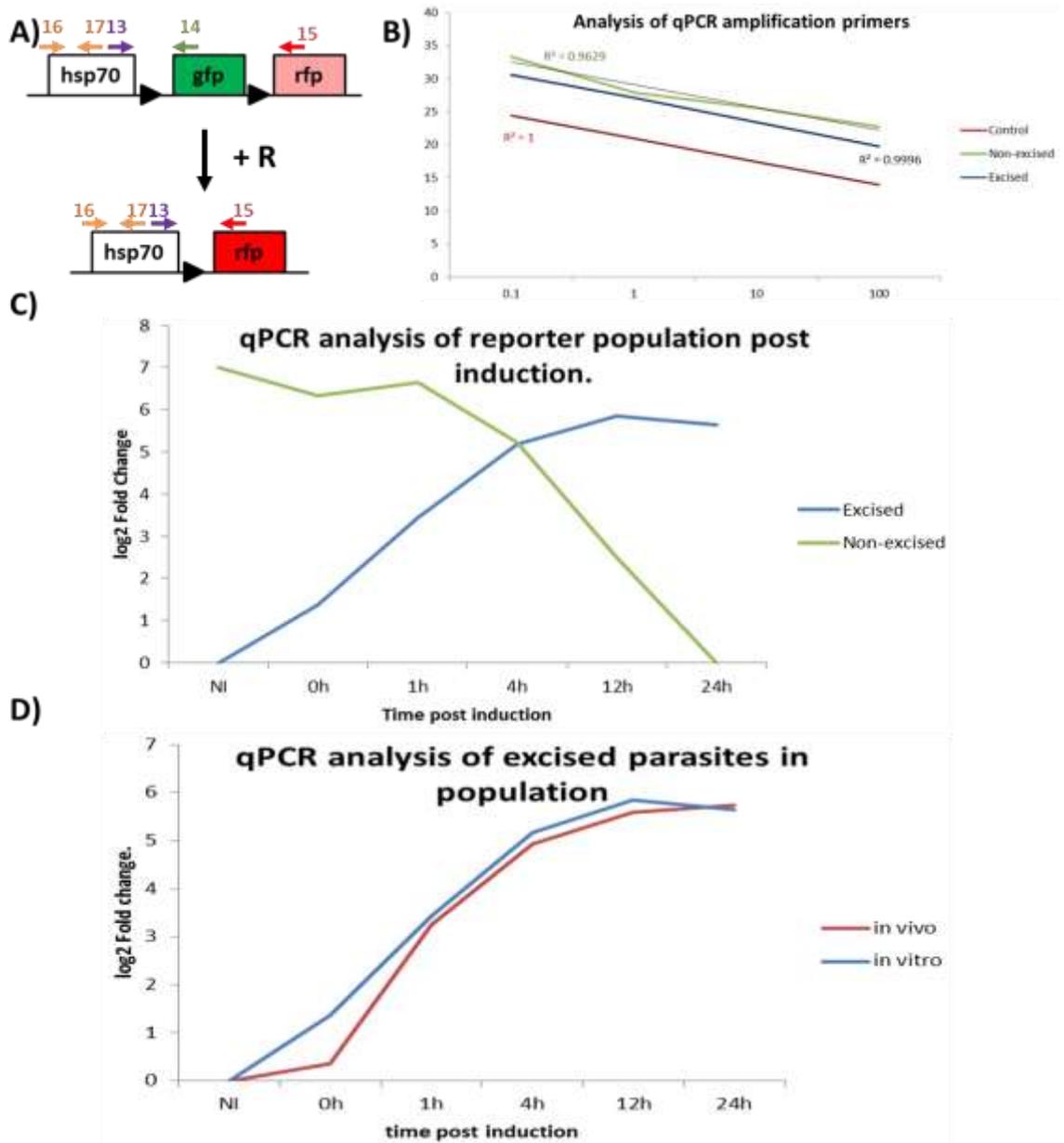
Figure 3.10 Rate of gene excision post induction *in vitro* and *in vivo*.

A) Schematic representation of the reporter plasmid pre and post induction with rapamycin (R). A forward primer (p13), specific to the HSP70 promoter was used in conjunction with a promoter that binds to the *gfp* cassette (non-excised, p14) or *rfp* cassette (excised, p15) to distinguish how much of the population is in each state. A region of the reporter was used as a control for quantification. This region will not change in the unexcised or excised populations (p16 & p17).

B) To ensure unbiased amplification the primers used for analysis were quality controlled for linear amplification.

C) *In vitro* synchronous cultures were induced with 200nM immediately after blood was harvested (2hpi). At each time point analysis was completed to determine the proportion of the population in the non-excised state and the excised state. At each time point proportions were normalised against the control gene and compared to the sample which contains the lowest proportion of the state being analysed (24hpi for non-excised and NI for excised). Over time more of the population is in the excised state and less in the non-excised.

D) To determine if excision rates are similar *in vivo* the experiment was repeated on synchronous parasites in mice. 2hpi mice were intraperitoneally injected with 4mg/kg rapamycin and samples collected at the same time points. The proportion of parasites in the flipped state was quantified by comparison with the non-induced sample and normalised to the same control gene as the *in vitro* assay. The initial excision (at 0hpi) is the only time point that shows a difference in the population ratios.

To ascertain if the rate of excision was the same *in vivo* as *in vitro,* the experiment was repeated *in vivo*. Two hours post infection of a synchronous population, mice were injected interperitoneally with 4mg/kg rapamycin resulting in induction. For the same time points as the *in vitro* assay (0h, 1h, 4h, 12h, 24h), genomic DNA was extracted and the rate of excision at the genomic level (Figure 3.10 A). *In vivo* excision, determined by qPCR, was shown to mirror that of the *in vitro* experiment (Figure 3.10 D). To determine if there were any differences in protein expression and turnover, the fluorescence of parasites was analysed by flow cytometry by sampling tail drops every four hours over a 24 hour period. The rate of change from GFP expressing parasites to RFP expressing parasites corresponds to that demonstrated *in vitro* (Figure 3.9 B).

Both *In vivo* and *In vitro* an increase in the population of parasites that had excised the *gfp* cassette was detectable by qPCR (p13 and p15, Figure 3.10 D). The unexcised locus was still detected (p13 and p14) in *in vivo* samples (Figure 3.11 C) by qPCR, however PCR amplification of the unexcised locus was not detected (Figure 3.11 B). The switch of parasite protein expression, from GFP to RFP, indicates that the excised population was more accurately estimated by PCR than qPCR.
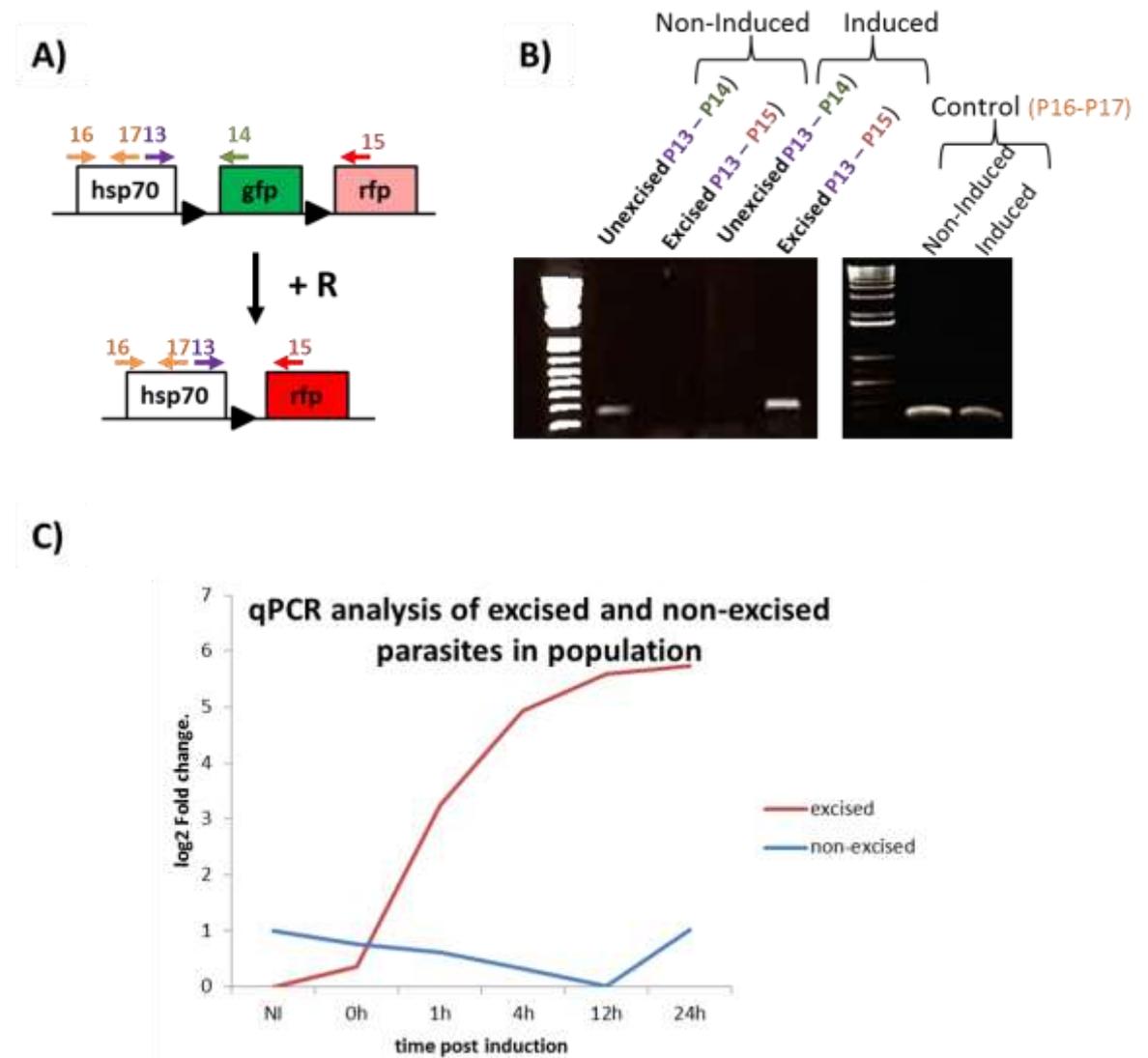
Figure 3.11 Comparison of PCR and qPCR in the reporter line.

A) Schematic representation of the primers used for the qPCR and PCR analysis. Control primers (p16 & p17) amplified in both the non-excised and excised populations. The non-excised population was amplified only from the non-induced population (p13 & p14) and the excised population was only amplified from the excised, induced population (p13 & p15).

B) PCR analysis of the primers used, carried out on genomic DNA extracted from *in vivo* samples with and without induction 24 hours post induction, showed amplification with the control primers for both samples. In the Non-Induced sample only the unexcised amplicon was detected and in the Induced sample only the excised amplicon was detected. PCR was completed on 50ng genomic DNA for 20 cycles.

C) qPCR analysis with the same primers, carried out on genomic DNA extracted from *in vivo* samples over time after induction, show an increase in the population with the excised genome makeup as expected. However, a non-excised product is being detected throughout the course of the experiment. The NI and 24h DNA samples in this qPCR are from the same preparation as the PCR (B).

For a longer term assessment of the change in protein expression a tail drop of blood (collected every 24 hours for 7 or 5 days) was analysed by flow cytometry to determine the rate at which RFP became detectable and the GFP signal was lost (Figure 3.12). It is clear that with a synchronous starting culture, excision is very rapid resulting in a detectable switch from GFP to RFP expression in 1 day post induction (Figure 3.12 A). If the starting population is non-synchronous, the switch from GFP to RFP expression can be less rapid, likely owing to the presence of different parasite stages within the asexual cycle, taking until day 2 to be clearly detectable by flow cytometry analysis of protein expression (Figure 3.12 B). Furthermore, the half-life of GFP is 26 hours therefore depletion of the protein to undetectable levels in individual parasites will not be immediate (Kitsera *et al.*, 2007).
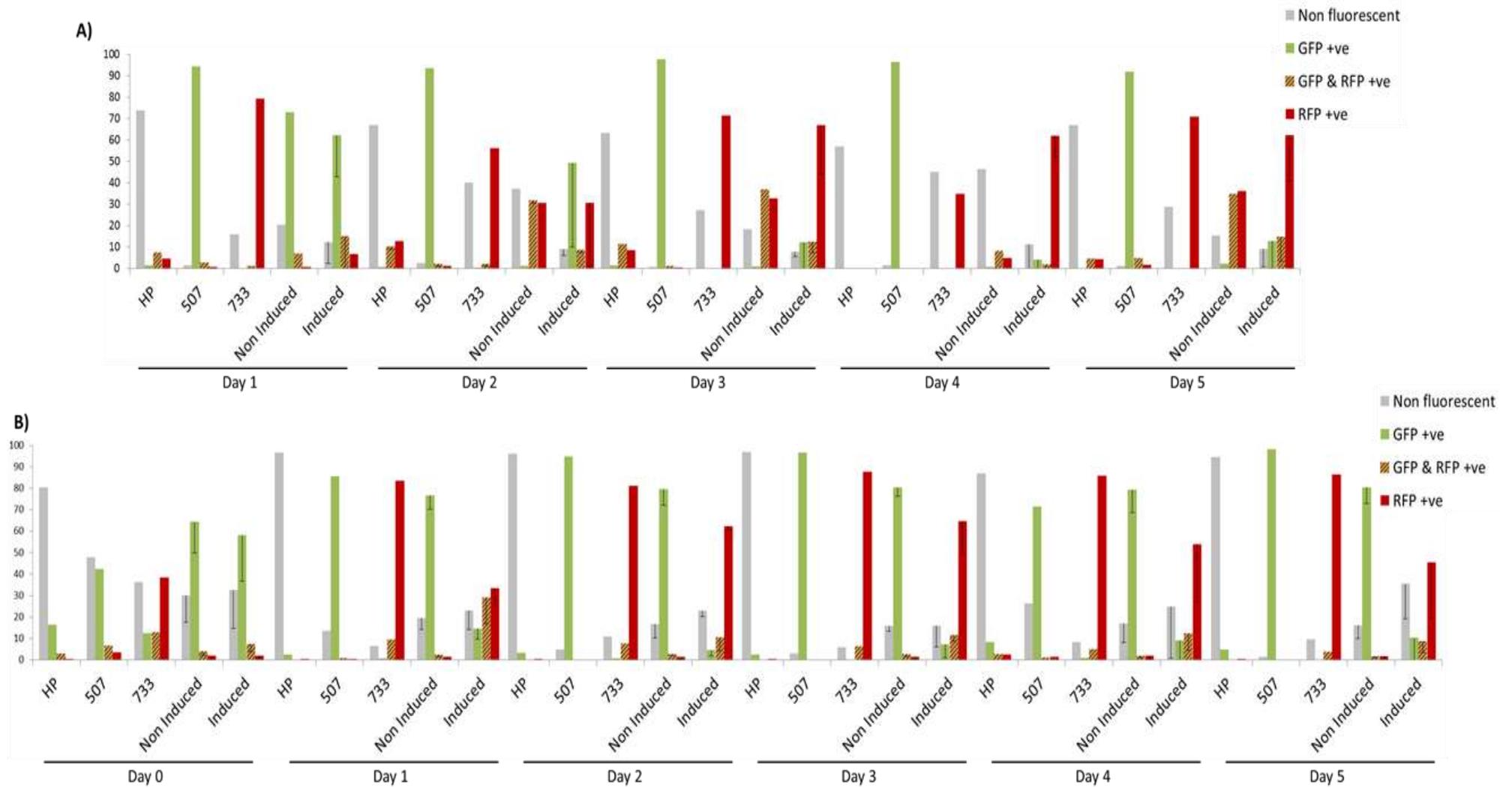
Figure 3.12 Long term flow cytometry analysis of fluorescent protein switching post induction

Figure 3.12 Long term flow cytometry analysis of fluorescent protein switching post induction

A) Flow cytometric analysis of a synchronous population of the reporter line (xxx) upon induction with 4mg/kg rapamycin. Excision, as determined by fluorescent protein expression, is very rapid and can be detected 1 day after induction

B) Flow cytometry analysis of a non- synchronous population. When a non-synchronous population is induced in the same way, the switch from GFP to RFP expression is evident on day 2 after induction, peaking on day 3.

Finally, as the non-synchronous population took multiple days to switch from a GFP expressing line to an RFP expressing line multiple inductions were applied to ascertain if additional inductions, achieved by subsequent interperitoneal injections of 4mg/kg rapamycin, improved the rate or percentage of gene excision, mice were induced once, twice or three times and the switch from GFP to RFP expression monitored by flow cytometry. No additional benefit was observed from these additional inductions (figure 3.1.10).

From these analyses we were able to demonstrate expression of the two inactive Cre subunits throughout the asexual cycle, activation of diCre by rapamycin and show that excision is both rapid and highly efficient both *in vivo* and *in vitro*. Finally, we were able to demonstrate that a single dose of 4mg/kg rapamycin was able to consistently induce rapid and efficient excision *in vivo*.
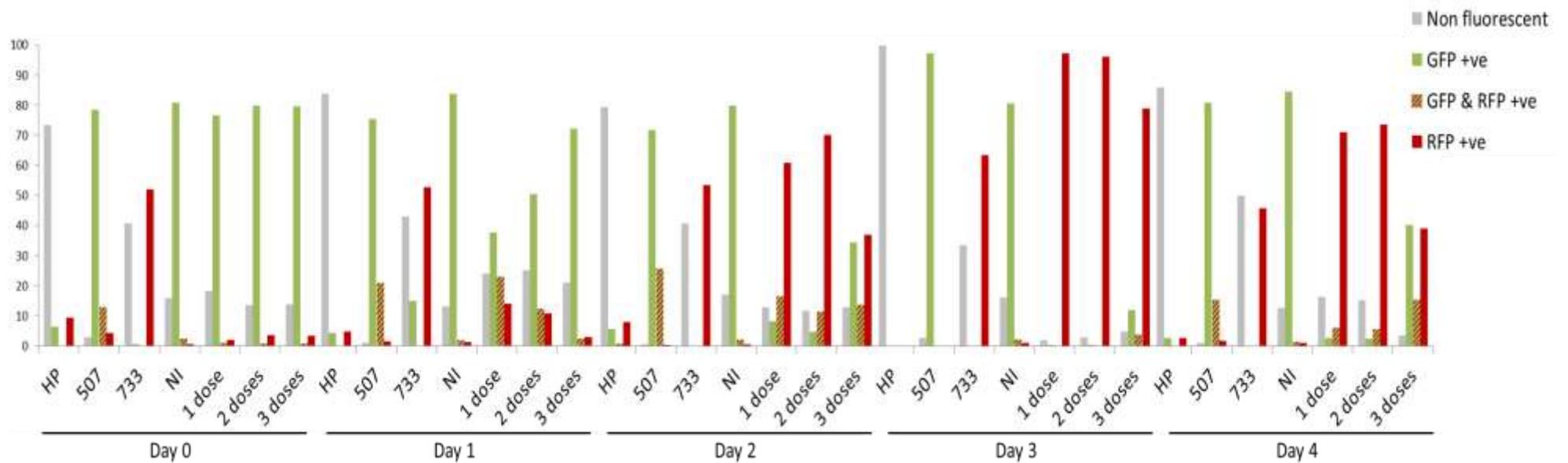
Figure 3.13 The effect of multiple inductions on a non-synchronous reporter population over time.

The percentage of parasites that excise GFP and therefore switch to RFP expression remains similar regardless of the number of doses of rapamycin administered *in vivo*. The rate at which the switch is observed was also consistent.

3.4 **Discussion**

With the generation of genome wide direct knockout vectors and the identification of the proportion of the genome that either causes detriment to or prevents asexual proliferation the need for conditional technologies to elucidate gene function becomes ever more clear (Aurrecoechea *et al.*, 2009). Further to this many studies have identified proteins expressed at multiple stages indicating multiple roles for single proteins throughout the life cycle (Khan *et al.*, 2005; Lasonder *et al.*, 2008; Philip & Waters, 2015). When comparing protein expression throughout the life cycle (asexual parasites, gametocytes, gametes, oocysts, oocyst sporozoites and salivary gland sporozoites) it was identified that no sampled stage expressed above 30% of their proteome uniquely. This means a minimum of 70% of the detectable proteome is shared between at least 2 life cycle stages (Lasonder *et al.*, 2008). If we wish to examine the role of any protein through gene deletion we will only be able to do so until its loss becomes detrimental to the parasite, anything downstream of this is not accessible with direct gene disruption. To understand all of these roles a protein plays throughout the life cycle only conditional systems can uncover the function at different stages.

In *T. gondii* the DiCre system to conditionally silence genes of interest has been used extensively to study genes previously designated essential for the replicative cycle. As well as giving vital insight into the intrinsic function of essential proteins (Andenmatten *et al.*, 2013; Harding *et al.*, 2016) the conditional dissection of protein function has led to the reclassification of essential genes as important but not essential. This underlies the identification of potential alternate mechanisms for motility and invasion of *T. gondii* in the lytic cycle (Andenmatten *et al.*, 2013; Egarter *et al.*, 2014). This lends to the idea that direct knockout of genes may not be allowing accurate and complete functional characterisation of genes and protein function.

Here we have demonstrated that the excision of a reporter gene resulting in the switch in expression from GFP to RFP is both rapid and highly efficient *in vivo* and *in vitro*. This indicates that if used to silence a gene of interest within the asexual cycle the gene would be lost within one cycle (see section 3.3) though this may be influenced by the gene targeted. The proportion of the population

that converted from GFP expressing to RFP expressing *in vitro* shows correlative qPCR, with an increase in the excised population and a corresponding decrease in the non-excised population. This, however, was not the case when completing the same analysis on DNA obtained from *in vivo* samples. In this case the increase in the excised population is evident and comparable to the *in vitro* samples (Figure 3.10) however there is no depletion of the non-excised template DNA. In further contrast to this qPCR analysis direct amplification, using the same primers to amplify the non-excised product never yield an amplicon from template DNA extracted 24hp-Ind (Figure 3.11). Combined, these data indicate that there is rapid excision resulting in the majority of the DNA providing the template only for the excised population and that the increased sensitivity of qPCR is detecting a different amplification product with the non-excised primers. This however is only possible when template gDNA is extracted from *in vivo* parasite material. Further work will be completed to understand this unusual result and endeavour to uncover what is being amplified and detected.

Expression of the two fragments of Cre (Cre59 and Cre60) was demonstrated in the blood stages and further work will confirm expression throughout the life cycle, which is expected due to the bidirectional *ef1a* promoter driving expression. While we have demonstrated rapid excision in asexual parasites the technology could also provide a novel means for excising a gene of interest at other stages not directly tractable to genetic manipulation. It would be beneficial to further test excision efficiencies in specific stages post the asexual proliferative stages, gametocytes, ookinetes and sporozoites to determine if *in vitro* rapamycin applications to these stages can result in excision to the same degree. To further demonstrate the use of the DiCre system in *P. berghei* its effect on later life cycle stages can be confirmed with the reporter system that has been established.

As this study was primarily focussed on the early stages of gametocyte commitment and development multiple DiCre lines were generated for the study of gametocytes and for the ready identification of mosquito stages. Firstly the HP::DiCre line was made, this was essential for the reporter work as this line would already have dual fluorescence in the form of the switching vector (GFP to RFP). We endeavoured to make a constitutively fluorescent line that also expressed DiCre and after several setbacks are on the brink of obtaining a

marker free CFP::DiCre line. This line will enable the identification of parasites in the mosquito stages of the life cycle and allow the purification of these stages. This will be invaluable for the quantification of transmissibility and to further allow study of fitness in conditional lines in the mosquito stages. The line primarily used in the remainder of this work was the 820::DiCre lines. This line expresses gametocyte specific proxies in the form of GFP (male gametocytes, driven by a putative dynein heavy chain promoter PBANKA_041610) and RFP (female gametocytes, driven by the CCP2 promoter PBANKA_131950) as well as constitutive expression of DiCre (driven by the bidirectional *ef1a* promoter). One of the issues that has been highlighted with the 820 parental line is that gametocytes, while detectable by their fluorescent proxies before morphology, are not identifiable until they are quite mature. To identify immature gametocytes a different fluorescent proxy that is detectable as early as 4 hours post invasion in gametocytes only (RL, unpublished data), but does not allow gender differentiation was combined with the HP::DiCre line. This will prove useful in further study with conditional lines as gametocytes will be identifiable from very early post commitment.

Chapter 4   Generation and characterisation of conditional gametocyte overexpression lines

4.1 Introduction

After the generation of parental DiCre expressing lines and the proof that the technology could be used within the asexual cycle to conditionally manipulate the genome in a rapid and highly efficient manner, we wanted to utilise these lines to control the onset of gametocytaemia. Previous work identified the transcription factor *ap2-g* as the master regulator of gametocytogenesis (see section 1.5.2 for full details (Kafsack *et al.*, 2014; Sinha *et al.*, 2014)) and a strategy was therefore designed to conditionally overexpress AP2-G using the inducible DiCre system. The hypothesis being that an overexpression of AP2-G would result in higher levels of commitment to gametocytes.

4.2 Design of a conditional overexpression system

The conditional overexpression system used in this study builds upon previous work in apicomplexa (Andenmatten *et al.*, 2013; Collins *et al.*, 2013) where gene excision using Cre recombinase has been shown to be successful. However, this study utilises LoxP sites in a head to head orientation and recombination between such sites would have a different outcome to that utilising LoxP sites in the same orientation (Albert *et al.*, 1995; Missirlis *et al.*, 2006; Oberdoerffer *et al.*, 2003). With LoxP sites in the same orientation the recombinase acts to loop out the DNA contained between the sites (for a full review of the system see Chapter 1.7 and Figure 4.1 A). However, with LoxP sites in a head to head orientation the recombinase results in the inversion of the sequence. Un-mutated head to head LoxP sites will invert the contained DNA whenever they encounter the recombinase (Figure 4.1 B). Previous work was carried out to identify mutations that would allow only one inversion making the flip irreversible. These mutated LoxP sites were identified (Oberdoerffer *et al.*, 2003) as Lox66 and Lox71. When used in combination they result in a single inversion of the contained DNA, even in the continued presence of the Cre recombinase (Figure 4.1 C). This single inversion is maintained because the LoxP sites resulting from the inversion are wild type LoxP and lox72 and these are not compatible (Oberdoerffer *et al.*, 2003).
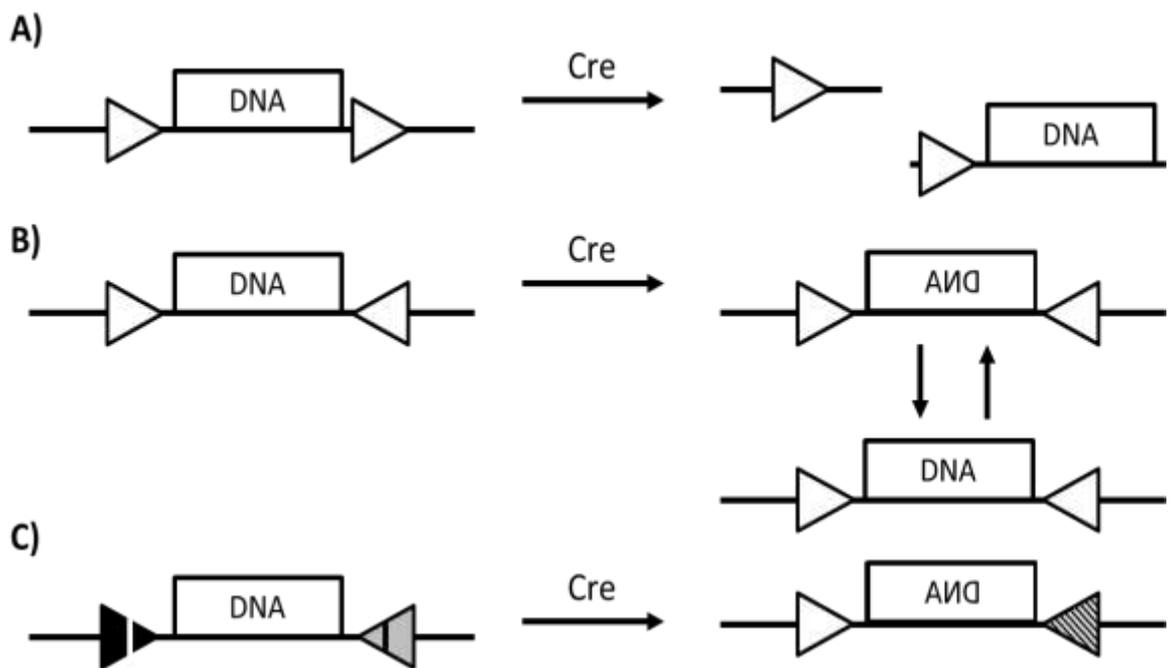
Figure 4.1 LoxP site activity in the presence of the Cre recombinase.

A) DNA flanked by LoxP sites (white) in the same orientation will excise the contained DNA upon Cre activity.

B) DNA flanked by LoxP sites (white) in a head to head orientation will invert the DNA upon Cre activity. Continued Cre activity will result in further inversions.

C) DNA flanked by the mutated LoxP sites Lox66 (black with white mutation) and Lox71 (grey with black mutation) in a head to head orientation will invert the contained DNA resulting in an upstream LoxP site (white) and a downstream Lox72 site (black and white). These sites are not compatible and will therefore no longer respond to Cre.

As parasites that do not express AP2-G are viable throughout the asexual cycle our conditional regulation strategy will disrupt the native *ap2-g* promoter region by insertion of an inducible promoter (Sinha *et al.*, 2014). This vector (designed and generated by Katarzyna Modrzynska (KM), Figure 4.2 A) contains an *hsp70* promoter flanked by mutated head to head LoxP sequences. Integration of the vector into the genome uses large arms homologous to the *ap2-g* promoter sequence and the *ap2-g* coding sequence. Before the addition of rapamycin this line expresses no *ap2-g,* because of the interruption of the promoter, and the *hsp70* promoter drives expression of *hdhfr/ts,* conferring resistance to pyrimethamine and allowing selection which is also contained on the integration vector used to introduce the *hsp70* promoter. Upon the addition of rapamycin, the Lox66 and Lox71 sites recombine inverting the contained DNA sequence and become LoxP and Lox72. This results in a halt in the expression of HDHFR/TS and expression of *ap2-g* by the *hsp70* promoter (Figure 4.2 B).
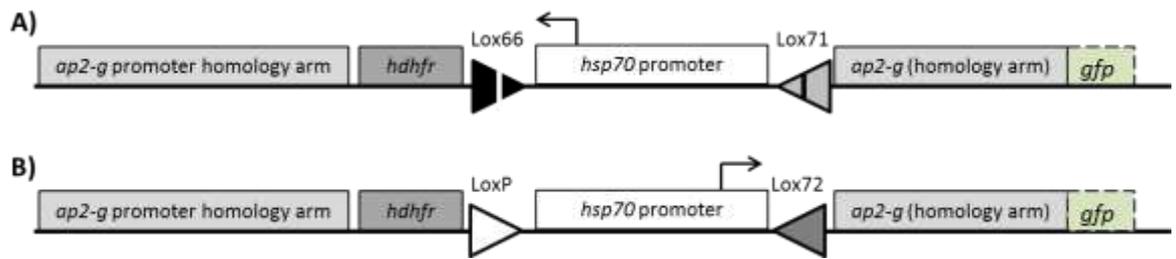
Figure 4.2 Vector for the conditional expression of *ap2-g*

A) The *hsp70* promoter is flanked by the mutated Lox66 (black and white) and Lox71 (grey and black) sites. Before Cre recombinase activity the *hsp70* promoter drives expression of *hdhfr* conferring resistance to pyrimethamine.

B) After Cre activity the *hsp70* promoter is driving *ap2-g* expression. The activity alters the Lox sites flanking the promoter to be LoxP (white) and Lox72 (grey) that are incompatible and therefore cannot react to further Cre exposure.

Further to this it was desirable to generate an inducible overexpression line whereby the expressed AP2-G was tagged. To obtain this, the existing vector was manipulated to also contain a C-terminal GFP tag (KM). Previous work has shown that a C-terminal tag on *ap2-g* does not impede the formation of gametocytes (Sinha *et al.*, 2014).

4.3 **Generation and kinetics of the conditional overexpression system**

The original overexpression vector was integrated into the previously generated parental parasite line, 820::DiCre. This line then allows the identification of male (GFP) and female (RFP) gametocytes based on their fluorescence as well as induction of recombinase activity upon addition of rapamycin. Integrants were selected by pyrimethamine and the population was confirmed by PCR (figure 4.3 A & B) and the loss of gametocytes (due to lack of *ap2-g* expression) determined by flow cytometry (figure 4.3 C). Although uncloned, the line showed only integrants by PCR and flow cytometry. As the parasites containing the conditional overexpression vector had a selective growth advantage over those that do not, the line was not cloned as it would self-enrich.
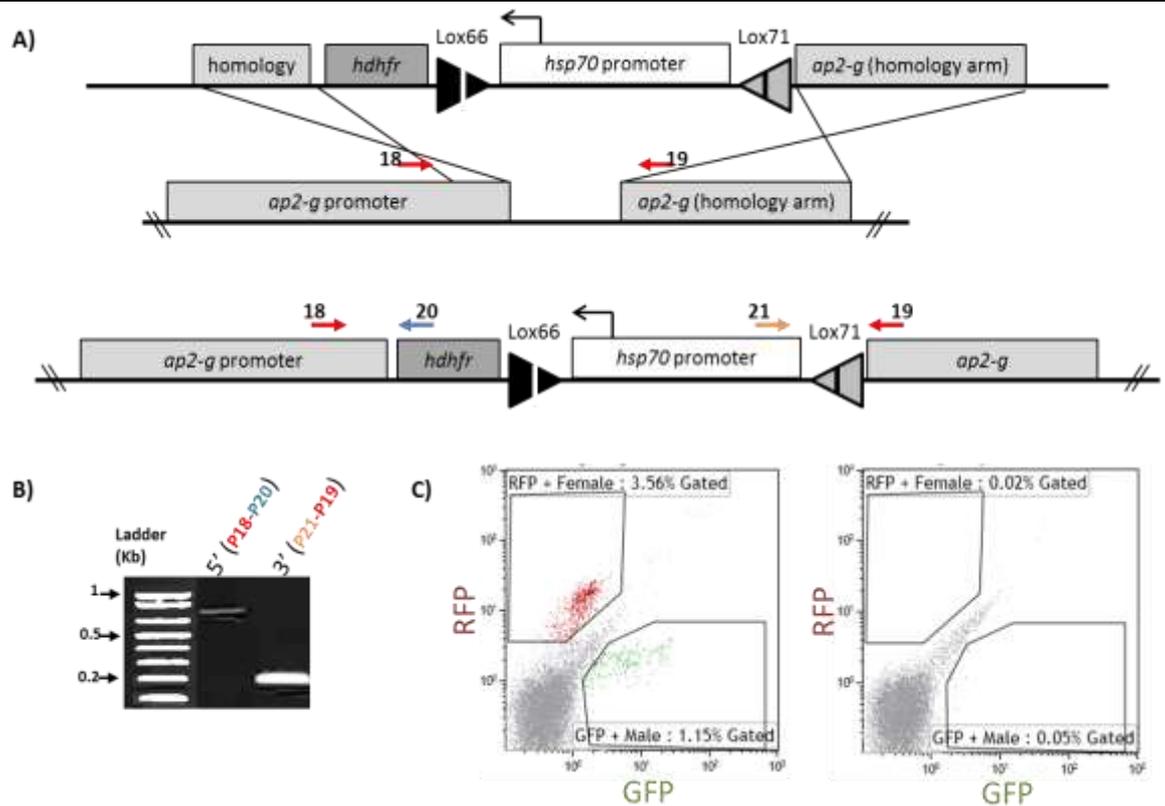
Figure 4.3 Generation of a conditional *ap2-g* overexpression line.

A) Schematic representation of integration of the conditional overexpression vector in the *ap2-g* locus. Primers p13 and p15 will demonstrate 5' integration and primers p16 and p14 will show 3' integration of the promoter disrupting inducible overexpression vector.

B) Amplification of the 5' integration product with primers p13 and p15 at 700bp and the 3' integration fragment with primers p16 and p14 at 211bp of the conditional overexpression vector show it has been replaced in the population.

C) Loss of both male (GFP) and female (RFP) gametocyte populations was identified by flow cytometry showing native *ap2-g* expression has been disrupted.

Although we had previously tested the speed and efficiency of recombinase activity post induction with rapamycin (Figure 3.10), this system utilised a variation of the LoxP sites resulting in an inversion of a promoter instead of a gene excision. Therefore the proportion of the population that responds to the recombinase resulting in an inversion of the promoter and the rate at which this maximum is reached was determined. As we had already determined excision rates *in vivo* and *in vitro* are comparable the rate of inversion was only determined *in vivo* as this would be the most representative to subsequent experiments.
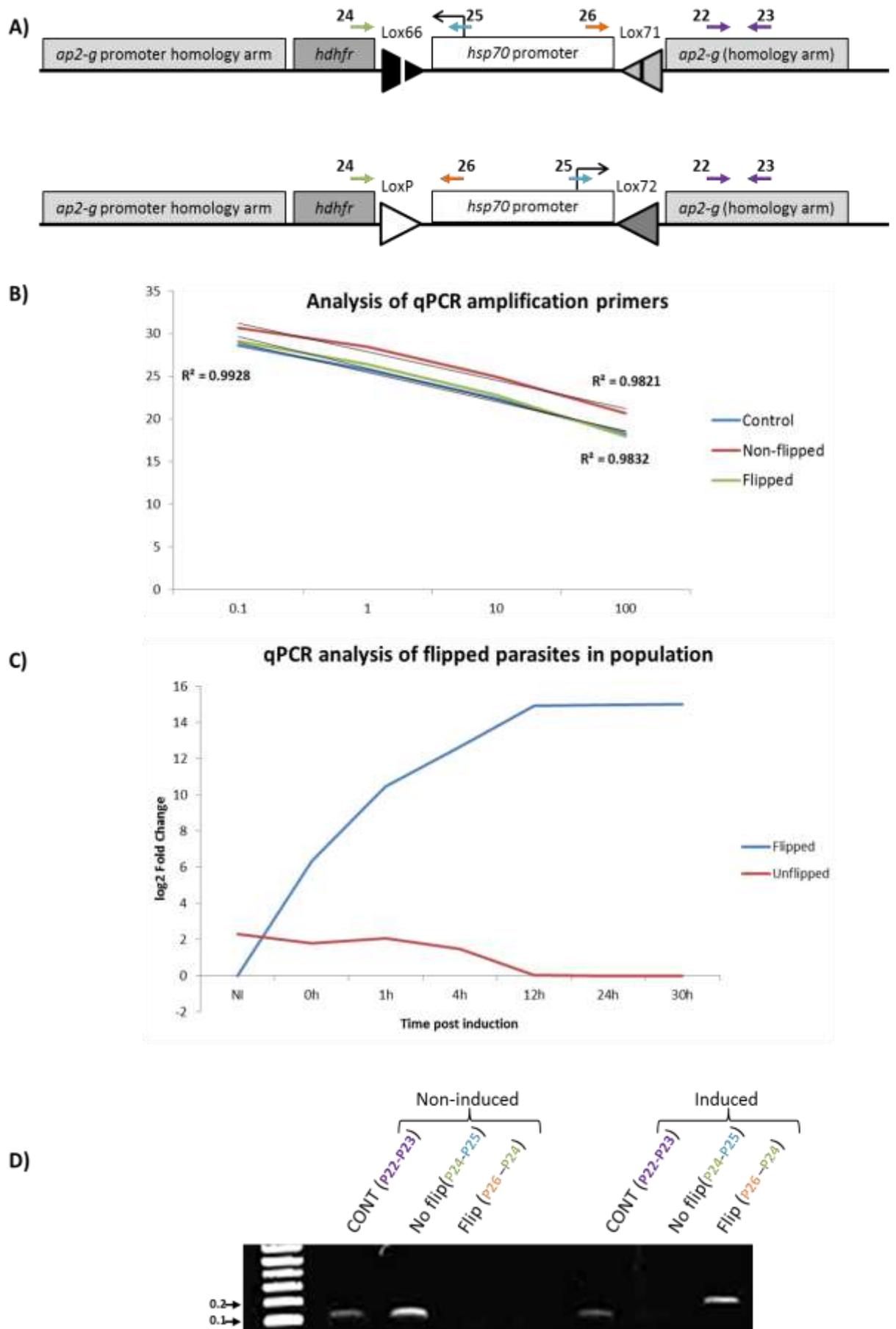
Figure 4.4 Quantitation of flip rate *in vivo*

Figure 4.4 Quantitation of flip rate in vivo

To determine the extent to which flipping of the promoter occurs *in vivo* quantitative PCR was carried out.

A) Schematic representation of the primers used to quantify the unflipped and flipped populations with the relevant control that does not alter in either population.

B) Analysis of the primers to quantify the different populations shows that they amplify with correlative kinetics.

C) Analysis of time points post induction shows that the proportion of parasites in the flipped genome orientation rapidly increases post induction with 4mg/kg rapamycin and reaches the maximal at between 4hpi and 12hpi (which correlates to the kinetics of the reporter excision (Figure 3.10). The proportion of the population that amplifies the unflipped population does not decrease with the corresponding decrease.

D) To show that the actual amplification with the primers used PCR was carried out on the uninduced line and the induced line where gDNA was harvested 30hpi. PCR analysis shows both the lines amplify the control product (110bp) to the same degree. The uninduced line does not amplify any product with the primers specific to the flipped locus but does amplify the unflipped locus (110bp). The induced line does not amplify any produce corresponding to the unflipped locus but does amplify the flipped locus product (210bp).

Synchronous infections established by intravenous infection with mature, purified schizonts, were induced with 4mg/kg rapamycin and the proportion of parasites in the unflipped to flipped orientation (Figure 4.4 A) were compared over time post induction (0 hours, 1 hour, 4 hours, 12 hours, 24 hours and 30 hours). After initial setup of the system to ensure that amplification efficiency was comparable for the three populations (control, flipped and unflipped) the populations were compared (Figure 4.4 B). As in the reporter system it is evident that the flip rate is not only very high but very rapid. The increase in flipped parasites in the population increases rapidly in the first hour post induction and between 4 and 12 hours reaches its maximum (Figure 4.4 C). However, like the reporter system when sampling *in vivo* the proportion of unflipped does not show a corresponding decrease. As we did not sample *in vitro* it is not possible to ascertain if this is an artefact of *in vivo* harvest of parasite material. To demonstrate that there was not actually still a high proportion, or the same proportion, of unflipped gDNA in the sample PCR was carried out on the uninduced sample and the 30hpi induced sample. No flipped DNA is detectable in the uninduced sample by PCR and no unflipped is detectable in the induced sample by PCR (Figure 4.4 D).

4.4 Viability of AP2-G overexpressing parasites throughout the lifecycle

After confirmation of the integration of the conditional overexpression vector it was essential to determine the level of commitment to gametocytogenesis achievable post induction and ascertain if the gametocytes that develop from this commitment are viable.

4.4.1 Commitment to gametocytogenesis

As the overexpression line generated forms of gametocytes that express fluorescent proxies identifying sexually committed parasites by their gender (GFP for male gametocytes and RFP for female), quantification of the proportion of gametocytes formed post induction is possible. These fluorescent proxies allow male and female gametocyte identification based on the expression of GFP under a male specific promoter (PBANKA_041610) and RFP under a female specific promoter (PBANKA_131950).

As the proportion of gametocytes generated from each asexual cycle can be quite variable, ranging from 5% - 15%, and the sex ratio in the population can vary, commitment and sex ratio are likely to be variable in the inducible line also. To determine if the degree in variability after induction was similar to that exhibited in a wild type line, the standard 820 line was compared with the 820::DiCre and the Induced Overexpresser line. To quantify gametocyte commitment, non-synchronous infections were sampled at multiple parasitaemias over subsequent days of infection and the proportion of gametocytes quantified by flow cytometry analysis of GFP (male) and RFP (female) signal. In all of these lines, expression of GFP and RFP were used to quantify the proportion of male and female gametocytes. In six independent inductions of the inducible AP2-G overexpression line, variation in commitment was observed, ranging from 11% to 71% (Figure 4.5 A). In a wild type 820 line the variability was from 2% to 5.5% (Figure 4.5 A). Similarly the variability in the 820::DiCre line was from 3.1% to 9.8% (Figure 4.5 A). While it was clear that overexpression of *ap2-g* results in an over-representation of gametocytes within the cycle it was also clear that there is variability in amount. To determine if the variability in each of these lines was similar, the coefficient of variance was calculated. This showed that the variability between each of the lines was much lower when considered independent of the mean of the datasets (Figure 4.5 A).

The parasites that have been induced but do not become gametocytes appear to be non-viable for multiple cycles. Giemsa analysis shows the parasites in the population that are not identifiable as gametocytes to be very dense and pyknotic in appearance (Figure 4.5 B). While no direct link with AP2-G levels has been demonstrated further work will be completed, with the AP2-G::GFP line, to determine if AP2-G levels across the population correlate to gender and unviable parasites. The maintained reduction in parasitaemia, quantified by flow cytometry, which follows the wave of commitment to gametocytogenesis confirms that the majority of these non-gametocyte parasites are not viable (Figure 4.5.C). This further analysis aims to elucidate if AP2-G expression is the only requirement for successful gametocyte commitment.

Previous work has identified *P. falciparum* parasites that extensively commit to gameyocytogenesis and appear to stall within the asexual cycle. In this line, which is a heterochromatin protein 1 (*Pf*HP1) knock down (using the DD system), the cycle after depletion occurs and levels of *Pf*HP1 become undetectable 52% of the population are Pfs16 positive (a marker for gametocytes) and nuclear division does not occur in the Pfs16 negative cells demonstrating a stall at the trophozoite stage (Brancucci *et al.*, 2014). While there are similarities in the arrested parasites that do not commit after *ap2-g* induction careful analysis of these parasites has not been completed to confirm any consistent responses. It would be interesting to carry out further analysis with *ap2-g* inducible overexpression to determine if, like in the HP1 depleted line, reduction in the levels of AP2-G (achieved by stabilisation of HP1) resulted in a return to characteristic commitment levels and the reactivation of the surviving arrested asexual parasites or if these deformed asexual parasites are in fact dying/dead. Due to the differing development times for asexual (48h in *P. falciparum* and 24h in *P. berghei*) and sexual (10 – 12 days in *P. falciparum* and 30h in *P. berghei*) it may be interesting to ascertain if the arrested parasites remain arrested for a similar length of time or if the shorter development period has an effect.
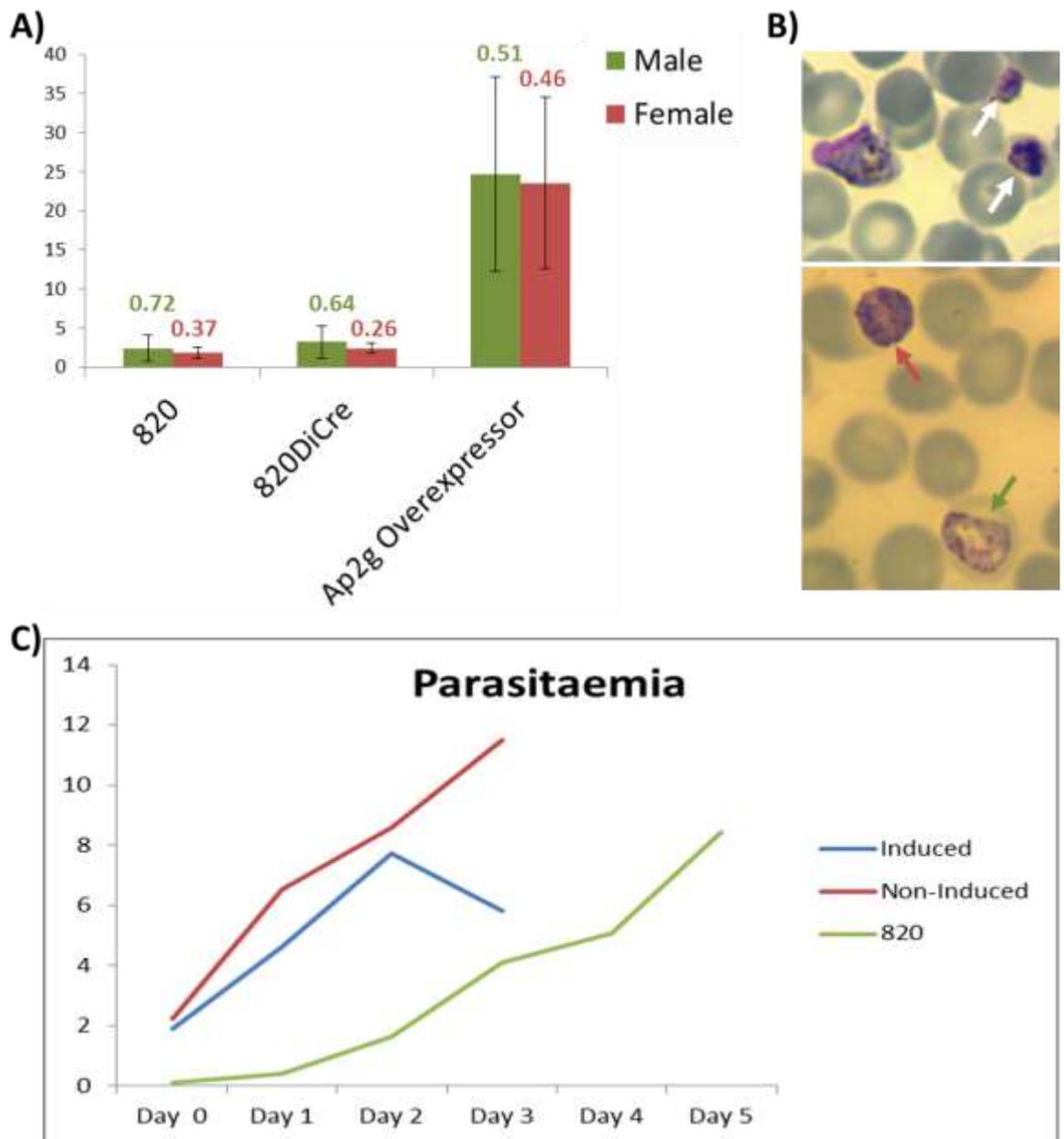
Figure 4.5 Variability and growth of the *ap2-g* overexpressing line.

A) The proportion of the population expressing GFP (male gametocytes) and RFP (female gametocytes) was quantified in 6 independent infections in the existing 820 line, the 820::DiCre line and the Induced over-expresser. While ratios and absolute proportions of male and female gametocytes vary the coefficients of variation are similar in all lines indicating similar variability in proportion and ratios.

B) Male (green arrow) and female (red arrow) gametocytes are readily identifiable in Giemsa stained blood smears after induction. Asexual parasites (white arrows) that do not commit to becoming gametocytes become dense and pyknotic in appearance.

C) The *ap2-g* over-expresser line without induction grows faster than the 820 line. After induction however within two cycles parasitaemia stalls and decreases as the asexuals are no longer viable and the gametocytes deplete.

Resurgence in parasitaemia does however happen 4 – 5 days after the wave of gametocytogenesis and the reduction in parasitaemia. These parasites contain the *ap2-g* inducible over-expression cassette but do not respond to additional inductions with rapamycin (Figure 4.6 C) ascertained by the lack of gametocytes after intraperitoneal injections of rapamycin. PCR analysis shows that a proportion of the population is refractory to rapamycin and even after addition remain in the un-induced orientation (Figure 4.6 A and B, Unflipped). However, a proportion are, at the genetic level, AP2-G over-expressers (Figure 4.6 A and B, Flipped).
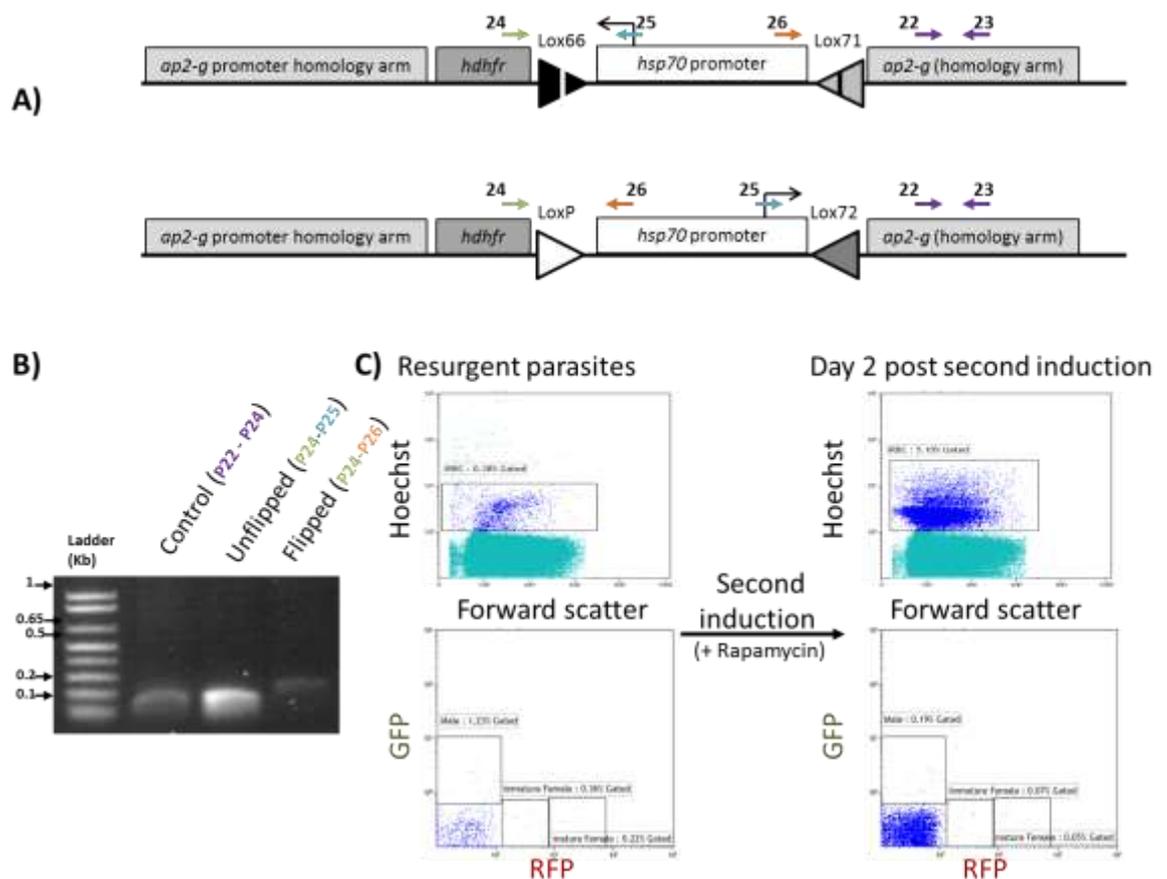
Figure 4.6 Analysis of the resurgent lines

A) Schematic representation of PCR to determine if the genomic locus in the population after resurgence of parasitaemia contains an hsp70 promoter in the unflipped (top panel) or flipped (bottom panel) orientation.

B) PCR amplification of the unflipped (p24 & p25, 110bp) and flipped orientation (p25 & p26, 210bp) indicates that the population is a mixture of parasites that have not responded to the rapamycin and remain unflipped and parasites that have flipped and overexpressed AP2-G.

C) Additional *in vivo* applications of rapamycin do not result in the subsequent flipping of the promoter after resurgence of the parasite line. If mice infected with the parasites that recolonise after rapamycin inductions are induced with a second dose of rapamycin (4mg/kg) no additional gametocytes are detected by their fluorescence profiles (male, GFP and female, RFP) by flow cytometry analysis.

Together these data indicate that a very high proportion of the population respond to the first dose of rapamycin resulting in a flip in the *hsp70* promoter to drive *ap2-g* and that this results in a wave of gametocyte commitment.

There appears to be a proportion of the population that are not able to survive the induction process. These become dense in appearance before disappearing from the blood stream. While these have not been directly shown to have the flipped locus the proportion of the parasites that have this appearance does not seem to correspond to the proportion that have not flipped. Therefore we assume these parasites are overexpressing *ap2-g* but cannot commit to gametocytogenesis and do not survive.

A small proportion survived the induction process but did not immediately go on to commit to gametocytogenesis. These parasites were able to re-establish an infection in mice after a reduction in parasitaemia. A portion of this resurgent population did not initially responded to the rapamycin induction (shown by amplification of the unflipped locus, Figure 4.6) and therefore remained AP2-G negative. An even smaller proportion of the population had flipped their *hsp70* promoter to drive *ap2-g* expression (shown by amplification of the flipped locus, Figure 4.6). As the entire resurgent population did not generate gametocytes (determined by flow cytometry analysis and further analysis of Giemsa smears) a second induction with 4mg/kg rapamycin was attempted to see if the refractory population could flip their promoter. This second induction was completed when the resurgent line had reached an approximately 1% parasitaemia and was monitored for two days post induction. No gametocytes were determined by flow cytometry analysis (Figure 4.6 C) therefore we concluded that the proportion of the population that did not initially flip was refractory. We also determined that the smaller population that had flipped the locus was not producing gametocytes.

4.4.2 Conversion to ookinetes

To determine if the gametocytes that arise from commitment in the *ap2-g* overexpresser are able to generate ookinetes, a non-synchronous infection was induced with 4mg/kg rapamycin and 30 hours after induction blood was incubated with ookinete media at 21°C. In this line, ookinetes were detectable and appear morphologically normal (Figure 4.7 A). Unfortunately, conversion rates were very low and therefore not quantified. This could indicate that only a proportion of the male and female gametocytes that mature from *ap2-g* overexpression are viable enough to form ookinetes within this time.
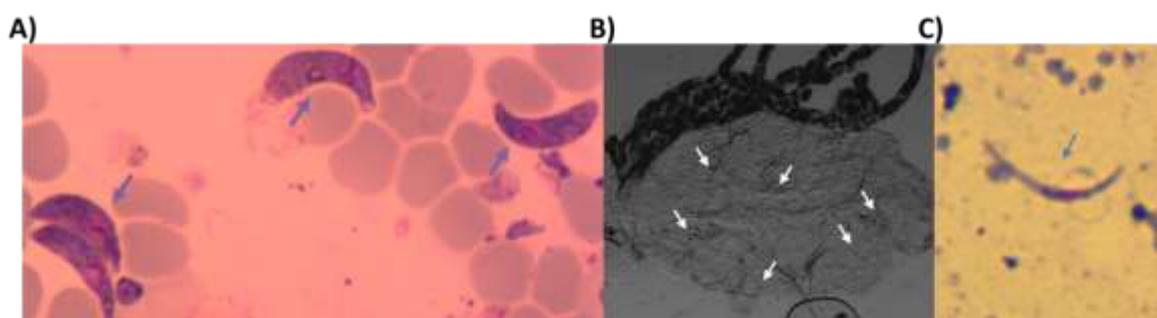
Figure 4.7 Post gametocyte stages of the *ap2-g* induced overexpresser.

A) Ookinetes that are identifiable by Giemsa smear appear morphologically normal. Conversion rates have currently not been quantied.

B) Non-fluorescent oocysts are visible in the mosquito midgut, however they are not quantifiable and comparable to the parental line.

C) 18 days post mosquito feed sporozoites are found in the salivary glands of infected mosquitoes. As the line is not fluorescent quantification of the proportion of mosquitoes infected and the intensity of infection within the salivary glands is not discernible.

4.4.3 Establishment of an infection in mosquitoes

To ascertain if the AP2-G overexpressing ookinetes are viable and able to colonise and replicate within the midgut of mosquitoes, cages of 200 mosquitoes were fed on mice infected with a non-synchronous induced AP2-G overexpresser line 30 or 50 hours post induction and colonisation of the midguts was checked 11 days post feeding. Oocysts were identifiable, however they are non-fluorescent and so quantification of the number of oocysts was not completed (Figure 4.7 B). Future work will utilise the constitutive CFP::DiCre parental line (Figure 3.5) in conjunction with the AP2-G inducible overexpression line to quantify the numbers of oocysts within the midguts of individual mosquitoes

4.4.4 Sporozoite establishment in the salivary glands

Mosquitoes fed on the non-synchronous AP2-G overexpresser line 30 or 50 hours post induction were maintained until day 18 when salivary glands were extracted. Salivary glands were ruptured and sporozoites stained with Giemsa (Figure 4.7 C), again showing their presence. Further work aims to quantify sporozoites viability and invasive ability however it has yet to be completed.

With the development of the CFP::DiCre line (Figure 3.5), quantification of transmission in the mosquito stages will be possible. While how transmissible parasites are after overexpression of AP2-G has occurred is not yet known it is clear than some of the gametocytes that mature from overexpression are able to form ookinetes and that these are able to colonise the midgut of mosquitoes before generation of sporozoites that migrate to the salivary glands of the infected mosquito.

4.4.5 Establishment of an infection in a naïve host

To determine if the AP2-G overexpression line could establish an infection in a new host, mosquitoes were fed on naïve mice 18 – 21days post feeding on a non-synchronous AP2-G overexpression line (30 hours post induction). From day 5 post blood feeding, mice were monitored for the onset of parasitaemia. From day 5 parasites could be observed in the bloodstream of naïve mice. On day 5 the parasitaemia of the AP2-G overexpressing line was approximately 10 fold reduced when compared to an 820 wild type control line. In the AP2-G overexpresser the parasites establishing the infection did not produce any gametocytes identifiable by flow cytometry analysis (Figure 4.8 A & B).
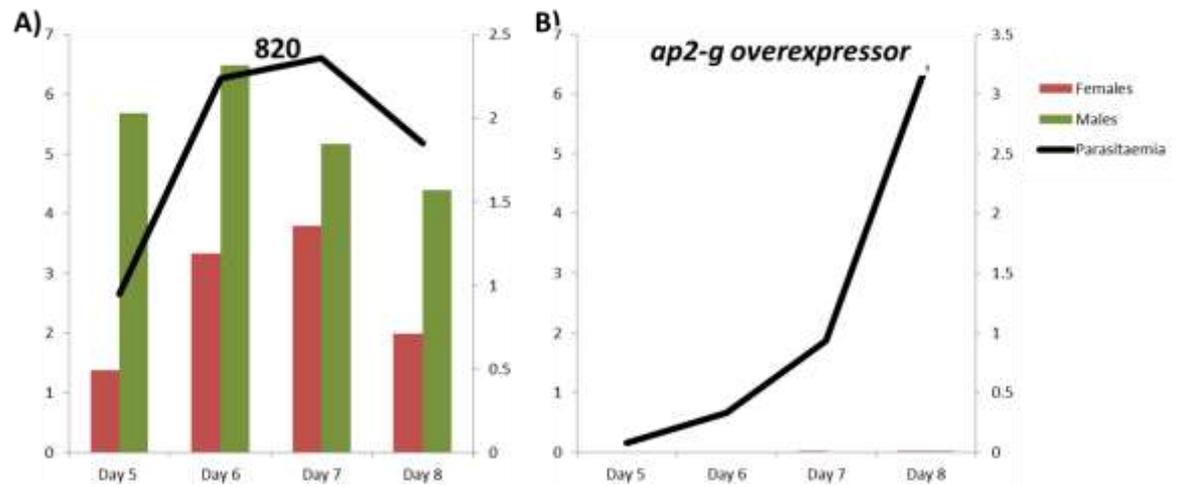
Figure 4.8 Establishment of infection in naïve mice

A) Parasitaemia is detectable on day 5 post mosquito feeding in the 820::DiCre control line with male (green) and female (ref) gametocytes detected by flow cytometry.

B) On day 5 post mosquito feeding, parasitaemia is detectable in the *ap2-g* overexpresser but is approximately 10x lower than the 820::DiCre control line. This line does not make any gametocytes identifiable by flow cytometry.

4.5   Detailed characterisation of gametocyte development in AP2-G overexpressing parasites

The *ap2-g* overexpression system, while generating gametocytes that are viable throughout the entire life cycle, will set off a cascade of commitment with such high levels of *ap2-g* that gametocytes may develop at different rates to wild type gametocytes. To ensure subsequent work was completed on representative gametocytes, the rate of development was monitored post induction to optimise induction timing.

4.5.1 Wild type gametocyte development

As *P. berghei* gametocytes are not morphologically distinguishable from asexual parasites until late in the blood stage cycle (approximately 18hpi when parasites are trophozoites), early development is impossible to monitor by morphology alone. The 820 parasite line allows us to identify male (GFP positive) and female (RFP positive) gametocytes based on their fluorescence before they are morphologically identifiable. However, the exact point of fluorescence detection has not been identified. To determine a normal development profile for gametocytes, naïve mice were infected with Nycodenz purified schizonts. To remove mature gametocytes, 2hpi blood was collected and gametocyte depleted by magnetic filtration. Parasites were then intravenously injected into naïve mice. Analysis of fluorescence showed distinct profiles in the onset of fluorescence for male and female gametocytes. In a cyclic pattern, GFP under the male specific promoter, PBANKA_041610 (putative, dynein heavy chain), was detectable in male gametocytes (Figure 4.9) between 12 and 16 hours post invasion each cycle. A decrease in the GFP signal was observed every 24 hours corresponding not to a decrease in male gametocytes but to an increase in asexual parasites post reinvasion. This decrease is due to the % male gametocytes in the population being calculated against parasitaemia. This pattern of fluorescence onset and decrease was clear for two cycles post synchronous infection however it becomes less well defined with invasions three and four. This is likely due to a loss of synchronicity in the infection and the added complication of gametocyte degradation when mature gametocytes from previous commitment cycles are no longer viable (Figure 4.9).

Onset of GFP fluorescent (male) marker expression
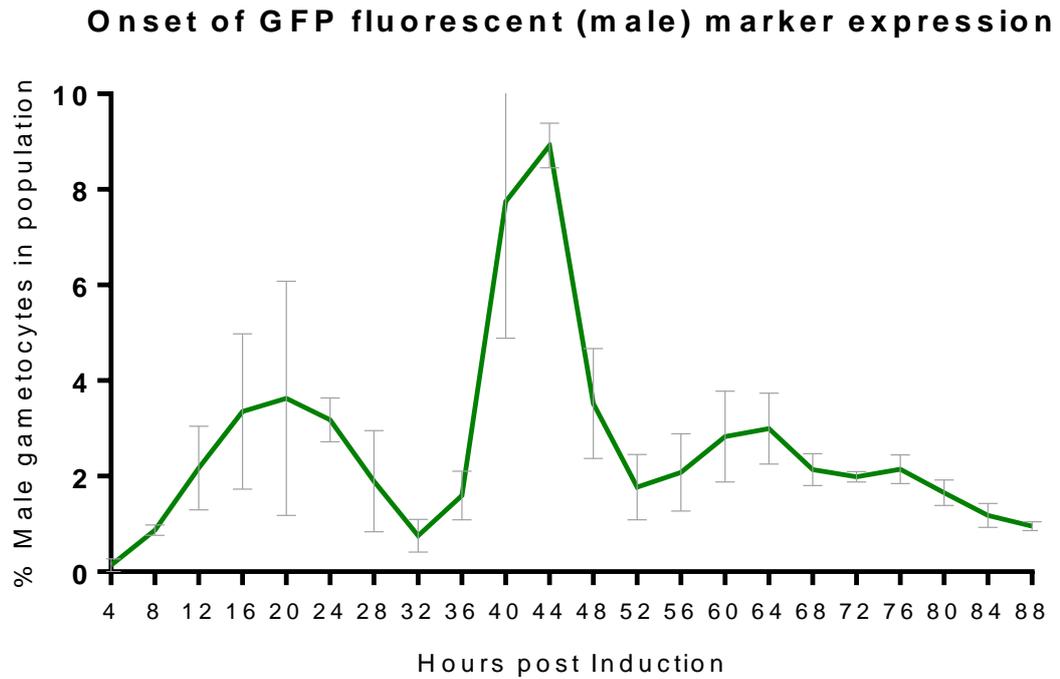
Figure 4.9 Onset of GFP expression in the parental line where male gametocytes express GFP and female gametocytes express RFP (820)

In a cyclic pattern, GFP signal was observed in maturing male gametocytes every 24 hours with detection beginning between 12 and 16 hours post merozoite invasion. For two cycles post synchronous infection, naïve mice showed a clear pattern of gametocyte commitment by fluorescence. By invasion round three, synchronicity began to fail and the proportion of gametocytes became less distinct due to degradation of earlier committed gametocytes

The fluorescent marker for female gametocytes is under the female specific promoter PBANKA_171950 (CCP2). This promoter, unlike the male specific promoter, does not express consistently in immature and mature female gametocytes. Female gametocytes show two distinct profiles of RFP expression low and high. The delay in detection of the high RFP signal female gametocytes led to the classification of low intensity RFP parasites as immature and high intensity RFP parasites as mature (Figure 4.10).

A) Parasite population, No additional population gating
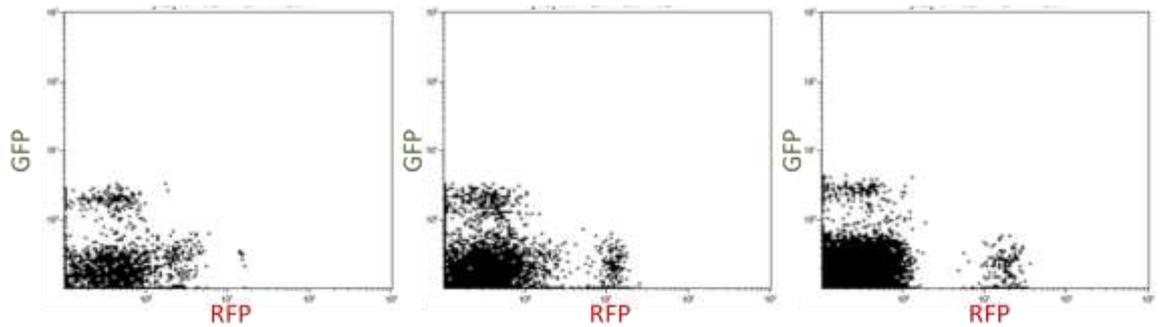
B) Parasite population, Male gametocytes gated by GFP, females by RFP.

Immature female gametocytes only

Immature & Mature female gametocytes

Mature female gametocytes only
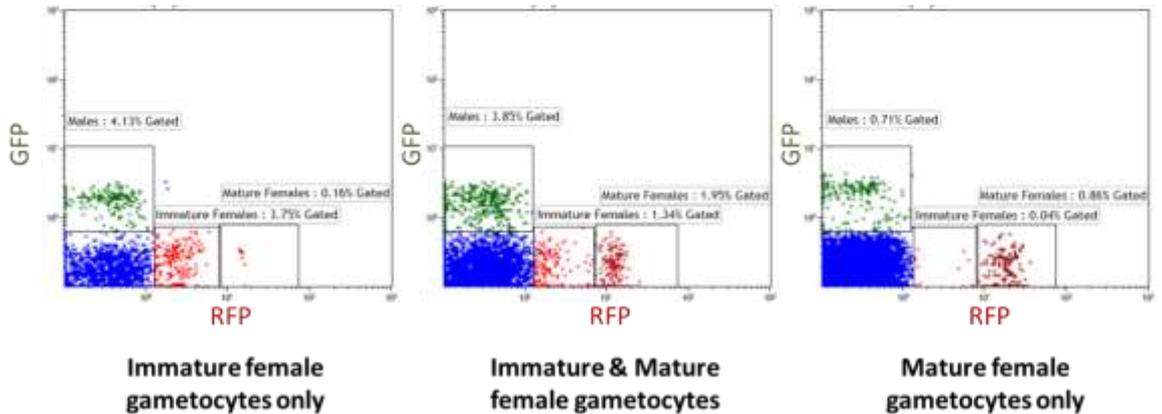
Figure 4.10 Immature and mature female gametocyte fluorescent profiles by flow cytometry analysis

Initially the RFP signal becomes distinct from the asexual or non-fluorescent population (left panel). As time goes on the intensity of the signal increases and a second wave of gametocyte commitment occurs. This gives rise to two populations distinct from the non-fluorescent asexuals. If no more commitment to gametocytogenesis occurs then all female gametocytes increase their RFP expression to the mature gametocyte intensity (right panel).

A) When the parasite population is plotted with RFP intensity against GFP intensity distinct populations are clear, even without gating.

B) When male gametocytes are gated based on their GFP fluorescence (green) and immature and mature females are gated based on their RFP intensity (red) the individual populations become clearer.

As with the fluorescent output driven by the male specific promoter, a cyclic pattern of expression of RFP in line 820 was demonstrated when examining the immature and mature female gametocytes independently. Initial detection of the RFP signal, in immature gametocytes (Figure 4.10) was seen from 20 – 22 hours post invasion of the merozoites (Figure 4.11) and every 24 hours, subsequent waves of immature female gametocytes was observed. By 36 hours post invasion, female gametocytes were what will be deemed as mature and identified by the increased intensity of the RFP fluorescence and its consistency (Figure 4.11). As the second wave of gametocytes was formed, it was clear to see the difference in populations of immature and mature females. The reduction in immature female gametocytes corresponded in timing to the increase in mature female gametocytes

as you would expect with a maturing population. In a similar manner to that observed for the male gametocyte GFP signal, after three cycles of invasion the synchronicity of the infection deteriorated and the mature gametocytes began to degrade making the populations less distinct.
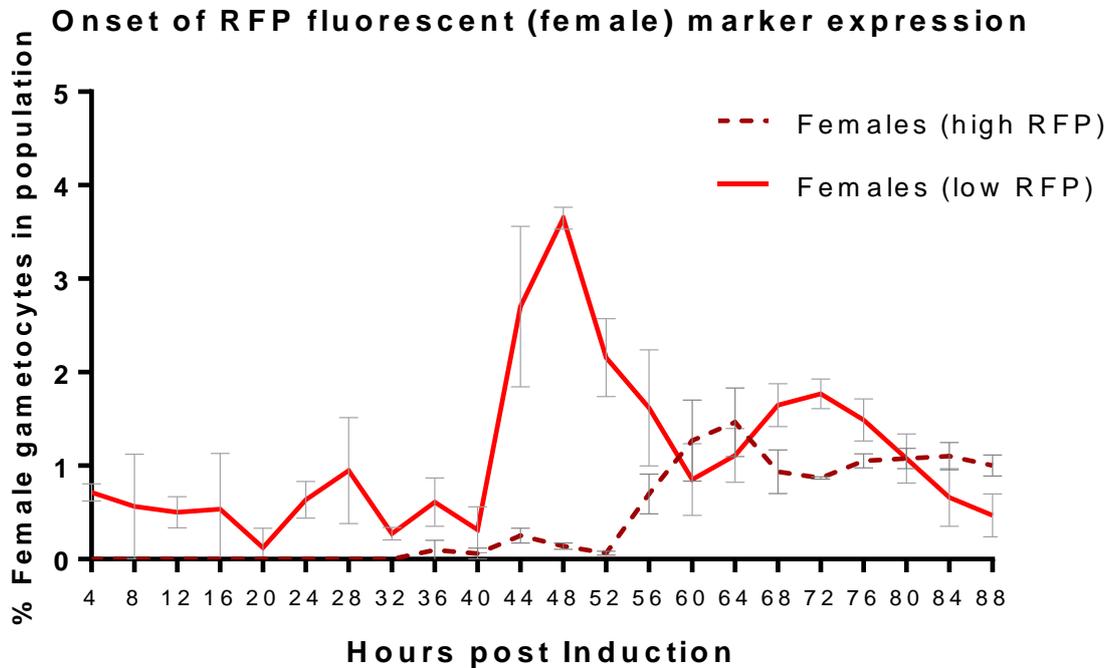
Figure 4.11 Onset of RFP expression in the parental line where male gametocytes express GFP and female gametocytes express RFP (820).

In a cyclic pattern, RFP signal is observed in maturing female gametocytes every 24 hours with detection beginning between 20 and 22 hours post merozoite invasion. This initial signal for the immature female gametocytes increased in intensity until the gametocytes matured at 36hpi. For two cycles post synchronous infection, naïve mice showed a clear pattern of gametocyte commitment by fluorescence. By invasion round three, synchronicity began to fail and the proportion of gametocytes became less distinct due to degradation of the previously committed gametocytes

4.5.2 Effect of *ap2-g* onset on parasite growth and development

Now that a base line of gametocyte maturation, defined by the onset of the fluorescent proxies, had been determined it was possible to see how overexpression of *ap2-g* affected the onset of these proxies and the rate at which gametocytes developed as determined by the onset of GFP (male gametocytes), the onset of RFP (immature female gametocytes) and the increase in intensity of RFP (mature female gametocytes).

To determine the affect *ap2-g* onset had on parasites when overexpressed from different asexual blood stages inductions were completed at different times post invasion to determine if there was an effect on commitment (total proportion and gender differentiation) and subsequent gametocyte development. Synchronous *in vivo* infections were induced at various time points and the infection monitered over time to determine parasite growth rates, the amount of commitment to gametocytogenesis, the ratio of male to female gametocytes and the rate at which these male and female gametocytes develop, as determined by the onset of the fluorescent proxies. A variety of induction time points were selected to cover the different life cycle stages when *ap2-g* may be required for commitment to occur. These were 2hpi, 6hpi, 8hpi, 12hpi, 18hpi and 22hpi (Figure 4.12).

As long term sampling was required *in vivo* time point samples were staggered across two cages of animals that had both their initial infections and variable inductions completed in parallel.
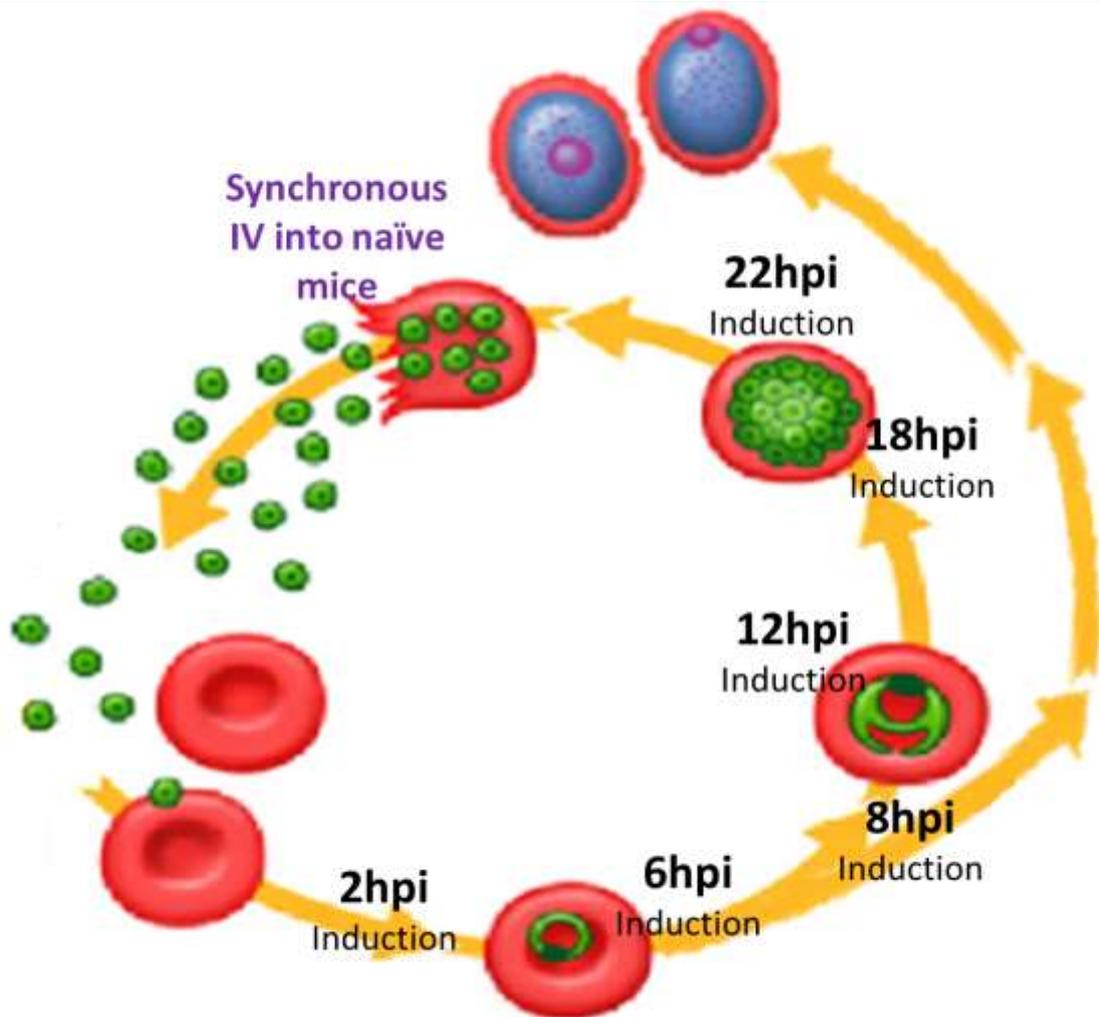
Figure 4.12 Schematic of varied *ap2-g* induction analysis

To determine the effect the time of *ap2-g* onset and overexpression had on asexual and sexual parasite growth synchronous infections (*in vivo*) were induced at different time points and asexual growth (as determined by parasitaemia increase) and commitment (quantified by fluorescent proxy expression) were analysed.

4.5.2.1 Effect of *ap2-g* onset on asexual parasite growth

First we wanted to determine the viability of asexual parasites if *ap2-g* was induced at various times post invasion. We previously established that after induction a proportion of the parasites were not viable. We therefore wanted to ascertain how long after induction of a*p2-g* overexpression this loss of viability occurred. To do this we monitored asexual parasite proliferation. To determine is parasites were still viable in the asexual cycle we monitored parasitaemia in *in vivo* infections post induction. In a synchronous population every 24 hours mature schizonts rupture, reinvasion occurs and the parasitaemia increases.

After induction the *hsp70* promoter rapidly flips (see section 4.3) and in place of driving expression of the selectable marker begins expression of *ap2-g*. After induction commitment to gametocytogenesis has been shown to occur. As gametocytes are a non-proliferative stage and we have seen the death of a proportion of the population after induction (see section 4.4) we wanted to monitor parasitaemia over several cycle to ascertain how many cycles post *ap2-g* induction and overexpression parasites remain asexually proliferative and if the time during development had an impact on this viability.

After induction to flip the *hsp70* promoter, resulting in overexpression of *ap2-g*, reinvasion occurs at approximately 24hpi as expected, however the second round of reinvasion, expected at approximately 48hpi does not occur in the induced lines. It is clear that this reinvasion does occur in the non-induced inducible *ap2-g* overexpresser (black line). This lack of parasitaemia increase is consistent for all induced lines indicating that the merozoites that invade with already high levels of AP2-G are not able to mature and replicate to generate viable merozoites for subsequent asexual proliferation (Figure 4.13). Combined with the previous data showing aberrant development of asexual parasites (4.4) it is probable that the initial parasitaemia increase at 24 - 32hpi is a result of the presence of these aberrant asexuals that are not able to complete schizogony and produce viable merozoites for a second reinvasion event.
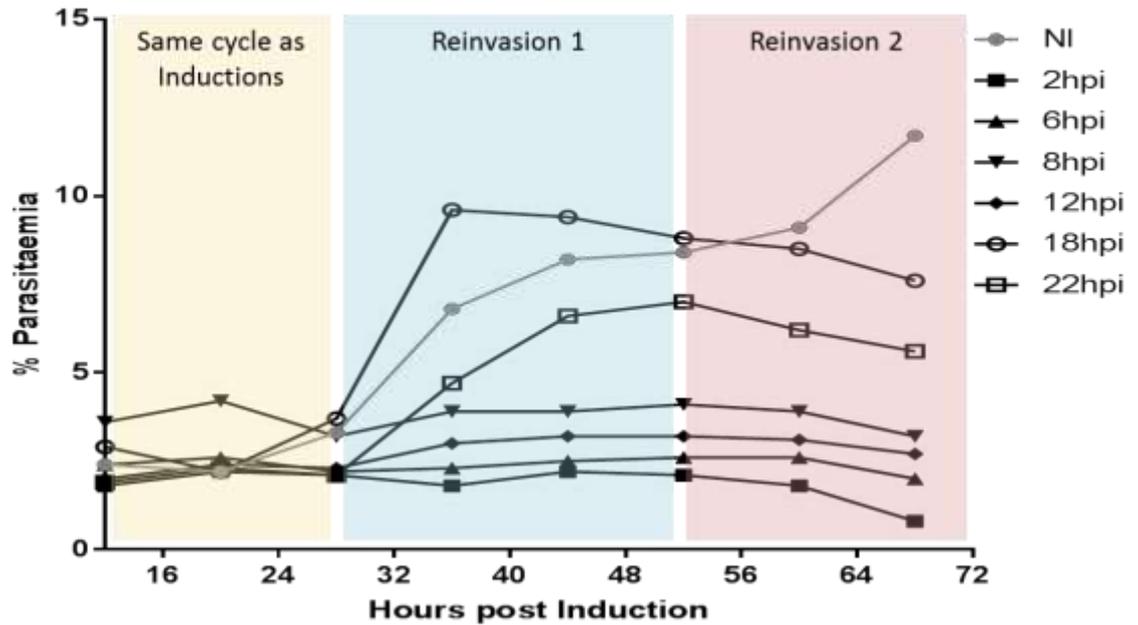
Figure 4.13 Tracking of parasitaemia after induction to overexpress *ap2-g*

Synchronous Inducible AP2g overexpression parasites were induced with 4mg/kg Rapamycin *in vivo* and the parasitaemia monitored. After induction only one cycle of reinvasion and parasitaemia increase occurred regardless of the time of induction. Clear increases in parasitaemia are evident at approximately 24hpi and 48hpi in the non-induced line (grey).

4.5.2.2 Identification of gametocytes post *ap2-g* induction

To determine if gametocytes can be formed without reinvasion, *ap2-g* was induced at various time points throughout the asexual cycle (2hpi, 6hpi, 8hpi, 12hpi, 18hpi and 22hpi) in a synchronous infection (Figure 4.12). In the same cycle as the inductions were completed we began monitoring parasite growth and the proportion of gametocytes in each induced mouse. It was expected that gametocytes would not be detected in the same cycle as inductions were completed as previous work indicated that in *P. falciparum,* invading merozoites were already pre-committed (Bruce *et al.*, 1990). However in *P. berghei* it had also been suggested that commitment was not fixed until the trophozoite stage which is the same cycle as the gametocytes develop (Mons, 1986b; Mons, 1986a).

Overall analysis of the detection of the proxies for male (GFP) and female (RFP) gametocytes showed that inductions resulting in overexpression of AP2-G led to gametocyte development without reinvasion for several induction time points (Figure 4.14). When inductions were completed at 2hpi, 6hpi, 8hpi or 12hpi, both GFP and RFP were detectable before reinvasion and maturation to the point of proxy expression could have occurred. This endpoint time was determined by GFP and RFP onset in the parental line; 36hpi for males, Figure 4.9, and 20hpi for females, Figure 4.11. When inductions were completed 18hpi or 22hpi no commitment was detectable without reinvasion.
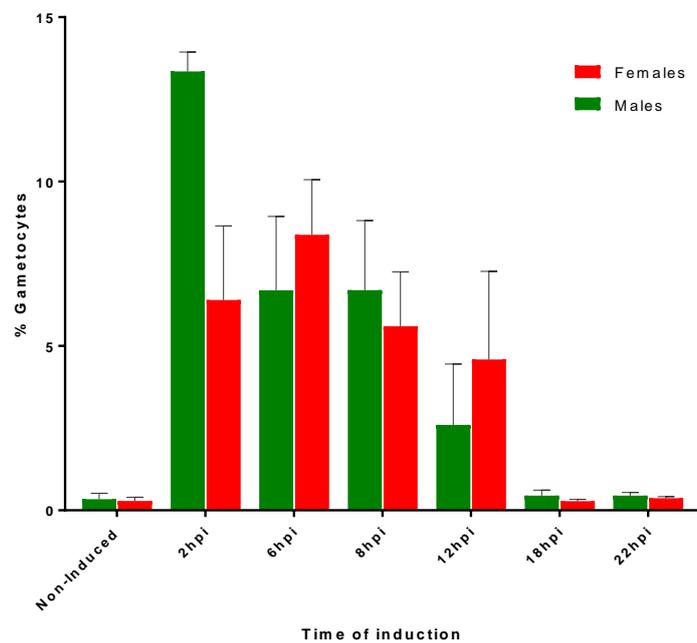
Figure 4.14 Gametocytes can be detected without reinvasion after ap2-g overexpression. Synchronous infections of an inducible *ap2-g* overexpressor were induced at multiple time points post invasion. Maximum gametocyte proportions were calculated based on GFP (male) and RFP (female) expression within the infected population. The proportion of gametocytes developing without reinvasion was calculated as the maximum percentage detected prior to possible detection of gametocytes formed after reinvasion (pre 36hpi for males and pre 40hpi for females). Early inductions, prior to 12hpi, resulted in gametocyte commitment and maturation without reinvasion. Late inductions, at 18hpi and 22hpi, resulted in no commitment prior to reinvasion.

In all induced lines, the population of gametocytes within the population increased after reinvasion (Figure 4.15). After this reinvasion, the gametocytes detected in the early inductions (2hpi, 6hpi, 8hpi and 12hpi) were a combination of gametocytes that had developed without reinvasion and those that had developed after reinvasion had occurred. The late inductions (18hpi and 22hpi) contained a single population of gametocytes maturing after reinvasion had occurred (Figure 4.15).
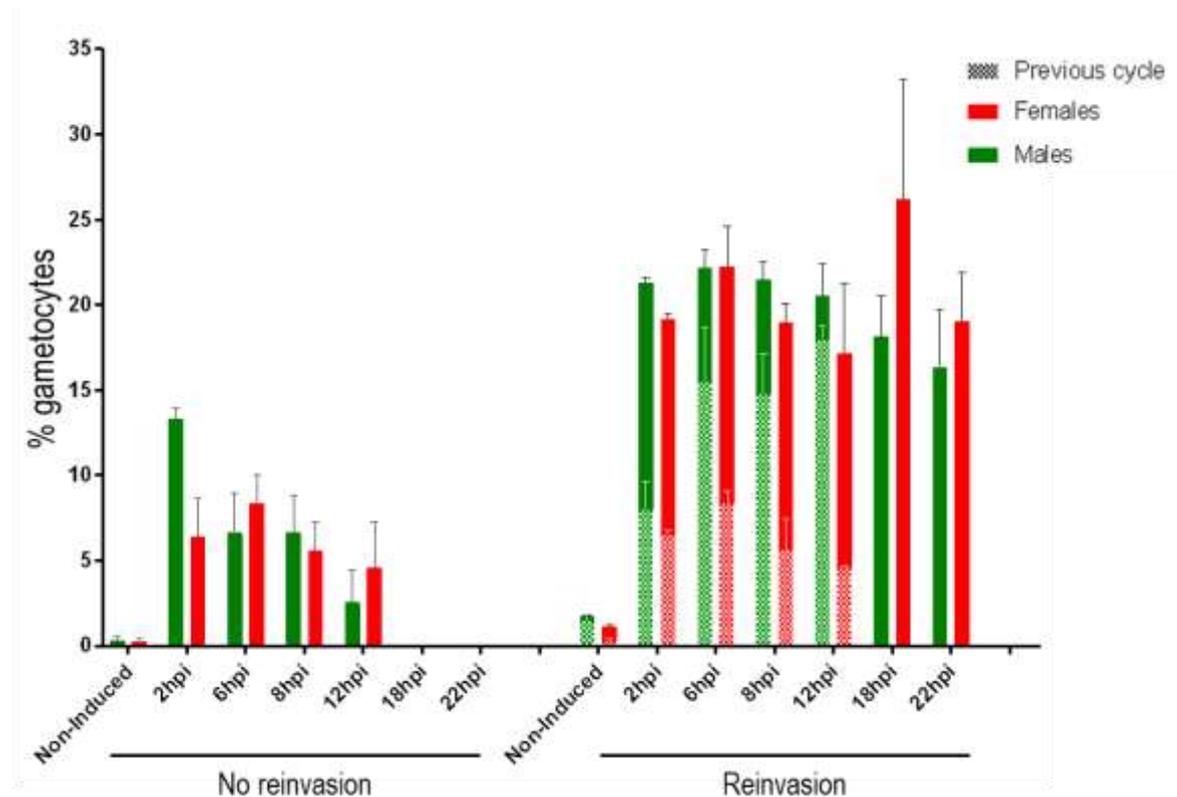
Figure 4.15 Gametocytes can be detected without reinvasion after *ap2-g* overexpression
Synchronous infections of an inducible *ap2-g* overexpressor were induced at multiple time points post invasion. Maximum gametocyte proportions were calculated based on GFP (male) and RFP (female) expression within the infected population. The proportion of gametocytes developing without reinvasion was calculated as the maximum percentage detected prior to possible detection of gametocytes formed after reinvasion (pre 36hpi for males and pre 40hpi for females). The proportion of gametocytes developing after reinvasion was calculated as the maximum percentage detected less the percentage of gametocytes remaining from the previous cycle (if commitment had occurred). In all induction time points until 12hpi gametocytes were readily detected without reinvasion. From 18hpi no commitment can occur without reinvasion. In the subsequent cycle additional gametocytes developed in all induction time points.

This primary analysis indicated that the point at which induction was completed could impact the populations of gametocytes being analysed. For example all inductions prior to 12hpi would contain two developmental stages after reinvasion whereas later inductions would not. To determine if there was any

variability in gametocyte maturation each of the induction time points was analysed to track the onset of the fluorescent proxies.

4.5.2.3 Characterisation of gametocyte maturation post *ap2-g* induction

Initial analysis showed that induction of the overexpression of *ap2-g* can lead to the development of gametocytes without reinvasion if induction was completed prior to 12hpi. To determine if any developmental defects were obvious in these maturing gametocytes, the time of onset of the gender specific proxies was analysed. Any delay in the detection of the onset of the proxy prior to reinvasion could indicate a delay in the maturation process indicating development may not be typical. After reinvasion the induced parasites who could not, or did not, commit in the same cycle as induction may also show growth defects and the proxies were also used to ascertain this.

Based on previous work identifying the onset of the fluorescent proxies we estimated that if no delay in development occurred then the previously determined proxy onsets would be maintained. Briefly, GFP fluorescence (specific to male gametocytes) should be detected from 12 – 16hpi, if reinvasion was not required for gametocyte development, or 36 – 40hpi, if reinvasion was necessary for gametocyte development; RFP fluorescence (low RFP signal specific to immature female gametocytes) should initially be detected from 20 – 22hpi, if reinvasion was not required for gametocyte development, or 40 – 44hpi, if reinvasion was necessary for gametocyte development; High-RFP fluorescence (high RFP specific to mature female gametocytes) intensity is reached, reclassifying parasites as mature females, at 36 – 40hpi with the second wave reaching the same intensity at 52 – 60hpi (Figure 4.16 A).

Each of the induction time points showed a degree of difference to the control parental line, not overexpressing *ap2-g*. To ascertain which induction time point would yield the most representative, gametocytes delay in proxy appearance was compared (Figure 4.16). As previously shown, induction of *ap2-g* expression at 2hpi resulted in gametocyte commitment and development without reinvasion.

Induction at 2hpi

Male gametocytes were initially detected 16hpi a delay from 12hpi in the parental line. Female gametocytes (low RFP) were initially detected 28hpi a delay from 20hpi. High RFP expressing female gametocytes showed no delay in detection after induction remaining similar to the parental line at 36hpi. The second wave of gametocytes, formed after merozoites had reinvaded, was detected in the population as a wave distinct to those matured from commitment without reinvasion. Male gametocytes showed a minor delay in detection from 36hpi (control) to 40hpi. Female gametocytes from this second wave were considerably delayed in their initial, low RFP, detection from 40hpi (control) to 60hpi. This delay reduced for high RFP female gametocytes from 52hpi (control) to 64hpi in the induced line (Figure 4.16 C).

Induction at 6hpi

Male gametocytes showed a more pronounced delay in same cycle gametocyte proxy onset than those induced at 2hpi, not being detected until 20hpi. Female gametocyte proxy detection, both low and high RFP, show responses consistent to induction at 2hpi, with the low RFP population not being detected until 28hpi and the high RFP population consistent with the control at 36hpi. Likewise after reinvasion the development of male and female gametocytes, as determined by the fluorescent proxies, reflects that of induction at 2hpi (Figure 4.16 D).

Induction at 8hpi

For gametocytes developing without reinvasion the onset of all fluorescent proxies was unchanged from an induction at 6hpi with the three proxies being detected at 20hpi, 28hpi and 36hpi respectively. The second wave of male gametocytes that developed after reinvasion were detectable and distinct from the first wave of male gametocytes developed without reinvasion at 32hpi which is before they are detectable in the control line (36hpi). The second wave of female gametocytes were significantly delayed in low RFP intensity signal and slightly delayed in high RFP intensity signal with both being detected from 60hpi compared to 40hpi and 52hpi in the control line respectively (Figure 4.16 E).

Induction at 12hpi

Male gametocytes developing without reinvasion when inductions were completed at 12hpi show the greatest delay in fluorescent proxy onset, not being discernible until 28hpi. Similarly initial detection of RFP (low RFP) is also the most delayed at 32hpi. However detection of high RFP female gametocytes was not delayed compared to the control line which is consistent with all other inductions. The second wave of male gametocytes, developed after reinvasion, mature at the same rate as the control line with the GFP signal being detectable from 36hpi. The second wave of female gametocytes showed a significant delay in the initial detection of the low RFP with detection not occurring until 56hpi compared to 40hpi in the control line. The delay in detection of the high RFP population was less, being detected at 60hpi instead of 52hpi in the control line (Figure 4.16 F).

Induction at 18hpi

Gametocytes of either gender were not detected with inductions completed at 18hpi. Only gametocyte development after reinvasion can be compared to development in the control line. With this induction there was no delay in the onset of the GFP signal or the low RFP signal. However, detection of the high RFP signal was significantly earlier than in the control line (52hpi) with these high RFP female gametocytes being detected from 44hpi (Figure 4.16 G)

Induction at 22hpi

Gametocytes of either gender were not detected with inductions completed at 22hpi. Only gametocyte development after reinvasion can be compared to development in the control line. With this induction there was no delay in the onset of the GFP signal or the high RFP signal. There was a minor delay in the initial detection of low RFP female gametocytes from 40hpi (control) to 44hpi (Figure 4.16 H).

% of parasite population

Hours post Induction
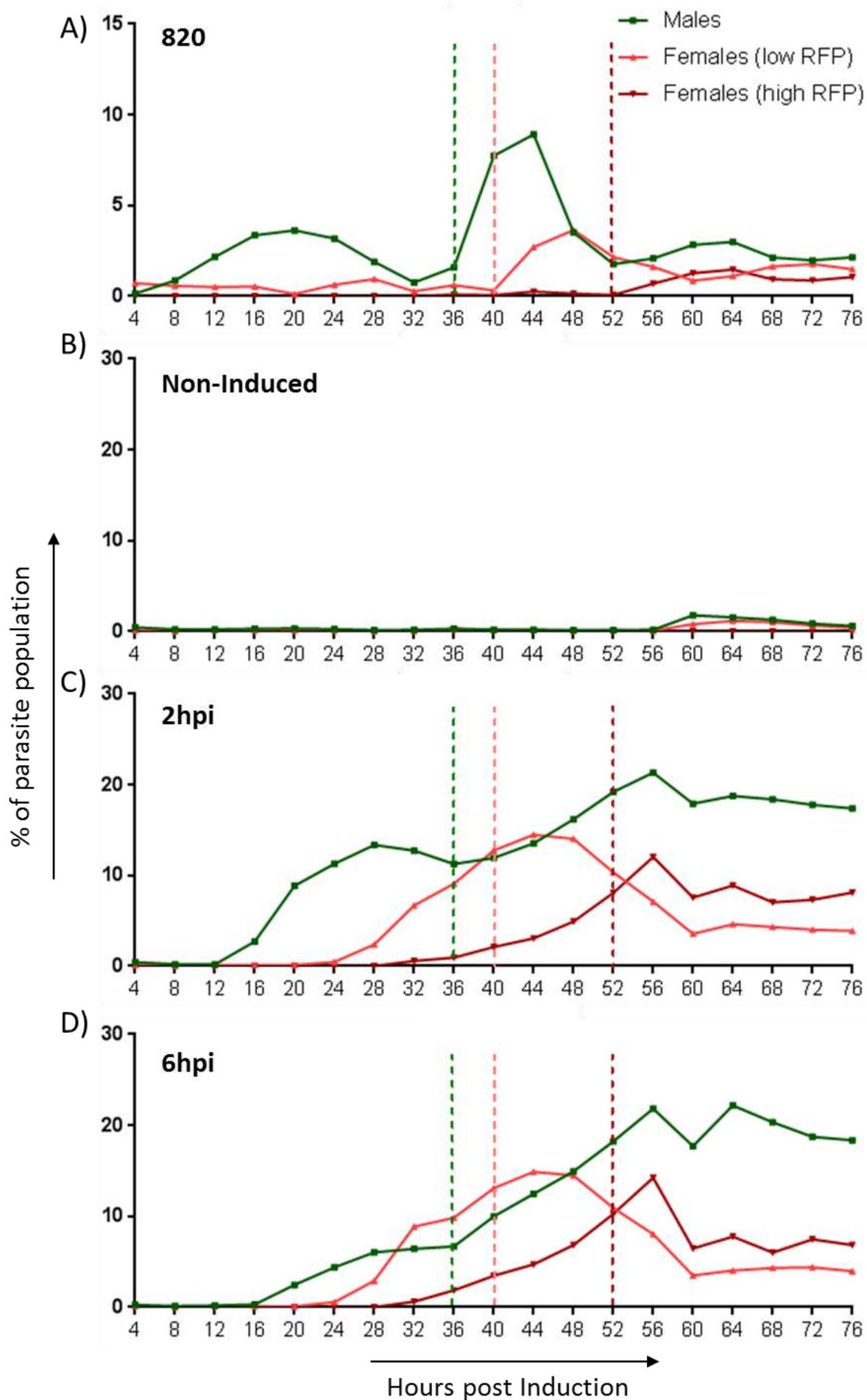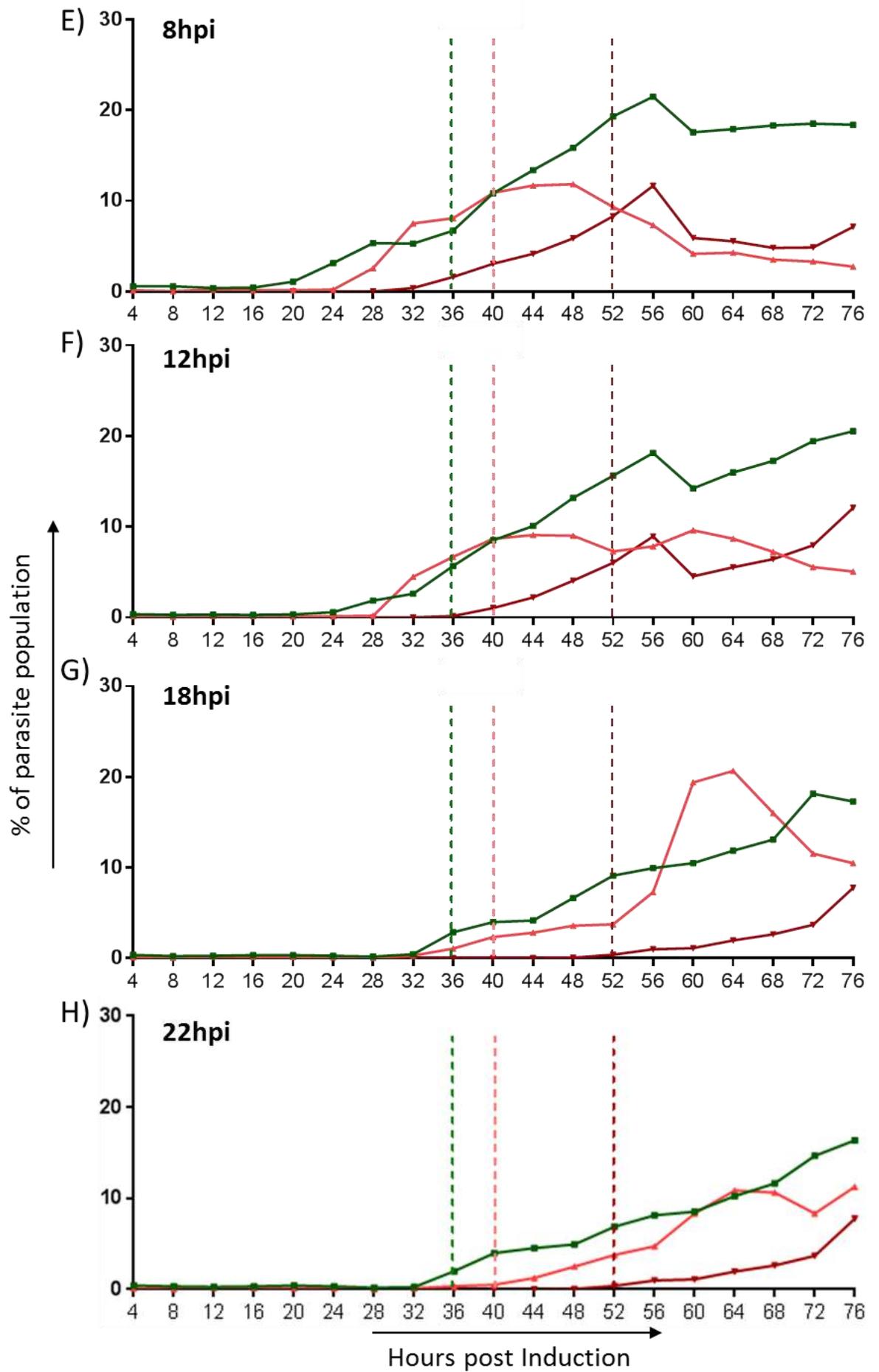
Figure 4.16 Fluorescent proxy detection in the induced *ap2-g* overexpressor

Synchronous inducible *ap2-g* overexpression parasites were induced with 4mg/kg Rapamycin *in vivo* at various time points post invasion (hpi). Every four hours post synchronous infection the proportion of parasites expressing GFP, low RFP and high RFP were calculated by flow cytometry analysis. The onset of detection of the fluorescent proxy and therefore the presence of gametocytes is classified as when the population exceeds 1%.

On each induction graph (A – H) the earliest gametocytes can be detected after reinvasion is indicated by a dashed line representing the fluorescent proxy.

Graphs A - H show the fluorescent populations, denoting male (GFP), female (low RFP) and mature female (high RFP), over time post induction. Each graph represents a specific induction time point after invasion (hpi).

Summary of induction

From the above analysis it is clear that the time at which *ap2-g* is overexpressed has a profound impact on commitment and gametocyte development both in the cycle that *ap2-g* begins expression and subsequent cycles post reinvasion. The overexpression of AP2-G is clearly detrimental to asexual development long term (Figure 4.13). High levels of AP2-G in early ring stages can result in commitment to male and female gametocytes without reinvasion (Figure 4.14), however these gametocytes do not mature with the same kinetics as parental or wild type gametocytes determined by the onset of known fluorescent protein proxies for both genders (Figure 4.16). Induction early in the initial cycle also has an effect on subsequent commitment and gametocyte maturation in the next cycle where the merozoites that reinvade have already been exposed to AP2-G. The variability in detection of gametocytes (Figure 4.5) varies significantly dependant on when induction has occurred. However it is clear from these analyses that induction at 22hpi results in a single wave of commitment. Post-reinvasion these gametocytes mature with similar kinetics to the 820 parental line, only with a higher proportion of the population committing to gametocytogenesis (Table 4.1).

sample		Male Gametocytes	Female gametocytes (low RFP)	Female gametocytes (high RFP)
820	Cycle 1	12	20	36
	Cycle 2	36	40	52
2hpi	No reinvasion	16	28	36
	Reinvasion	40	60	64
6hpi	No reinvasion	20	28	36
	Reinvasion	40	60	64
8hpi	No reinvasion	20	28	36
	Reinvasion	32	60	60
12hpi	No reinvasion	28	32	36
	Reinvasion	36	56	60
18hpi	No reinvasion	n/a	n/a	n/a
	Reinvasion	36	40	44
22hpi	No reinvasion	n/a	n/a	n/a
	Reinvasion	36	44	52

Table 4.1 Sumary of gametocyte development after induction
Synchronous inducible *ap2-g* overexpression parasites were induced with 4mg/kg Rapamycin *in vivo* at various time points post invasion. The time at which GFP could first be detected was considered the onset of male gametocyte development (Males). The time at which RFP could be first detected was considered the onset of female gametocyte development (Immature females). When the RFP signal reached maximum intensity and plateaued, the female gametocytes were considered mature (Mature females). Each time point of induction was compared with the 820 parental line, Maturation 1 of this line was compared to the gametocytes arising from the same cycle (no reinvasion) and Maturation 2 of the parental line was compared to the gametocytes formed after reinvasion and commitment. The most similar, or representative time point of induction was 22hpi which gave rise to one round of commitment post reinvasion with similar kinetics of fluorescent onset as the 820 parental line.

To ensure that when considering gametocyte commitment and development we are analysing the most representative population, with only one stage, all subsequent work will be completed with schizonts induced with rapamycin at 22hpi. This will lead to the development of a single wave of gametocytes with development kinetics that resemble most closely and consistently the 820 parental line.

4.6 **Discussion**

Generation of the inducible *ap2-g* expresser was the foremost requirement for this project. Testing of the line has shown it to correlate to the speed and efficiency of the reporter line. The rapidity and efficiency of the flip of the promoter demonstrates that the system will give us stringent control over the onset of gametocyte commitment through rapid flipping of the *hsp70* promoter to drive the transcription factor (*ap2-g*) responsible for initiating gametocytogenesis.

This rapid and highly efficient promoter flipping was demonstrated by qPCR where the proportion of flipped DNA obtained from the population increased rapidly after induction and reached the predominant form in the population quickly (around 4 hours) indicating the onset of gametocytogenesis will be relatively synchronous within the population. However, it is worth noting that while the proportion of the donor template DNA (obtained directly from *in vivo* parasite infection) for the flipped population rapidly increased we also noted no substantial decrease in the proportion of donor template allowing amplification of the unflipped product in a qPCR setting. In contrast to this, when using the same template DNA and amplification primers no actual PCR product can be amplified for the unflipped product from the same template DNA that was harvested 30hp-Ind. Likewise the flipped product cannot be amplified from DNA harvested from a line that has never been induced. These contradictory pieces of evidence, combined with similar contradictory results obtained with the reporter line (*in vivo* only), most likely indicate that no specific product was amplified by the primers specific to the unflipped DNA but the increased sensitivity of qPCR detects a non-specific product and the template for this non-specific product is only found in DNA harvested directly from *in vivo* parasite samples (Figure 4.4).

The evidence that the induction was very successful at inducing AP2-G expression becomes even clearer when quantifying commitment to gametocytogenesis based on the fluorescent proxies (GFP for males gametocytes and RFP for female gametocytes). After induction, variability in the generation of gametocytes was evident with low parasitaemias giving rise to lower commitment and higher variability in male to female ratios. This variability with low parasitaemias also occurs in the parental line (Figure 4.5). However, gametocyte commitment has been recorded as high as 70% after induction and is on average 48% (with an approximately equal distribution of males and females at 24.6% and 23.5% respectively). Though some of the variability in gametocyte commitment can be explained by the parasitaemia at the time of induction there also exists an innate degree of variability. In rodent and human malarias the variability in commitment throughout an infection and between strains and clones has been evaluate. It has been shown that commitment is a dynamic process and exhibits phenotypic plasticity within an infection. This plasticity means that responses to stimuli and environment as well as developmental cues can result in modulated commitment each cycle. In a complex life cycle that relies on a balance between commitment to gametocytogenesis for transmission to naïve hosts and asexual proliferation to maintain an infection in an immunocompetent (often naturally infected with multiple genotypes of the same species and potentially a multitude of different pathogens) host this plasticity in response provides mechanisms ensuring survival.

Some factors shown to influence gametocyte commitment should not be relevant for laboratory environment infections including; competition between parasites with different genotypes, because infections are seeded with a single genotype; infection with other species leading to an altered immune response, while animals are not housed in a sterile environment they are clean and should be infection free; sub-curative drug treatments, no additional drug administrations are given to animals post selection (Dixon *et al.*, 2008; Pollitt *et al.*, 2011).

Other factors have been shown to elicit a specific response in rodent and human malarias that may not intuitively translate to *P. berghei*. Primarily this would be the response of *P. falciparum* to increased reticulocytes in anaemic patients and *P. chabaudi* to increased reticulocytaemia in mice treated with the hormone erythropoietin (EPO). In infections with more reticulocytes in the bloodstream

an increase in gametocyte commitment has been observed (Reece *et al.*, 2008; Reece *et al.*, 2005). In extreme cases the ratio of male to female gametocytes can also alter, presumably to facilitate successful fertilisation and therefore transmission through the mosquito vector (Reece *et al.*, 2008). While these data suggest the response to severe anaemia is to increase gametocyte commitment, in a potential terminal investment strategy, the overall preference of *P. berghei* for reticulocyte invasion may preclude this response (Reece *et al.*, 2009).

Potentially contributing to the variability apparent in the parental strain and the induced *ap2-g* inducible overexpresser are variant gametocyte commitment due to the immune response, the resources available, parasite burden and the nutritional status of the host.

A potential simulation of increased parasite density in the host was first achieved by the replacement of *P. falciparum* culture media with conditioned media. This resulted in an increased commitment to gametocytaemia which indicated sensing of parasite density was achieved by monitoring of cell derived factors secreted or released in the culture media (Williams, 1999). A similar induction of gametocyte commitment was achieved by the culture of parasites with lysed parasite infected red blood cells (Schneweis *et al.*, 1991). Taken together these indicate that after red cell lysis (by asexual parasites egress) parasite derived factors are released and elements of this are perceived by intracellular parasites to indicate density of the hosts infection which in turn informs the decisions made for commitment. This phenomenon was further exacerbated in mixed infections where multiple genotypes are present (Carter *et al.*, 2013; Reece *et al.*, 2008).

The immune system of rodents has been shown to have an effect on *P. chabaudi* and *P. yoelii* gametocyte commitment. Infection of pre-immunised animals with either of these rodent infective species increased early gametocyte commitment in mice. In *P. yoelli* this increased gametocyte commitment was in fact correlated to an increase in hepatic schizonts releasing gametocytes compared to the wild type (Motard *et al.*, 1995). In *P. chabaudi* pre-immunisation reduced parasite density in the blood stream of the host giving a similar magnitude effect on both gametocytes and asexual blood stages. However, the immunisation did not have a strain specific effect on gametocytes and the increase in gametocyte

commitment in pre-immunised mice occurred faster (day 7) than in non-immunised mice (day 11) implying pre-immunisation may give rise to an initial increase in gametocyte commitment (Buckling & Read, 2001). In *P. falciparum* cultures increased gametocyte commitment can also be induced by the addition of steroid hormones to culture (Lingnau *et al.*, 1993) or by the addition of serum and lymphocytes obtained from infected hosts (Smalley & Brown, 1981). While the exact factors driving increased gametocyte commitment in pre-immunised animals or cultures containing immune serum has not been demonstrated for *P. berghei* the consistent response in two rodent malarias and *P. falciparum* indicate that immune modulation during an infection may have an impact on gametocyte commitment.

While in the inducible system we are controlling the onset of gametocytaemia by preventing AP2-G expression until induction, when we are modulating the system to express extremely high levels of the transcription factor, we cannot completely control the environment in which the infection resides. Factors that have influenced the cell before and after the onset of AP2-G expression are as variable for the inducible overexpresser as they are for a wild type population. It is therefore not surprising that the degree of variability in commitment remains in the inducible system as well as in natural commitment which has never been shown to occur in a synchronous manner.

In all cases after induction of AP2-G increased commitment was observed (Figure 4.5). In this system we are not only artificially initiating expression of *ap2-g* potentially at a stage when it is not normally expressed but we are also expressing it to a level which (based on transcription levels) is likely between 100 fold and 6,800 fold increase in expression levels dependant on the stage (Hall *et al.*, 2005). This of course insights concerns about the viability of any gametocytes generated after such an over expression of the transcription factor initiating the cascade of commitment. While we make no claims that all gametocytes generated from overexpression of AP2-G are viable, transmission through mosquitoes and the establishment of an infection in a naïve host demonstrates at least a proportion of the gametocytes are viable. Future work using the constitutively fluorescent lines will allow full quantification of ookinete conversion rates (which have been indicated to be relatively low *in vitro*), oocyst development in mosquitoes, salivary gland colonisation and

infection of hepatocytes *in vivo*. Currently our understanding is limited to possible but far less efficient than a wild type infection.

In an attempt to analyse transcriptional differences in synchronous and representative gametocytes when compared to asexual parasites we have characterised the maturation of gametocytes post induction carefully (see section 4.5.2). To hone in on the factors influencing commitment immediately after AP2-G sets off the cascade responsible for commitment it was necessary to obtain a single wave of gametocytes that matured in a similar way to wild type gametocytes. While morphologically indistinguishable from asexual parasites until late in the cycle gametocytes are identifiable based on fluorescent proxy expression in the 820 parental line (GFP expression driven by the male specific PBANKA_041610 promoter and RFP expression driven by the female specific PBANKA_131950 promoter) earlier. In depth characterisation of the onset of these fluorescent proxies (see section 4.5.1) gave a point of reference for gametocyte development with overexpression of AP2-G. Not only was this analysis useful for characterising developmental differences between the induced *ap2-g* overexpresser but may provide standard developmental timings (in the 820 parental line) to be used in subsequent phenotyping analysis. One observation that was reassuring to the direct impact of AP2-G overexpression on the actual cascade resulting in commitment was the correspondence of the expression levels of the fluorescent proxies in the gametocytes formed after overexpression. The intensity of the fluorescent proxies was consistent in the induced lines and the parental line. While this doesn't confirm that proteins downstream of *ap2-g* in the commitment cascade are not dysregulated it does demonstrate that these gametocyte specific genes seem consistently expressed in individual parasites in the overexpresser.

While we have not actively investigated the reasons for the delays in maturation as determined by the onset of the proxies reasons for these delays can be hypothesised with some support from the initial experiments. For example, while *ap2-g* induction can result in commitment in the same cycle for wild type development kinetics to be observed AP2-G must be present in the cell the cycle before commitment. This would explain why the parasites unable to commit in the same cycle as induction show the most similar onsets of both proxies to the control line. The pattern of male fluorescent proxy onset delay could also

support this as the delay in onset seems related to the time after invasion that induction was completed. In this case it could be suggested that the pathways resulting in male gametocyte development require a fixed amount of time to complete and AP2-G begins this cascade when the ring invades, as would be the case for reinvading parasites with wild type AP2-G expression, pre-induced AP2-G oxer-expressors or when AP2-G begins expression after induction.

The female gametocyte specific promoter does not show this correlative delay in onset, being always delayed to onset at 28hpi when induction same cycle results in gametocyte development. This could indicate that when induction occurred post merozoites invasion, regardless of when *ap2-g* expression was induced, the promoter driving RFP cannot become active until this time point (28hpi). Additionally, there was a reduced time between RFP onset (low RFP) and the increasing intensity to high RFP, which remained consistent with the control regardless of RFP onset.

While we endeavoured to optimise induction of AP2-G to give rise to synchronous and representative gametocytes, the necessity of these gametocytes to mature into viable mature gametocytes is also not fundamental to this work. As outlined in the aims the control of the onset of gametocyte commitment and the over-production of gametocytes was attempted to facilitate study of the initial stages of commitment a stage where very little is actually known. Comparatively, much is already known about transcription and protein expression in mature gametocytes obtaining data from more natural systems (Hall *et al.*, 2005; Khan *et al.*, 2005; Silvestrini *et al.*, 2005).

One of the current questions, arising from this work, which we are interested in is the difference between the parasites able to commit and those that seem to respond adversely to the signal to commit (Figure 4.5). Understanding the differences in this response may give understanding to the mechanism that underlie commitment and so called bet-hedging that prevents absolute commitment occurring allowing all parasites to terminally differentiate into the sexual stages. It would be of interest to identify potential differences in not only gene expression in the aberrant parasites, formed after induction, that are not viable but also their genome structure. It would also be interesting to establish if alterations to the environment that that a parasite has been in before and

after commitment to see if this alters the proportion of parasites able to commit to gametocytogenesis versus the proportion that cannot respond appropriately and therefore do not completely develop into gametocytes.

Another caveat that is interesting is the observation that no matter when they are induced AP2-G overexpressing parasites seem only able to complete one reinvasion to generate viable gametocytes (Figure 4.13). If a ring stage parasite (2hpi) is induced to overexpress AP2-G a proportion of the parasites become gametocytes that same cycle (without reinvasion) and a portion go on to make schizonts that release merozoites to reinvade reticulocytes and a second wave of commitment occurs. In this cycle however schizonts do not develop normally and parasitaemia does not increase (Figure 4.13). If a schizont (22hpi) is induced to overexpress AP2-G no commitment in the same cycle is observed, as would be expected, and the already almost mature schizonts rupture releasing merozoites that invade reticulocytes and a high proportion of these commit to gametocytogenesis. The remaining parasites that have not committed also seem unviable and a subsequent increase in parasitaemia does not occur. While this could simply be an artefact of the overexpression of AP2-G it could point to a regulatory role of downstream factors that control asexual proliferation and development as well as providing signals to differentiate into gametocytes and control this cascade.

Furthermore of interest are the parasites that make it through the mosquito. These appear to be gametocyte non-producers in the naïve host. However, to transmit they must have been gametocytes. To have been gametocytes they must also have been expressing AP2-G at high levels (having flipped the *hsp70* promoter). It would be interesting to determine factors that are subsequently regulating expression to make the line appear *ap2-g* null when *hsp70* should be driving expression.

Finally the resurgent parasites that arise after the vast majority of the population has both committed and failed to transmit or not-committed and not progressed through the asexual life cycle could also provide insight into the alternate mechanisms of control of commitment or *ap2-g* expression. We have identified that this resurgent population is a gametocyte non-producing line (based on lack of gametocyte specific fluorescent proxy expression) that is made

up of a mixture of non-responsive parasites (those that have not flipped the *hsp70* promoter to drive AP2-G expression) and parasites that should be overexpressing *ap2-g* based on the promoter orientation. It would be interesting to determine if AP2-G is actually being expressed by these parasites that do not appear to generate gametocytes and if not how the parasite is silencing a constitutively expressing promoter.

The development of this line has provided the novel tool required for the study of the early stages of commitment after induction of AP2-G expression. It has also raised many questions about gametocyte development and alternate mechanisms employed by the parasite to seemingly prevent 100% commitment to the terminally differentiated sexual stages that cannot maintain an infection in the current host.

To facilitate future work two methods are being employed to obtain a conditionally overexpressing AP2-G line in which the AP2-G is C-terminally tagged with a fluorescent protein. Previous work identified that AP2-G::GFP parasites are able to produce normal amounts and transmissible gametocytes indicating it does not impede function but that the GFP signal was not detectable at any life cycle stage (Sinha *et al.*, 2014). One method we are using to generate this inducible line (Figure 4.2) adds the *gfp* tag to the C-terminus of the *ap2-g* gene that is used as the homology arm for integration of the floxed *hsp70* promoter. Upon successful integration the promoter of *ap2-g* was interrupted with the inducible promoter and the *gfp* tag was integrated at the c-terminus of *ap2-g*. While we have been able to obtain parasites that express GFP to detectable levels in morphologically identifiable gametocytes (KM) we do not obtain populations that are gametocyte non producers before induction or that generate the overrepresentation of gametocytes (to the same degree) as we get with the inducible overexpresser (Supplementary Figure 1 A). This is likely due to the size of the 3' homology arm integration required to correctly integrate both the inducible promoter and the *gfp* tag (> 7kb). We postulate that in the population after integration of this vector there contains a mixture of parasites with a GFP tagged AP2-G but the native promoter, a proportion with the native promoter and no GFP tag, a proportion with no GFP tag and the inducible, interrupting, promoter and a proportion with a GFP tag and the inducible promoter. The final population are identifiable as gametocytes when they are

mature and a distinct GFP signal is observed in two foci (Supplementary Figure 1 B) one of which co-localises with the Hoechst stain of the nucleus. However this mix of different loci means we do not have the same level of control over commitment and the gametocytes generated after induction will be both AP2-G overexpresser and AP2-G expressers. To combat this a second method is now being pursued to initially *gfp* tag the native *ap2-g* before recycling the selectable marker and integrating the inducible promoter. It is expected that this will yield induction rates equivalent to the non-tagged line generated and that all expressed AP2-G will be tagged (AP2-G::GFP).

This line is necessary for further study of direct interactions of AP2-G after initiation of commitment and may provide valuable insight into protein levels in different life stages after induction.

Chapter 5   Identification of novel gametocyte specific factors

5.1 Introduction

Several studies have been completed to evaluate gene expression at different P. berghei life stages (Hall *et al.,* 2005; Otto *et al.,* 2014). In these studies synchronous samples of asexual and sexual stage parasites were transcriptionally profiled to quantify gene expression at each of the stages. Hall et al sampled synchronous populations at ring stage, young trophozoites, mature trophozoites, immature schizonts and mature schizonts. They also purified immature (24hpi) and mature (30hpi) gametocytes. Microarray analysis was carried out and combined with proteomics data obtained from mixed asexual stages, purified gametocytes, ookinetes and ooycsts (day 9 – 12 post mosquito infection) to characterise function and expression specificity of the genes analysed. Otto et al sampled synchronous blood stage parasites at the ring, trophozoite and schizont stage as well as gametocytes and ookinetes. Here RNAseq data was analysed to compare gene expression levels at each of the analysed stages.

In these studies 1,094 transcripts and 733 proteins were obtained from gametocyte samples. Of these 127 proteins were uniquely detected in the gametocyte samples. One of the limitations with these studies is the methodologies that exist for gametocyte purification which is not possible before 24hpi. Analysis of immature gametocytes and a comparison of these to the other stages might elucidate transient protein expression that is not present in the mature stages.

Previous work on protein expression differences in P. berghei blood stages has identified specific proteins expressed in the mature male and female gametocytes (Khan et al., 2005). In this study highly pure populations of male and female gametocytes were purified by their differing GFP intensities in different reference lines. In the reference lines asexual parasites and male and female gametocytes express gfp with different fluorescence intensities (Franke-Fayard et al., 2004). Mature gametocytes were initially separated using a 48% Nycoprep gradient, then highly pure male and female gametocytes populations were purified by flow cytometry. When gfp is expressed by the ef1α promoter it

is four times more intense in female gametocytes (compared to males), when it is driven by the α-tubulin-II promoter male gametocytes display fluorescence of approximately six times more intense (than female gametocytes) (Khan et al., 2005). While this study elucidated proteins expressed in gametocytes along with asexual parasites and unique gametocyte, male or female proteins the gametocytes sampled would always be mature. This is a limitation of the purification methodologies.

While these studies have identified many genes expressed by the mature gametocytes and proteins important for their maturation, a limiting factor has always been the ability to identify and purify early stage gametocytes from the blood stage population. Identification of *ap2-g* as the transcription factor that initiates the cascade resulting in commitment to gametocytogenesis gives the opportunity to control the onset of differentiation (Sinha et al., 2014). Overexpression of this transcription factor has been shown to vastly increase commitment in a synchronous manner (see section 4.5) which will allow transcriptional profiling of gametocytes from the onset of commitment and throughout maturation, stages that have previously been unavailable for study.

5.2 Transcriptional comparison of gametocytes

5.2.1 Time course of RNA sampling

As changes in gene expression during commitment and gametocyte development are likely to be quite dynamic a pilot time course was initially completed to evaluate key times that should be analysed (Figure 5.1). All infections in this time course were established from the same synchronous expansion of an inducible *ap2-g* overexpressor line. Half of the mice were intraperitoneally injected with 4mg/kg rapamycin 22 hours after the synchronous intravenous infection to induce *ap2-g* overexpression and trigger commitment to gametocytogenesis. A 22hpi time point for induction was chosen as this gave rise to a single wave of gametocyte commitment whereby both male and female gametocytes appear to develop in a manner consistent with the 820 parental line (see Chapter 4 for full analysis of the *ap2-g* overexpressor). This consistent maturation was determined by the onset of two fluorescent proxies for gametocyte development. In these lines, male gametocytes express GFP driven

by the male specific dynein promoter (PBANKA_041610) and in female gametocytes RFP expression is driven by the female specific LCCL (CCP2) promoter (PBANKA_131950). If the *ap2-g* overexpressor line is induced with rapamycin *in vivo* at 22hpi, merozoites reinvade and in the next development cycle the onset of these gametocyte proxies mirrors the 820 parental line. As reinvasion occurs in a manner consistent with the parental line and gametocyte development is the same, as far as we can currently tell, we hypothesise that the initial stages of commitment, gender assignation and gametocyte development will be representative of wild type gametocytes.

At each time point two mice (one rapamycin induced and one non-induced) were harvested and the blood processed. The pilot time points were chosen as representative for development stages within the asexual life cycle and while gametocytes are developing (Figure 5.1). As stage specific RNAseq datasets are available in the lab (prepared by Agnieszka Religa (AR)) it was possible to compare these time points (where the stage can be determined or estimated) with raw RNAseq datasets obtained from samples of rings, trophozoites, schizonts, gametocytes and ookinetes (Otto *et al.*, 2014). This would allow us to determine which further time points would be of value to sample to elucidate novel transcripts and potential points of commitment.
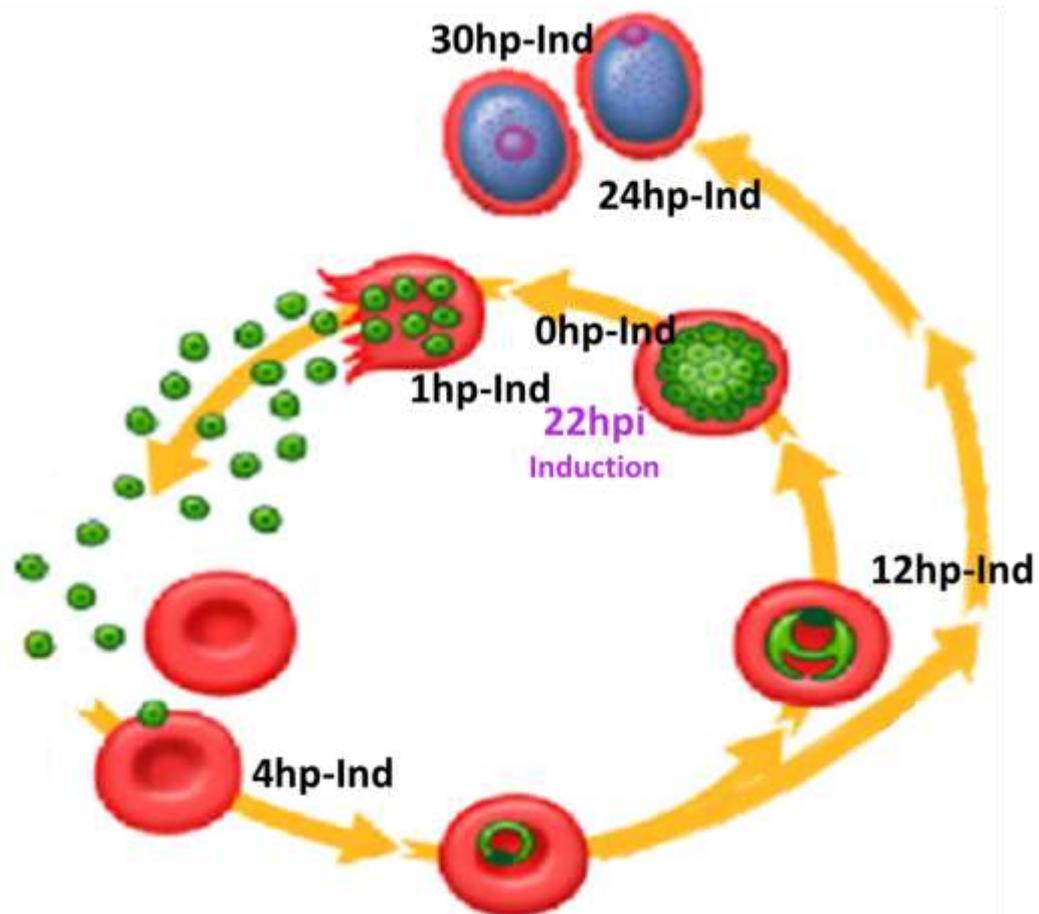
Figure 5.1 Pilot RNAseq time course to identify gametocyte specific transcripts.

All inductions were completed at 22hpi with a single IP injection of 4mg/kg rapamycin. Blood was harvested by cardiac puncture at each time point from one mouse who had received a rapamycin induction and one that had not. All subsequent analysis was completed on these paired samples.

At each time point post induction (hpi) Giemsa smears were analysed to determine the stage of parasite development in the population and to ensure the populations were synchronous (Figure 5.2). As expected at 0hp-Ind very few parasites were visible in the circulating blood of the mice. The few parasites that were visible in the thin blood smear were segmented schizonts that had not sequestered. Likewise at 1hp-Ind schizonts could be seen still circulating and few ring stage parasites could be seen in both the – and + rapamycin populations. By 4hp-Ind all parasites were ring stage in the blood and a clear increase in parasitaemia was evident. At 12hp-Ind the population was comprised of trophozoites in both populations. By 24hp-Ind differences in the stages are clear between the non-induced (- R) and the induced (+ R) samples. Without induction, circulating schizonts were clear within the bloodstream along with young ring stage parasites that had recently invaded. However in the induced sample, late trophozoites/early gametocytes were clearly visible with very few schizonts being identified. The clearest difference came when examining the

30hp-Ind parasites. By this point in the non-induced samples rings and early trophozoites were evident in the smears but in the induced sample gametocytes were morphologically identifiable with very few asexual stage parasites detected in the smears.
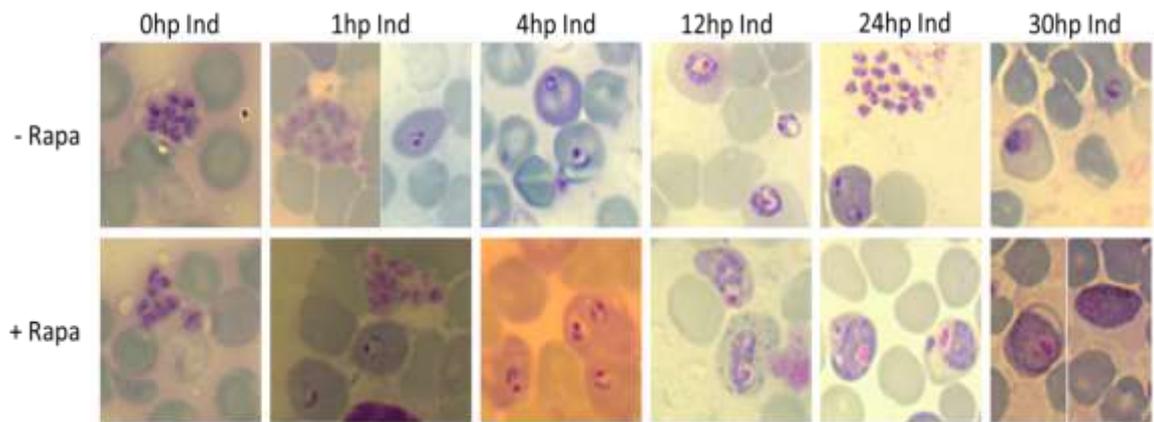
Figure 5.2 Giemsa blood smears of time points analysed in the pilot time course.

All inductions were carried out at 22hpi with an IP injection of 4mg/kg rapamycin. At the point of induction (0hp-Ind) schizonts are visible in the thin blood smears but are rare due to the high proportion sequestered. At 1hp-Ind few parasites have ruptured from schizonts and reinvaded but schizonts that have not ruptured are still identifiable. By 4hp-Ind reinvasion has completed. A dramatic increase in parasitaemia is clear and all infected red blood cells (+ and – rapamycin) contain ring stage parasites. At 12hp-Ind trophozoites are evident in both the + and – rapamycin samples and rare ring stage parasites persist in both. By 24hp-Ind differences in the – and + rapamycin smears become evident. Without induction of *ap2-g,* schizonts and newly invaded ring stage parasites are clear but with induction what appear to be late stage trophozoites (likely immature gametocytes) are the predominant form with occasional schizonts present. By 30hp-Ind the non-induced sample contains only rings and young trophozoites but the induced sample contains clear gametocytes with few asexual parasites being found.

Based on these giemsa smears it appears that the induced and non-induced samples are developing within the normal time range (24 hour cycle) and that there is significant commitment to gametocytogenesis post induction. This correlates with the previous fluorescence profiling time course, which showed that male and female gametocytes induced at 22 hour schizonts develop, after reinvasion, at a corresponding rate to the parental. Based on the proportion of the population expressing these fluorescent proxies it was already clear there was a significant increase in the population committing to gametocytes (Figure 4.5).

5.2.2 Quality control (FastQC)

With high throughput pipelines, like RNAseq, it is imperative that data obtained is checked for quality before proceeding with analysis. To this end all raw data files obtained from the pilot time course were analysed by FastQC (Bioinformatics, 2011). For this initial analysis the primary aim was to ensure that the total number of sequences obtained was sufficient for good genome coverage, that the overall sequence quality was high and that the sequence read length obtained was the expected 100bp. All samples obtained were of high quality (Table 5.1) and the minimum number of sequences obtained was almost 10 million (0h + R) with the mean Phred score being 38 or 39 meaning the base call accuracy is between 99.9% and 99.99% accurate (full FastQC analysis available in supplementary 3, pre-trim FastQC).

Sample	Total sequences (Pair 1)	Total sequences (Pair 2)	Sequence length	Mean Phred score
0h – R	19,045,647	19,045,647	100	39
0h + R	9,957,654	9,957,654	100	39
1h – R	15,692,499	15,692,499	100	39
1h + R	19,613,512	19,613,512	100	39
4h – R	21,684,346	21,684,346	100	39
4h + R	21,638,915	21,638,915	100	39
12h – R	18,146,275	18,146,275	100	39
12h + R	23,809,122	23,809,122	100	39
24h – R	20,247,480	21,017,016	100	39
24h + R	19,018,757	19,018,757	100	38
30h – R	21,160,470	21,160,470	100	39
30h + R	22,939,854	22,939,854	100	38

Table 5.1 Raw RNAseq data file analysis

At each time point analysed the total number of sequences ranged from ~9.5 million to ~23.5million reads. Sequence read length distribution was 100bp in all samples. The mean Phred score was 39 for all samples except 24h + R and 30h + R, where was it 38.

5.2.3 Read quality trimming

To prevent the inclusion of reads in the analysis that might be inaccurately called and therefore inaccurately aligned, all raw files were subjected to a quality control and trimming step using Trim Galore! (Krueger, 2015). This procedure removes any adapter sequences on the reads obtained, removes any reads with a Phred quality score below 20 (selected to give minimum 99% call accuracy), removes any samples that are shorter than 20bp in length (this would

result in inaccurate alignments) and (because in our case we had paired end samples) removes any pair in which the forward or reverse sequence read was shorter than 20bp in length. As the raw data files were of good quality (see supplementary file 3. Pre-trim FastQC) sequencing adapters were the predominant bases removed from the sequence reads during the trimming process. Occasional bases were trimmed due to quality scoring. For each paired read in each sample the pre-trim sequences total (Table 5.1) was very similar to the post-trim sequence total (Table 5.2) and at least 99.4% of the original reads were retained for analysis.

Sample	Post-trim sequences (Pair 1)	Post-trim sequences (Pair 2)	Sequences retained Post-trim (%)
0h − R	18,958,096	18,958,096	99.5
0h + R	9,896,644	9,957,362	99.4
1h − R	15,596,022	15,691,995	99.4
1h + R	19,522,014	19,612,945	99.5
4h − R	21,583,056	21,683,616	99.5
4h + R	21,537,395	21,638,260	99.5
12h − R	18,060,306	18,145,713	99.5
12h + R	23,696,647	23,808,472	99.5
24h − R	20,115,375	21,016,404	99.5
24h + R	18,925,994	18,925,994	99.5
30h − R	21,060,758	21,159,837	99.5
30h + R	22,829,551	22,939,171	99.5

Table 5.2 Post-trimming RNAseq data file analysis

At each time point analysed the total number of sequences ranges from ~9.5 million to ~23.5million reads. Each pair was trimmed and compared to the original number of reads (Table 5.1) A small number of sequences haves been lost in all samples due to adapter trimming and some lower quality called bases in the sequence.

5.2.4 Quality control on trimmed transcripts (FastQC)

To ensure post-trimming read quality was still high a second quality control step was completed using FastQC. While there was a reduction in the number of reads in each sample and the length distribution was no longer uniformly 100bp (adapter trimming and removal of lower Phred quality base pairs results in some sequence length reduction) the variability was minimal. Phred scores remained above 38 (full FastQC analysis available in supplementary 4, post-trim FastQC).

5.2.5 Transcript alignment

Paired transcripts from each time point were aligned to the *P. berghei* ANKA reference genome (version 10.0), obtained from the PlasmoDB resource (Aurrecoechea *et al.*, 2009; Bahl *et al.*, 2002) using HiSat2 (Kim *et al.*, 2015; Pertea *et al.*, 2016). As the samples were paired the aligner positioned two reads of a pair on opposing strands. To ensure repetitive regions were not overrepresented the number of positions a transcript can be aligned to with the same confidence was limited to 10. To appropriately orient paired samples the maximum intron length for *P. berghei* was set to 50kb. In all samples the overall alignment rate was high (Table 5.3; minimum 84.67%) meaning there was little contamination from mouse material in the prepared samples. In the majority of cases the concordant alignments, where both pairs in the sample align in opposite orientations with a distance that does not violate the maximum intron length specified, were the predominant alignment types. In all cases some of the pairs aligned discordantly or their partner did not align at all. In RNAseq analysis the alignment of discordant or non-paired alignments does not hinder the analysis therefore discordant and unpaired alignments are retained and our overall alignment is the most important data to consider (Table 5.3).

Sample	Total reads	Concordantly aligned	Discordantly aligned	Aligned unpaired	Overall alignment (%)
0h − R	42,232,200	36,048,948	2,007,600	5,638	90.13
0h + R	21,690,155	18,928,930	968,072	3,153	91.75
1h − R	34,852,295	29,654,776	1,698,166	5,704	89.98
1h + R	43,140,632	37,228,692	1,964,340	6,206	90.86
4h − R	51,156,755	40,389,002	2,912,224	11,090	84.67
4h + R	50,267,531	40,458,094	2,755,120	9,684	85.99
12h − R	40,126,077	34,309,402	1,951,754	3,843	90.38
12h + R	52,115,646	45,360,636	2,209,132	5,222	91.29
24h − R	10,840,848	314,544	9,994,278	287,604	97.75
24h + R	11,219,416	406,732	10,284,706	407,287	98.92
30h − R	46,892,311	40,148,506	2,128,044	6,363	90.17
30h + R	47,120,237	44,364,690	1,501,054	3,020	97.34

Table 5.3 Alignment statistics for the pilot time course.

At each time point analysed the total number of sequences that aligned to the *Plasmodium berghei* genome was high, ranging from 84.67 – 98.92%). Concordant alignments (where both pairs align conforming to maximum intron length and orientation) and discordant alignments (where either the maximum intron length was exceeded or the orientation of the alignment was not as expected) accounted for majority of the alignments. In all cases some transcripts could not be successfully paired but could be aligned singly.

5.2.6 Expression quantifications

Quantification of the number of transcripts aligned to each gene in the genome is fundamental to the identification of genes differentially expressed (DE) between the two conditions (+ and – rapamycin at each time point). Quantification of transcripts was achieved with CuffQuant (Trapnell *et al.*, 2012). By default multi-mapped reads align to all positions proportionally but equally (eg. If a transcript maps twice each position is scored ½). To more accurately assign transcripts that map to more than one position in the genome, weighting of the proportion assigned to each position was altered depending on how highly expressed each gene a transcript maps to was expressed in all other (non-multi-mapping) transcripts (eg. a transcript that maps to two genes, one that is expressed 10x higher than the other will have a weighting of 9/10 and 1/10 instead of equal scoring). During the library preparation steps PCR bias can be introduced for transcripts. Transcripts under- or over-represented occurring from library prep and PCR amplification bias can be identified by sequence bias at the ends of the transcripts. Normalisation to the size of each transcript is utilized to minimise quantification errors due to PCR bias. Finally, to ensure that there is no bias in transcript alignment based on the reads obtained for each individual sample the FPKM (<u>F</u>ragments <u>Pe</u>r <u>K</u>ilobase of transcript per <u>M</u>illion mapped reads) values are calculated, using the package Cuffdiff (Trapnell *et al.*, 2012), so the number of transcripts aligned to a gene is normalised to the relative size of that gene in the genome. FPKM values of each gene in the *P. berghei* genome for each sample can be found in Appendix B.

5.2.7 Reanalysis of RNAseq datasets

The stage specific transcriptomes were sequenced using Sanger capillary, Illumina and 454 sequencing whereas all new datasets were sequenced using Illumina sequencing only. Reads for the reference stage samples were obtained as non-paired analysis and were 76bp in length. To best ensure that the datasets compared had been quantified in the same way the stage specific raw RNAseq datasets available within the lab were processed in the same way as the raw files obtained from the *ap2-g* induction time course. Files were subjected to trimming as described above to remove adapters, poor quality called bases (Phred < 20) and to remove any resulting fragments shorter than 20bp. As these

samples are not paired end samples there was no need to filter to ensure both partners of a pair were removed if a single transcript was too short or too low quality (Table 5.4). The transcripts were then aligned to the *P. berghei* ANKA reference genome (version 10.0) using HiSat2 (Kim *et al.*, 2015; Pertea *et al.*, 2016). In most cases overall alignment was high (above 88%) indicating samples contained little contaminating, non-parasite RNA (Table 5.4). However, the ring sample showed a much lower alignment percentage (51.5%) which is likely a carry-over from sample preparation as the parasite RNA contained within ring stage parasites is significantly less than other stages often resulting in higher contamination with non-parasite material. This reanalysis of the existing datasets was completed to control for variability introduced from analysis methods as variability in experimental design could not be altered.

Sample	Number of reads	Reads post trimming	Remaining reads (%)	Mean Phred score	% Genome Alignment
Ring	20,770,103	15,638,553	75.3	34	51.5
Trophozoite	26,541,543	24,869,155	93.7	39	88.66
Schizont	24,280,713	23,548,843	97.0	38	92.85
Ookinete	27,050,650	23,522,975	87.0	36	94.17
Gametocyte1	18,365,383	18,247,880	99.4	31	96.13
Gametocyte2	31,951,285	31,894,533	99.8	39	90.42
Gametocyte3	26,212,054	24,339,386	92.9	39	94.02

Table 5.4 Quality control and alignment statistics for stage specific samples.

For each of the stage specific samples FastQC was used to analyse read numbers and visualise quality of the reads. After trimming a small proportion of the reads were discarded due to low quality or size (due to trimming of low quality called bases). After low quality reads were discarded the remainder were aligned to the *Plasmodium berghei ANKA* genome. The majority of the samples showed very high alignment. The ring sample is likely to have higher mouse material contamination due to processing as their DNA content when compared to all other stages is reduced.

5.2.8 Overall RNAseq analysis

To determine that the samples obtained were comparable in terms of how much of the total RNA was sampled the fpkm ranges obtained for each of the samples was compared. To visualise this box plots were generated with each gene in the genomes fpkm value being plotted. This shows us the range of fpkm values obtained. Considering all samples were obtained from the same organism with similar coverage we expected to see similar mean fpkm values with ranges that are comparable. Though the non-paired samples for the single stages are slightly

different to the time course samples overall the mean and range is comparable. This was a good indication that the samples would be comparable to each other and that the amount of the transcriptome sampled in each condition was similar to all others.
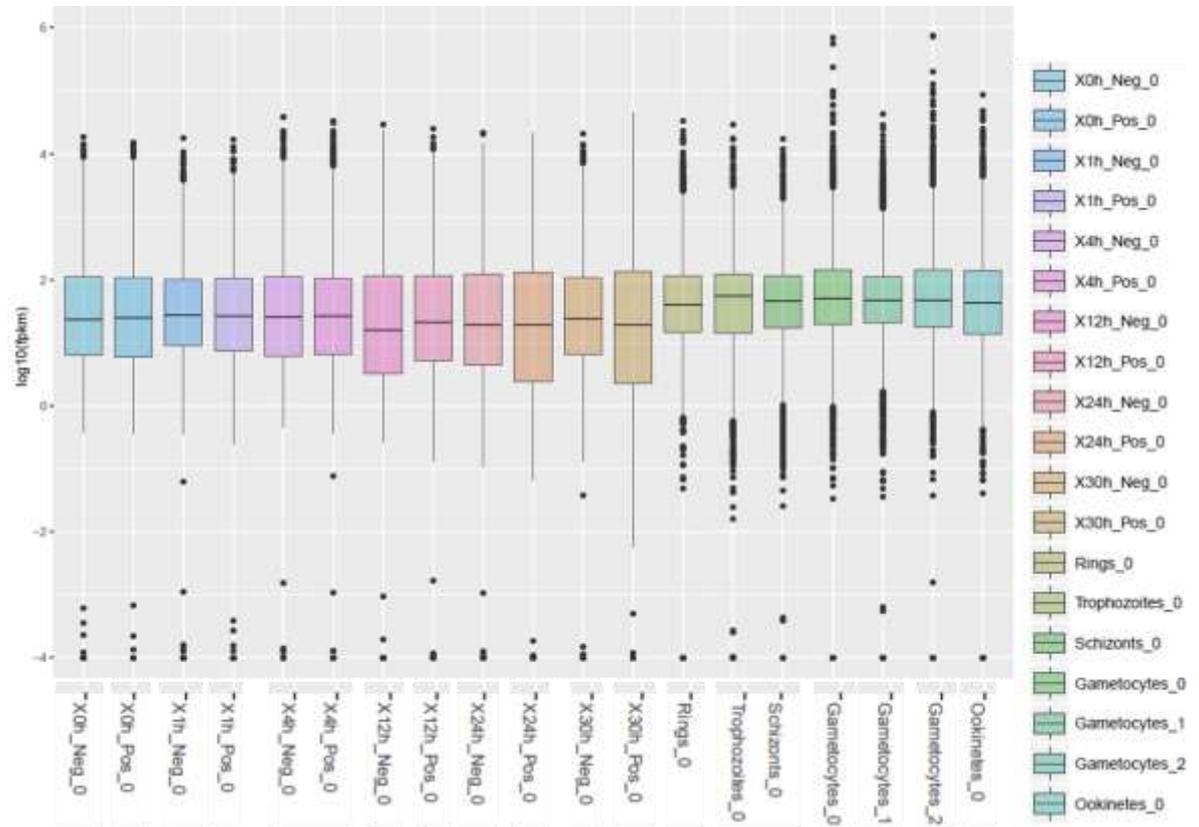
Figure 5.3 fpkm distribution for all samples.

For all samples the fpkm values for each gene in the genome was plotted to show the sampling distribution for each. Comparable samples have similar mean fpkm values and their distribution ranges were alike. While there are differences in the range for the stage specific samples their overall distribution is similar to that of the time course and therefore comparisons were possible.

Initial analysis focussed on identifying how each of the time points clustered when compared to the single stage specific transcription profiles. Comparisons were based on the FPKM expression of each gene within the genome. To ascertain the variability and similarity of the samples, principal components analysis (PCA) was completed using the R (Team, 2013) package FactoMineR (Lê *et al.*, 2008). Unfortunately, it was clear after this analysis that the variability in the way the RNA samples were generated, RNAseq runs unpaired vs. paired, meant that comparing FPKM values only identified the differences in sample run (Figure 5.4 Principal component analysis of time course plus stage-specific datasets.). Three distinct clusters are clearly evident when looking at the

predominant sources of variability. Firstly, (red, green and grey clusters) the stage-specific unpaired samples, that all cluster closely and quite distinctly the other samples, principally along dimension 2. Secondly, (orange cluster) the two late stage induced samples (24hp-Ind and 30hp-Ind). Finally, the remaining time course of paired-end samples that cluster very distinctly from all the other samples, but cluster very tightly together.
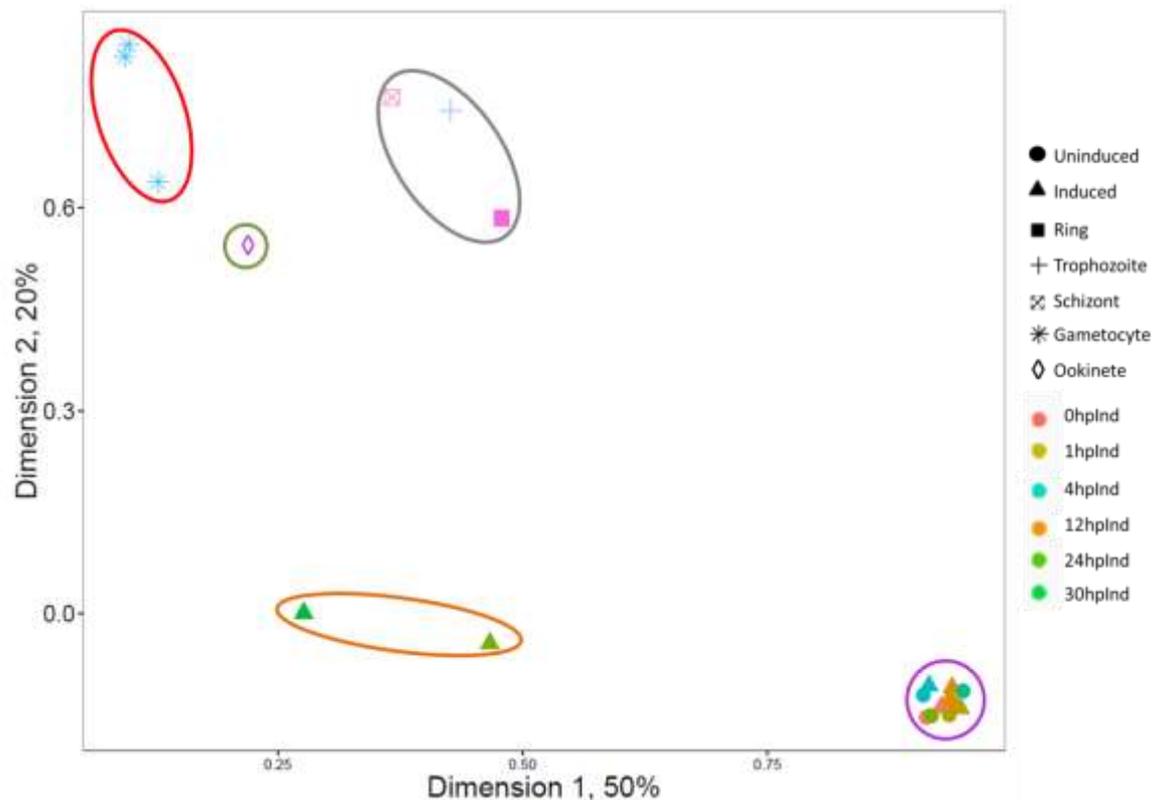
Figure 5.4 Principal component analysis of time course plus stage-specific datasets.

The two components accounting for ~70% of the variability between samples were plotted to show the distance between each. It is clear from this analysis they the unpaired samples for the stage specific RNA transcripts (red, green and grey circles) cluster closely although the asexual parasites (grey circle) are distinct from the sexual differentiated gametocytes (red) and ookinetes (green). Only the latest two induced samples in the time course (orange circle) are distinct from them rest of the time course (purple). This indicates that the dimensions plotted, that account for the largest proportion of the variability, are focussed on genes which appear differently expressed based on the sampling variation.

To ascertain if including the entire time course but investigating the other dimensions of variability would sufficiently describe the data all other combinations and dimensions of variability were analysed by PCA (Figure 5.5). We can see that the 4hp-Ind 12hp-Ind and non-induced 30hp-Ind samples (pink cluster) cluster closely with ring stage parasites if we consider dimensions 3 and 5 that account for 14% of the variability between samples. Separate to this group clusters the 0hp-Ind, 1hp-Ind and non-induced 24hp-Ind samples (blue cluster).

These samples cluster with the schizont sample and the ookinete. Very distinct from these populations are the induced 24hp-Ind and 30hp-Ind samples that cluster similarly. Although not directly clustering these two samples appear to cluster most similarly with the gametocyte populations (blue stars). One of the problems with this analysis, however, is that the variability in the analysis that includes the stage specific samples is predominantly attributed to sample collection, preparation and analysis variability and not the variability exhibited because of the time course.
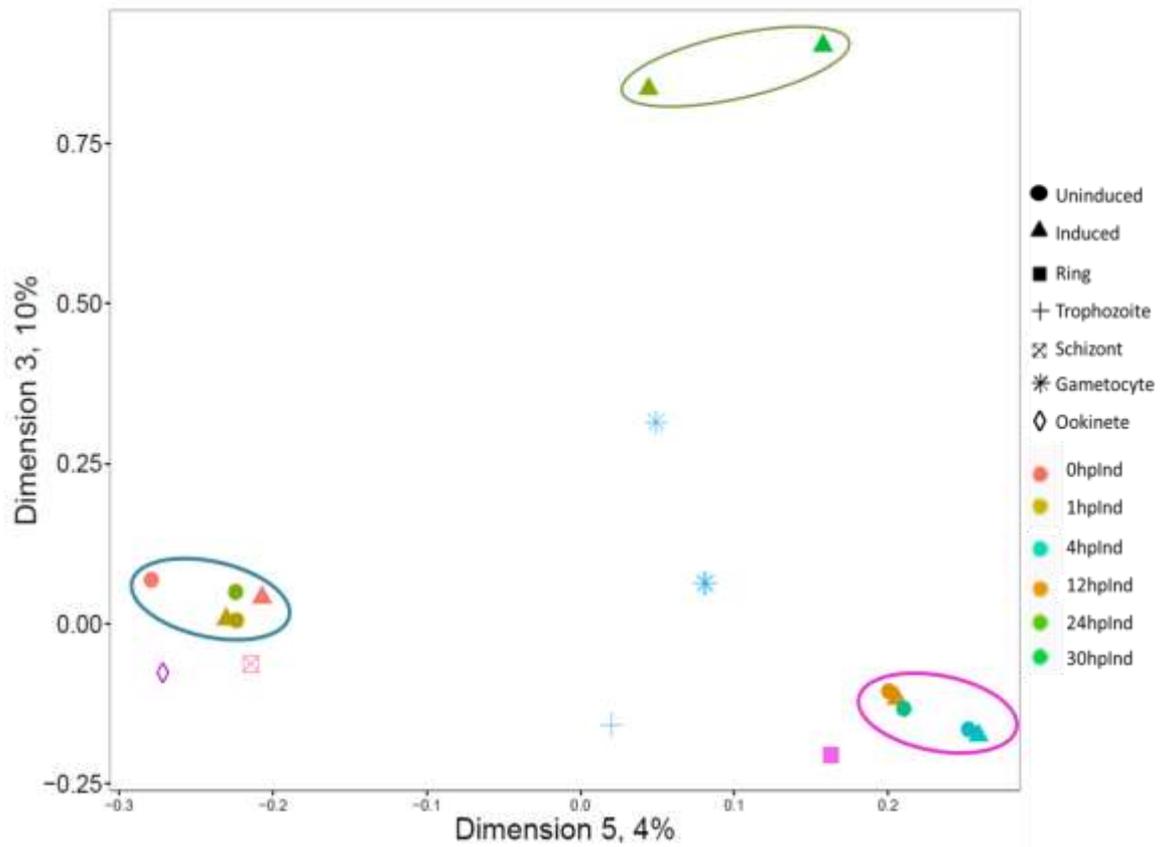
Figure 5.5 Principal component analysis of the time course plus stage specific datasets (time course dimensions).

The two components that seemed to hone in on the variability between samples (not based on preparation method) were plotted (dimensions 3 and 5) to show the distance between each. It is clear from this analysis that there are three distinct populations post induction. Firstly, the 0hp-Ind, 1hp-Ind and 24hp-Ind (non-induced) samples (Blue cluster). These cluster closely with the schizont sample and the ookinete sample. As the age of these parasites would correspond to schizonts this seems to correlate. The second, containing the 4hp-Ind, 12hp-Ind and 30hp-Ind (Pink cluster) samples. This cluster correlates most closely to the ring stage purified parasites. Initially this was surprising as a 12hp-Ind time point is effectively 10hpi parasites that would be considered trophozoites. However, the trophozoite stage specific sample was generated from mature trophozoites (18hpi) and is therefore likely quite distinct. The final cluster is that of the 24hp-Ind and 30hp-Ind induced samples (Green cluster). These are quite distant from all other samples indicating they are quite different. They do not closely cluster with any of the stage specific samples, however they are most closely related to the gametocyte samples (blue stars).

Therefore, to get an overview of the variation in samples in just the time course samples, the stage-specific unpaired samples were excluded. This allowed the variability between the time points and conditions to be the only factors influencing the component analysis. In this second analysis it is clear that two samples are very different from all the others (Figure 5.6). These are the rapamycin induced samples from 24hp-Ind and 30hp-Ind which were the only samples containing immature (24hp-Ind) and mature (30hp-Ind) gametocytes. All other samples (Purple cluster) contained only life cycle stages that would be morphologically identified as asexual (Figure 5.2).

While this more clearly clusters the two different populations from the time course it seems to use only the variation that accounts for differences in the two mature induced populations. It also appears that dimension 1 is not explaining any variation between samples within the time course excepting the induced 24hp-Ind and 30hp-Ind samples.
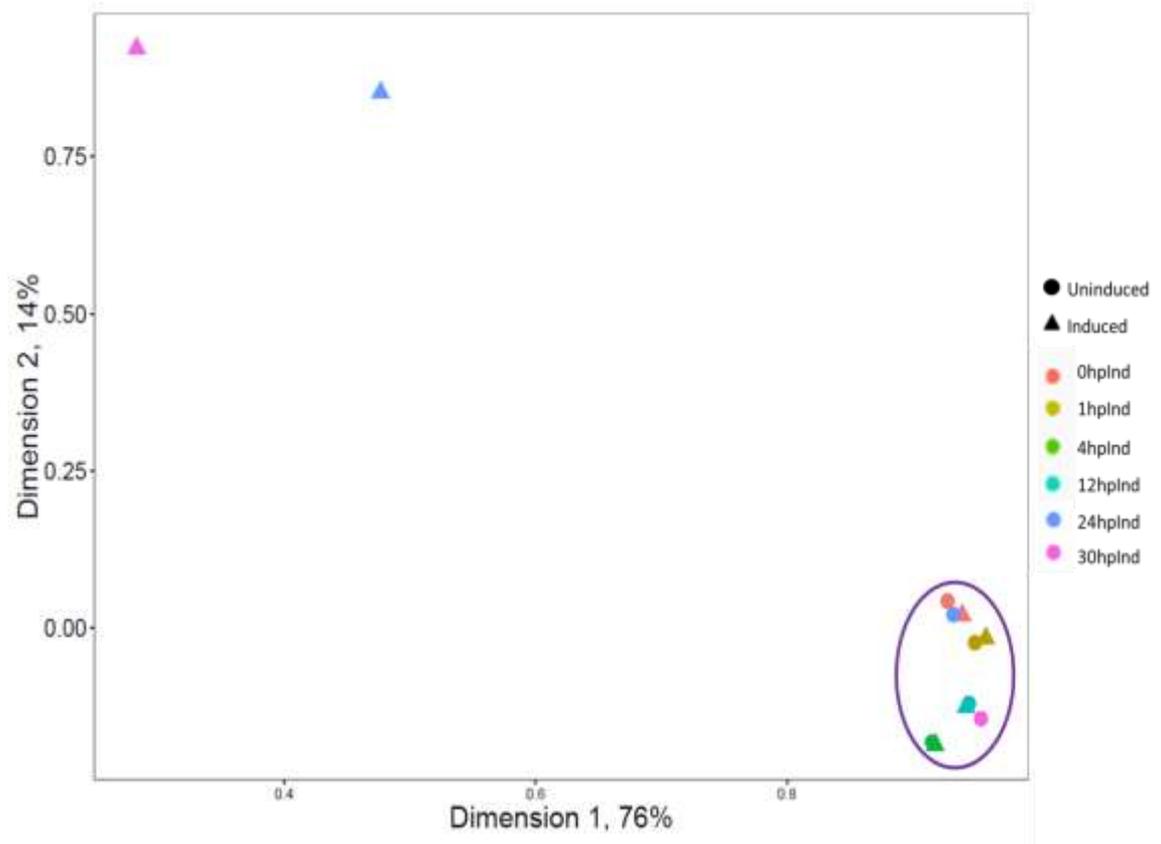
Figure 5.6 Principal component analysis of the time course

The two components accounting for ~90% of the variability between samples were plotted to show the distance between each sample. It is clear from this analysis that two samples cluster quite distinctly from all others. These are the two samples which morphologically (Figure 5.2) contain gametocytes. The more tightly clustered group (purple circle) would all be morphologically identified as asexual stages.

To determine if any of the other dimensions explain more the variability of the earlier time points sampled they were all compared. Dimension 2 was clearly explaining some of the variability between the time course samples and the late gametocyte stage samples (Figure 5.6). Therefore this variability dimension was plotted with dimension three (Figure 5.7). Together they account for 25% of the variability and show distinct populations. The two induced late stage samples (pink and blue triangles) cluster very separately from the rest of the time course. The time course is split in to two clusters by dimension three. Here we see a population (green cluster) that contains the very early time points, 0hp-Ind and 1hp-Ind along with the 24hp-Ind sample (Figure 5.7). The clustering of the 24hp-Ind non-induced sample with these early time points is reassuring as effectively the 24hp-Ind sample and the 0hp-Ind samples are the same life cycle stage because of the 24 hour life cycle. Likewise with the population of slightly later time points (red cluster) that contains the 4hp-Ind and 12hp-Ind samples along with the 30hp-Ind non-induced sample (Figure 5.7).
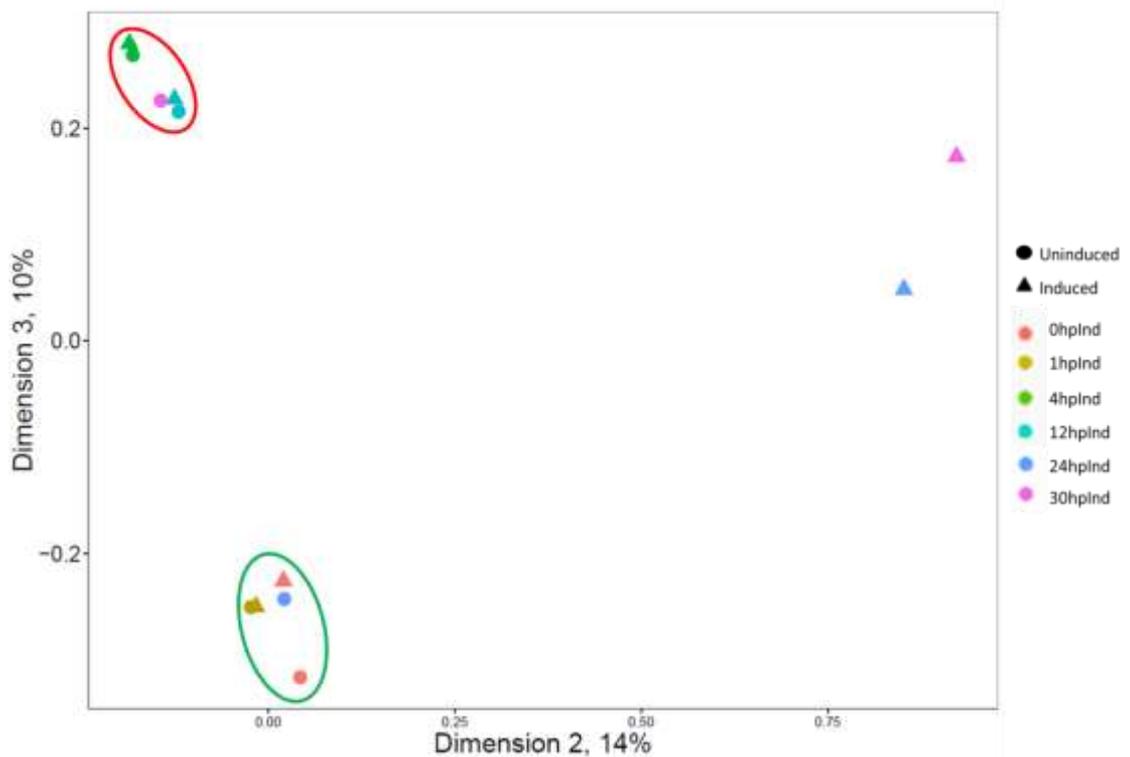
Figure 5.7 Principal component analysis of the time course (early time point dimensions).

The two components that offer the best separation of time points in the time course that account for ~24% of the variability between samples were plotted to show the distance between each sample. It is clear from this analysis that two samples cluster quite distinctly from all others. These are the two samples which morphologically (Figure 5.2) contain gametocytes. However, by looking at the variability accounted for by dimension 3 we see two other distinct populations in the time course. These are (green cluster) the 0hp-Ind, 1hp-Ind and 24hp-Ind (uninduced) samples. These would morphologically be schizonts. Secondly, the (red cluster) 4hp-Ind, 12hp-Ind and 30hp-Ind (uninduced) samples that would morphologically be considered ring and trophpzoite parasites.

As the single ended samples reanalysed as controls were not broadly comparable to the time course due to sampling and RNAseq methodology we decided to compare the stages in a more focussed manner.

Firstly, we performed a principal component analysis (PCA) on only the top 20% most differentially expressed genes of the 30hp-Ind samples (when comparing Non-Induced with Induced). In this analysis there was still one dimension (dimension 2) that solely separated the paired end samples from the single ended samples but the other dimensions (dimensions 1 and 3) as clearly as when analysing all genes if not more clearly separated the time course with its closest single time point sample in terms of development (Supplementary Figure 2). Secondly we calculated R^2 coefficients of variance for all time points compared to the mature gametocyte transcriptome. The earliest time points post induction do not correlate at all with the gametocyte population ($R^2 < 0.003$) however the later time points correlate more closely ($R^2 \sim 0.5$). If we plot the most highly expressed and lowly expressed gametocyte genes (top/bottom 100) we see that the post induction population begins to express more similarly to gametocytes over time (Supplementary Figure 3). Finally, the transcription profile of the 24hp-Ind time point with and without induction was compared to the schizont and gametocyte samples to see if transcription of known asexual stage and known gametocyte specific genes were expressed consistently. These time points and stage specific samples were chosen as gametocytes and late schizonts should be the predominant stages in the populations at 24hp-Ind. To ensure that expression values from each sample were comparable the gene of interests expression was normalised to a gene known to be not differentially expressed at any stage of the life cycle (a 40S ribosomal protein, PBANKA_132500). If we consider genes known to be expressed highly in asexual stages (Figure 5.8 A) and lowly expressed in gametocyte stages it was evident that the purified schizont sample (green bar) had high expression and purified mature gametocytes (purple bar) have low. The non-induced 24hp-Ind time point (which from giemsa smears, Figure 5.2, was made up of young rings and late stage schizonts) has high expression of all asexually expressed genes analysed. The induced sample (which by giemsa analysis, Figure 5.2, was predominantly immature gametocytes) showed low expression of all the asexually expressed genes. The genes, selected based on mRNA expression levels from existing transcriptomics datasets (Bahl *et*

al., 2002; Otto *et al.*, 2014), indicated to be specifically expressed in mature gametocytes (Figure 5.8 B) were highly expressed in the gametocyte sample (purple bar) and lowly expressed in the schizont sample (green bar). The Induced 24hp-Ind time point in this case showed very high expression of all of the gametocyte specific genes and the non-induced time point shows little expression of any.

Taken together, these data indicate that the non-induced population closely resembled a schizont population expressing known asexual genes and not gametocyte specific genes and that the induced population resembled that of a mature gametocyte population expressing known gametocyte specific genes and not genes highly expressed by schizont stage parasites.
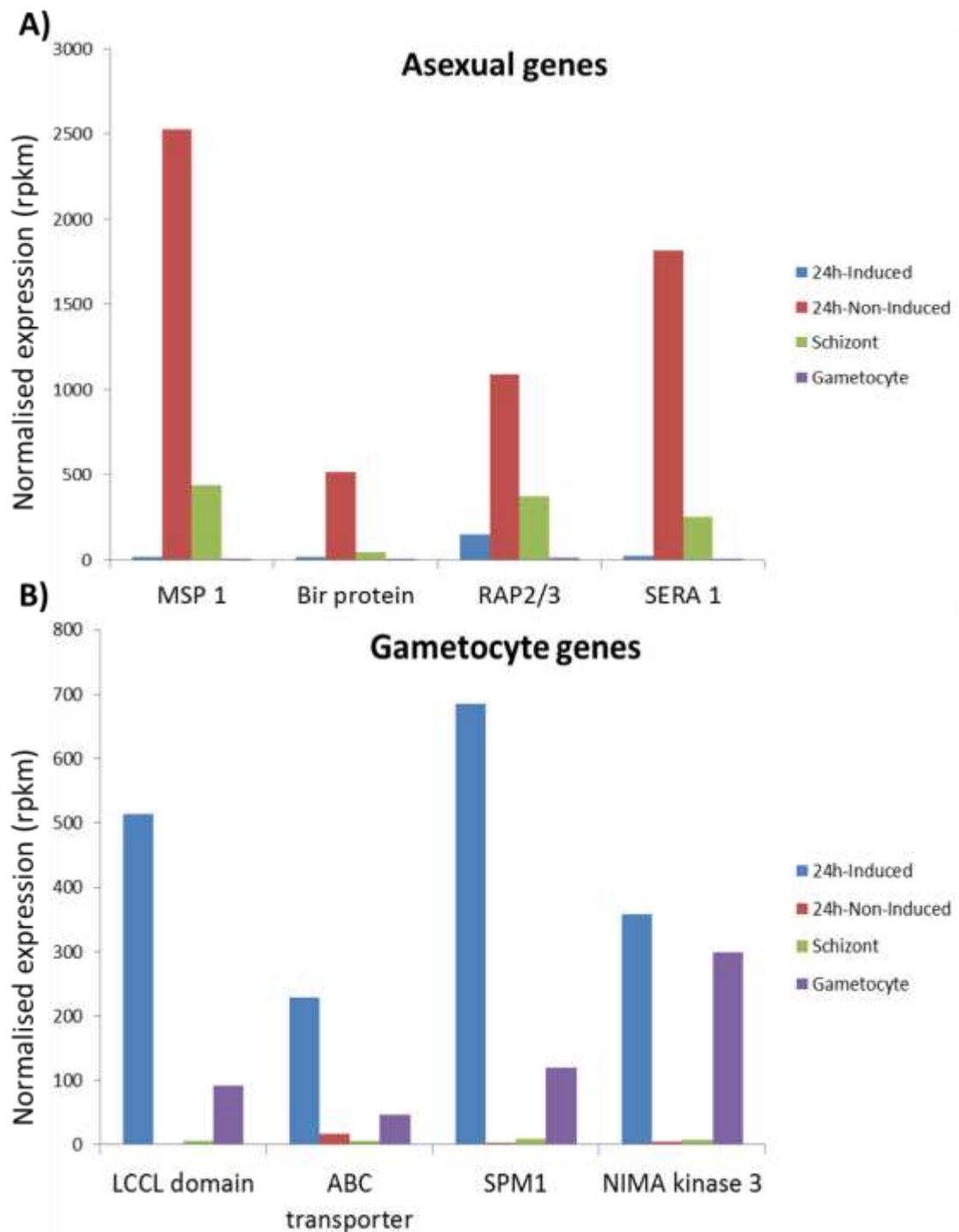
Figure 5.8 Comparison of transcription in induced and non-induced samples at 24hp-Ind with purified schizont and gametocyte samples.

All expression values were normalised to a housekeeping gene (PBANKA_132500) to ensure comparability between samples.

A) Four genes known to be highly expressed in asexual parasites (particularly schizonts) were compared in purified schizont gametocyte samples along with the induced and non-induced 24hp-Ind samples from the time course. All four genes were highly expressed in the schizont sample and the non-induced sample but lowly expressed in the induced and gametocyte samples.

B) Four genes known to be expressed specifically in gametocytes but not in asexual stage parasites (especially low in schizonts) were compared. All four genes were lowly expressed in the schizont sample and the 24hp-Ind non-induced sample. They were highly expressed in the purified gametocyte sample and the 24hp-Ind induced sample.

5.2.9 Analysis of differential gene expression

Genes that are differentially expressed within this pilot time course were initially analysed to ensure that the method utilised was yielding genes of interest that would seem to be related to gametocyte commitment. From the FPKM expression values for each gene throughout the time course differential expression was calculated between the non-induced *ap2-g* non-expressing line and the induced *ap2-g* overexpressor (Supplementary 5. Counts and Differential expression pilot RNAseq).

From this analysis, three time points of interest were selected to identify potential genes of interest. As the main focus of interest was the initial events of gametocyte commitment we elected to evaluate genes differentially expressed (2 fold increased in expression in *ap2-g* overexpressor vs *ap2-g* null sample) at 4hp-Ind. These would correlate to recently invaded ring stage parasites that express AP2-G and those that do not. This identified 21 genes. We were also interested in correlating the genes of interest we have obtained with existing data looking at levels of transcription at different life cycle stages (Hall *et al.*, 2005; Otto *et al.*, 2014) and with the published proteome of gametocytes (Khan *et al.*, 2005) and therefore also investigated genes of interest that were differentially expressed (5 fold increased in expression in *ap2-g* overexpressor vs *ap2-g* null sample) at 24hp-Ind. This time point is the first where gametocytes become morphologically distinguishable from asexual parasites (see Figure 5.2). This accounted for 471 genes. Further to this, genes differentially expressed (5 fold increase in the *ap2-g* overexpressor vs the *ap2*-null sample) in the 12hp-Ind time point were also selected for analysis. The 12hp-Ind time point yielded 130 genes of interest. There is significant overlap in the genes of interest with many being upregulated at multiple time points (Figure 5.9).
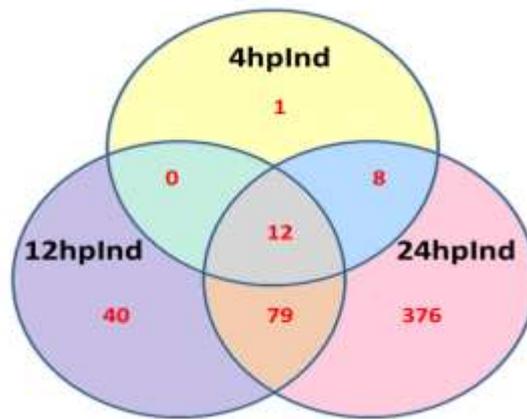
Figure 5.9 Venn diagram showing the differentially expressed (DE) genes of interest.

Genes of interested were selected based on DE at 4hp-Ind (2-fold), 12hp-Ind (5-fold) and 24hp-Ind (5-fold). Many genes of interest were upregulated in the induced *ap2-g* overexpression line at multiple time points when compared to the non-induced *ap2-g* non-expresser.

To determine whether the genes identified as differentially expressed at these time points were gametocyte specific, we compared these with two known gametocyte specific datasets. Firstly DE genes were compared with the stage specific transcriptomics data generated for rings, trophozoites, schizonts, gametocytes and ookinetes. To determine if our genes correlated to existing transcriptome data the stage with the highest expression for each gene was classified (Figure 5.10). A total of 415, or 80%, of the differentially expressed genes from the three selected time points were expressed most highly at the gametocyte stage. The majority of the other genes were expressed most highly at the ookinete stage (69 genes, or 13%). This indicated that many of the genes identified from our analysis of differential expression after induction, in parasites expressing high levels of *ap2-g* vs. those expressing no *ap2-g*, were specific to the sexual stages. None of the genes identified were genes specifically expressed or upregulated in the asexual stages of the life cycle (Hall *et al.*, 2005).
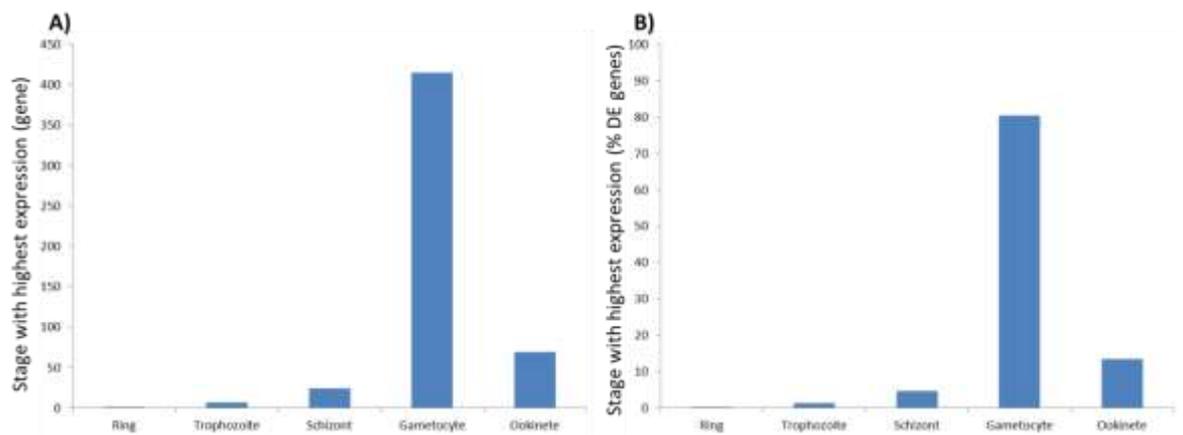
Figure 5.10 Analysis of differentially expressed (DE) genes based on their stage specific expression.

The majority of genes identified as differentially expressed in the pilot time course were most highly expressed at the gametocyte stage. The second most common stage for highest expression was the ookinete stage. Actual numbers of genes is shown in A and % of the total in B.

Secondly, proteome data has been generated for male and female gametocytes as well as purified mixed gender gametocytes. A comparison with this data was completed to identify what proportions of the differentially expressed genes identified were present in the specific populations. In this study asexual blood stage parasites, mixed gametocyte, male gametocyte and female gametocyte proteins were sampled. In total 1413 proteins were present in one or more of these samples at detectable levels. In total 150 of the genes identified as differentially expressed genes in this work were also detected in one or more of these proteome samples. Reassuringly only 5 of the differentially expressed genes identified were detected in only the asexual proteome samples. More than 95% of the differentially expressed genes identified from the time course analysis were expressed in one or more of the gametocyte proteomes (Figure 5.11).
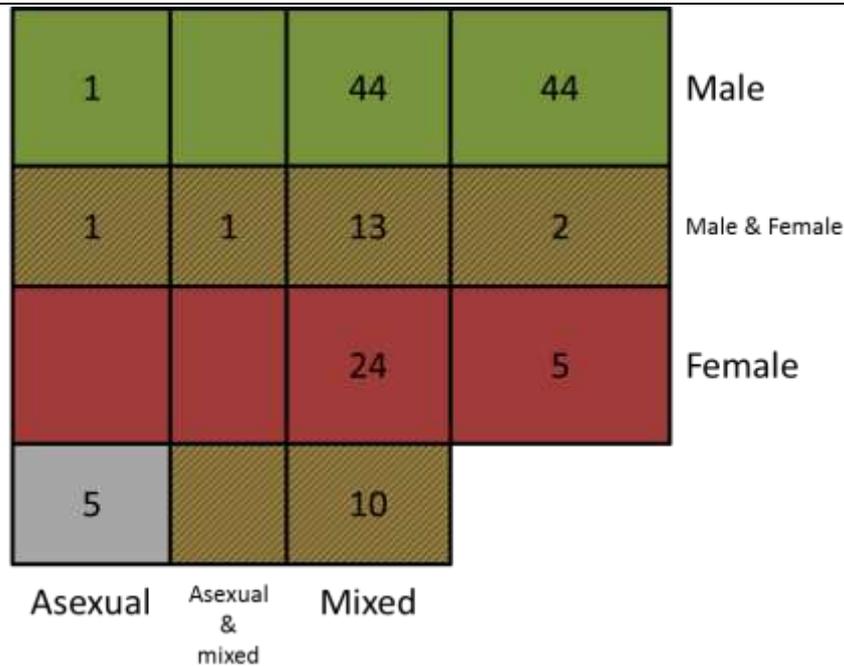
Figure 5.11 Venn diagram showing the differentially expressed genes of interest also present in gametocyte specific proteome samples.

Of the 516 DE genes identified in the three selected time course samples 150 were also detected in one or more of the blood stage proteome samples. Only 5 of these were unique to the asexual stage proteome. All others were present in one or more of the gametocyte proteome datasets (Khan *et al.*, 2005).

Upon satisfaction that the pilot RNAseq time course was yielding gametocyte specific transcription profiles for three different time points, a greater number of time points were sampled at times that had previously been inaccessible for investigation.

One of the key aims for this work was to identify factors involved in the very early stages of commitment to gametocytogenesis and any proteins involved in gender assignation. These crucial steps have been impossible to study previously as commitment to gametocytogenesis is both variable and low in each cycle (5 – 15%) and young gametocytes have not been distinguishable from asexual parasites.

To try and elucidate some of the proteins critical for these key processes additional time points were added to the time course. The majority of these additional time points sampled sexual stage parasites before they are morphologically identifiable as gametocytes (6hp-Ind, 8hp-Ind, 18hp-Ind) and one sampled fully mature gametocytes formed after commitment (44hp-Ind). Additionally, to eliminate any genes that may be differentially expressed due to

the parasites' response to the drug used to induce *ap2-g* overexpression (rapamycin), control induced and non-induced lines were also sampled at each of these time points (Figure 5.12). The full time course would allow comparison between the non-induced *ap2-g* inducible overexpressor (AP2-G null) and the induced *ap2-g* inducible overexpressor (AP2-G overexpressing) along with control parasites. These control parasites consisted of a parental line that does not overexpress *ap2-g* in response to induction by rapamycin but does express the fluorescent proxies for gender discrimination. This control would allow the identification of genes that are differentially expressed in response to the induction drug and not the overexpression of *ap2-g*.
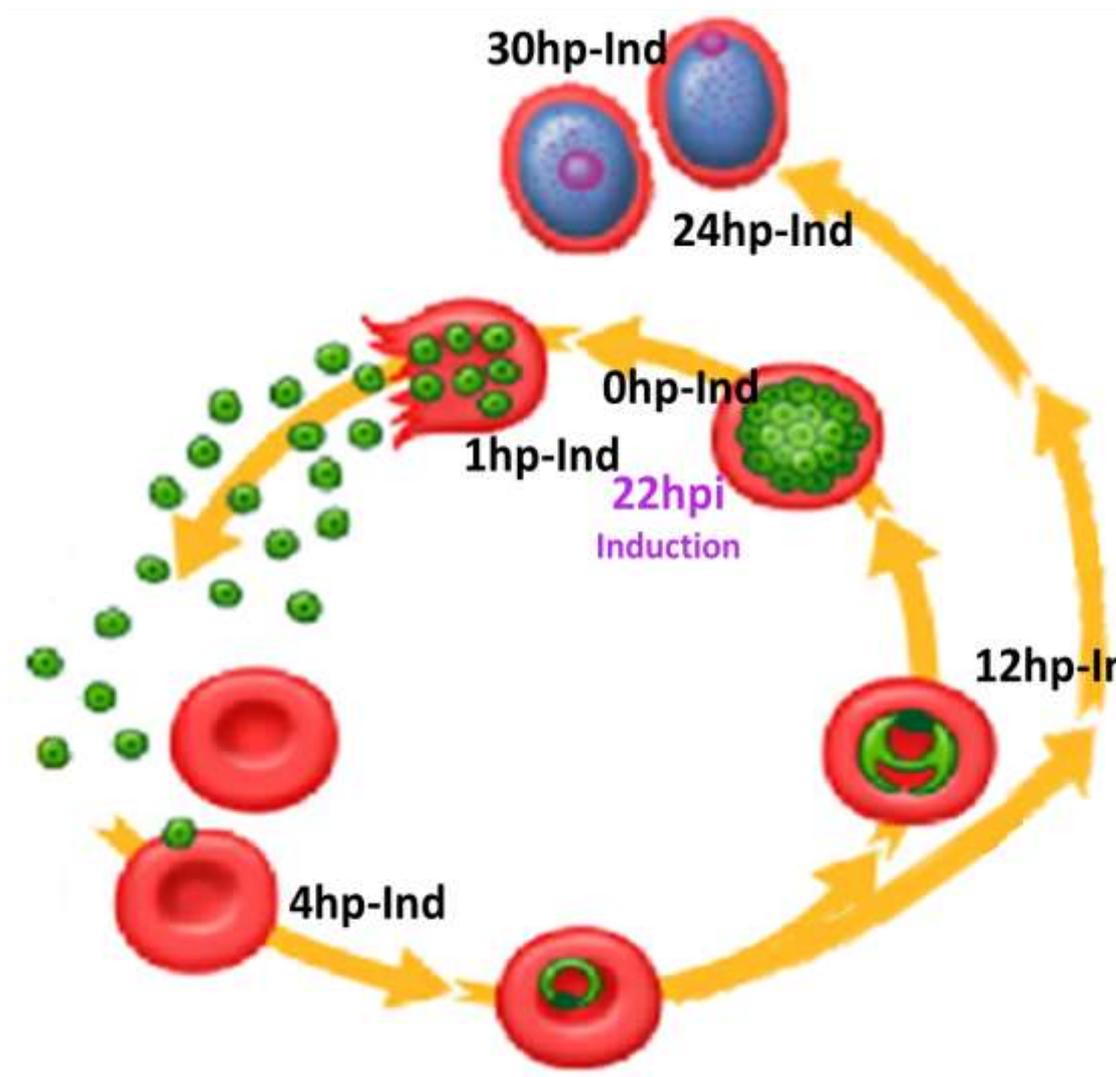
Figure 5.12 Full RNAseq time course to identify gametocyte specific transcripts.

All inductions were completed at 22hpi with a single intraperitoneal injection of 4mg/kg rapamycin. Blood was harvested by cardiac puncture at each time point. At each time point samples were always collected paired to ensure minimal variation. At each time point a total of 8 samples were collected, 3x induced (*ap2-g* overexpressor), 3x non-induced (*ap2-g* null), 1x parental 820 non-induced, 1x parental 820 induced.

Due to the magnitude of the time course, samples have not been fully analysed to complete the identification of novel candidate genes of interest (Figure 5.13). RNA for all time points and all conditions has been collected (Figure 5.13 A). Almost half of the libraries of the RNA samples have been prepared and paired end sequencing has been completed on an Illumina HiSeq2500 system (Figure 5.13 B). All samples that have been sequenced have also been quality controlled (FastQC and quality trimmed with TrimGalore! (Bioinformatics, 2011; Krueger, 2015)), aligned to the genome and the number of reads per gene quantified (Figure 5.13 C). The pilot time course initially completed as a check for the method has also been fully analysed to identify potential gametocyte specific genes of interest (Figure 5.13 D) though this analysis must be re-performed when all samples have been completed to appropriately normalise gene expression.
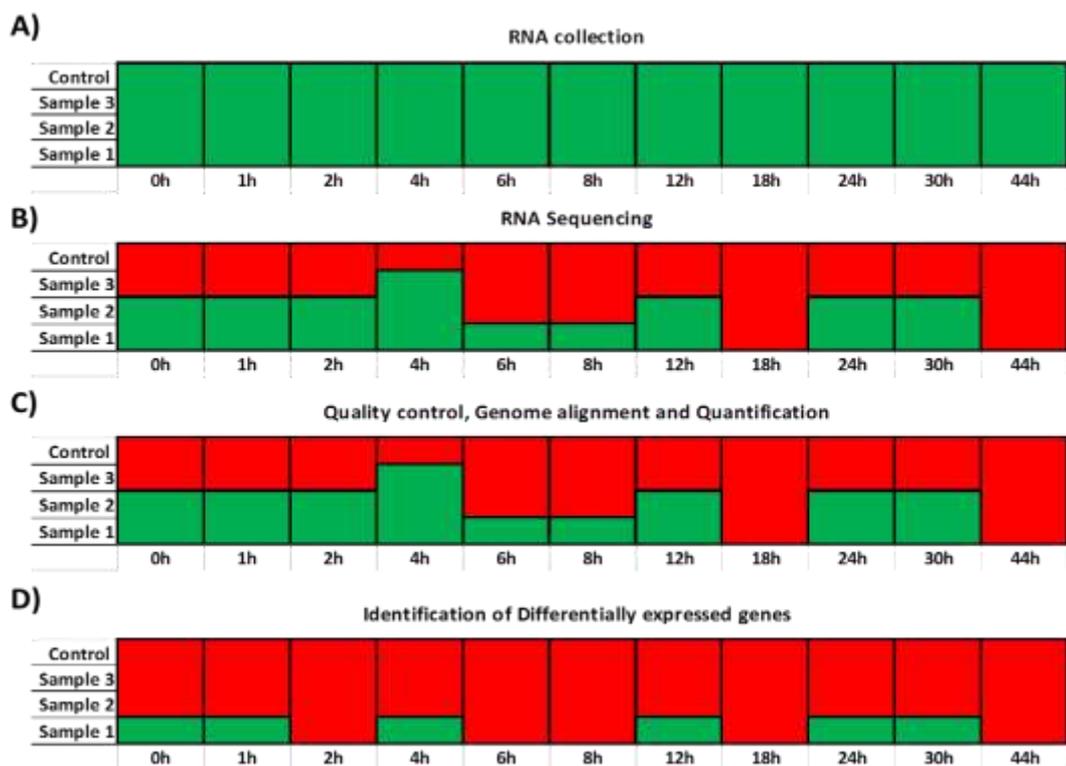
Figure 5.13 Full RNAseq time course progress

In all representations Green = Complete and Red = Incomplete

A) RNA for all time points and conditions has been collected.

B) For most time points (except 18hp-Ind and 44hp-Ind) at least one paired sample has been sequenced. No Control samples have been sequenced.

C) All samples that have been sequenced have been quality controlled, aligned to the *Plasmodium berghei* genome and number of reads in each condition has been quantified.

D) The samples that made up the pilot time course have been analysed to identify genes of interest that are differentially expressed between the non-induced (*ap2-g* null) and the induced (*ap2-g* overexpresser).

5.3 **Discussion**

While the pilot time course completed during this work has been fully analysed and used to identify genes of interest for further study the real power of the transcriptional profiling has yet to be realised with the extensive time course of transcriptional profiling intended (Figure 5.12).

We hypothesise that the changes in gene expression occurring immediately after AP2-G expression are the initial factors involved in commitment and potentially gametocyte gender assignation. These would be the earliest responders but their differential expression will only just have been initiated and therefore identification of these as differentially expressed within a subset of the population might have been neglected in previous untargeted studies where the subset of interest (committed gametocytes) is small (lower, natural % commitment) or the extensiveness of results and replicates does not overcome natural variability which may include different commitment onsets within the population.

The experimental design employed here has endeavoured to overcome the challenges by employing the most sensitive sampling methods for detecting changes in gene expression (RNAseq), attempting to increase the proportion of commitment to its maximum (overexpression of the transcription factor initiating the cascade of commitment), synchronously initiating commitment (with induction of overexpression) to sample only gametocytes of the same maturation and using paired replicates for each time point analysed (to reduce variability) and analyse triplicate samples of each time point.

RNAseq analysis was utilised to provide the untargeted view of transcription in the populations examined as it is a highly sensitive technique able to detect changes in gene expression as soon as mRNA is polyadenylated and can be selected for. This may be of benefit when looking at the initial stages of commitment where a cascade of activity occurs resulting in very different male and female gametocytes. It has further benefits with a wide range for detection limits that can be compared, low background noise of detection and it is not restrained to the annotated genome and known splicing of introns.

Synchronous initiation of commitment to a high level offers an advantage over previous work that has been limited by the proportion of gametocytes committing naturally and the added influence that the first synchronous cycle, where parasite density is at its lowest is the only cycle that contains a single wave of committing gametocytes. While we face the new issue that we are artificially inducing AP2-G expression when it is not normally expressed as highly we have demonstrated that the time points post induction that we are sampling appear similar to those of wild type gametocytes (see section 4.5). We further demonstrated that compared to single stage purified samples (single ended RNAseq) our datasets cluster with their relevant stage (Figure 5.5). In this analysis we saw that the 0hp-Ind, 1hp-Ind (induced and non-induced) and 24hp-Ind (non-induced only) samples where the predominant life cycle stage would be mature schizonts cluster very distinctly with the schizont transcriptome. We also saw that the 4hp-Ind, 12hp-Ind (induced and non-induced) and 30hp-Ind (non-induced only) samples clustered between the ring stage transcriptome and the trophozoite stage transcriptome. This pattern is feasible as there is a considerable time difference, post invasion, between sample collection for these two stage specific transcriptomes (Hall *et al.*, 2005). Trophozoite samples were not harvested for their transcriptome analysis until 18hpi whereas ring stage samples were harvested within 2 hours of merozoites invasion. The least closely clustered samples were the gametocyte transcriptomes and the 24hp-Ind (induced) and 30hp-Ind (induced) samples. There seem to be two possible reasons for this disparity. Firstly, that the gametocytes maturing after induction are not as close in their transcription profile (overall) as the asexual parasites sampled. Secondly, that we are looking at the maturing gametocyte population at 24hp-Ind (a 22hpi gametocyte) and 30hp-Ind (a 28hpi gametocyte) whereas the transcriptome for gametocytes was obtained for fully mature gametocytes at 30hpi which may not contain the same gene expression. Additional sampled time points (44hp-Ind) may elucidate the cause of this disparity as these gametocytes would be fully mature. It is likely that a combination of factors explain this difference in gene expression as known specific gametocyte expression is occurring to a higher degree than in mature gametocyte populations and asexual gene expression appears to be downregulated (Figure 5.5 and Figure 5.8). If we consider only the most differentially expressed genes (top 20%) then we see more clearly a separation of the late stage induced samples with the

gametocytes and ookinetes versus the early stage induced samples with all uninduced samples (Supplementary Figure 2). However, it is worth noting that even when considering only the top 20% differentially expressed genes one of the dimensions for analysis (in this case dimension 2) seems to divide the purified single stage samples (rings, trophozoites, schizonts, gametocytes and ookinetes) from the time course post induction. Furthermore if we look at the correlation of gene expression we see a shift towards a mature gametocyte population over time. This analysis shows a low correlation between mature gametocyte expression and the early time points post induction (from 0hp-Ind to 12hp-Ind the R^2 is below 0.003) with increases of an R^2 of 0.5 indicating the mature gametocyte population and the induced 24hp-Ind and 30hp-Ind have 50% of their variance explained as a mature gametocyte population (Supplementary Figure 3).

Constrained as our differential expression analysis was, being solely based on the pilot time course and therefore limited in the power of analysis to identify differentially expressed genes, we were able to identify a selection of genes of interest purely based on fold-change of expression between non-induced and induced at three selected time points. These genes were identified based on a minimum of a 2-fold increase in expression within the induced samples at 4hp-Ind or a minimum of a 5-fold increase in expression within the induced samples at 12hp-Ind and 24hp-Ind (Figure 5.9).

Comparisons of these identified genes with existing datasets were reassuring. Many (80%) were most highly transcribed in gametocytes with the remaining majority being most highly expressed in ookinetes (13%) indicating transcription at the stage of interest (Bahl *et al.*, 2003; Hall *et al.*, 2005). Existing proteomics data for male, female and mixed gender gametocytes (Khan *et al.*, 2005), which identified 828 gametocyte specific (only when compared asexual stages) proteins, showed an overlap of 150 genes identified in the transcriptional pilot time course. More than 95% of the genes identified and contained in both datasets were detected in the gametocytes samples only.

While this enrichment of gametocyte specific transcripts is likely highlighting only genes of interest involved in the pathways responsible for gametocyte development and preparation for development within the mosquito the extensive and sensitive transcriptome profiling aims to elucidate more subtle and early responses to *ap2-g*.

As we aim to analyse the entire time course with a view to elucidating the early players in gametocyte commitment the transcriptional similarity between these early stages is reassuring. The ring and trophozoite samples will contain both morphologically indistinguishable committed gametocytes as well as stages remaining in the asexual pathway and the transcriptome is likely to reflect this.

When the remainder of the transcriptional data has been sequenced, quality controlled and analysed we believe that the transcriptome data will provide a wealth of information about genes differentially expressed immediately after AP2-G expression considered the initiation of gametocyte commitment. Already we have identified 21 genes of interest that appear upregulated at 4hp-Ind. Many of these remain upregulated throughout the gametocyte development cycle (12) and those that aren't continually upregulated are still differentially expressed in more mature gametocytes (24hp-Ind, 8). Only one gene considered differentially expressed at 4hp-Ind is not considered differentially expressed at the other time points considered in the differential expression analysis (PBANKA_050720). This gene however, is again considered differentially expressed at 30hp-Ind. Interestingly, though annotated as only a conserved hypothetical protein, this gene shows high expression in ookinetes and its ortholog in *P. falciparum* (PF3D7_1023000) is most highly expressed in stage 5 gametocytes (López-Barragán *et al.*, 2011).

We anticipate that with higher statistical power from the biological replicates and the additional controls for variation caused by the induction drug rapamycin that we will be able to detect more subtle changes in expression early post induction and that these genes will provide insight into developmental pathways associated with differentiation to gametocytogenesis.

Chapter 6   Screen of potential gametocyte specific factors by gene disruption

6.1 Knockout screen of differentially expressed genes of interest identified in the pilot time course.

To determine if any of the genes of interest identified from the pilot analysis have specific gametocyte phenotypes, a selection were selected to be knocked out and their phenotype monitored to determine if any effect on gametocyte commitment, maturation or gender assignation was discernible.

6.1.1 Design of the screening assay

To ensure a large selection of genes of interest could be screened, a simple flow cytometry based assay, that could be completed directly after parasite transfection and selection, was devised to determine if any effect on gametocytes could be seen. This assay utilised the parental line (820) in which male (GFP driven by the male specific promoter PBANKA_041610) and female (RFP driven by the female specific promoter PBANKA_131950) gametocytes are distinguishable by their fluorescence proxy profiles. When analysing this line parasites were identified first by their positive nuclear stain (Figure 6.1 A). Parasites were distinguishable in blood samples by flow cytometry as only infected cells stain with the Hoechst nuclear dye, non-nucleated red blood cells did not. The Hoechst positive cells were then gated and analysed based on their GFP and RFP fluorescence (Figure 6.1 B). Three distinct populations were identified in this analysis; the Hoechst positive cells that display no additional fluorescence, asexual parasites (or gametocytes too immature to express the fluorescent proxies); GFP positive parasites, male gametocytes and RFP positive parasites, female gametocytes (Figure 6.1).
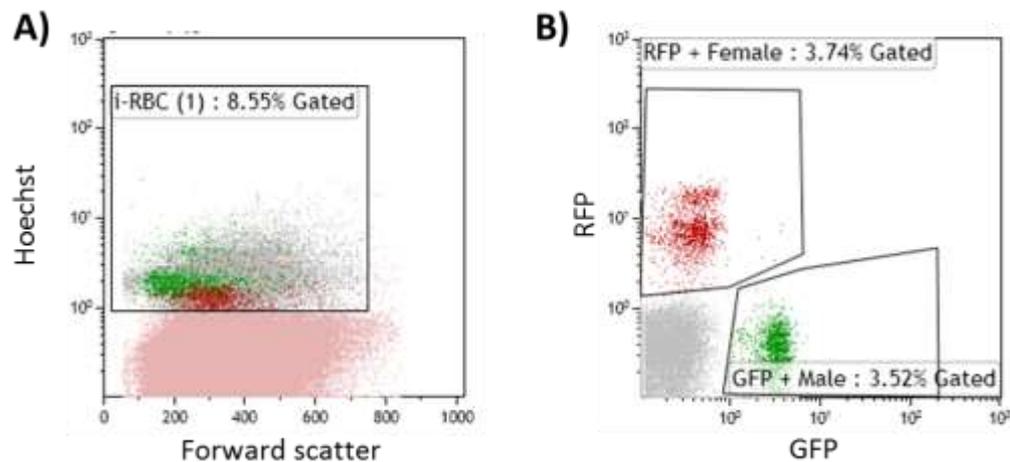
Figure 6.1 820 control flow cytometry analysis.

Figure shows a representative flow cytometry analysis of the parental 820 parasite line.

A) Flow cytometry analysis of the parental 820 line shows initially identifies parasite infected red blood cells by the Hoechst nuclear stain and takes these positive cells only to analyse GFP and RFP fluorescence.

B) The Hoechst positive population is analysed for the proportion of male (GFP positive) and female (RFP positive) gametocytes. While published data suggests commitment varies from 5% - 15% previous analysis with these fluorescent proxies has determine the range to be 3% - 10%.

Previous analysis of the 820 parental line identified some standard parameters to which the line adheres. Firstly, while commitment to gametocytes can be variable analysis of 10 independent infections, monitored over several days with increasing parasitaemias, showed total gametocyte percentage to be between 3.15% and 9.85% with the average being 5.58% (see section 4.4.1 and Figure 4.5). Secondly, the ratio of male to female gametocytes in a wild type infection displayed no more than 2.5x as many of one gender than the other (with these proxies it was observed more commonly with increased representation of male gametocytes) when parasitaemia was above 0.5%. Thirdly, if an *in vivo* infection was collected and allowed to mature *in vitro* overnight the ratio of males to females tended towards equal numbers. This is likely due to the initial male bias being an artefact of female proxy onset being significantly later in the development than the male. This balancing of male and female gametocytes likely occurs as the female gametocyte specific proxy (RFP driven by the female specific promoter PBANKA_131950) is not detectable until later than the male gametocyte specific proxy (GFP driven by the male specific promoter PBANKA_041610) and this additional maturation time, without reinvasion and subsequent commitment allows detection and accurate measurement of the

female gametocytes. Finally, as overall parasitaemia increases so does commitment to gametocytaemia.

To identify potential gametocyte specific phenotypes these parameters were used to define some factors that would allow unbiased classifications of phenotypes. Table 6.1 defines the 9 different classifications used to characterise knockout lines. It is worth noting that the final classification, likely asexual essential, may be premature in its definition as a single transfection for each vector was attempted and only one knockout vector was used. Furthermore we note that the classifications are entirely based on the onset of a proxy that is expressed from a single gender specific promoter and that this may indicate a difference in the population in question it by no means will identify every phenotype that could be exhibited. For example; a GFP negative line could still generate male gametocytes and just not express PBANKA_041610 which would eliminate expression of the GFP signal also; likewise a RFP positive line could result in the development of aberrant female gametocytes but because the proxy is expressed their presence is indicative of no phenotype by these measures.

Classification	Code	Description/Criteria
Gametocyte non-producer	GNP	No proxies are detectable for male or female gametocytes
Reduction in Gametocyte commitment	< G	Gametocyte commitment has fallen below the minimum demonstrated in the parental line (3%) with a parasitaemia above 1%
Increase in Gametocyte commitment	> G	Gametocyte commitment has exceeded the maximum demonstrated in the parental line (12%) irrespective of parasitaemia
Loss of Males	M-NP	No male specific (GFP) proxy is detected
Reduction in Males	< M	Gametocytes of both genders are detected but the ratio of males is at least 3x lower than females
Loss of Females	F-NP	No female specific (RFP) proxy is detected
Reduction in Females	< F	Gametocytes of both genders are detected but the ratio of females is at least 3x lower than males
No phenotype	N-Ph	Gametocyte commitment is within the range of the parental line (820)
Asexual essential	E	No parasites were recovered after transfection though controls in the batch were successful

Table 6.1 Classification of gametocyte specific phenotypes in the pilot screen

To ensure consistency of phenotype calling 9 classifications were defined based on comparisons to the 820 parental line and its variability as defined over 10 independent infections (Chapter **).

6.1.2 Selection of potential gametocyte specific genes

To ensure high integration rates which would allow for direct screening of phenotypes after transfection knockout recombineering vectors from the PlasmoGEM resource (Gomes *et al.*, 2015; Schwach *et al.*, 2015) were obtained for 40 genes of interest. These vectors use large homology arms for high efficiency integration and replace all or part of the gene of interest with a selectable marker cassette (Table 6.2, Vector maps in Supplementary 1. Recombineering vector maps) The 40 genes selected for this analysis were all identified in the RNAseq pilot time course as differentially expressed. Of the 40 selected, 7 of the candidates were more than 2 fold upregulated by 4hp-Ind, 14 were upregulated by 5 fold by 12hp-Ind and the remainder were a minimum of 5 fold upregulated by 24hp-Ind (and come from the analysis summarised in Figure 5.9). In an attempt to identify novel candidates as well as phenotypes specific to gametocytes and even to individual gender gametocytes, several candidates from each of the time points selected were annotated as conserved hypotheticals and several were previously identified, by proteomics data, as male (green box), female (red box) or gametocyte (orange box) specific (Table 6.2).

6.1.3 Summary of the screen

All knockouts screened were classified into one of the 9 groups previously described (Table 5.1). The knockout phenotypes have been summarised in Table 6.2.

Seven of the knockouts attempted were not recovered within two weeks of the transfection and selection and were therefore deemed likely essential (E) for the asexual cycle. Twenty of the knockouts showed no phenotypic difference (N-Ph) to the 820 parental line. In these cases the proportion of male and female gametocytes was not more than 3 fold different between the genders and 3 – 12% of the population expressed the fluorescent proxies for gametocytes. However, it should be noted that though these 20 knockouts showed no phenotypic difference to the parental line, only the effect on gametocyte percentage and the fluorescence ratio (as a proxy for gender) was analysed. They have also not all been shown to have correctly integrated the

recombineering vector or to have lost the gene of interest (Table 6.2). Further work could use different assays to elucidate potential functions for these genes in later processes such as ookinete conversion or transmission.

The remaining thirteen knockouts were split in to two categories, those with clear and distinct gametocyte specific phenotypes (five) and those with potential gametocyte specific phenotypes (eight) and will be discussed further.

Gene	Annotation	Upregulated from	Line made	Int?	Phenotype?					Fig
					Parasitaemia %	Gametocytes %	Male %	Female %	Classification	
PBANKA_090850	myosin heavy chain subunit	24hpInd	Yes	Yes	3.54	4.49	1.83	2.66	N-Ph	6.2 A
PBANKA_101830	dynein-associated protein	12hpInd	Yes	No	6.06	4	1.97	2.03	N-Ph	6.2 B
PBANKA_121830	CPW-WPC family protein	12hpInd	Yes	No	4.75	4.24	2.84	1.4	N-Ph	6.2 C
PBANKA_070470	conserved Plasmodium protein	12hpInd	Yes	No	1.65	4.46	2.58	1.88	N-Ph	6.2 D
PBANKA_142920	conserved Plasmodium protein	12hpInd	Yes	No	7.27	9.73	6.1	3.63	N-Ph	6.2 E
PBANKA_081180	actin-like protein	12hpInd	Yes	Yes	2.82	3.3	1.54	1.76	N-Ph	6.2 F
PBANKA_093360	dynein light chain 2	24hpInd	Yes	No	5.83	11.13	6.7	4.43	N-Ph	6.2 G
PBANKA_021390	dynein light chain	24hpInd	Yes	Yes	1.13	3.03	1.13	1.96	N-Ph	6.2 H
PBANKA_111500	conserved Plasmodium protein	12hpInd	Yes	No	3.37	9.18	4.22	4.96	N-Ph	6.3 A
PBANKA_050830	conserved Plasmodium protein	24hpInd	Yes	No	2.28	5.91	3.69	2.22	N-Ph	6.3 B
PBANKA_133500	conserved Plasmodium protein	24hpInd	Yes	No	2.33	5.95	3.78	2.17	N-Ph	6.3 C
PBANKA_133810	conserved Plasmodium protein	24hpInd	Yes	Yes	7.35	7	3.5	3.5	N-Ph	6.3 D
PBANKA_010510	conserved Plasmodium protein	24hpInd	Yes	No	9.87	5.74	2.64	3.1	N-Ph	6.3 E
PBANKA_122550	conserved Plasmodium protein	12hpInd	Yes	Yes	8.37	6.15	3.01	3.14	N-Ph	6.3 F
PBANKA_122990	conserved Plasmodium protein	24hpInd	Yes	No	8.37	9.5	4.5	5	N-Ph	6.3 G
PBANKA_121260**	HAP2	12hpInd	Yes	Yes	2.64	3.02	1.81	1.21	N-Ph	6.3 H
PBANKA_050500**	dihydrolipoamide	4hpInd	Yes	No	5.34	4.75	3.04	1.71	N-Ph	6.4 A
PBANKA_082800**	conserved Plasmodium protein	24hpInd	Yes	Yes	4.19	9.89	6.45	3.44	N-Ph	6.4 B
PBANKA_041720**	conserved Plasmodium protein	24hpInd	Yes	No	3.22	11.32	6.48	4.84	N-Ph	6.4 C
PBANKA_050500(HA)**	dihydrolipoamide acyltransferase	24hpInd	Yes	No	6.11	11.07	5.31	5.76	N-Ph	6.4 D
PBANKA_112040	Pfs77 homologue	12hpInd	Yes	Yes	3.03	4.88	3.82	1.06	< F	6.5 A
PBANKA_060120	dynein heavy chain	24hpInd	Yes	Yes	3.81	4.14	3.2	0.94	< F	6.5 B
PBANKA_113860	conserved Plasmodium protein	24hpInd	Yes	No	5.99	7.68	6.18	1.5	< F	6.5 C
PBANKA_041340**	conserved Plasmodium protein	24hpInd	Yes	No	6	2.42	0.31	2.11	< M	6.5 D
PBANKA_112510(HA)**	FabB/FabF	24hpInd	Yes	Yes	8.13	2.34	0.93	1.41	< G	6.5 E
PBANKA_136040	conserved Plasmodium protein	24hpInd	Yes	No	4.88	12.45	7.59	4.86	> G	6.5 F
PBANKA_143520**	conserved Plasmodium protein	4hpInd	Yes	No	4.05	15.13	8.42	6.71	> G	6.5 G
PBANKA_101830**	dynein associated protein	24hpInd	Yes	No	2.22	13.15	5.15	8	> G	6.5 H
PBANKA_050440	conserved Plasmodium protein	12hpInd	Yes	Yes	3.11	0.09	0.02	0.02	GNP	6.6 A
PBANKA_090720	conserved Plasmodium protein	24hpInd	Yes	Yes	3.97	0.1	0.04	0.03	GNP	6.6 B
PBANKA_010240**	conserved Plasmodium protein	4hpInd	Yes	No	8.85	4.36	0.06	4.3	M-NP	6.7
PBANKA_031270	DEAD/DEAH helicase	12hpInd	Yes	No	3.63	3.13	0.55	2.58	< M	6.8
PBANKA_090240**	conserved Plasmodium protein	24hpInd	Yes	Yes	4.91	2.1	2.03	0.07	F-NP	6.9
PBANKA_141810**	conserved Plasmodium protein	4hpInd	No	N/A	N/A	N/A	N/A	N/A	E	
PBANKA_041720**	RNA-binding protein	12hpInd	No	N/A	N/A	N/A	N/A	N/A	E	
PBANKA_135250	conserved Plasmodium protein	24hpInd	No	N/A	N/A	N/A	N/A	N/A	E	
PBANKA_083040**	conserved Plasmodium protein	24hpInd	No	N/A	N/A	N/A	N/A	N/A	E	
PBANKA_112510**	FabB/FabF	4hpInd	No	N/A	N/A	N/A	N/A	N/A	E	
PBANKA_051060**	conserved Plasmodium protein	12hpInd	No	N/A	N/A	N/A	N/A	N/A	E	
PBANKA_121260**	conserved Plasmodium protein	4hpInd	No	N/A	N/A	N/A	N/A	N/A	E	

Table 6.2 Summary of the knockout screen

Table 6.2 Summary of the knockout screen

Recombineering gene knockout vectors were obtained for each of the genes of interest. In these vectors a part, or the entire gene of interest is replaced with a selectable marker. To ensure efficient recombination large homology arms to the upstream and downstream gene regions are used to integrate.

A subset of genes that were upregulated from 4, 12 or 24 hours post Induction were selected for knock out. Of these several were selected with conserved hypothetical gene annotations. From existing proteomics data (Khan *et al.*, 2005) genes isolated in specific gametocyte fractions were selected. Two male (green gene number), two female (red gene number) and one mixed (orange gene number) gametocyte population genes were also chosen.

If after transfection parasites were not recovered, but transfections controls were obtained, the line was considered asexually essential and not generated (E).

Integration PCRs specific to the gene of interest were attempted for all lines obtained, however due to the length of homology arms used for the generation of the recombineering vectors integration was not always detectable, even in lines displaying clear phenotypes. However, only one integration pair was attempted per knockout.

Phenotypes were determined on the knockout population from 7 days post transfection. For a phenotype to have been recorded it must have been consistent over multiple days (minimum of 2) and retained the phenotype when cultures *in vitro* for the final day.

All knockouts that were obtained were compared to an 820 parental line though the classifications were based on criteria defined over multiple infections.

Genes denoted ** were subsequently re-annotated due to a mis-annotation of a plate of recombineering vectors. Two previously denoted KO vectors were in fact HA-tagging vectors.

6.1.4 Knockouts with no gametocyte specific phenotype

As summarised above (Table 6.2), seven of the knockouts attempted were not recovered. While this indicated there may be a role for these seven genes in the asexual proliferative cycle the data is by no means complete. Further attempts to disrupt these genes with recombineering vectors will be attempted. If no direct knockouts are obtainable then conditional systems to disrupt gene expression may be attempted to elucidate their function.

Twenty of the knockouts screened showed no phenotypic difference in gametocyte commitment and gender segregation to the parental 820 line. This was defined as committing to between 3% and 12% gametocytes as determined by the presence of the fluorescent proxies for male (GFP) and female (RFP) gametocytes and to have a ratio of males to females that did not differ by more than 3 fold biased to either gender. As a standard all lines were analysed on multiple days post transfection and the parasitaemia had to exceed 1% before gametocytaemia was determined. This parasitaemia threshold was set as lower parasitaemias are more prone to low gametocyte commitment and variable gender proportions.

A summary of the knockouts can be seen in Table 6.2 and representative flow cytometry analysis for the knockouts screened, and characterised as having no gametocyte specific phenotype can be seen below. In each figure a representative 820 parental control line is shown for reference only (Figure 6.2, Figure 6.3 & Figure 6.4).
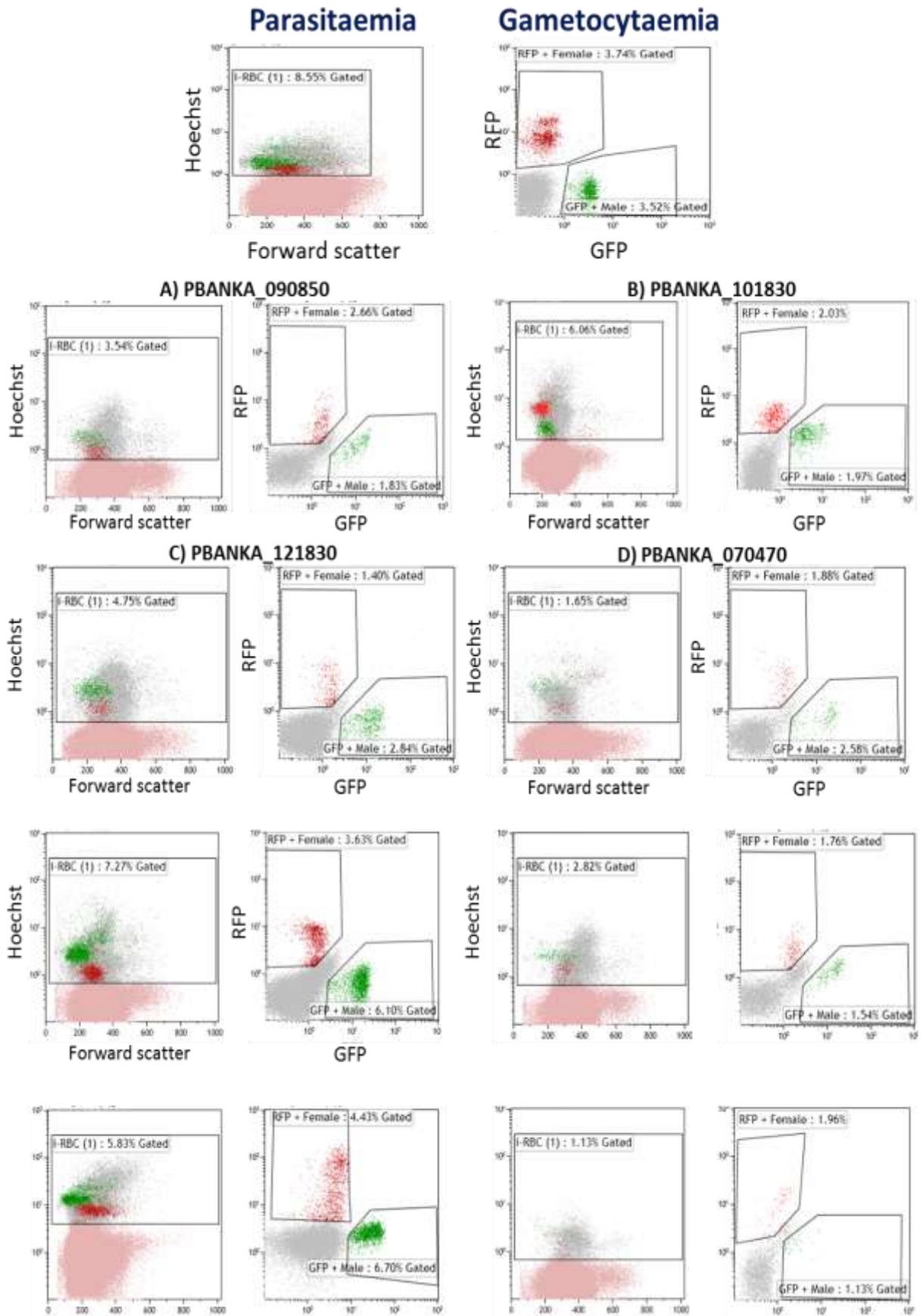
Figure 6.2 Flow cytometry screen showing knockouts (1 - 8) with no gametocyte specific phenotype.

Figure 6.2 Flow cytometry screen showing knockouts (1 - 8) with no gametocyte specific phenotype

For definitions of gametocyte specific phenotype classifications see Table 6.1

A) PBANKA_090850, myosin heavy chain subunit, shows no classified differences to the parental 820 line.

B) PBANKA_101830, dynein-associated protein, shows no classified differences to the parental 820 line.

C) PBANKA_121830, CPW-WPC family protein, shows no classified differences to the parental 820 line.

D) PBANKA_070470, conserved plasmodium protein, shows no classified differences to the parental 820 line.

E) PBANKA_142920, conserved plasmodium protein, shows no classified differences to the parental 820 line.

F) PBANKA_081180, actin-like protein, shows no classified differences to the parental 820 line.

G) PBANKA_093360, dynein light chain 2, shows no classified differences to the parental 820 line.

H) PBANKA_021390, dynein light chain, shows no classified differences to the parental 820 line.
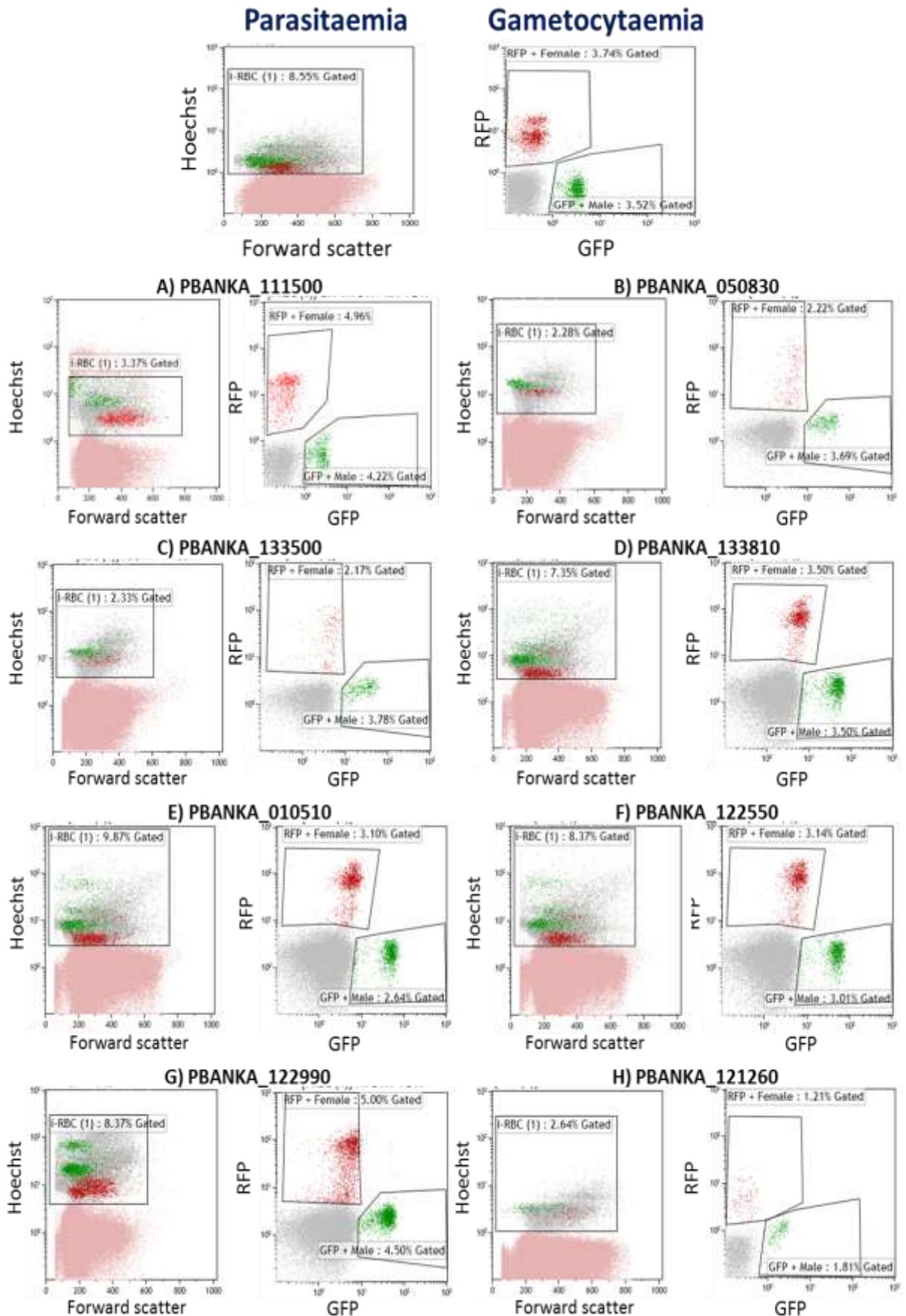
Figure 6.3 Flow cytometry screen showing knockouts (9 - 16) with no gametocyte specific phenotype.

Figure 6.3 Flow cytometry screen showing knockouts (9 - 16) with no gametocyte specific phenotype.

For definitions of gametocyte specific phenotype classifications see Table 6.1

A) PBANKA_111500, conserved plasmodium protein, shows no classified differences to the parental 820 line.

B) PBANKA_050830, conserved plasmodium protein, shows no classified differences to the parental 820 line.

C) PBANKA_133500, conserved plasmodium protein, shows no classified differences to the parental 820 line.

D) PBANKA_133810, conserved plasmodium protein, shows no classified differences to the parental 820 line.

E) PBANKA_010510, conserved plasmodium protein, shows no classified differences to the parental 820 line.

F) PBANKA_122550, conserved plasmodium protein, shows no classified differences to the parental 820 line.

G) PBANKA_122990, conserved plasmodium protein, shows no classified differences to the parental 820 line.

H) PBANKA_121260, male gametocyte fusion factor HAP2, shows no classified differences to the parental 820 line.
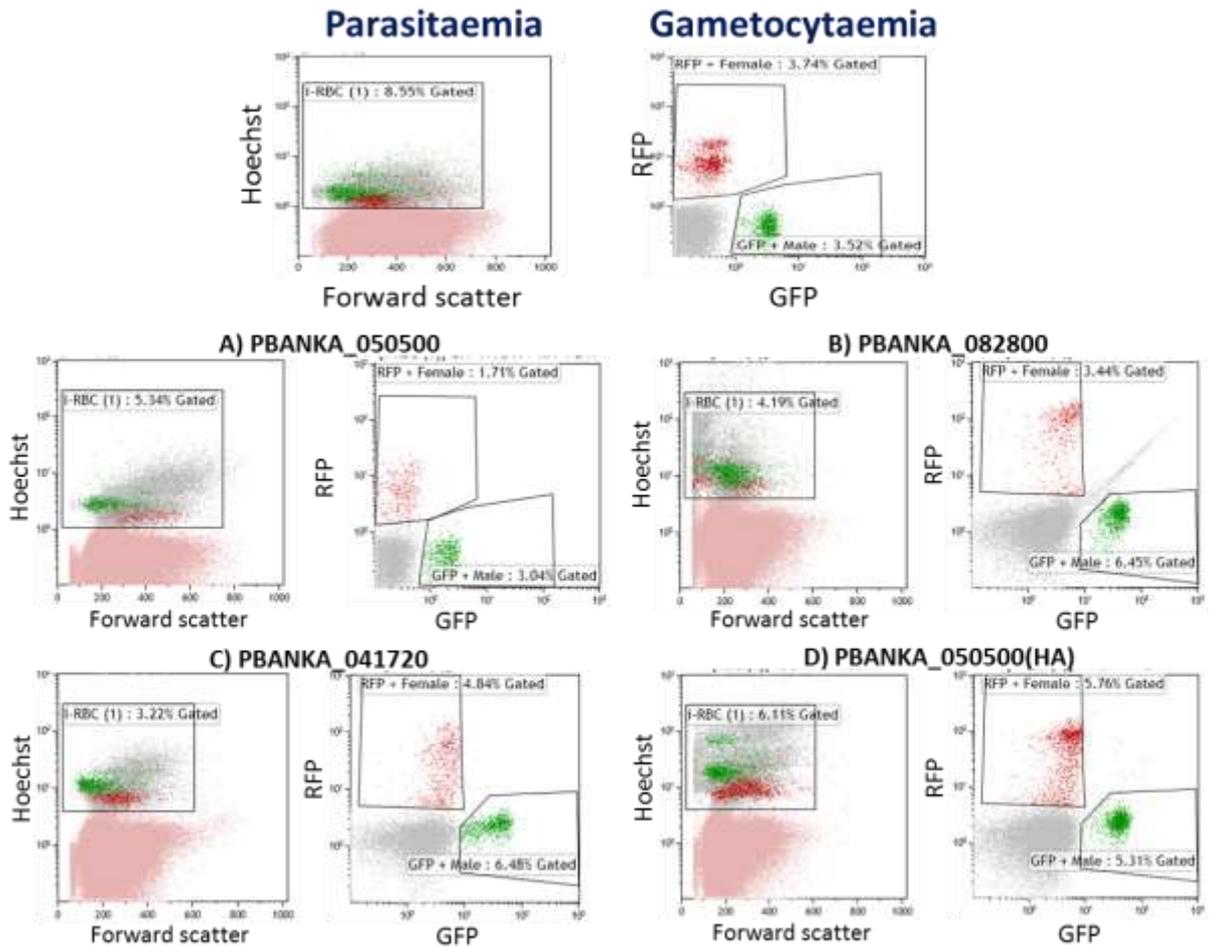
Figure 6.4 Flow cytometry screen showing knockouts (17 - 20) with no gametocyte specific phenotype.

For definitions of gametocyte specific phenotype classifications see Table 6.1.

A) PBANKA_050500, dihydrolipoamide acyltransferase, shows no classified differences to the parental 820 line.

B) PBANKA_082800, conserved plasmodium protein, shows no classified differences to the parental 820 line.

C) PBANKA_041720, conserved plasmodium protein, shows no classified differences to the parental 820 line.

D) PBANKA_050500(HA), dihydrolipoamide acyltransferase, shows no classified differences to the parental 820 line.

6.1.5 Potential gametocyte specific phenotypes

The eight knockouts with potential gametocyte specific phenotypes showed consistency in their phenotype over several days of monitoring *in vivo* and retained this phenotype when allowed to fully mature *in vitro*, however the phenotypes were not as distinct as the five identified and classified as having definite gametocyte specific phenotypes.

Of these eight potentials three showed a significant reduction in female gametocytes compared to males. This classification was only applied to lines where the proportion of female gametocytes was at least three fold less than male gametocytes. Two lines (PBANKA_112040 and PBANKA_060120, Figure 6.5 A & B) produced only 1/3 the percentage female (RFP positive) gametocytes when compared to males (GFP positive). Another (PBANKA_113860, Figure 6.5 C) contained only 1/4 the proportion of females when compared to males.

One of the gene knockouts appeared to have a reduction in the proportion of male gametocytes when compared to females. This classification was only applied to lines where the proportion of male gametocytes was at least three fold less than female gametocytes. This line (PBANKA_041340, Figure 6.5 D) generated approximately 1/7 the proportion of male gametocytes (GFP positive) when compared to female (RFP positive).

One of the gene tags showed a potential reduction in the proportion of parasites that commit to becoming gametocytes (PBANKA_050500, Figure 6.5 E). This classification was only applied if the overall commitment to gametocytogenesis was below 3% when parasitaemia was above 1% and if the reduction was consistent over multiple cycles. This line showed less than 2.5% commitment to gametocytes, even when the parasitaemia had reached high levels (above 8%). This reduction with a tagged line could indicate an inhibited role for the protein due to the presence of the HA-tag.

Three of the knockout lines appeared to show an increase in gametocyte commitment when compared to the parental line. This classification was only imposed if gametocytaemia exceeded 12% consistently over multiple cycles. One knockout (PBANKA_136040, Figure 6.5 F) consistently showed commitment of

approximately 12.5% to gametocytes with a gender ratio within the normal parameters for the parental line. A second knockout (PBANKA_143520, Figure 6.5 G) consistently showed gametocyte commitment levels of approximately 15% based on the fluorescent proxies. The final knockout classified as a gametocyte over-producer (PBANKA_101830, Figure 6.5 H) showed commitment of more than 13% in multiple cycles.

While these phenotypes could provide valuable insight into gametocyte commitment and gender differentiation time constraints have prevented further analysis on any of these lines. To be confident that their phenotypes were authentic and consistent it would also be advisable to screen over a longer period of time and with multiple independent generations of the mutants. Initial screening was completed directly from transfection of the knockout vector to reduce animal usage and initially identify potential candidates from a large pool of possibilities.
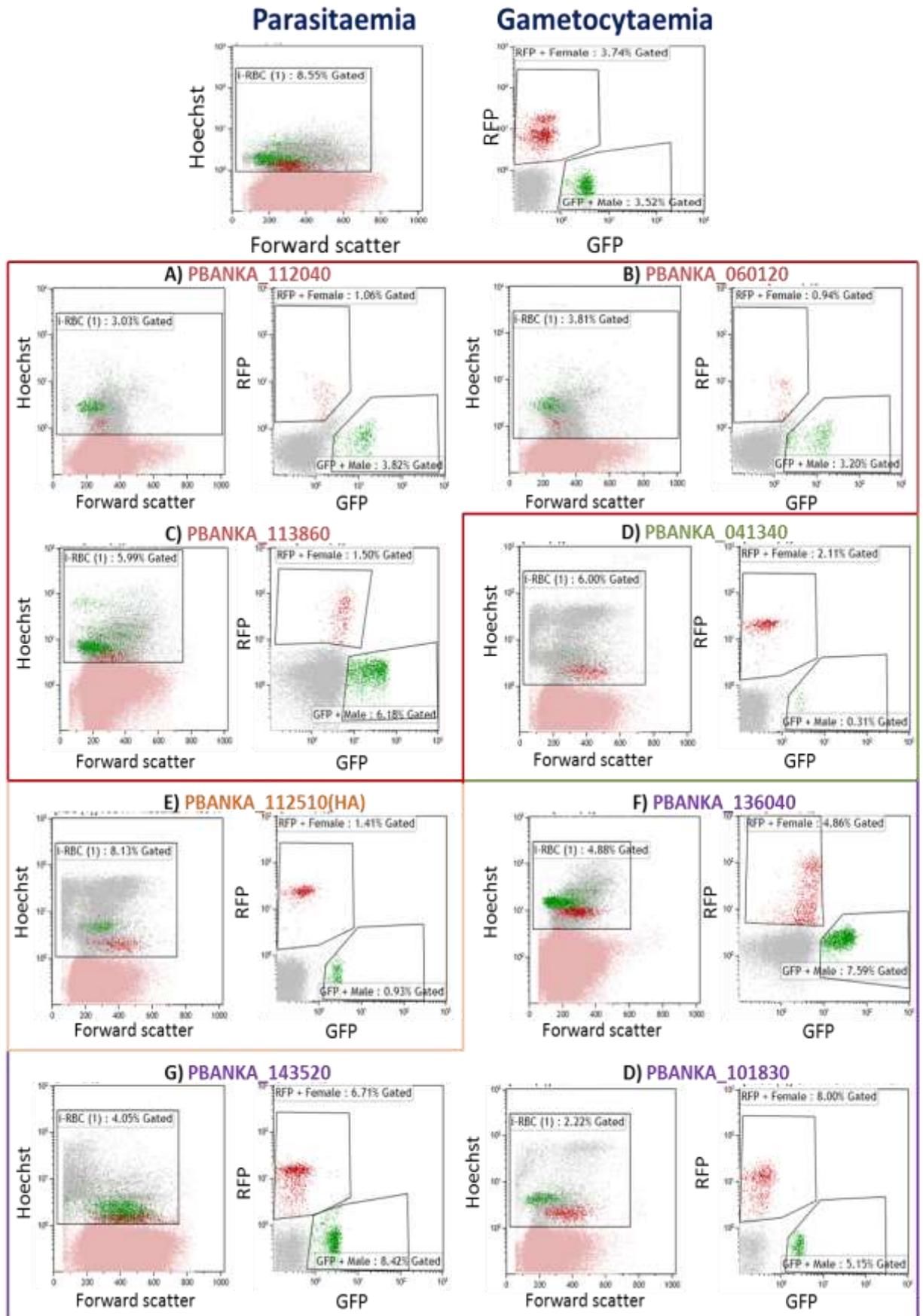
Figure 6.5 Flow cytometry screen showing the eight knockouts with potential gametocyte specific phenotypes.

Figure 6.5 Flow cytometry screen showing the eight knockouts with potential gametocyte specific phenotypes.

For definitions of gametocyte specific phenotype classifications see Table 6.1

Figure is coded based on classifications, red boxed are all lines with a potential reduction in female gametocytes, greed boxed is the line demonstrating a potential reduction in male gametocytes, orange boxed is the line showing a potential reduction in gametocyte commitment and purple boxed are the lines showing potential increases in gametocyte commitment.

A) PBANKA_112040, Pfs77 homologue, shows a 3-fold reduction in female gametocytes related to males when compared to the parental 820 line.

B) PBANKA_060120, dynein heavy chain, shows a 3-fold reduction in female gametocytes related to males when compared to the parental 820 line.

C) PBANKA_113860, conserved plasmodium protein, shows a 4-fold reduction in female gametocytes related to males when compared to the parental 820 line.

D) PBANKA_041340, conserved plasmodium protein, shows a 7-fold reduction in male gametocytes related to females when compared to the parental 820 line.

E) PBANKA_112510(HA), FabB/FabF shows a reduction in gametocytes (<2.5%) compared to the parental 820 line.

F) PBANKA_136040, conserved plasmodium protein, shows an increase in gametocytes when compared to the parental 820 line. Commitment appears to be approximately 12.5%.

G) PBANKA_143520, conserved plasmodium protein, shows an increase in gametocytes when compared to the parental 820 line. Commitment appears to be approximately 15%

H) PBANKA_101830, dynein associated protein, shows an increase in gametocytes when compared to the parental 820 line. Commitment appears to be approximately 13%.

.

6.1.6 Gametocyte specific phenotypes

While the eight knockout lines described above (6.1.5 Potential gametocyte specific phenotypes) were classified as having a phenotypic difference when compared to the normal range of characteristics exhibited by the parental line (see Table 6.1 for details of classifications) their classification was less robust than the five defined as having definite gametocyte specific phenotypes.

As these five were the most definite candidates they were investigated more thoroughly and will eventually be characterised fully. Their classifications were; two gametocyte non-producers (GNP), one male non-producer (M-NP), one reduced male producer (< M) and one female non-producer (F-NP).

6.1.6.1 Gametocyte non-producers

Of the five genes that were identified as having distinct gametocyte specific phenotypes, two (PBANKA_ 050440 & PBANKA_090720) gave rise to lines that no longer produced gametocytes as determined by the loss of GFP (male gametocytes) and RFP (female gametocytes) fluorescence (Figure 6.6). To confirm this phenotype knockout lines were generated again with both vectors independently. These repeats showed the same loss of gametocyte production. To determine that there was actually a loss of gametocytes, and not just a loss of the fluorescence proxies, giemsa smears were also analysed over subsequent days of infection. No gametocytes were identifiable over multiple cycles and with extended culture of blood *in vitro* to allow full maturation of any gametocytes present. Furthermore, these gametocyte non-producing lines were cloned (all clones displaying the same GNP phenotype) and transmission through mosquitoes was attempted. In two independent experiments transmission of these lines was unsuccessful but the control 820tbb parental line was transmitted to naïve mice with the usual kinetics (blood stage parasites detected 4 – 5 days post mosquito feeding).
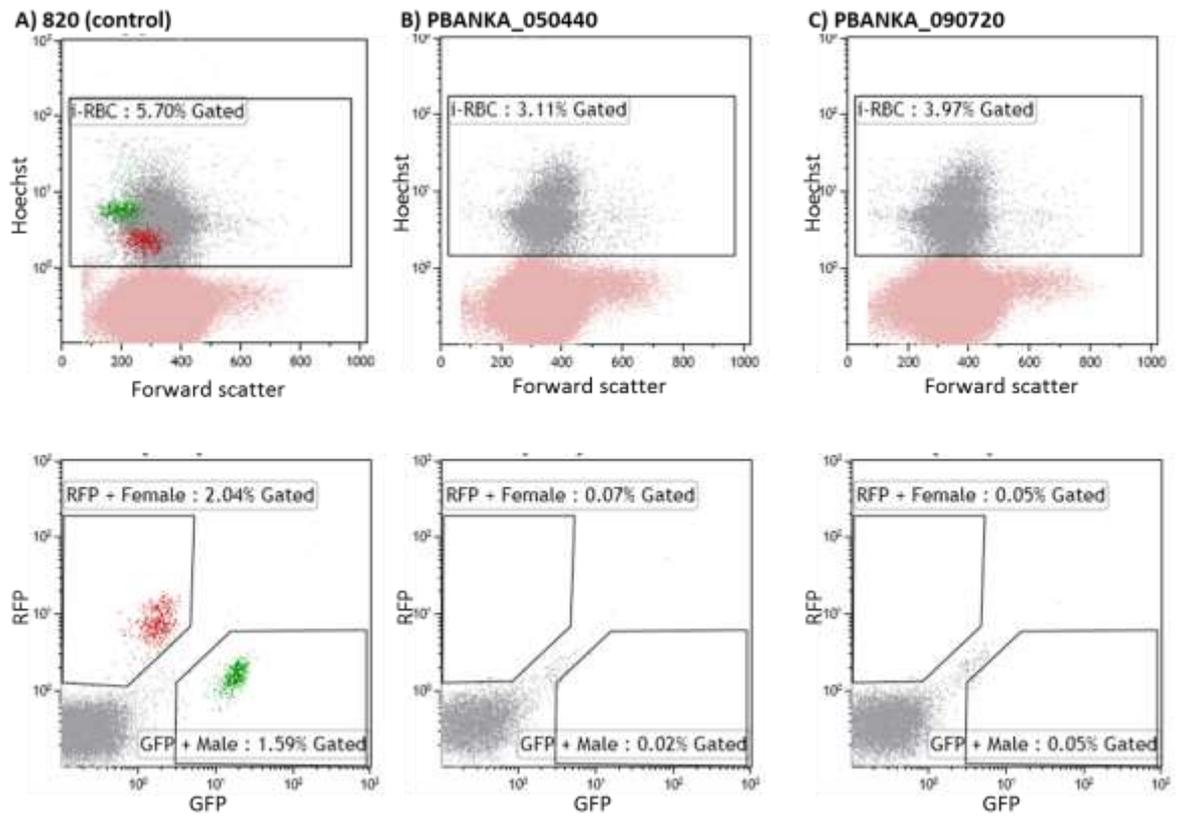
Figure 6.6 Flow cytometry analysis of the two gametocyte non-producer lines (PBANKA_050440-KO and PBANKA_080720-KO).

A) The parental control line, 820, shows clear fluorescence of approximately 3.6% gametocytes with almost equal proportions of male (GFP) and female (RFP) gametocytes.

B) The PBANKA_050440, conserved hypothetical protein; knockout line no longer shows any fluorescence signal for gametocytes. This phenotype is consistent in clonal lines.

C) The PBANKA_080720, conserved hypothetical protein; knockout line no longer shows a fluorescence signal for gametocytes. This phenotype is consistent in clonal lines.

6.1.6.2 Male specific phenotypes

Two of the lines appeared to affect only the production of male gametocytes. The first of these, PBANKA_010240, a conserved hypothetical protein, results in a complete loss of male gametocytes (initially determined by a loss of the GFP proxy signal in flow cytometric analysis). Over multiple cycles and with extended culture to allow for maturation of gametocytes, no male gametocytes were identifiable by their fluorescent proxy (Figure 6.7). To ensure that the loss was not just of the fluorescent proxy, mature *in vitro* blood cultures were smeared and giemsa stained. While female gametocytes were morphologically identifiable, males were not present. It was also clear from this longer term analysis that the proportion of female gametocytes generated in each cycle appears normal and the fluorescence intensity and morphology mirror that of the 820 control line.
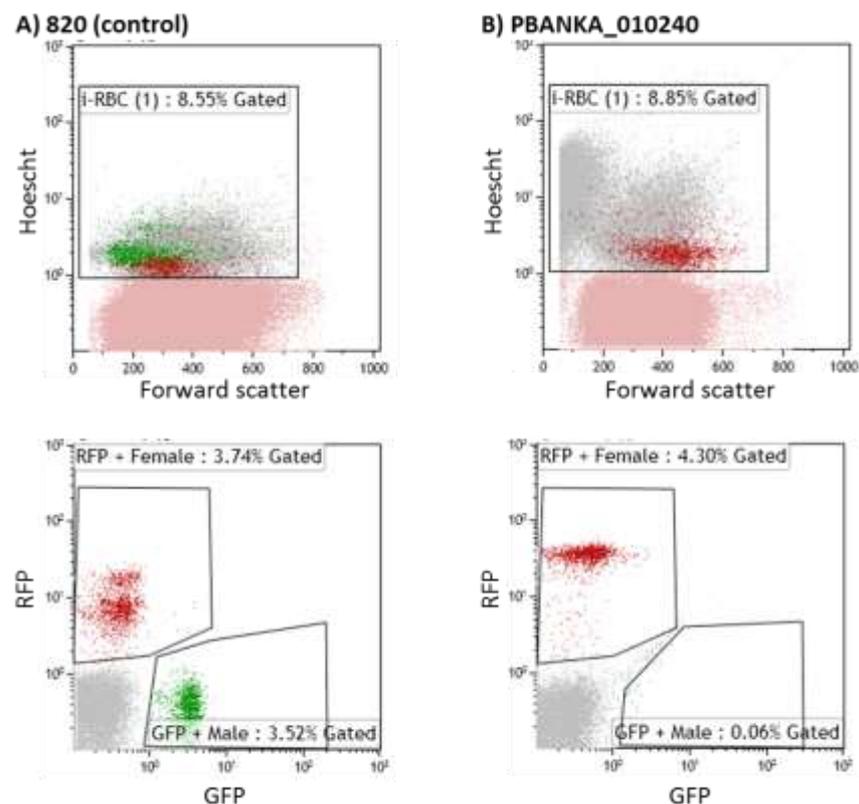
Figure 6.7 Flow cytometry analysis of the male non-producer line (calmodulin-like protein, PBANKA_010240-KO).

A) The parental control line, 820, shows clear fluorescence of approximately 7.5% gametocytes with similar proportions of male (GFP) and female (RFP) gametocytes.

B) The PBANKA_010240, conserved hypothetical protein, knockout contains female gametocytes (RFP) in the population but no male gametocytes (GFP). Over subsequent days analysis male gametocytes are never detected. Even if *in vitro* cultures are retained no emergence of mature males appears.

The second male specific gametocyte phenotype was the DEAD/DEAH helicase, PBANKA_031270. In this line there appears to be a marked decrease in the number of male gametocytes, determined by the reduction in the population expressing a GFP signal.

In addition to this, it was also noticed that the GFP signal intensity of the males that were identified was weaker than males in the parental line. When a blood culture was allowed to mature *in vitro* the gametocytes in the culture mature and GFP (driven by the male specific promoter) and RFP (driven by the female specific promoter) expression reach their maximum intensities. The few males that were present in the population, and detectable by GFP fluorescence, in theory should reach their highest GFP expression intensity when allowed to mature *in vitro.* However, in this line the GFP signal did not increase to that of the parental line (Figure 6.8) even in the extended culture. In addition to this both *in vivo* and these extended *in vitro* cultures were analysed by giemsa smears. Flow cytometry analysis indicated that male gametocytes were formed but did not reach maturity based on their GFP signal, this was corroborated by the giemsa morphology analysis. Mature male gametocytes were never identifiable in the giemsa smears whereas female gametocytes were readily identified.

As both male and female gametocytes were present in this knockout line it was pertinent to check if the gametocytes formed were viable. To achieve this we attempted to generate ookinetes *in vitro* and checked to determine if transmission could be achieved *in vivo.* To ensure transmissions were not enabled by the presence of wild type gametocytes the knockout line was first cloned. In three independent ookinete conversion experiments, using the clonal knockout population, no successful conversion to ookinetes was detected. Furthermore, in three separate transmission experiments no infection was detected in mosquitoes and transmission to naïve mice was not achieved. This indicated that the line was not viable for transmission and while it is tempting to attribute this to the lack of morphologically identifiable mature male gametocytes further work using genetic crosses will have to be completed to robustly demonstrate this.
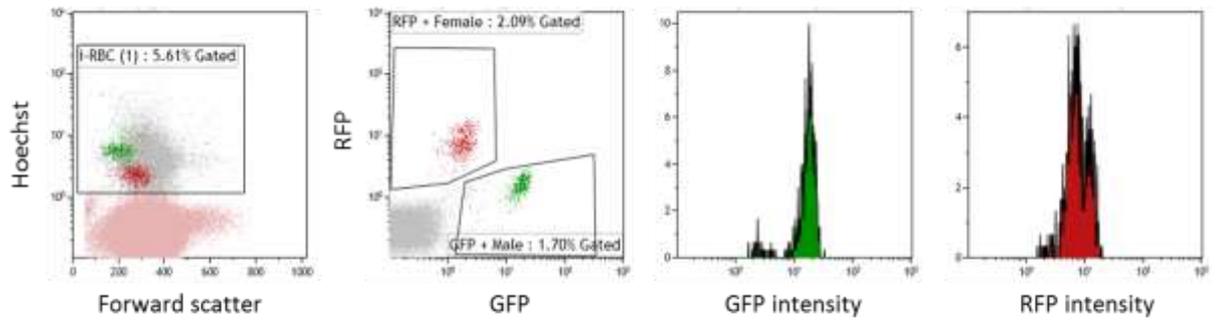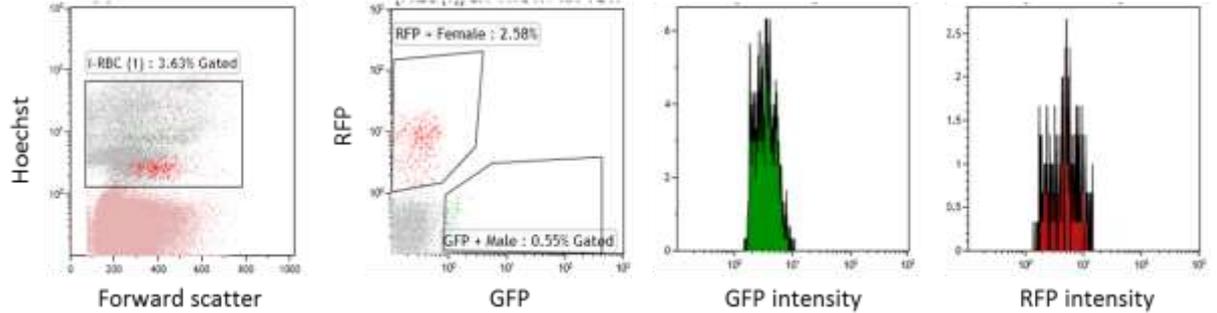
Figure 6.8 Flow cytometry analysis of the reduced male producing line (DEAD/DEAH helicase, PBANKA_031270-KO)

A) The parental control line 820, shows a fluorescence signal for both male and female gametocytes. The relative intensities of GFP and RFP are shown when a culture has matured and therefore fluorescence intensity has reached a consistent maximum.

B) In the knockout line both male and female gametocytes are detectable. After *in vitro* culture, the female (RFP) intensity seems similar to that of the parental line. The male (GFP) intensity was markedly decreased in the knockout when compared to the control line. This reduced intensity was consistent over multiple experiments.

6.1.6.3 Female specific phenotypes

The disruption of the conserved hypothetical protein, PBANKA_090240, resulted in a phenotype of a lack of production of female gametocytes *in vivo*. Over subsequent days tail drops were monitored and female gametocytes were not detected in the blood stream, determined by a lack of fluorescence signal in the RFP channel (Figure 6.9 B).

However, the standard protocol for this assay required the maturation of a blood culture to allow full development of the gametocytes. When infected blood was cultured, both male (determined by the GFP fluorescence) and female (determined by the RFP fluoresence) gametocytes become detectable after extended culture (Figure 6.9 C). While we began to see a population of RFP expressing parasites these remained quite indistinct. This is unusual in a matured culture, where all gametocytes should have reached full maturity and therefore express the proxies to their highest intensity.

Taken together these data could implicate many reasons for the lack of female gametocytes *in vivo,* including clearance by the host immune system. However it is also interesting that even with extended *in vitro* culture that should allow complete maturation of gametocytes the fluorescence intensity is potentially 10-fold less intense than the control line. Furthermore when looking at smears of the extended culture mature female gametocytes were not obvious in the culture whereas males were. This could indicate that while present the females were not mature enough to be morphologically distinguishable from asexual blood stages.
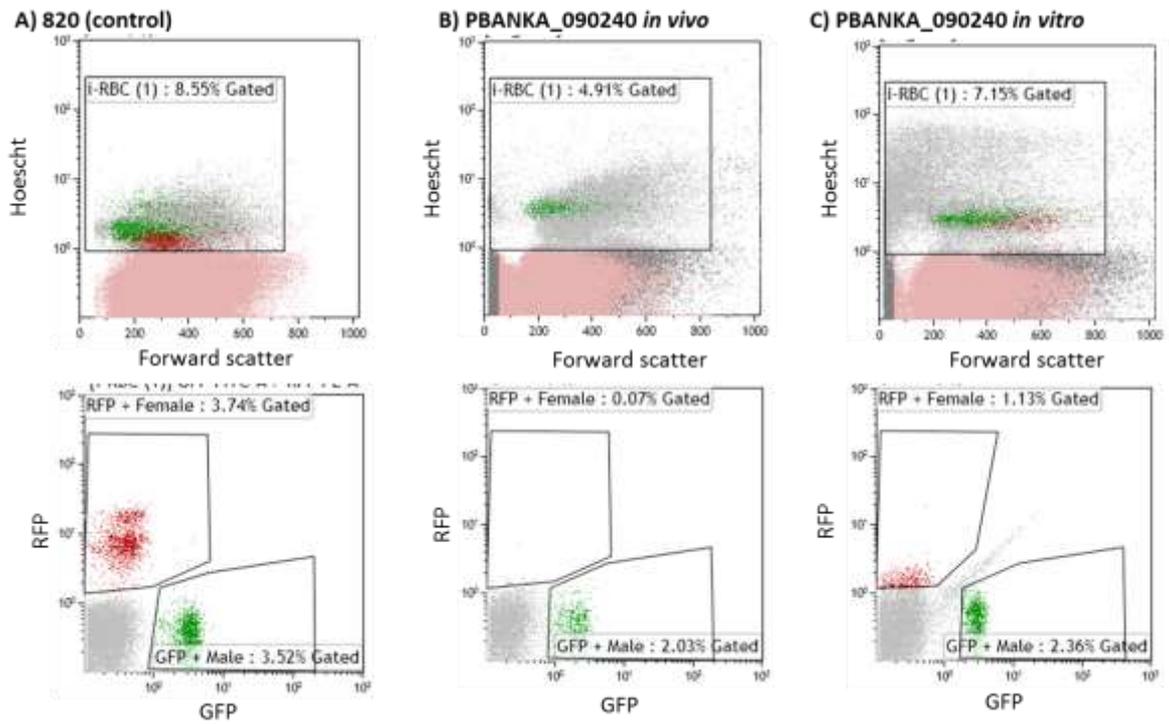
Figure 6.9 Flow cytometry analysis of the female non-producing line (Conserved hypothetical protein, PBANKA_090240-KO).

A) The parental control line, 820, shows a fluorescence signal for both male and female gametocytes. This fluorescence is similar fin proportion for both males and females.

B) In the PBANKA_090240 knockout line *in vivo* male gametocytes are clearly detectable with a percentage within the normal range for the parental line. However, female gametocytes are not observed based on their RFP signal.

C) When parasites are cultured for an additional 24 hours *in vitro* then the male population shows a very compact range of GFP intensities. In the RFP channel it appears female gametocytes become detectable. The fluorescence signal observed for these females, while distinct to the asexual, non-fluorescent population, is not as intense as an 820 signal even though the gametocytes have been cultured *in vitro* and therefore should be mature.

6.2 **Discussion**

As a screening methodology for gametocyte specific phenotypes the flow cytometry based assay designed here appears to yield reliable assignation of phenotypes. As the screen can be carried out directly from transfection it reduces animal usage and can identify potential candidates rapidly. It has been proposed that this assay be combined with the recent screening method used at the sanger institute (Gomes *et al.*, 2015) where pools of barcoded knockout vectors were screened for growth phenotypes in asexual stages in a high throughput manner. If this was employed pools of barcoded knockout with no asexual growth defect would be transfected and intravenously injected into naïve mice. After selection male (GFP driven by the gametocyte specific promoter PBANKA_041610) and female (RFP driven by the gametocyte specific promoter PBANKA_131950) gametocytes would be sorted and sequencing used to identify barcodes present in the samples. Any barcode missing from male, female or both gender gametocytes would be identifiable as having a potential impact on gametocyte formation or development (Figure 6.10).

In a non-high-throughput manner this screening method can and will be used in subsequent experiments with additional genes of interest identified in the full transcriptome screen (Figure 5.12).

It has already been utilised to screen 40 knockouts resulting in the identification of 7 genes with an essential role for asexual proliferation, 8 with potential roles in gametocyte commitment and maturation and 5 with confirmed roles. Of the confirmed roles two knockouts resulted in lines that were unable to generate gametocytes, one that generated no male gametocytes and one that generated no females and finally one that generated fewer males that seem not to mature fully.
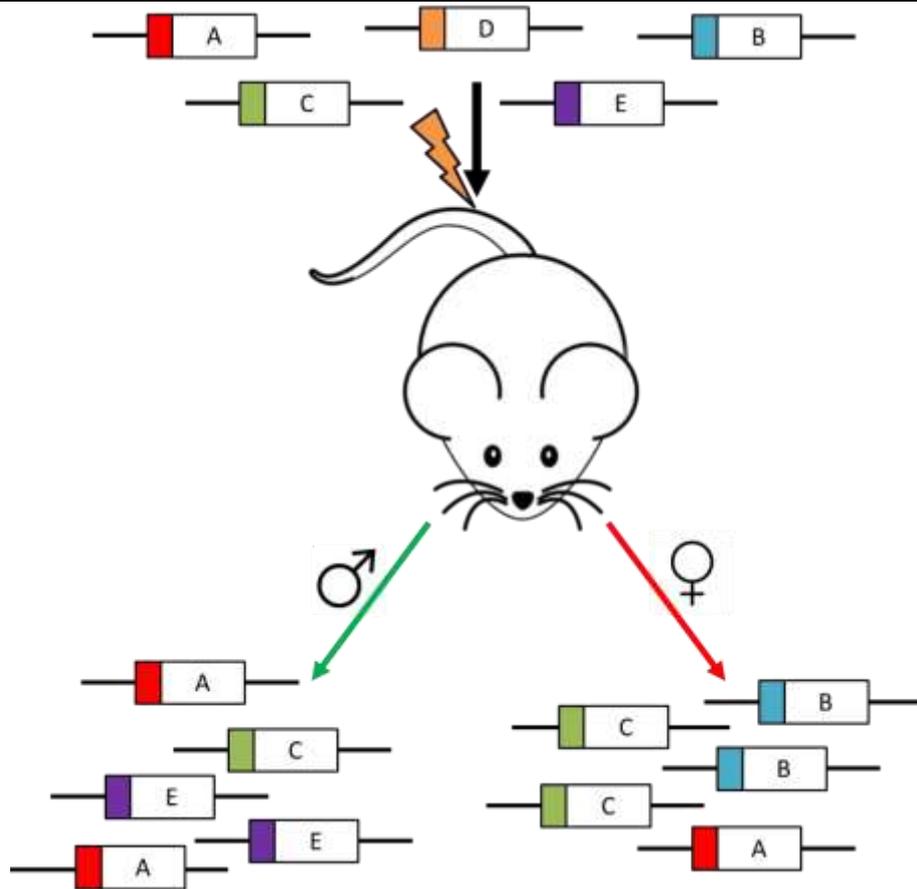
Figure 6.10 Schematic representation of high throughput adaptation of the gametocyte specific phenotype screen.

Pools of barcoded knockouts known not to have a growth phenotype in asexual stages would be transfected into *P. berghei* schizonts and intravenously injected into naïve mice. After selection gametocytes would be sorted based on their fluorescent proxies (GFP for males and RFP for females) and sequencing used to identify the presence or absence of barcodes in each population. In this example barcode D gives a knockout that does not generate any fluorescently identifiable gametocytes (of either gender), barcode E is a knockout that does not generate mature female gametocytes and barcode B is a knockout that does not generate mature male gametocytes.

The 5 genes that have been preliminarily studied to elucidate their function will be investigated further but in some cases their annotation or annotations of orthologous genes gives some indication of potential function.

The two GNP lines obtained in this work were annotated as conserved hypothetical proteins. However, orthologous proteins in other *Plasmodium* species identified orthologs of PBANKA_050440 as putative flagellar dyneins. In previous proteome analysis in *P. berghei* (Khan *et al.*, 2005; Lasonder *et al.*, 2002) and *P. falciparum* (Silvestrini et al., 2010) dyneins have largely been identified in male gametocyte samples. This has been attributed to a role in the formation of the axoneme required for motile male gametes though a single

dynein has also been associated with female gametocytes. This is a novel identification for a dynein as neither gender gametocyte is detectable and only one dynein has previously been identified in female gametocytes. To uncover the exact role of this putative dynein in gametocyte biology work is underway to generate tagged versions of the protein, both HA and GFP tagged. This should allow us to discover the function of the protein in gametocytes.

The second conserved hypothetical protein that generated a GNP line when knocked out contains no orthologs with more descriptive annotations therefore its function cannot be estimated. Domain searches also yield no further information about a potential role for the protein. We are also generating GFP and HA tagged versions of this protein to try and uncover its biological function.

The two male specific phenotypes uncovered give us a clearer path to follow to elucidate their functions.

While annotated as a conserved hypothetical protein PBANKA_010240 the protein contains two transmembrane domains and an N-terminal ARID (AT-rich interaction domain) domain (Figure 6.11). Very few transcription factors have been identified in *Plasmodium spp.* therefore the identification of an ARID/BRIGHT DNA binding domain containing protein with a gender specific phenotype is intriguing. In higher eukaryotes ARID proteins make up a large family of transcription factors involved in cell proliferation, development and differentiation (Iyer *et al.*, 2008). In the protozoan parasite *Giardia lamblia* ARID/BRIGHT domain proteins are conserved and have been shown to transactivate the *cwp1* (cyst wall protein 1) promoter. This protein is involved in cyst wall formation (Wang *et al.*, 2007). The upregulation of this transcript from 4hpi could identify as it a potential gender specific transcription factor involved in the onset of a cascade that allows differentiation and development of male gametocytes.

Figure 6.11 Domain architecture of PBANKA_010240
Sequence analysis reveals the presence of two transmembrane domains (Blue boxes) and an N-terminal ARID domain (orange hexagon) along with many low complexity regions (pink boxes) and a coiled coil domain (green box). These domains help elucidate function (Schultz *et al.*, 1998).

The second male specific phenotype was found with the knockout of a DEAD/DEAH helicase (PBANKA_031270). This family of helicases has been implicated throughout the life cycle of *plasmodium* and in many different roles in many organisms. The helix structure of the proteins is involved in the ATP-dependant unwinding of DNA and RNA. Several DEAD-helicases and many DEAH-helicases have been implicated in pre-mRNA splicing, others DEAD-helicases have been implicated in translation initiation, spermatogenesis, cellular differentiation, carcinogenesis and RNA synthesis. The most well studied DEAD-box helicase in *Plasmodium* is the DDX6 ortholog DOZI (development of zygote inhibited, PBANKA_121770). This acts as a translational repressor in the female gametocyte. The loss of DOZI impedes the ability of the female gametocyte to store and stabilise multiple mRNA necessary for zygote development. Several of the mRNAs translationally repressed are detected as proteins as soon as 2 hours after female gamete activation and fertilisation (P25 and P28) indicating a rapid release and translation after fertilisation (Mair *et al.*, 2006; Mair *et al.*, 2010).
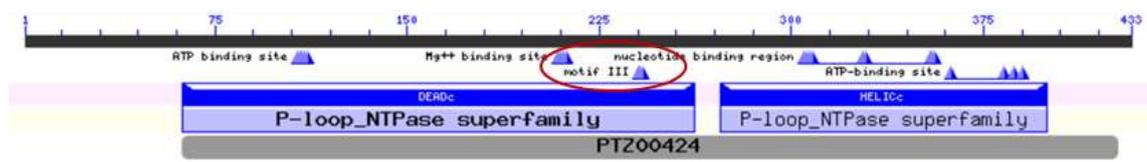
Figure 6.12 Domain comparison of the DOZI helicase and the putative DEAD/DEAH helicase

Using NCBIs conserved domains software (Marchler-Bauer *et al.*, 2014; Marchler-Bauer *et al.*, 2011) we can see two of the DEAD/DEAH motifs of DOZI are also conserved in the putative helicase here. However the motif III (red circle) is not in the putative helicase. Several ATP binding domains and the Mg^{++} binding site are also conserved between the two.

While purely speculative it could proposed that the DEAD/DEAH-helicase that prevents maturation of male gametocytes plays a role in differentiation of male gametocytes specifically (as females seem to develop normally) or that its absence prevents translation of other specific factors required for male gametocyte development (Abdelhaleem *et al.*, 2003; Tuteja & Pradhan, 2006).

While a possibility is that this helicase could play a corresponding role to DOZI in male gametocytes no evidence currently exists for the translational repression of mRNA in male gametocytes and the motif III domain found in DOZI (associated with single stranded-RNA binding) is not found in this putative helicase (Banroques *et al.*, 2010). Work is underway to generate GFP and HA tagged versions of this protein to facilitate study of its role in gametocyte development.

While female gametocytes of this line generally appear normal and male gametocytes do not appear to mature correctly (morphologically by giemsa smears and through flow cytometry analysis) it has not been possible to generate ookinetes (*in vitro)* or transmit the line through mosquitoes. This indicates that one or both of the gender gametocytes are not viable. It will be necessary to carry out genetic crosses with both a male defective line and a female defective line to demonstrate the viability or lack thereof for both gender gametocytes. These crosses will be carried out with *Pb47-ko* parasites and *Pb48/45-ko* parasites to test the viability of the female and male gametocytes respectively. It is still possible that the female gametocytes are also arrested or non-viable and expression of the proxy is just not affected. The elucidation of the role of this helicase will be guided by the results from this as both gender gametocytes are committed to so the phenotype is likely one of development and maturation.

Finally the female specific phenotype discovered in the screen proves the most difficult to unravel. The targeted protein, PBANKA_090240, a conserved plasmodium protein, is only disrupted by deleting 1.4kb of the coding sequence at the 3' end of the gene (Figure 6.13). In the most recent genome annotation this gene has also been combined with the gene PBANKA_090230 (adopting its gene reference) increasing the coding sequence even further. The deletion of this region may be insufficient for the removal of the 2709 amino acid protein and could result in spurious phenotype assignation based on a truncated protein in the cells.
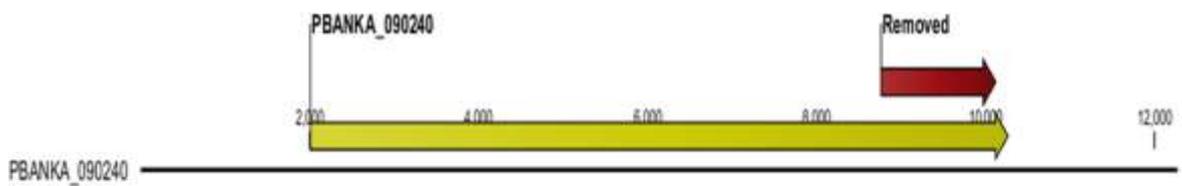
Figure 6.13 Gene disruption PBANKA_090230/090240

Schematic representation of the gene disruption used to target PBANKA_090240.

This protein shows no annotated domain structure and syntenic orthologs identify any potential functional components. Further work will aim to elucidate its function by directly knocking out the gene to confirm a functional knockout and to tag the protein to identify localisation and interacting partners.

It is also worth noting that several of the knockouts with potential gametocyte phenotypes are consistently expressed in gametocytes in *P. falciparum* and *P. berghei*. Some have annotations that could provide even more insight into biological differences between asexual parasites and the sexually differentiated gametocytes.

The PBANKA_112040 knockout, that results in a reduction in female gametocytes, was identified as a gametocyte specific marked in *P. falciparum* (*pf77*) and has been shown to only be expressed by female gametocytes (Baker *et al.*, 1995).

PBANKA_041340 knockout results in a reduction in male gametocytes and encodes a zinc-finger protein that also has multiple coiled coil regions. Zinc finger proteins have been implicated in many biological processes such as transcriptional regulation and DNA/RNA binding in many organisms. Its potential gender specific role would need further characterisation including tagging to identify interactors and localisation.

The PBANKA_112510 knockout has been identified as essential in this screen however the HA-tagged version of the gene appears to reduce the proportion of gametocytes. Due to this disparity it may be beneficial to conditionally knockout the gene or degrade the protein to determine if its role in asexual parasites and gametocytes is different.

The PBANKA_050500 knockout showed a potential reduction in the proportion of gametocytes. This protein is putatively annotated as a part of the pyruvate dehydrogenase complex (dihydrolipoamide acyltransferace) whose role is to link glycolysis and the citric acid cycle. Other knockouts of components of this complex have been investigated in *P. yoelii* and shown to be non-essential for asexual proliferation and sexual stages but indispensable for liver development and merozoites delivery to the blood stream. However, its expression profile shows high expression in *P. falciparum* (which is not the case in *P. yoelii)* indicating there may be a differing role. This enzyme might be dispensable to the parasites but may also play a role in metabolism in the gametocytes. Its loss may not completely abolish gametocyte development but may have an impact on the the fitness of gametocytes and reduce the proportion that survive until morphological identification or detection by the fluorescent proxies (Foth *et al.*, 2005; Pei *et al.*, 2010).

All three of the potential gametocyte over-producing knockouts were annotated with potentially interesting functions. They were all detected in gametocytes either in *P. falciparum* or *P. berghei*.

PBANKA_136040 encodes a conserved hypothetical protein with an interesting structure. This small protein contains 13 exons at a size of only 270 amino acids (very similar in *P. falciparum)* it also contains a transmembrane domain at its N-terminus. While it could be suggested the differentially spliced forms of this protein could be important for different stages or even genders biology it is also worth noting that during the asexual cycle this protein is poorly spliced (Sorber *et al.*, 2011) indicating its splicing is not tightly regulated.

PBANKA_143520 encodes another zinc-finger protein that also has multiple coiled coil regions. Zinc finger proteins have been implicated in many biological processes such as transcriptional regulation and DNA/RNA binding in many organisms. The potential that the loss of this zinc finger increases gametocyte commitment could indicate a negative regulatory role.

PBANKA_101830 encodes a protein designated a dynein associated protein but protein structure analysis reveals the presence of 2 WD40 domains. In *Plasmodium* asexual stages and gametocytes a WD40 protein has been identified involved in the formation and anchoring of protein adhesion complexes (von Bohl *et al.*, 2015).
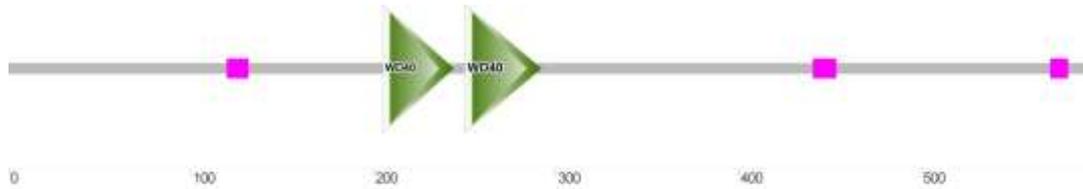
Figure 6.14 Domain architecture of PBANKA_101830
Sequence analysis reveals the presence of two WD40 domains (Green arrowheads) along with several low complexity regions (pink boxes). The identification of these WD40 domains may help elucidate function (Schultz *et al.*, 1998)

Chapter 7   Screen of potential gametocyte specific factors by promoter expression

7.1 Identification of candidates

Twelve of the genes identified in the pilot time course (Figure 5.9) as differentially expressed had been previously investigated in the lab (Generation of lines and primary expression analysis carries out by Dr K. Hughes and published in (Sinha *et al.*, 2014)). In this work 17 genes were identified as potentially gametocyte specific when *ap2-g* was identified as the transcription factor responsible for beginning the cascade of sexual commitment (Sinha *et al.*, 2014). Selection of the genes of interest in that work was based on two criteria. A selection of genes of interest (11) were chosen based on their differential expression in gametocyte non-producer lines (GNP9, GNP8, GNP7 generated through serial passage of asexual parasites and the *ap2-g* knockout) compared to the parental line 820. A second set were chosen based on stage specific transcriptomes generated from the *ap2-g* knockout line compared to the parental line (rings, trophozoites and schizonts). In an attempt to identify genes involved in the early stages of commitment a selection of genes (6) were chosen based on differential expression at the ring or trophozoite stage. Finally a constitutively expressed gene (*hsp70)* was used as a control. All of these candidate genes (except PBANKA_143220 & PBANKA_051910) contained at least one AP2-G binding motif upstream of the transcription start site (Sinha *et al.*, 2014).

To generate the lines for analysis, 2kb of the promoter of interest was cloned into a *cfp* expression vector. This vector was then integrated into the silent P230p locus in the 820 parental line (all parasite lines generated by KH). This parental line already expresses fluorescent proxies for mature male (GFP under the male gametocyte specific PBANKA_041610 promoter) and mature female (RFP under the female specific PBANKA_131950 promoter) gametocytes. Integration of each *cfp* line resulted in expression of CFP under the control of the promoter of interest thereby allowing for analysis of its stage expression. The initial promoter strategy used to determine stage and gender specific expression was carried out on non-synchronous, uncloned lines using a flow cytometry based analysis (Figure 7.1). Briefly, infected red blood cells were identified based on their nuclear DNA stain with DyeCycle Ruby® (Figure 7.1 A). The infected red blood cells were then

characterised by their fluorescence as asexual (non-fluorescent), male gametocytes (GFP fluorescence driven by a male gametocyte specific promoter) or female gametocytes (RFP fluorescence driven by a female gametocyte specific promoter). It is worth noting that until mature, stage specific expression of GFP or RFP immature gametocytes would always be classified as asexual due to their lack of fluorescence (Figure 7.1 B). Separately the total population of CFP positive parasites was identified (Figure 7.1 C). Finally the CFP positive population was back-gated on to the flow cytometry analysis that classified the population as asexual, male or female (Figure 7.1 D).
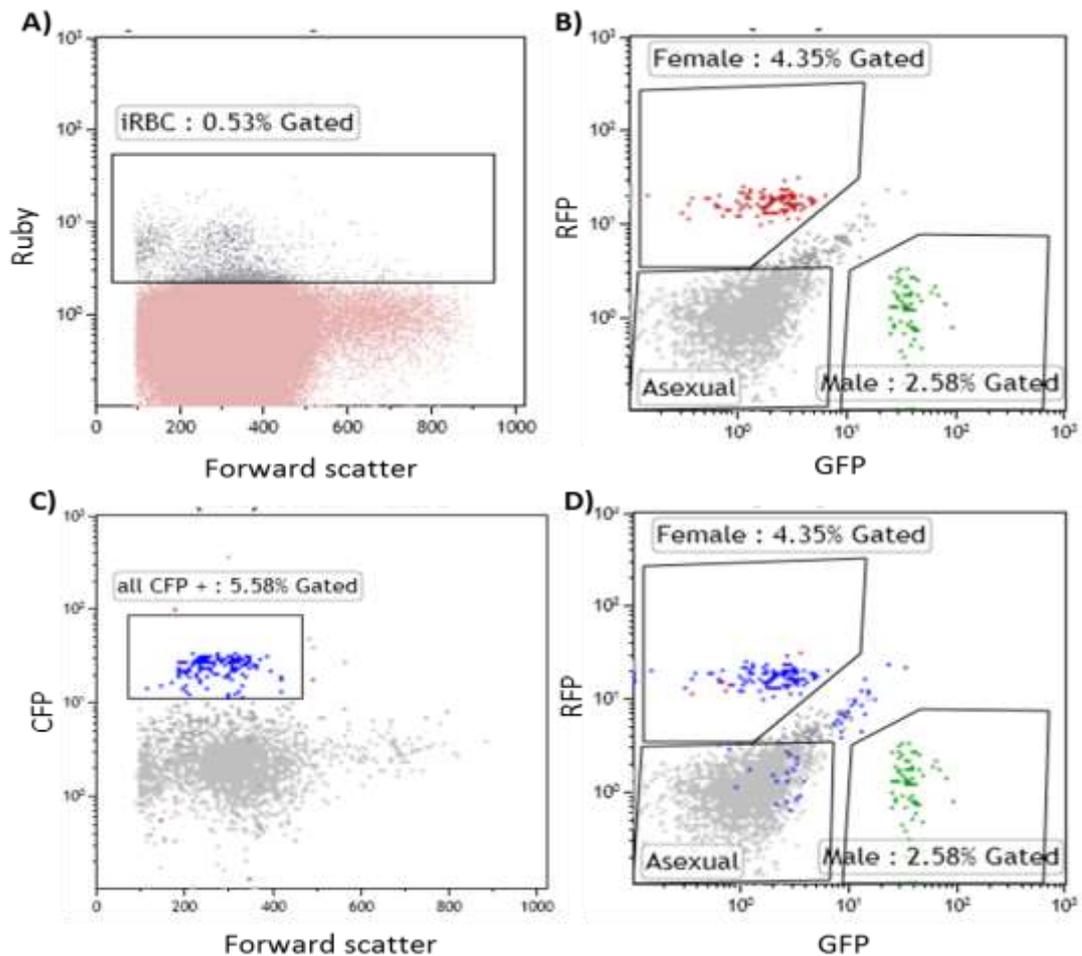
Figure 7.1 Stage specific expression flow cytometry assay

The entire method shown in this figure is used to demonstrate the stage and gender specific expression of a gene based on the use of 2kb of its promoter to drive expression of a fluorescent proxy (CFP).

A) Infected red blood cells are initially identified based on the nuclear stain with DyeCycle Ruby®.

B) Infected red blood cells are then classified as either asexual parasites (those that do not express any fluorescent proxies) male gametocytes (those expressing GFP under the male specific promoter) or female gametocytes (those expressing RFP under the female specific promoter).

C) The proportion of the parasites expressing CFP is also quantified and gated.

D) Back-gating of the CFP positive population (C) onto the sexually classified populations (B) shows if there is stage or gender specific expression of CFP with the promoter of interest.

Using this strategy the 18 selected promoters were analysed (all assays and analysis were carried out by KH) and their expression classified (Table 7.1). Of the 17 experimentally selected promoters of interest 12 were also classified as of interest in the differential gene analysis from the pilot RNAseq experiment (Figure 5.9).

Of the 17 selected promoters analysed four showed expression in only female gametocytes, five showed expression in both male and female gametocytes (one of which shows expression in a small subset of asexual parasites), three showed no expression in blood stages at all (though one showed expression in ookinetes), two showed expression in all blood stages (all of the population or a subset) and three showed expression in male gametocytes (one only in a subset with a subset of the asexual population showing expression also). The constitutively expressed control (hsp70) showed expression at all staged (Table 7.1)

Gene	Description	Why?	Why? Sinha *et al*	CFP expression analysis			
				Asexual	F	M	Ookinete
PBANKA_133370	phosphodiesterase delta (PDEdelta)	24h FC	GNP vs. WT	N	Y	N	NT
PBANKA_110800	transcription factor IIb, putative	24h FC	GNP vs. WT	N	Y	N	NT
PBANKA_143220	male development gene 1 (MDV1)	24h FC	GNP vs. WT	N	Y	Y	NT
PBANKA_081070	subpellicular microtubule protein 1 (SPM1)	12h FC	GNP vs. WT	N	faint	N	NT
PBANKA_144900	aspartyl protease, putative	24h FC	GNP vs. WT	N	Y	Y	Y
PBANKA_050440	conserved Plasmodium protein	12h FC	GNP vs. WT	N	N	N	faint
PBANKA_103790	conserved Plasmodium protein	24h FC	GNP vs. WT	N	N	N	NT
PBANKA_143520	zinc finger domain	4h FC	ring/troph	N	Y	N	Y
PBANKA_051500	25 kDa ookinete surface antigen precursor	24h FC	GNP vs. WT	subset	Y	subset	Y
PBANKA_140300	small heat shock protein	12h FC	ring/troph	subset	N	subset	NT
PBANKA_142770	"RuvB-like protein 1 (RUVB1)	4h FC	GNP vs. WT	subset	N	Y	N
PBANKA_134040	oxidoreductase	24h FC	GNP vs. WT	N	N	Y	NT
PBANKA_101870	conserved Plasmodium protein	n/a	ring/troph	subset	Y	Y	Y
PBANKA_082670	conserved Plasmodium protein	n/a	ring/troph	N	N	N	NT
PBANKA_061520	calcium dependent protein kinase 4	n/a	GNP vs. WT	N	Y	Y	NT
PBANKA_051910	CCAT-binding transcription factor-like protein	n/a	ring/troph	Y	Y	Y	N
PBANKA_144230	conserved Plasmodium protein	n/a	ring/troph	N	Y	Y	NT
PBANKA_071190	hsp70	n/a	control	Y	Y	Y	Y

Table 7.1 Summary of previous CFP promoter analysis

The eighteen promoters selected (2kb) were cloned into a *cfp* expression vector and transfected into the silent P230p locus of the 820 parental line that expresses fluorescent proxies for male (GFP) and female (RFP) gametocytes. Infected red blood cells were initially identified based on their dyecycle ruby nuclear stain. In parallel all parasites were stage/gender classified as male (GFP), female (RFP) or asexual (no fluorescence) parasites and a CFP positive population was defined. The CFP positive population was then back-gated onto the stage/gender classified population to determine if stage and gender specific expression of CFP was occurring (Figure 7.1). To show gender specificity the *cfp* expression vector was also transfected into a gametocyte non-producer line and the absence of CFP noted in all but the control line. To determine if CFP was expressed in ookinetes some of the lines were cultured to produce ookinetes and the presence or absence of a CFP signal determined by microscopy. The top twelve genes in this table correspond to those that were also identified in the RNAseq pilot time course as potentially gametocyte specific. All experiments pertaining to this promoter analysis was completed by KH (Sinha *et al.*, 2014).

To extend this assay we wanted to determine if we could use these lines, expressing CFP under gametocyte or gender specific promoters, to more precisely define the onset of the promoter of interest and by association the gene. Generation of all lines used, and their primary analysis, identifying them as gametocyte specific genes, was completed by Dr K. Hughes.

7.1.1 Promoter onset determination by mean CFP expression

To ascertain if this was possible three of the candidate genes identified (Sinha *et al.*, 2014) were selected. Two of these candidates were also identified in the pilot RNAseq screen as differentially expressed from 4hp-Ind (4 hours post induction) indicating they may be involved in the early stages of commitment. One of the mutual candidates showed male specific expression (PBANKA_142770, RuvB1) and the other showed female specific expression (PBANKA_143520, zinc-finger). To get the most out of developing this assay we also selected the gender unspecific promoter that appeared to be expressed earliest. This gene (PBANKA_101870, conserved hypothetical) was not identified in the RNAseq pilot time course but was identified as having reduced expression at the trophozoite stage in the stage specific microarray comparisons between the *ap2-g* knockout compared to the wild-type (Sinha *et al.*, 2014).

To ascertain when the promoter of interest becomes active enough to detect CFP expression, the flow cytometry assay developed previously (initially developed by KH) was extended upon. Briefly, parasites, identified through nuclear staining with DyeCycle Ruby, were categorised as asexual (no additional fluorescence signal), male gametocytes (GFP positive) or female gametocytes (RFP positive). These individual populations were then analysed to determine the proportion that was CFP positive and the intensity of the CFP expression (if present).

If, for example, CFP expression was driven by a constitutive promoter all populations (asexual, male gametocytes and female gametocytes) would have detectable CFP fluorescence. If, however, the promoter was driving CFP expression in a gametocyte population but not asexual parasites CFP expression would only be detected in the gametocytes. If the onset of the promoter driving CFP expression was earlier than the current male or female specific promoter driving GFP and RFP respectively then as soon as male or female gametocytes were identifiable (by their GFP/RFP fluorescence) CFP would also be detected, however, prior to expression of GFP/RFP the CFP signal would be classified as

asexual. If expression of CFP occurred in a subset of the asexual population, for example those not yet classifiable as gametocytes, we hoped to detect an increase in the mean CFP intensity and use it to ascertain when the promoter had become active.

The three promoter lines selected for this analysis were first synchronised by intravenous injection of purified schizonts into naïve mice. Two hours after this synchronous infection parasites were harvested and mature gametocytes that had been injected along with the purified schizonts were depleted using a magnetic column. These synchronous, gametocyte depleted parasite cultures were then monitored from 4 hours post initial intravenous injection into the naïve mice (4hpi) until 44 hours post initial intravenous injection into the naïve mice (44hpi).

CFP expression in four populations (all events, asexual parasites, male gametocytes and female gametocytes) was analysed using flow cytometry to determine onset of CFP expression. CFP intensity was compared in each of the populations in two control lines, a non-fluorescent line (HP) and the parental line for the *cfp* vectors (820) and the three experimental lines. The experimental lines were lines where CFP expression was driven by a gender non-specific promoter (PBANKA_101870), a female specific promoter (PBANKA_143520), or a male specific promoter (PBANKA_142770).

As the *cfp*-vectors were integrated into a line that expresses male and female gametocyte specific proxies (GFP driven by the male specific promoter, PBANKA_041610 and RFP driven by the female specific promoter, PBANKA_131950) gametocytes were identifiable. From 20hpi male specific GFP fluorescence allowed classification of male gametocytes and from 32hpi female specific RFP fluorescence allowed classification of female gametocytes.

To quantify CFP expression in each population the mean CFP intensity was divided by the proportion of the population that was classified as CFP positive.

Figure 7.2 shows the CFP intensity in each of the 4 populations, in each of the *cfp*-promoter lines at key time points analysed.

This analysis showed that the gender non-specific promoter (PBANKA_101870, green/red striped bar Figure 7.2) was detected in male and female gametocytes as soon as they were identified. When male gametocytes were identified at 20hpi they were already expressing CFP. When female gametocytes were identified, at 32hpi, they were also already expressing CFP. However, in the time points leading up to gender specific identification of gametocytes no change in fluorescence intensity was determinable within the asexual population. This means we were

unable to determine the onset of the promoter as anything more specific than before the current proxies were expressed.

The male specific promoter (PBANKA_142770, RuvB1, green bar Figure 7.2) driving CFP expression showed a specific CFP signal in only male gametocytes. However the increased CFP intensity was not detected until 2 – 4 hours after the GFP signal identified the population as male gametocytes. This indicated that RuvB1 expression while restricted to male gametocytes, does not begin until after expression of the dynein heavy chain gene that is used as the current male specific proxy (PBANKA_041610).

The female specific promoter (PBANKA_143520, zinc finger, red bar Figure 7.2) driving CFP expression showed a specific CFP signal in only female gametocytes. As soon as female gametocytes were identified (32hpi) by their RFP fluorescence CFP was also detected. Unfortunately, as with the gender unspecific promoter (PBANKA_101870), no increase in CFP intensity was discernible in the mean fluorescence of the asexual population before identification of the female gametocytes.

While this method of analysis gave us some information about the time of expression of genes based on when 2kb of their promoters drive CFP expression it did not allow the identification of onset as hoped.
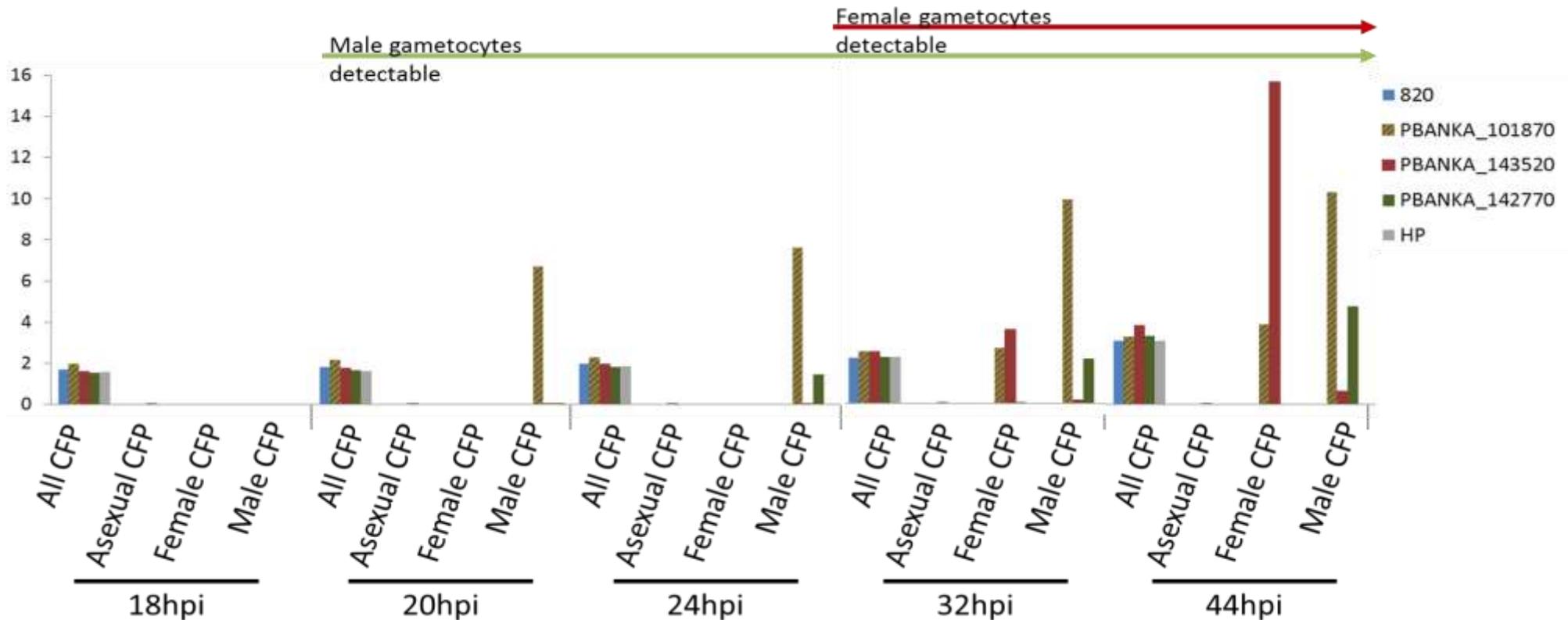
Figure 7.2 Promoter onset and stage and gender specificity defined by cfp detection in different populations

Synchronous ring stage parasites were depleted of mature gametocytes and cultured. At key time points samples were analysed by flow cytometry. Parasitaemia was determined by dye-cycle ruby nuclear staining. Asexual parasites were identified as those with no fluorescence signal. Male gametocytes were identifiable in this assay at 20hpi by their GFP signal and female gametocytes were identifiable by their RFP signal by 32hpi. From this analysis we can clearly identify the PBANKA_101870 promoter (green/red bar) drives CFP expression in male and female gametocytes as soon as they are detectable, indicating it has become active before the promoters driving GFP and RFP expression. The PBANKA_143520 promoter (red bar) does not express in male gametocytes, only females. It is also detectable in the population as soon as female gametocytes are distinguishable from the asexual population. The PBANKA_142770 promoter driving CFP expression is detectable in only male gametocytes. It appears that CFP detection is slightly delayed compared to the current male specific promoter which allows males to be detected at 20hpi with the CFP signal only becoming clear at 24hpi

7.1.2 Promoter onset determination by CFP intensity distribution

As overall analysis of mean CFP expression in the different populations did not allow identification of the onset of the promoter we wanted to analyse in more detail the intensity of CFP expression in the asexual population only. As the male specific promoter (PBANKA_142770, RuvB1) was shown to be active after classification of male gametocytes based on their GFP signal its expression was not analysed further (Figure 7.2).

To identify when and if a subset of the asexual population began expressing CFP the distribution of the intensity of CFP expression was analysed in the asexual population only. When a subset of the population began expressing CFP to a higher degree than the majority of the population a clear, distinct peak became evident (Figure 7.3). This was seen when plotting the intensity of the CFP signal against the number of parasites expressing that intensity.

First we analysed the gametocyte non-sex-specific promoter (PBANKA_101870) as this would result in the largest subset of the parasites classified as asexual having a CFP signal, arising from both immature male and immature female gametocytes. As this promoter seemed to be detectable before our current gender specific markers, we now refer to it as the early gametocyte (EG) promoter.

When analysing CFP intensity in the asexual parasite population a clear distinct more intense CFP population was identified separate from the main population 18hpi (Figure 7.3 A). This population was clear 2 hours before the male gametocyte marker (PBANKA_041610 driving GFP), and 14 hours before the female specific marker (PBANKA_131950 driving RFP) expressed in the parental line were detectable (Figure 7.3 A). We also saw that the population became less distinct at 20hpi which we attributed to the proportion of the asexual population that was then classified as male gametocytes and therefore no longer classified in this asexual population.

Next we analysed the female specific promoter (PBANKA_143520). Here it was clear from analysis of the intensity of the CFP fluorescence in the asexual population that a new distinct more intense CFP population was visible from 20hpi (Figure 7.3 B).

This onset was identified as the same time as the current male specific promoter (PBANKA_041610) which drives GFP expression in the parental line becomes active and 12 hours before the current female specific promoter in the parental line (PBANKA_131950) becomes active.
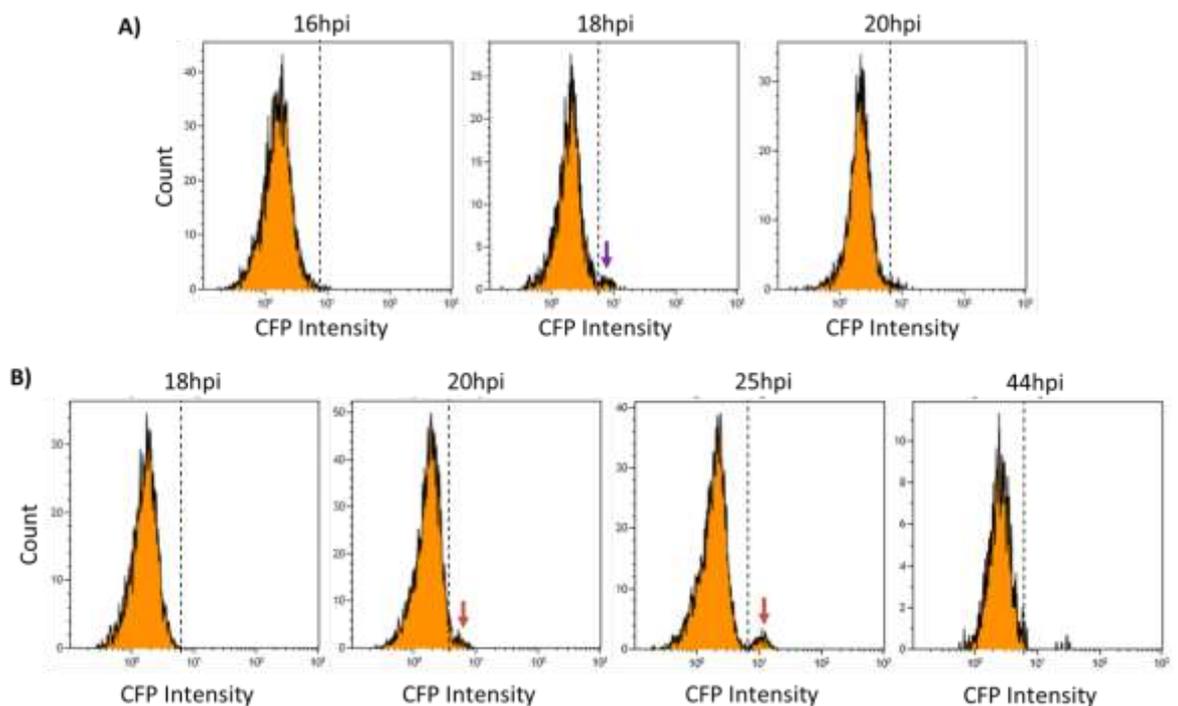
Figure 7.3 Flow cytometry analysis of CFP onset with the early gametocyte promoter and the female specific promoter.

A) CFP expression is driven by the early gametocyte promoter (PBANKA_101870**)**. At 16hpi CFP intensity formed a single expression peak indicating a single population existed within the asexual parasites analysed. At 18hpi a second peak appeared, distinct from the majority of the population. This peak had a clear increased CFP expression profile when compared to the rest of the population, which could indicate that these would develop into gametocytes. At 20hpi when the male gametocytes became identifiable by their GFP expression and therefore no longer appeared in the asexual population, there was a reduction in this separate population. The subset that remain distinct from the main population are likely immature female gametocytes not yet identified based on their RFP expression.

B) PBANKA_143520 expresses CFP driven by the female specific promoter. At 18hpi there was a single CFP expression peak. By 20hpi a secondary peak became visible and was clearly distinct by 25hpi. This peak disappeared from the asexual population by 44hpi when female gametocytes were no longer part of this population because of their RFP expression.

These extensions to the previous promoter assay (Sinha *et al.*, 2014) allowed us to identify the time post invasion that promoters became active. This can then be related to the onset of expression of the gene. This time identification could be determined whether their onset was after (RuvB1, PBANKA_142770 Figure 7.2) the current gametocyte specific proxies or before (early gametocyte promoter, PBANKA_101870 and zinc finger, PBANKA_143520 Figure 7.3) using analysis of mean CFP expression or CFP intensity distribution.

7.2 **Discussion**

While this screening method had previously allowed us to confirm gametocyte specific expression in several hypothetically gametocyte specific genes (Table 7.1) extensions to this assay have now allowed us to determine onset of promoter activity within a subset of the population. This will provide information about gender specific expression and some expression timing data which may be particularly useful for genes that cannot be successfully tagged. However as the actual gene of interest is not tagged the information is limited.

There are several limitations and caveats to this system that may make it less informative. Firstly, the onset of gene expression does not give us any information about its role, just when it is expressed and if its expression is restricted to a specific life cycle stage (for example male or female gametocytes). Secondly, we have used an approximately 2kb region of the 5' upstream region of the gene of interest as the promoter. While this may be sufficient for accurate promoter activity for some genes it is unlikely to be sufficient for others. For example the *ap2-g* gene has a promoter region that has been shown to be negatively affected if disruptions occur up to 5kb upstream of the transcription start site (Sinha *et al.*, 2014). Furthermore CFP is driven by the promoter region and the 3'UTR is always the Pb48/45 3'UTR which might mean we are missing important regulatory roles played by the genes endogenous 3'UTR (Caro *et al.*, 2014).

While this method of analysing gene expression does provide information that identify specific expression patterns it will always be used in combination with other screening methods and will be complemented with many other functional assays to uncover gene function.

Chapter 8   Discussion

This work aimed to develop a novel system that would allow the dissection of the early stages of gametocyte commitment with the overall aim of elucidating players involved in key mechanisms such as commitment and differentiation to gametocytogenesis, gender assignation and early development of male and female gametocytes. While some progress has been made towards realising these goals a large amount of the data that will aid in this identification is not yet available. With the data currently available advancements have been made to identify genes that appear to play a role in gametocyte commitment and maturation of both male and female gametocytes.

Overall we believe that the data generated from the time course of transcript expression after induction of high levels of *ap2-g* will allow us to isolate different stages in the commitment process uncovering key players involved. Though we have identified potential caveats with the population resulting from such high expression of AP2-G the early stages of gametocyte development seem to represent that of wild-type gametocytes (see section 4.5).

The current candidates we have uncovered may offer insight into cascades involved in development of both male and female gametocytes and further investigation of their roles is underway.

In depth analysis of synchronously induced gametocytes has highlighted some interesting characteristics that may impact the way we think about the entire process of commitment. For example the dogma widely held that invading merozoites are pre-committed to becoming gametocytes and that these have already become committed too to a gender.

With the variety of data that has been generated here, and in many other studies it seems opportune to re-evaluate key factors and unite the information into a potential model of gametocyte commitment.

In all systems where proliferation and terminal differentiation occur from a single population the regulation of the two states is controlled in numerous ways often responding to many different stimuli. These appear to prevent terminal

differentiation of the entire population able to proliferate and differentiate and likewise prevent complete loss of the differentiated cells. Mathematical modelling has indicated that this is not simply a stochastic process as this results in the loss of one population over time (Sun & Komarova, 2015).

Current data from *P. falciparum* and *P. berghei* suggests that complete loss of gametocyte commitment occurs only when the population mutates its *ap2-g* gene (Kafsack *et al.*, 2014; Sinha *et al.*, 2014), though mature gametocytes are deficient in several other mutant lines (Eksi *et al.*, 2012). Some research suggests that reductions in the proportion of gametocytes generated by a line is not in fact a reduction in the capability of gametocyte production but the out-growth of the non-producers within a population of parasites that have and have not mutated their *ap2-g* (Janse et al., 1992). Reports have suggested that before *ap2-g* expression proteins important for gametocyte differentiation and development are expressed without detriment to remaining in the asexual proliferative cycle (Tiburcio *et al.*, 2015). Meanwhile many stimuli have been shown to increase gametocyte commitment (Carter *et al.*, 2013). Additionally to this epigenetic regulation of *ap2-g* expression has been intrinsically linked with *var* gene expression in *P. falciparum*. In the natural state the *ap2-g* locus is tightly bound in a heterochromatic state demarked by histone 3 lysine 9 trimethylation (H3K9me3). When bound to heterochromatin protein 1 (HP1) this heterochromatic state is maintained and expression is silenced due to inaccessibility of the promoter to RNA polymerase II which catalyses transcription of mRNA. In *P. falciparum* the conditional downregulation of HP1 results in the opening of the promoter regions of all *var* genes and *ap2-g* to a euchromatic state where RNA polymerase II can access the promoter and mRNA transcription can occur. This resulted in expression of all *var* genes that are usually expressed in a mutually exclusive manner (single gene choice) and an increase in gametocytes in the population (approximately 52% compared to approximately 2% in the HP1 expressing line (Brancucci *et al.*, 2014)). A second study looked at the histone deacetylase *Pf*Hda2 which is considered a global silencer of *var* gene expression. While dysregulation of this protein by destabilisation resulted in a loss of mutually exclusive *var* gene expression (almost 86% of all *vars* were present) it only increased gametocyte prevelance around 3-fold (HP1 dysregulation results in approximately 25-fold increase).

While this could suggest that the mechanism for *var* switching and *ap2-g* expression are not linked it could also indicate that the role of HP1, maintaining the silencing marks surrounding *ap2-g,* is more fundamental to preventing excessive differentiation to gametocyte commitment than HDA2 deacetylation to allow methylation that will silence an active gene (Coleman *et al.*, 2014)

I propose that integrating these factors gives rise to two independent methods for gametocyte commitment (Figure 8.2). Firstly, those where onset of the *ap2-g* transcription factor is controlled by the same or closely related mechanisms as *var* gene switching. In this case the stochastic switching from expressing one *var* gene to another would result in the expression of *ap2-g*. This switch on of *ap2-g* expression could then result in the development of pre-committed schizonts that generate merozoites that will all become gametocytes and gametocytes of a specific gender (Figure 8.1). This theory may be supported by the way *var* genes are expressed in merozoites released from liver schizonts. It has been shown that the variety in *var* gene expression within a population when determined immediately after release from the liver is very different to populations in chronic infections with a large range if not all *var* genes present in the population (Wang *et al.*, 2009). Furthermore the minimal generation of gametocytes unless a single gene is mutated (*ap2-g*) might demonstrate why below a certain level intermediate gametocyte lines, where gametocyte commitment consistently reaches 1 – 5% are not observed. Previous identifications of observations of intermediate gametocyte producers have been attributed to mixed genotype lines with a proportion of GNP and a proportion of gametocyte producers (Janse *et al.*, 1989b; Janse *et al.*, 1992). Support is also offered for this in the form of pre-committed and single gender plaque formation from single schizont rupture and reinvasion (Bruce *et al.*, 1990). It could also explain some generational variability in gametocyte commitment rates and differences between strains and species. It appears that modelling of switching rates and *in vivo* and *in vitro* experimental data suggest that switching varies between *var* genotypes (Cortes *et al.*, 2012; Duffy *et al.*, 2005; Paget-McNicol *et al.*, 2002).

If we consider the mutually exclusive expression of *clag3* in *P. falciparum,* whose single expression can be temporarily overridden by selective pressure there may be biologically relevant mechanisms existing that will increase switching or potentially force the expression of two genes considered mutually exclusively expressed simultaneously (Rovira-Graells *et al.*, 2015)
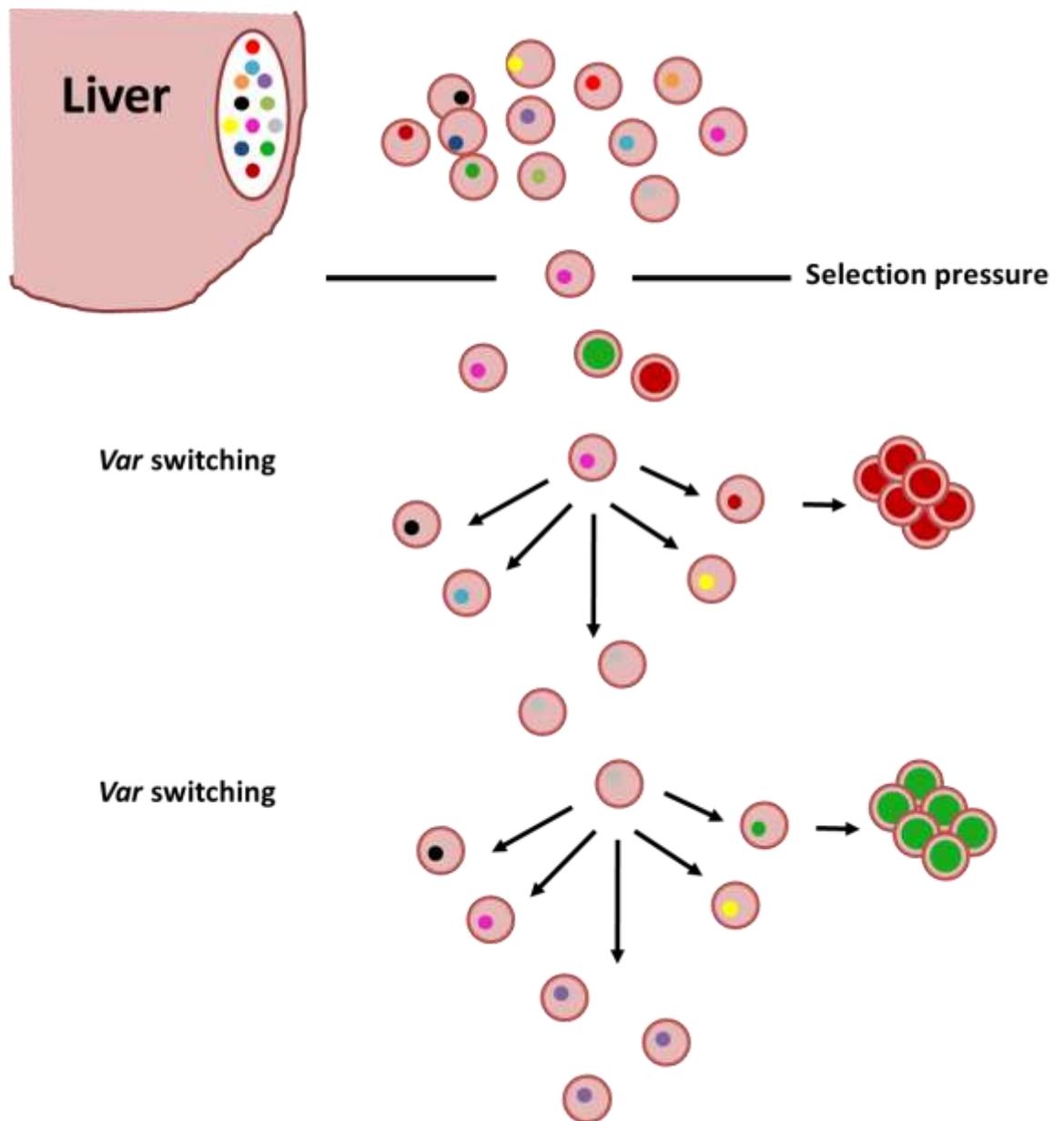
Figure 8.1 *var* gene controlled gametocytes

It has previously been shown that infection from the liver releases a range of *var* gene expressing parasites not normally seen in chronic infections where the predominant population is a single *var* type and stochastic switching results in a change in the population. Here I propose that one *var* option is infact the expression of *ap2-g* and that schizonts developing with this expression generate merozoites that all terminally differentiate into gametocytes of one gender. The release of gametocytes directly from liver schizonts may provide support for this theory.

While evidence seems to exist supporting this basal pre-commitment to the sexual pathway in *P. falciparum* the *var* gene family is not conserved across species and is not found in *P. berghei*. However, similar mechanisms of antigenic variation, with different multigene families showing tightly regulated expression, have been identified in *P. vivax (vir), P. chabaudi (cir), P. yeolii (yir)* and *P. berghei (bir)* which could indicate a conserved function with different proteins but a conserved mechanism (Deitsch & Hviid, 2004; del Portillo *et al.*, 2001; Janssen *et al.*, 2002)

These have not been as extensively studied as the *var* family in *P. falciparum* however the mechanism that regulates antigenic variation could be conserved and acting upon different virulent multigene families that have diversified as species have evolved. If this mechanism of clonally variant gene expression was maintained, with similar parameters (such as pattern switching between members and continual switching per cycle) then the basal level of commitment would be conserved across species.

The second mechanism by which gametocytes may arise does not rely upon pre-commitment of invading merozoites but instead uses environmental sensing to determine the proportion of intracellular parasites that will differentiate (Figure 8.2). Once invasion has occurred signals are interpreted by the parasites and can result in a higher proportion of the population committing to gametocytogenesis. However, as terminal differentiation of the entire population has not been seen, even in autopsy samples, it seems likely that there is a defence mechanism against this outcome. I suggest that there are two populations of asexual stage parasites, those that are receptive or permissive to gametocyte commitment and those that are not. Those that are permissive may respond to factors such as host health (related to reticulocytaemia) or parasite density to differentiate into gametocytes or remain in the asexual proliferative cycle. These permissive cells may begin expression of gametocyte specific factors in readiness for becoming gametocytes but if they are not signalled to do so (by *ap2-g* expression) they can revert or remain in the asexual cycle (Tiburcio *et al.*, 2015).

Previous work suggests *P. berghei* commitment is not fixed until the trophozoite stage (same cycle (Mons, 1986b; Mons, 1986a)). The data indicating that 12hpi is the point at which overexpression of *ap2-g* does not give rise to a high proportion of gametocytes in the same cycle might corroborate this hypothesis.

Extensive work has been undertaken in recent years to demonstrate and characterise environmental sensing by *P. falciparum* in culture conditions and *in vivo*. Key has been the identification of extracellular vesicles released both by infected red blood cells and host cells. A range of these vesicles exist with different makeup and cargo that influence the behaviour of the parasite population as well as the immune response of the infected host. The ability of the infected red blood cell to not only release vesicles but uptake them as well was key to demonstrating sensing could occur at all (Mantel *et al.*, 2013). Subsequent identification of potential members of a sensing pathway have supported the response of *P. falciparum* to a multitude of environmental stimulus and communications from other *P. falciparum* infected red blood cells (Wu *et al.*, 2016). It has been demonstrated that these vesicles can transfer not DNA, miRNA and protein between cells as a means of stimulation and communications (LaMonte *et al.*, 2012; Regev-Rudzki *et al.*, 2013). A plethora of vesicle derived proteins have been identified and several have been characterised for their inductive phenotype. Although immunomodulatory functions for these proteins have been identified of import to this work is the inductive effect of these vesicles to gametocyte commitment (Mantel *et al.*, 2013; Mantel & Marti, 2014).

This dual commitment potential idea could also be supported by the variability in gametocyte commitment with different environmental stimulus (discussed in section 4.5.2.2). This secondary reservoir for potential gametocytes could explain some questions that have yet to be answered about gametocyte commitment. For example the indistinct expression profile for *ap2-g* that does not give an indication as to when its expression results in commitment, if it was expressed variably dependant on the conditions inducing gametocyte commitment it would not always be expressed at a specific stage. Secondly, the data presented here (that has been similarly shown in *P. falciparum* (Oriol Llora-Batlle, 2016)) showing that gametocytes, when induced to express high levels of *ap2-g,* can differentiate into gametocytes without reinvasion.
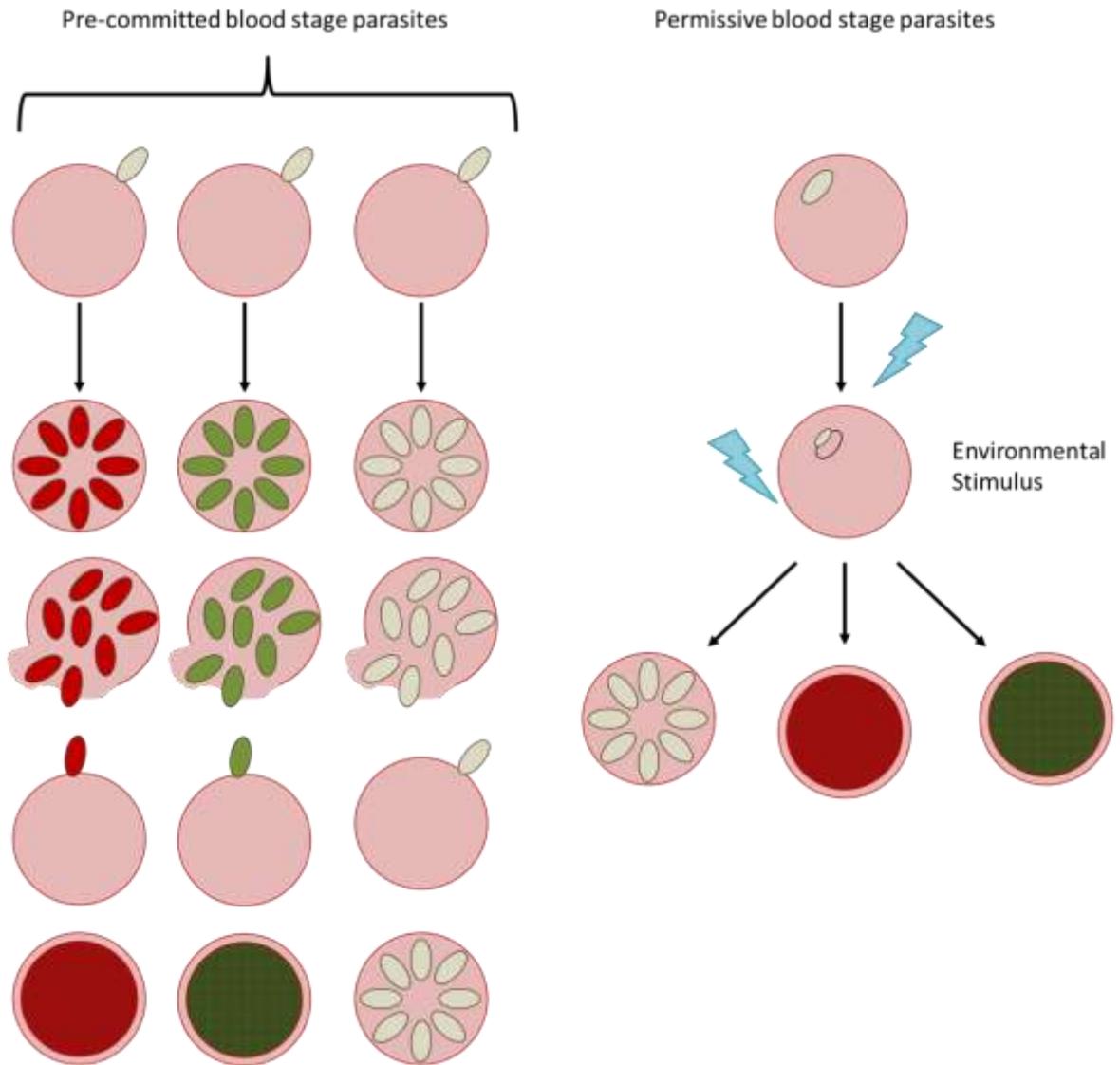
Figure 8.2 Proposed model of gametocyte commitment
In this proposed model there are two different ways for gametocytes to commit. Firstly, an invading merozoites generates a schizont that is either pre-committed to a single gender or to remaining in the asexual proliferative cycle. I propose that this mechanism is stochastically regulated in the same was as *var* gene switching. Secondly, intracellular parasites respond to environmental stimulus to induce gametocyte commitment without reinvasion. A proportion of the intracellular parasites will also be non-permissive to environmental stimuli and therefore will not differentiate into gametocytes but ensure proliferation in the current host.

If we consider a natural infection it is necessary for the parasites to maintain an infection in the current host, avoiding the immune system to ensure population survival and to moderate the population as to not kill the host and ensure successful transmission of the differentiated sexual stages. This all takes place in a host with a functioning immune system and variability of multiple factors over time. With such a long maturation time before *P. falciparum* gametocytes can be transmitted a dynamic response to environmental perturbations would be beneficial *in vivo*. Observational studies have shown that identification and quantification of gametocyte burden can be inaccurate over time (Akim *et al.*, 2000). It is feasible that the response of the infection to drug treatment and environmental stimulus is dynamic and gametocyte burden alters rapidly and that sequestration of different stages of maturing gametocytes reduces the sensitivity of detection methods. The multi-layered model of gametocyte commitment proposed would facilitate this rapid response while preventing a complete commitment to either asexual proliferation or sexual differentiation when environmental stimulus are extreme.

Many additional experiments would be required to establish the validity of the model and to identify the controlling factors. It would be interesting to establish if commitment rates vary over the course of an infection *in vivo* with differing environmental cues. I believe this would be highly variable in regions where transmission is not possible continually.

What is clear is that the process of gametocyte commitment is highly complex and likely involves many different layers of regulation to ensure the proliferation of parasites within the infected host and to facilitate transmission to new hosts. The factors influencing the balance between proliferation and terminal differentiation are poorly understood in eukaryotic systems but have been shown to be influenced by many factors and that the responses are dynamic and controlled to ensure survival. This work has uncovered some novel candidates involved in gametocyte development and holds the potential to elucidate further candidates early in the process that has been so difficult to study.

Chapter 9   References

(2001). PlasmoDB: An integrative database of the Plasmodium falciparum genome. Tools for accessing and analyzing finished and unfinished sequence data. The Plasmodium Genome Database Collaborative. *Nucleic Acids Res, 29*, 66-69.

Abdelhaleem, M., Maltais, L. and Wain, H. (2003). The human DDX and DHX gene families of putative RNA helicases. *Genomics, 81*, 618-622.

Absalon, S., Robbins, J. A. and Dvorin, J. D. (2016). An essential malaria protein defines the architecture of blood-stage and transmission-stage parasites. *Nat Commun, 7*. doi: 10.1038/ncomms11449.

Akim, N. I., Drakeley, C., Kingo, T., Simon, B., Senkoro, K. and Sauerwein, R. W. (2000). Dynamics of P. falciparum gametocytemia in symptomatic patients in an area of intense perennial transmission in Tanzania. *Am J Trop Med Hyg, 63*, 199-203.

Albert, H., Dale, E. C., Lee, E. and Ow, D. W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J, 7*, 649-659.

Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F. and Menard, R. (2006). Quantitative imaging of Plasmodium transmission from mosquito to mammal. *Nat Med, 12*, 220-224. doi: 10.1038/nm1350.

Andenmatten, N., Egarter, S., Jackson, A. J., Jullien, N., Herman, J. P. and Meissner, M. (2013). Conditional genome engineering in Toxoplasma gondii uncovers alternative invasion mechanisms. *Nat Methods, 10*, 125-127. doi: 10.1038/nmeth.2301.

Armstrong, C. M. and Goldberg, D. E. (2007). An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. *Nat Methods, 4*, 1007-1009. doi: 10.1038/nmeth1132.

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Jr., Treatman, C. and Wang, H. (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res, 37*, D539-543. doi: 10.1093/nar/gkn814.

Autino, B., Noris, A., Russo, R. and Castelli, F. (2012). Epidemiology of malaria in endemic areas. *Mediterr J Hematol Infect Dis, 4*, e2012060. doi: 10.4084/MJHID.2012.060.

Bahl, A., Brunk, B., Coppel, R. L., Crabtree, J., Diskin, S. J., Fraunholz, M. J., Grant, G. R., Gupta, D., Huestis, R. L., Kissinger, J. C., Labo, P., Li, L., McWeeney, S. K., Milgram, A. J., Roos, D. S., Schug, J. and Stoeckert, C. J., Jr. (2002). PlasmoDB: the Plasmodium genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished). *Nucleic Acids Res, 30*, 87-90.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., Li, L., Mailman, M. D., Milgram, A. J., Pearson, D. S., Roos, D. S., Schug, J., Stoeckert, C. J., Jr. and Whetzel, P. (2003). PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Res, 31*, 212-215.

Bakar, N. A., Klonis, N., Hanssen, E., Chan, C. and Tilley, L. (2010). Digestive-vacuole genesis and endocytic processes in the early intraerythrocytic stages of Plasmodium falciparum. *J Cell Sci,* **123**, 441-450.

Baker, D. A., Thompson, J., Daramola, O. O., Carlton, J. M. and Targett, G. A. (1995). Sexual-stage-specific RNA expression of a new Plasmodium falciparum gene detected by in situ hybridisation. *Mol Biochem Parasitol,* **72**, 193-201.

Balaji, S., Babu, M. M., Iyer, L. M. and Aravind, L. (2005). Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Research,* **33**, 3994-4006. doi: 10.1093/nar/gki709.

Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G. and Wandless, T. J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell,* **126**, 995-1004. doi: 10.1016/j.cell.2006.07.025.

Banroques, J., Doère, M., Dreyfus, M., Linder, P. and Tanner, N. K. (2010). Motif III in Superfamily 2 "Helicases" Helps Convert the Binding Energy of ATP into a High-Affinity RNA Binding Site in the Yeast DEAD-Box Protein Ded1. *Journal of Molecular Biology,* **396**, 949-966. doi: http://dx.doi.org/10.1016/j.jmb.2009.12.025.

Baum, J., Gilberger, T., Frischknecht, F. and Meissner, M. (2008). Host-cell invasion by malaria parasites: insights from Plasmodium and Toxoplasma. *Trends Parasitol,* **24**. doi: 10.1016/j.pt.2008.08.006.

Baum, J., Papenfuss, A. T., Mair, G. R., Janse, C. J., Vlachou, D., Waters, A. P., Cowman, A. F., Crabb, B. S. and de Koning-Ward, T. F. (2009). Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res,* **37**, 3788-3798. doi: 10.1093/nar/gkp239.

Belmont, B. J. and Niles, J. C. (2010). Engineering a direct and inducible protein-RNA interaction to regulate RNA biology. *ACS Chem Biol,* **5**, 851-861. doi: 10.1021/cb100070j.

Belshaw, P. J., Ho, S. N., Crabtree, G. R. and Schreiber, S. L. (1996). Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins. *Proc Natl Acad Sci U S A,* **93**, 4604-4607.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briet, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., Lynch, M., Lindgren, F., Cohen, J. M., Murray, C. L. J., Smith, D. L., Hay, S. I., Cibulskis, R. E. and Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature,* **526**, 207-211. doi: 10.1038/nature15535 http://www.nature.com/nature/journal/v526/n7572/abs/nature15535.html#supplementary-information.

Billker, O., Lindo, V., Panico, M., Etienne, A., Paxton, T., Dell, A., Rogers, M., Sinden, R. and Morris, H. (1998). Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature,* **392**, 289-292.

Billker, O., Miller, A. J. and Sinden, R. E. (2000). Determination of mosquito bloodmeal pH in situ by ion-selective microelectrode measurement: implications for the regulation of malarial gametogenesis. *Parasitology,* **120 (Pt 6)**, 547-551.

Billker, O., Shaw, M., Margos, G. and Sinden, R. (1997). The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of Plasmodium berghei in vitro. *Parasitology,* **115**, 1-7.

Bioinformatics, B. (2011). FastQC A quality control tool for high throughput sequence data. *Cambridge, UK: Babraham Institute.*

Blackman, M. J. and Carruthers, V. B. (2013). Recent insights into apicomplexan parasite egress provide new views to a kill. *Current opinion in microbiology,* **16**, 459-464.

Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J. and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol,* **1**, e5.

Braks, J. A., Franke-Fayard, B., Kroeze, H., Janse, C. J. and Waters, A. P. (2006). Development and application of a positive-negative selectable marker system for use in reverse genetics in Plasmodium. *Nucleic Acids Res,* **34**, e39. doi: 10.1093/nar/gnj033.

Brancucci, N. M., Bertschi, N. L., Zhu, L., Niederwieser, I., Chin, W. H., Wampfler, R., Freymond, C., Rottmann, M., Felger, I., Bozdech, Z. and Voss, T. S. (2014). Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe,* **16**, 165-176. doi: 10.1016/j.chom.2014.07.004.

Bruce, M. C., Alano, P., Duthie, S. and Carter, R. (1990). Commitment of the malaria parasite Plasmodium falciparum to sexual and asexual development. *Parasitology,* **100 Pt 2**, 191-200.

Buchholz, F., Angrand, P. O. and Stewart, A. F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol,* **16**, 657-662. doi: 10.1038/nbt0798-657.

Buckling, A. and Read, A. F. (2001). The effect of partial host immunity on the transmission of malaria parasites. *Proceedings of the Royal Society of London B: Biological Sciences,* **268**, 2325-2330.

Campbell, T. L., De Silva, E. K., Olszewski, K. L., Elemento, O. and Llinas, M. (2010). Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog,* **6**, e1001165. doi: 10.1371/journal.ppat.1001165.

Caro, F., Ahyong, V., Betegon, M. and DeRisi, J. L. (2014). Genome-wide regulatory dynamics of translation in the Plasmodium falciparum asexual blood stages. *eLife,* **3**, e04106. doi: 10.7554/eLife.04106.

Carter, L. M., Kafsack, B. F., Llinas, M., Mideo, N., Pollitt, L. C. and Reece, S. E. (2013). Stress and sex in malaria parasites: Why does commitment vary? *Evol Med Public Health,* **2013**, 135-147. doi: 10.1093/emph/eot011.

Carter, R. and Miller, L. H. (1979). Recent developments in production and purification of malaria antigens: Evidence for environmental modulation of gametocytogenesis in Plasmodium falciparum in continuous culture. *Bulletin of the World Health Organization,* **57**, 37-52.

Carvalho, T. G., Thiberge, S., Sakamoto, H. and Menard, R. (2004). Conditional mutagenesis using site-specific recombination in Plasmodium berghei. *Proc Natl Acad Sci U S A,* **101**, 14931-14936. doi: 10.1073/pnas.0404416101.

Chomczynski, P. (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques,* **15**, 532-534, 536-537.

Coleman, B. I., Skillman, K. M., Jiang, R. H., Childs, L. M., Altenhofen, L. M., Ganter, M., Leung, Y., Goldowitz, I., Kafsack, B. F., Marti, M., Llinas, M., Buckee, C. O. and Duraisingh, M. T. (2014). A Plasmodium

falciparum histone deacetylase regulates antigenic variation and gametocyte conversion. *Cell Host Microbe,* **16**, 177-186. doi: 10.1016/j.chom.2014.06.014.

Collins, C. R., Das, S., Wong, E. H., Andenmatten, N., Stallmach, R., Hackett, F., Herman, J. P., Muller, S., Meissner, M. and Blackman, M. J. (2013). Robust inducible Cre recombinase activity in the human malaria parasite Plasmodium falciparum enables efficient gene deletion within a single asexual erythrocytic growth cycle. *Mol Microbiol,* **88**, 687-701. doi: 10.1111/mmi.12206.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X. and Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology,* **17**, 1-19. doi: 10.1186/s13059-016-0881-8.

Cortes, A., Crowley, V. M., Vaquero, A. and Voss, T. S. (2012). A view on the role of epigenetics in the biology of malaria parasites. *PLoS Pathog,* **8**, e1002943. doi: 10.1371/journal.ppat.1002943.

Das, S., Hertrich, N., Perrin, Abigail J., Withers-Martinez, C., Collins, Christine R., Jones, Matthew L., Watermeyer, Jean M., Fobes, Elmar T., Martin, Stephen R., Saibil, Helen R., Wright, Gavin J., Treeck, M., Epp, C. and Blackman, Michael J. (2015). Processing of Plasmodium falciparum Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs. *Cell host & microbe,* **18**, 433-444. doi: http://dx.doi.org/10.1016/j.chom.2015.09.007.

Day, K., Karamalis, F., Thompson, J., Barnes, D., Peterson, C., Brown, H., Brown, G. and Kemp, D. (1993). Genes necessary for expression of a virulence determinant and for transmission of Plasmodium falciparum are located on a 0.3-megabase region of chromosome 9. *Proceedings of the National Academy of Sciences,* **90**, 8292-8296.

de Koning-Ward, T. F., Fidock, D. A., Thathy, V., Menard, R., van Spaendonk, R. M., Waters, A. P. and Janse, C. J. (2000). The selectable marker human dihydrofolate reductase enables sequential genetic manipulation of the Plasmodium berghei genome. *Mol Biochem Parasitol,* **106**, 199-212.

De Silva, E. K., Gehrke, A. R., Olszewski, K., León, I., Chahal, J. S., Bulyk, M. L. and Llinás, M. (2008). Specific DNA-binding by Apicomplexan AP2 transcription factors. *Proceedings of the National Academy of Sciences,* **105**, 8393-8398. doi: 10.1073/pnas.0801993105.

Deitsch, K. W. and Hviid, L. (2004). Variant surface antigens, virulence genes and the pathogenesis of malaria. *Trends Parasitol,* **20**, 562-566. doi: 10.1016/j.pt.2004.09.002.

del Portillo, H. A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C. P., Schneider, N. K., Villalobos, J. M., Rajandream, M. A., Harris, D., Pereira da Silva, L. H., Barrell, B. and Lanzer, M. (2001). A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax. *Nature,* **410**, 839-842. doi: 10.1038/35071118.

Dixon, M. W., Thompson, J., Gardiner, D. L. and Trenholme, K. R. (2008). Sex in Plasmodium: a sign of commitment. *Trends Parasitol,* **24**, 168-175. doi: 10.1016/j.pt.2008.01.004.

Duffy, M. F., Byrne, T. J., Elliott, S. R., Wilson, D. W., Rogerson, S. J., Beeson, J. G., Noviyanti, R. and Brown, G. V. (2005). Broad analysis reveals a consistent pattern of var gene transcription in Plasmodium falciparum repeatedly selected for a defined adhesion phenotype. *Mol Microbiol,* **56**, 774-788. doi: 10.1111/j.1365-2958.2005.04577.x.

Egarter, S., Andenmatten, N., Jackson, A. J., Whitelaw, J. A., Pall, G., Black, J. A., Ferguson, D. J. P., Tardieux, I., Mogilner, A. and Meissner, M. (2014). The Toxoplasma Acto-MyoA Motor Complex Is Important but Not Essential for Gliding Motility and Host Cell Invasion. *PLoS One,* **9**, e91819. doi: 10.1371/journal.pone.0091819.

Eksi, S., Morahan, B. J., Haile, Y., Furuya, T., Jiang, H., Ali, O., Xu, H., Kiattibutr, K., Suri, A., Czesny, B., Adeyemo, A., Myers, T. G., Sattabongkot, J., Su, X.-z. and Williamson, K. C. (2012). *Plasmodium falciparum* Gametocyte Development 1 (*Pfgdv1*) and Gametocytogenesis Early Gene Identification and Commitment to Sexual Development. *PLoS Pathog,* **8**, e1002964. doi: 10.1371/journal.ppat.1002964.

Flueck, C., Bartfai, R., Niederwieser, I., Witmer, K., Alako, B. T., Moes, S., Bozdech, Z., Jenoe, P., Stunnenberg, H. G. and Voss, T. S. (2010). A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog,* **6**, e1000784. doi: 10.1371/journal.ppat.1000784.

Foth, B. J., Stimmler, L. M., Handman, E., Crabb, B. S., Hodder, A. N. and McFadden, G. I. (2005). The malaria parasite Plasmodium falciparum has only one pyruvate dehydrogenase complex, which is located in the apicoplast. *Molecular microbiology,* **55**, 39-53.

Francia, M. E. and Striepen, B. (2014). Cell division in apicomplexan parasites. *Nat Rev Micro,* **12**, 125-136. doi: 10.1038/nrmicro3184.

Franke-Fayard, B., Janse, C. J., Cunha-Rodrigues, M., Ramesar, J., Büscher, P., Que, I., Löwik, C., Voshol, P. J., den Boer, M. A. M., van Duinen, S. G., Febbraio, M., Mota, M. M. and Waters, A. P. (2005). Murine malaria parasite sequestration: CD36 is the major receptor, but cerebral pathology is unlinked to sequestration. *Proceedings of the National Academy of Sciences of the United States of America,* **102**, 11468-11473. doi: 10.1073/pnas.0503386102.

Franke-Fayard, B., Trueman, H., Ramesar, J., Mendoza, J., van der Keur, M., van der Linden, R., Sinden, R. E., Waters, A. P. and Janse, C. J. (2004). A Plasmodium berghei reference line that constitutively expresses GFP at a high level throughout the complete life cycle. *Mol Biochem Parasitol,* **137**, 23-33. doi: 10.1016/j.molbiopara.2004.04.007.

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

Gilson, P. R. and Crabb, B. S. (2009). Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites. *International Journal for Parasitology,* **39**, 91-96.

Goldfless, S. J., Wagner, J. C. and Niles, J. C. (2014). Versatile control of Plasmodium falciparum gene expression with an inducible protein-RNA interaction. *Nat Commun,* **5**, 5329. doi: 10.1038/ncomms6329.

Gomes, Ana R., Bushell, E., Schwach, F., Girling, G., Anar, B., Quail, Michael A., Herd, C., Pfander, C., Modrzynska, K., Rayner, Julian C. and Billker, O. (2015). A Genome-Scale Vector Resource Enables High-Throughput Reverse Genetic Screening in a Malaria Parasite. *Cell host & microbe,* **17**, 404-413. doi: http://dx.doi.org/10.1016/j.chom.2015.01.014.

Gossen, M. and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A,* **89**, 5547-5551.

Hall, N., Karras, M., Raine, J. D., Carlton, J. M., Kooij, T. W., Berriman, M., Florens, L., Janssen, C. S., Pain, A., Christophides, G. K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M. A., Ormond, D., Doggett, J., Trueman, H. E., Mendoza, J., Bidwell, S. L., Rajandream, M. A., Carucci, D. J., Yates, J. R., 3rd, Kafatos, F. C., Janse, C. J., Barrell, B., Turner, C. M., Waters, A. P. and Sinden, R. E. (2005). A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science,* **307**, 82-86. doi: 10.1126/science.1103717.

Harding, C. R., Egarter, S., Gow, M., Jimenez-Ruiz, E., Ferguson, D. J. and Meissner, M. (2016). Gliding Associated Proteins Play Essential Roles during the Formation of the Inner Membrane Complex of Toxoplasma gondii. *PLoS Pathog,* **12**, e1005403. doi: 10.1371/journal.ppat.1005403.

Hebenstreit, D., Fang, M., Gu, M., Charoensawan, V., van Oudenaarden, A. and Teichmann, S. A. (2011). RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol Syst Biol,* **7**, 497. doi: 10.1038/msb.2011.28.

Herm-Gotz, A., Agop-Nersesian, C., Munter, S., Grimley, J. S., Wandless, T. J., Frischknecht, F. and Meissner, M. (2007). Rapid control of protein level in the apicomplexan Toxoplasma gondii. *Nat Methods,* **4**, 1003-1005. doi: 10.1038/nmeth1134.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature,* **329**, 219-222. doi: 10.1038/329219a0.

Hillyer, J. F., Barreau, C. and Vernick, K. D. (2007). Efficiency of salivary gland invasion by malaria sporozoites is controlled by rapid sporozoite destruction in the mosquito hemocoel. *International Journal for Parasitology,* **37**, 673-681. doi: 10.1016/j.ijpara.2006.12.00.

Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D. and Wandless, T. J. (2010). A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem Biol,* **17**, 981-988. doi: 10.1016/j.chembiol.2010.07.009.

Iwanaga, S., Kaneko, I., Kato, T. and Yuda, M. (2012). Identification of an AP2-family protein that is critical for malaria liver stage development. *PLoS One,* **7**, e47557. doi: 10.1371/journal.pone.0047557.

Iwanaga, S., Khan, S. M., Kaneko, I., Christodoulou, Z., Newbold, C., Yuda, M., Janse, C. J. and Waters, A. P. (2010). Functional identification of the Plasmodium centromere and generation of a Plasmodium artificial chromosome. *Cell Host Microbe,* **7**, 245-255. doi: 10.1016/j.chom.2010.02.010.

Iyer, L. M., Anantharaman, V., Wolf, M. Y. and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol,* **38**, 1-31. doi: 10.1016/j.ijpara.2007.07.018.

Janse, C., Van der Klooster, P., Van der Kaay, H., Van der Ploeg, M. and Overdulve, J. (1986). Rapid repeated DNA replication during microgametogenesis and DNA synthesis in young zygotes of Plasmodium berghei. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **80**, 154-157.

Janse, C. J., Boorsma, E. G., Ramesar, J., Grobbee, M. J. and Mons, B. (1989a). Host cell specificity and schizogony of Plasmodium berghei under

different in vitro conditions. *International Journal for Parasitology,* **19**, 509-514. doi: http://dx.doi.org/10.1016/0020-7519(89)90080-5.

Janse, C. J., Boorsma, E. G., Ramesar, J., van Vianen, P., van der Meer, R., Zenobi, P., Casaglia, O., Mons, B. and van der Berg, F. M. (1989b). Plasmodium berghei: gametocyte production, DNA content, and chromosome-size polymorphisms during asexual multiplication in vivo. *Exp Parasitol,* **68**, 274-282.

Janse, C. J., Franke-Fayard, B., Mair, G. R., Ramesar, J., Thiel, C., Engelmann, S., Matuschewski, K., van Gemert, G. J., Sauerwein, R. W. and Waters, A. P. (2006). High efficiency transfection of Plasmodium berghei facilitates novel selection procedures. *Mol Biochem Parasitol,* **145**, 60-70. doi: 10.1016/j.molbiopara.2005.09.007.

Janse, C. J., Kroeze, H., van Wigcheren, A., Mededovic, S., Fonager, J., Franke-Fayard, B., Waters, A. P. and Khan, S. M. (2011). A genotype and phenotype database of genetically modified malaria-parasites. *Trends Parasitol,* **27**, 31-39. doi: 10.1016/j.pt.2010.06.016.

Janse, C. J., Ramesar, J., van den Berg, F. M. and Mons, B. (1992). Plasmodium berghei: in vivo generation and selection of karyotype mutants and non-gametocyte producer mutants. *Exp Parasitol,* **74**, 1-10.

Janssen, C. S., Barrett, M. P., Turner, C. M. and Phillips, R. S. (2002). A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proc Biol Sci,* **269**, 431-436. doi: 10.1098/rspb.2001.1903.

Josling, G. A. and Llinas, M. (2015). Sexual development in Plasmodium parasites: knowing when it's time to commit. *Nat Rev Microbiol,* **13**, 573-587. doi: 10.1038/nrmicro3519.

Jullien, N., Sampieri, F., Enjalbert, A. and Herman, J. P. (2003). Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res,* **31**, e131.

Kafsack, B. F., Rovira-Graells, N., Clark, T. G., Bancells, C., Crowley, V. M., Campino, S. G., Williams, A. E., Drought, L. G., Kwiatkowski, D. P., Baker, D. A., Cortes, A. and Llinas, M. (2014). A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature,* **507**, 248-252. doi: 10.1038/nature12920.

Kahle, D. and Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *R Journal,* **5**.

Kaneko, I., Iwanaga, S., Kato, T., Kobayashi, I. and Yuda, M. (2015). Genome-Wide Identification of the Target Genes of AP2-O, a Plasmodium AP2-Family Transcription Factor. *PLoS Pathog,* **11**, e1004905. doi: 10.1371/journal.ppat.1004905.

Khan, S. M., Franke-Fayard, B., Mair, G. R., Lasonder, E., Janse, C. J., Mann, M. and Waters, A. P. (2005). Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. *Cell,* **121**, 675-687. doi: 10.1016/j.cell.2005.03.027.

Khan, S. M., Kroeze, H., Franke-Fayard, B. and Janse, C. J. (2013). Standardization in generating and reporting genetically modified rodent malaria parasites: the RMgmDB database. *Methods Mol Biol,* **923**, 139-150. doi: 10.1007/978-1-62703-026-7_9.

Kim, D., Langmead, B. and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Meth,* **12**, 357-360. doi: 10.1038/nmeth.3317

http://www.nature.com/nmeth/journal/v12/n4/abs/nmeth.3317.html#supplementary-information.

Kitsera, N., Khobta, A. and Epe, B. (2007). Destabilized green fluorescent protein detects rapid removal of transcription blocks after genotoxic exposure. *Biotechniques,* **43**, 222-227.

Kobayashi., M. N. S. I. I. I. K. H. A. M. Y. S. K. F. (2016). Dual disruption of nucleoside transporter 1 and purine nucleoside phosphorylase leads to attenuation of the growth and virulence of Plasmodium berghei.

Kolb, A. F. (2002). Genome engineering using site-specific recombinases. *Cloning Stem Cells,* **4**, 65-80. doi: 10.1089/153623002753632066.

Kozarewa, I., Ning, Z., Quail, M. A., Sanders, M. J., Berriman, M. and Turner, D. J. (2009). Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods,* **6**, 291-295. doi: 10.1038/nmeth.1311.

Krueger, F. (2015). Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

Lacroix, C., Giovannini, D., Combe, A., Bargieri, D. Y., Spath, S., Panchal, D., Tawk, L., Thiberge, S., Carvalho, T. G., Barale, J. C., Bhanot, P. and Menard, R. (2011). FLP/FRT-mediated conditional mutagenesis in pre-erythrocytic stages of Plasmodium berghei. *Nat Protoc,* **6**, 1412-1428. doi: 10.1038/nprot.2011.363.

LaMonte, G., Philip, N., Reardon, J., Lacsina, J. R., Majoros, W., Chapman, L., Thornburg, C. D., Telen, M. J., Ohler, U., Nicchitta, C. V., Haystead, T. and Chi, J. T. (2012). Translocation of sickle cell erythrocyte microRNAs into Plasmodium falciparum inhibits parasite translation and contributes to malaria resistance. *Cell Host Microbe,* **12**, 187-199. doi: 10.1016/j.chom.2012.06.007.

Langhorne, J., Buffet, P., Galinski, M., Good, M., Harty, J., Leroy, D., Mota, M. M., Pasini, E., Renia, L., Riley, E., Stins, M. and Duffy, P. (2011). The relevance of non-human primate and rodent malaria models for humans. *Malaria Journal,* **10**, 1-4. doi: 10.1186/1475-2875-10-23.

Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G. and Mann, M. (2002). Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature,* **419**, 537-542. doi: 10.1038/nature01111.

Lasonder, E., Janse, C. J., van Gemert, G.-J., Mair, G. R., Vermunt, A. M. W., Douradinha, B. G., van Noort, V., Huynen, M. A., Luty, A. J. F., Kroeze, H., Khan, S. M., Sauerwein, R. W., Waters, A. P., Mann, M. and Stunnenberg, H. G. (2008). Proteomic Profiling of <italic>Plasmodium</italic> Sporozoite Maturation Identifies New Proteins Essential for Parasite Development and Infectivity. *PLoS Pathog,* **4**, e1000195. doi: 10.1371/journal.ppat.1000195.

Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De la Vega, P., Holder, A. A., Batalov, S., Carucci, D. J. and Winzeler, E. A. (2003). Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. *Science,* **301**, 1503-1508. doi: 10.1126/science.1087025.

Lê, S., Josse, J. and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of statistical software,* **25**, 1-18.

Li, C., Sanni, L. A., Omer, F., Riley, E. and Langhorne, J. (2003). Pathology of Plasmodium chabaudi chabaudi infection and mortality in interleukin-10-deficient mice are ameliorated by anti-tumor necrosis factor alpha and exacerbated by anti-transforming growth factor ß antibodies. *Infection and immunity,* **71**, 4850-4856.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, **25**, 2078-2079. doi: 10.1093/bioinformatics/btp352.

Lin, J. W., Annoura, T., Sajid, M., Chevalley-Maurel, S., Ramesar, J., Klop, O., Franke-Fayard, B. M., Janse, C. J. and Khan, S. M. (2011). A novel 'gene insertion/marker out' (GIMO) method for transgene expression and gene complementation in rodent malaria parasites. *PLoS One, **6**, e29289. doi: 10.1371/journal.pone.0029289.

Lingnau, A., Margos, G., Maier, W. and Seitz, H. (1993). The effects of hormones on the gametocytogenesis of Plasmodium falciparum in vitro. *Applied parasitology, **34**, 153-160.

López-Barragán, M. J., Lemieux, J., Quiñones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K. and Su, X.-z. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. *BMC Genomics, **12**, 1-13. doi: 10.1186/1471-2164-12-587.

Lumb, V. and Sharma, Y. D. (2011). Novel K540N mutation in Plasmodium falciparum dihydropteroate synthetase confers a lower level of sulfa drug resistance than does a K540E mutation. *Antimicrob Agents Chemother, **55**, 2481-2482. doi: 10.1128/AAC.01394-10.

Maier, A. G., Braks, J. A., Waters, A. P. and Cowman, A. F. (2006). Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in Plasmodium falciparum for targeted gene deletion by double crossover recombination. *Mol Biochem Parasitol, **150**, 118-121. doi: 10.1016/j.molbiopara.2006.06.014.

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

Mair, G. R., Lasonder, E., Garver, L. S., Franke-Fayard, B. M., Carret, C. K., Wiegant, J. C., Dirks, R. W., Dimopoulos, G., Janse, C. J. and Waters, A. P. (2010). Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development. *PLoS Pathog, **6**, e1000767. doi: 10.1371/journal.ppat.1000767.

MALWEST (2016). Malaria: Plasmodium life cycle. Vol. 2016.

Mantel, P. Y., Hoang, A. N., Goldowitz, I., Potashnikova, D., Hamza, B., Vorobjev, I., Ghiran, I., Toner, M., Irimia, D., Ivanov, A. R., Barteneva, N. and Marti, M. (2013). Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system. *Cell Host Microbe, **13**, 521-534. doi: 10.1016/j.chom.2013.04.009.

Mantel, P. Y. and Marti, M. (2014). The role of extracellular vesicles in Plasmodium and other protozoan parasites. *Cell Microbiol, **16**, 344-354. doi: 10.1111/cmi.12259.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M. and Hurwitz, D. I. (2014). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, gku1221.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C. and Gonzales, N. R. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research, **39**, D225-D229.

Matera, A. G. and Wang, Z. (2014). A day in the life of the spliceosome. *Nat Rev Mol Cell Biol,* **15**, 108-121. doi: 10.1038/nrm3742.

Meissner, M., Brecht, S., Bujard, H. and Soldati, D. (2001). Modulation of myosin A expression by a newly established tetracycline repressor-based inducible system in Toxoplasma gondii. *Nucleic Acids Res,* **29**, E115.

Meissner, M., Krejany, E., Gilson, P. R., de Koning-Ward, T. F., Soldati, D. and Crabb, B. S. (2005). Tetracycline analogue-regulated transgene expression in Plasmodium falciparum blood stages using Toxoplasma gondii transactivators. *Proc Natl Acad Sci U S A,* **102**, 2980-2985. doi: 10.1073/pnas.0500112102.

Meissner, M., Schluter, D. and Soldati, D. (2002). Role of Toxoplasma gondii myosin A in powering parasite gliding and host cell invasion. *Science,* **298**, 837-840. doi: 10.1126/science.1074553.

Menard, R., Tavares, J., Cockburn, I., Markus, M., Zavala, F. and Amino, R. (2013). Looking under the skin: the first steps in malarial infection and immunity. *Nat Rev Micro,* **11**, 701-712. doi: 10.1038/nrmicro3111.

Mendis, K., Sina, B. J., Marchesini, P. and Carter, R. (2001). THE NEGLECTED BURDEN OF PLASMODIUM VIVAX MALARIA. *American Journal Tropical Medicine and Hygeine,* **64**, 97-106.

Missirlis, P. I., Smailus, D. E. and Holt, R. A. (2006). A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination. *BMC Genomics,* **7**, 73. doi: 10.1186/1471-2164-7-73.

Mons, B. (1986a). Intra erythrocytic differentiation of *Plasmodium berghei*. pp. 124. Leiden University.

Mons, B. (1986b). Intra erythrocytic differentiation of Plasmodium berghei. *Acta Leiden,* **54**, 1-124.

Mons, B., Janse, C., Boorsma, E. and Van der Kaay, H. (1985). Synchronized erythrocytic schizogony and gametocytogenesis of Plasmodium berghei in vivo and in vitro. *Parasitology,* **91**, 423-430.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* **5**, 621-628. doi: 10.1038/nmeth.1226.

Mota, M. M., Pradel, G., Vanderberg, J. P., Hafalla, J. C. R., Frevert, U., Nussenzweig, R. S., Nussenzweig, V. and Rodríguez, A. (2001). Migration of Plasmodium Sporozoites Through Cells Before Infection. *Science,* **291**, 141-144. doi: 10.1126/science.291.5501.141.

Motard, A., Marussig, M., Rénia, L., Baccam, D., Landau, I., Mattei, D., Targett, G. and Mazier, D. (1995). Immunization with the malaria heat shock like protein hsp70-1 enhances transmission to the mosquito. *International immunology,* **7**, 147-150.

Muralidharan, V., Oksman, A., Iwamoto, M., Wandless, T. J. and Goldberg, D. E. (2011). Asparagine repeat function in a Plasmodium falciparum protein assessed via a regulatable fluorescent affinity tag. *Proc Natl Acad Sci U S A,* **108**, 4411-4416. doi: 10.1073/pnas.1018449108.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods,* **6**, 917-922. doi: 10.1038/nmeth.1401.

O'Donnell, R. A., Freitas-Junior, L. H., Preiser, P. R., Williamson, D. H., Duraisingh, M., McElwain, T. F., Scherf, A., Cowman, A. F. and Crabb, B. S. (2002). A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of Plasmodium falciparum chromosomes. *EMBO J,* **21**, 1231-1239. doi: 10.1093/emboj/21.5.1231.

O'Neill, M. T., Phuong, T., Healer, J., Richard, D. and Cowman, A. F. (2011). Gene deletion from Plasmodium falciparum using FLP and Cre recombinases: implications for applied site-specific recombination. *Int J Parasitol,* **41**, 117-123. doi: 10.1016/j.ijpara.2010.08.001.

Oberdoerffer, P., Otipoby, K. L., Maruyama, M. and Rajewsky, K. (2003). Unidirectional Cre-mediated genetic inversion in mice using the mutant loxP pair lox66/lox71. *Nucleic Acids Res,* **31**, e140.

Oriol Llora-Batlle, C. B., Alfred Cortes (2016). A conditional over-expression system for pfap2-g: a tool for the study of sexual stages.

Orr, R. Y., Philip, N. and Waters, A. P. (2012). Improved negative selection protocol for Plasmodium berghei in the rodent malarial model. *Malar J,* **11**, 103. doi: 10.1186/1475-2875-11-103.

Otto, T. D., Bohme, U., Jackson, A. P., Hunt, M., Franke-Fayard, B., Hoeijmakers, W. A., Religa, A. A., Robertson, L., Sanders, M., Ogun, S. A., Cunningham, D., Erhart, A., Billker, O., Khan, S. M., Stunnenberg, H. G., Langhorne, J., Holder, A. A., Waters, A. P., Newbold, C. I., Pain, A., Berriman, M. and Janse, C. J. (2014). A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol,* **12**, 86. doi: 10.1186/s12915-014-0086-0.

Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., Lemieux, J., Barrell, B., Pain, A. and Berriman, M. (2010). New insights into the blood–stage transcriptome of Plasmodium falciparum using RNA–Seq. *Molecular microbiology,* **76**, 12-24.

Paget-McNicol, S., Gatton, M., Hastings, I. and Saul, A. (2002). The Plasmodium falciparum var gene switching rate, switching mechanism and patterns of parasite recrudescence described by mathematical modelling. *Parasitology,* **124**, 225-235.

Painter, H. J., Campbell, T. L. and Llinás, M. (2011). The Apicomplexan AP2 family: Integral factors regulating Plasmodium development. *Molecular and Biochemical Parasitology,* **176**, 1-7. doi: http://dx.doi.org/10.1016/j.molbiopara.2010.11.014.

Pei, Y., Tarun, A. S., Vaughan, A. M., Herman, R. W., Soliman, J., Erickson–Wayman, A. and Kappe, S. H. (2010). Plasmodium pyruvate dehydrogenase activity is only essential for the parasite's progression from liver infection to blood infection. *Molecular microbiology,* **75**, 957-971.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protocols,* **11**, 1650-1667. doi: 10.1038/nprot.2016.095 http://www.nature.com/nprot/journal/v11/n9/abs/nprot.2016.095.html#supplementary-information.

Peyron, F., Burdin, N., Ringwald, P., Vuillez, J. P., Rousset, F. and Banchereau, J. (1994). High levels of circulating IL-10 in human malaria. *Clinical & Experimental Immunology,* **95**, 300-303. doi: 10.1111/j.1365-2249.1994.tb06527.x.

Pfander, C., Anar, B., Schwach, F., Otto, T. D., Brochet, M., Volkmann, K., Quail, M. A., Pain, A., Rosen, B., Skarnes, W., Rayner, J. C. and Billker, O. (2011). A scalable pipeline for highly effective genetic modification of a malaria parasite. *Nat Meth,* **8**, 1078-1082. doi: http://www.nature.com/nmeth/journal/v8/n12/abs/nmeth.1742.html#supplementary-information.

Philip, N. and Waters, A. P. (2015). Conditional Degradation of Plasmodium Calcineurin Reveals Functions in Parasite Colonization of both Host and Vector. *Cell Host Microbe,* **18**, 122-131. doi: 10.1016/j.chom.2015.05.018.

Pieperhoff, M. S., Pall, G. S., Jimenez-Ruiz, E., Das, S., Melatti, C., Gow, M., Wong, E. H., Heng, J., Muller, S., Blackman, M. J. and Meissner, M. (2015). Conditional U1 Gene Silencing in Toxoplasma gondii. *PLoS One,* **10**, e0130356. doi: 10.1371/journal.pone.0130356.

Pino, P., Sebastian, S., Kim, E. A., Bush, E., Brochet, M., Volkmann, K., Kozlowski, E., Llinas, M., Billker, O. and Soldati-Favre, D. (2012). A tetracycline-repressible transactivator system to study essential genes in malaria parasites. *Cell Host Microbe,* **12**, 824-834. doi: 10.1016/j.chom.2012.10.016.

Pollitt, L. C., Mideo, N., Drew, D. R., Schneider, P., Colegrave, N. and Reece, S. E. (2011). Competition and the Evolution of Reproductive Restraint in Malaria Parasites. *The American naturalist,* **177**, 358-367. doi: 10.1086/658175.

Prudencio, M., Rodriguez, A. and Mota, M. M. (2006). The silent path to thousands of merozoites: the Plasmodium liver stage. *Nat Rev Microbiol,* **4**, 849-856. doi: 10.1038/nrmicro1529.

Reece, S. E., Drew, D. R. and Gardner, A. (2008). Sex ratio adjustment and kin discrimination in malaria parasites. *Nature,* **453**, 609-614. doi: http://www.nature.com/nature/journal/v453/n7195/suppinfo/nature06954_S1.html.

Reece, S. E., Duncan, A. B., West, S. A. and Read, A. F. (2005). Host cell preference and variable transmission strategies in malaria parasites. *Proceedings of the Royal Society B: Biological Sciences,* **272**, 511-517. doi: 10.1098/rspb.2004.2972.

Reece, S. E., Ramiro, R. S. and Nussey, D. H. (2009). SYNTHESIS: Plastic parasites: sophisticated strategies for survival and reproduction? *Evolutionary applications,* **2**, 11-23.

Regev-Rudzki, N., Wilson, D. W., Carvalho, T. G., Sisquella, X., Coleman, B. M., Rug, M., Bursac, D., Angrisano, F., Gee, M., Hill, A. F., Baum, J. and Cowman, A. F. (2013). Cell-cell communication between malaria-infected red blood cells via exosome-like vesicles. *Cell,* **153**, 1120-1133. doi: 10.1016/j.cell.2013.04.029.

Risco-Castillo, V., Topçu, S., Marinach, C., Manzoni, G., Bigorgne, A. E., Briquet, S., Baudin, X., Lebrun, M., Dubremetz, J.-F. and Silvie, O. (2015). Malaria sporozoites traverse host cells within transient vacuoles. *Cell host & microbe,* **18**, 593-603.

Robinson, M. S., Sahlender, D. A. and Foster, S. D. (2010). Rapid inactivation of proteins by rapamycin-induced rerouting to mitochondria. *Dev Cell,* **18**, 324-331. doi: 10.1016/j.devcel.2009.12.015.

Rovira-Graells, N., Crowley, V. M., Bancells, C., Mira-Martinez, S., Ribas de Pouplana, L. and Cortes, A. (2015). Deciphering the principles that govern mutually exclusive expression of Plasmodium falciparum clag3 genes. *Nucleic Acids Res,* **43**, 8243-8257. doi: 10.1093/nar/gkv730.

Schneweis, S., Maier, W. and Seitz, H. (1991). Haemolysis of infected erythrocytes—a trigger for formation ofPlasmodium falciparum gametocytes? *Parasitology research,* **77**, 458-460.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A,* **95**, 5857-5864.

Schwach, F., Bushell, E., Gomes, A. R., Anar, B., Girling, G., Herd, C., Rayner, J. C. and Billker, O. (2015). PlasmoGEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites. *Nucleic Acids Research,* **43**, D1176-D1182. doi: 10.1093/nar/gku1143.

Sen, G. L. and Blau, H. M. (2006). A brief history of RNAi: the silence of the genes. *FASEB J,* **20**, 1293-1299. doi: 10.1096/fj.06-6014rev.

Sidjanski, S. and Vanderberg, J. P. (1997). Delayed migration of Plasmodium sporozoites from the mosquito bite site to the blood. *Am J Trop Med Hyg,* **57**, 426-429.

Silvestrini, F., Alano, P. and Williams, J. L. (2000). Commitment to the production of male and female gametocytes in the human malaria parasite Plasmodium falciparum. *Parasitology,* **121 Pt 5**, 465-471.

Silvestrini, F., Bozdech, Z., Lanfrancotti, A., Di Giulio, E., Bultrini, E., Picci, L., Derisi, J. L., Pizzi, E. and Alano, P. (2005). Genome-wide identification of genes upregulated at the onset of gametocytogenesis in Plasmodium falciparum. *Mol Biochem Parasitol,* **143**, 100-110. doi: 10.1016/j.molbiopara.2005.04.015.

Silvestrini, F., Lasonder, E., Olivieri, A., Camarda, G., van Schaijk, B., Sanchez, M., Younis Younis, S., Sauerwein, R. and Alano, P. (2010). Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics,* **9**, 1437-1448. doi: 10.1074/mcp.M900479-MCP200.

Sinden, R. E. and Billingsley, P. F. (2001). Plasmodium invasion of mosquito cells: hawk or dove? *Trends in Parasitology,* **17**, 209-211. doi: http://dx.doi.org/10.1016/S1471-4922(01)01928-6.

Sinha, A., Hughes, K. R., Modrzynska, K. K., Otto, T. D., Pfander, C., Dickens, N. J., Religa, A. A., Bushell, E., Graham, A. L., Cameron, R., Kafsack, B. F., Williams, A. E., Llinas, M., Berriman, M., Billker, O. and Waters, A. P. (2014). A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. *Nature,* **507**, 253-257. doi: 10.1038/nature12970.

Smalley, M. and Brown, J. (1981). Plasmodium falciparum gametocytogenesis stimulated by lymphocytes and serum from infected Gambian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **75**, 316-317.

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. and Hay, S. I. (2005). The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature,* **434**, 214-217. doi: 10.1038/nature03370
10.1029/2002GB001991
10.1029/2001GL014076.

Sorber, K., Dimon, M. T. and Derisi, J. L. (2011). RNA-Seq analysis of splicing in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic Acids Res,* **39**. doi: 10.1093/nar/gkq1223.

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

Suhrbier, A., Janse, C., Mons, B., Fleck, S. L., Nicholas, J., Davies, C. S. and Sinden, R. E. (1987). The complete development in vitro of the vertebrate phase of the mammalian malarial parasite Plasmodium berghei. *Transactions of the Royal Society of Tropical Medicine and*

Hygiene, **81**, 907-909. doi: http://dx.doi.org/10.1016/0035-9203(87)90346-4.

Sun, Z. and Komarova, N. L. (2015). Stochastic control of proliferation and differentiation in stem cell dynamics. *J Math Biol,* **71**, 883-901. doi: 10.1007/s00285-014-0835-2.

Talisuna, A. O., Bloland, P. and D'Alessandro, U. (2004). History, Dynamics, and Public Health Importance of Malaria Parasite Resistance. *Clinical Microbiology Reviews,* **17**, 235-254. doi: 10.1128/cmr.17.1.235-254.2004.

Tan, X., Calderon-Villalobos, L. I., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M. and Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature,* **446**, 640-645. doi: 10.1038/nature05731.

Tarun, A. S., Peng, X., Dumpit, R. F., Ogata, Y., Silva-Rivera, H., Camargo, N., Daly, T. M., Bergman, L. W. and Kappe, S. H. (2008). A combined transcriptome and proteome survey of malaria parasite liver stages. *Proceedings of the National Academy of Sciences,* **105**, 305-310.

Tavares, J., Formaglio, P., Thiberge, S., Mordelet, E., Van Rooijen, N., Medvinsky, A., Ménard, R. and Amino, R. (2013). Role of host cell traversal by the malaria sporozoite during liver infection. *The Journal of Experimental Medicine,* **210**, 905-915. doi: 10.1084/jem.20121130.

Team, R. C. (2013). R: A language and environment for statistical computing.

Templeton, T. J., Iyer, L. M., Anantharaman, V., Enomoto, S., Abrahante, J. E., Subramanian, G., Hoffman, S. L., Abrahamsen, M. S. and Aravind, L. (2004). Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome research,* **14**, 1686-1695.

Tiburcio, M., Dixon, M. W., Looker, O., Younis, S. Y., Tilley, L. and Alano, P. (2015). Specific expression and export of the Plasmodium falciparum Gametocyte EXported Protein-5 marks the gametocyte ring stage. *Malar J,* **14**, 334. doi: 10.1186/s12936-015-0853-6.

Trager, W. and Jensen, J. (1976). Human malaria parasites in continuous culture. *Science,* **193**, 673-675. doi: 10.1126/science.781840.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols,* **7**, 562-578.

Triglia, T., Menting, J. G., Wilson, C. and Cowman, A. F. (1997). Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in Plasmodium falciparum. *Proc Natl Acad Sci U S A,* **94**, 13944-13949.

Tuteja, R. and Pradhan, A. (2006). Unraveling the 'DEAD-box' helicases of Plasmodium falciparum. *Gene,* **376**, 1-12. doi: http://dx.doi.org/10.1016/j.gene.2006.03.007.

van Dijk, M. R., van Schaijk, B. C., Khan, S. M., van Dooren, M. W., Ramesar, J., Kaczanowski, S., van Gemert, G. J., Kroeze, H., Stunnenberg, H. G., Eling, W. M., Sauerwein, R. W., Waters, A. P. and Janse, C. J. (2010). Three members of the 6-cys protein family of Plasmodium play a role in gamete fertility. *PLoS Pathog,* **6**, e1000853. doi: 10.1371/journal.ppat.1000853.

von Bohl, A., Kuehn, A., Simon, N., Ngongang, V. N., Spehr, M., Baumeister, S., Przyborski, J. M., Fischer, R. and Pradel, G. (2015). A WD40-repeat protein unique to malaria parasites associates with adhesion protein complexes and is crucial for blood stage progeny. *Malaria Journal,* **14**, 435. doi: 10.1186/s12936-015-0967-x.

Voss, T. S., Kaestli, M., Vogel, D., Bopp, S. and Beck, H. P. (2003). Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters. *Mol Microbiol,* **48**, 1593-1607.

Wang, C. H., Su, L. H. and Sun, C. H. (2007). A novel ARID/Bright-like protein involved in transcriptional activation of cyst wall protein 1 gene in Giardia lamblia. *J Biol Chem,* **282**, 8905-8914. doi: 10.1074/jbc.M611170200.

Wang, C. W., Hermsen, C. C., Sauerwein, R. W., Arnot, D. E., Theander, T. G. and Lavstsen, T. (2009). The Plasmodium falciparum var gene transcription strategy at the onset of blood stage infection in a human volunteer. *Parasitol Int,* **58**, 478-480. doi: 10.1016/j.parint.2009.07.004.

Wang, J., Sarov, M., Rientjes, J., Hu, J., Hollak, H., Kranz, H., Xie, Y., Stewart, A. F. and Zhang, Y. (2006). An improved recombineering approach by adding RecA to λ Red recombination. *Molecular Biotechnology,* **32**, 43-53. doi: 10.1385/mb:32:1:043.

Wang, P., Read, M., Sims, P. F. G. and Hyde, J. E. (1997). Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. *Molecular microbiology,* **23**, 979-986. doi: 10.1046/j.1365-2958.1997.2821646.x.

White, N. J., Turner, G. D. H., Medana, I. M., Dondorp, A. M. and Day, N. P. J. (2010). The murine cerebral malaria phenomenon. *Trends in Parasitology,* **26**, 11-15. doi: http://dx.doi.org/10.1016/j.pt.2009.10.007.

WHO (2014). Malaria report.

WHO (2015). Malaria Report.

WHO (2016). Children: Reducing mortality.

Williams, J. L. (1999). Stimulation of Plasmodium falciparum gametocytogenesis by conditioned medium from parasite cultures. *Am J Trop Med Hyg,* **60**, 7-13.

Wu, Y., Cruz, L. N., Szestak, T., Laing, G., Molyneux, G. R., Garcia, C. R. and Craig, A. G. (2016). An external sensing system in Plasmodium falciparum-infected erythrocytes. *Malar J,* **15**, 103. doi: 10.1186/s12936-016-1144-6.

Wu, Y., Sifri, C. D., Lei, H.-H., Su, X.-z. and Wellems, T. E. (1995). Transfection of Plasmodium falciparum within human red blood cells. *Proceedings of the National Academy of Sciences,* **92**, 973-977.

Yuda, M., Iwanaga, S., Kaneko, I. and Kato, T. (2015). Global transcriptional repression: An initial and essential step for Plasmodium sexual development. *Proceedings of the National Academy of Sciences,* **112**, 12824-12829. doi: 10.1073/pnas.1504389112.

Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T. and Kaneko, I. (2010a). Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol Microbiol,* **75**, 854-863. doi: 10.1111/j.1365-2958.2009.07005.x.

Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T. and Kaneko, I. (2010b). Transcription factor AP2–Sp and its target genes in malarial sporozoites. *Molecular microbiology,* **75**, 854-863.

Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G. R., Janse, C. J., Waters, A. P., Kato, T. and Kaneko, I. (2009). Identification of a transcription factor in the mosquito–invasive stage of malaria parasites. *Molecular microbiology,* **71**, 1402-1414.

Zhang, Y., Buchholz, F., Muyrers, J. P. and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet,* **20**, 123-128. doi: 10.1038/2417.

Zhang, Z. and Lutz, B. (2002). Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. *Nucleic Acids Res,* **30**, e90.

Chapter 10   Supplementary Information
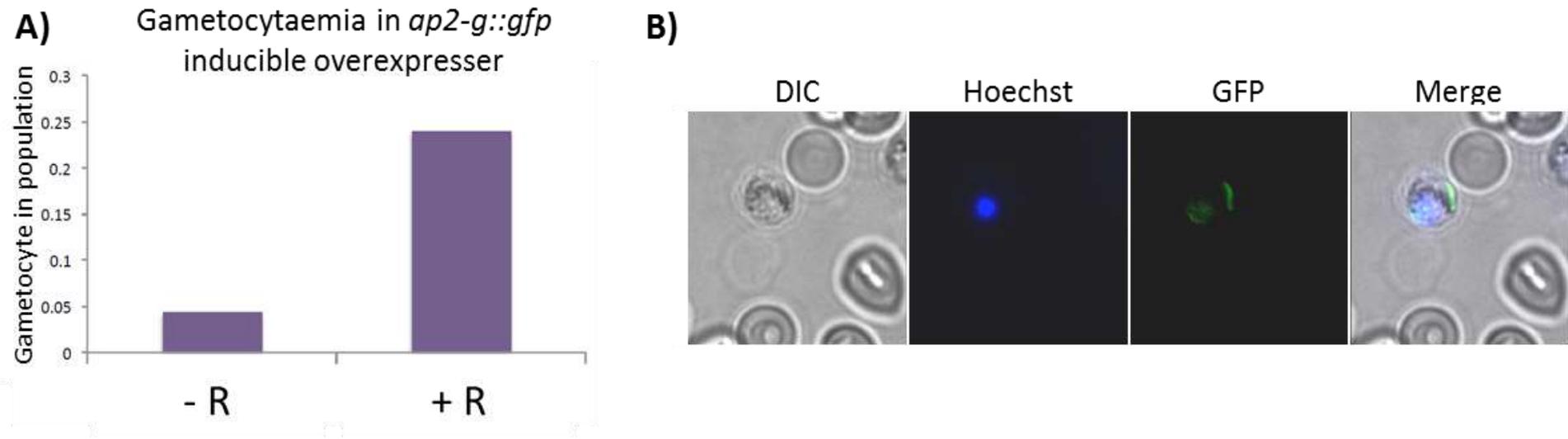
Supplementary Figure 1 Response of the *ap2-g::gfp* inducible overexpresser

A) In the *ap2-g::gfp* inducible overexpresser the line retains the ability to generate gametocytes (approximately 5%) before induction with rapamycin. Upon addition of rapamycin a clear increase in commitment is observed with gametocytaemia reaching around 25%.

B) After induction and maturation of the culture overnight mature gametocytes are discernible and contain two distinct foci of GFP expression. One of these co-localises with the nuclear dye (Hoechst).
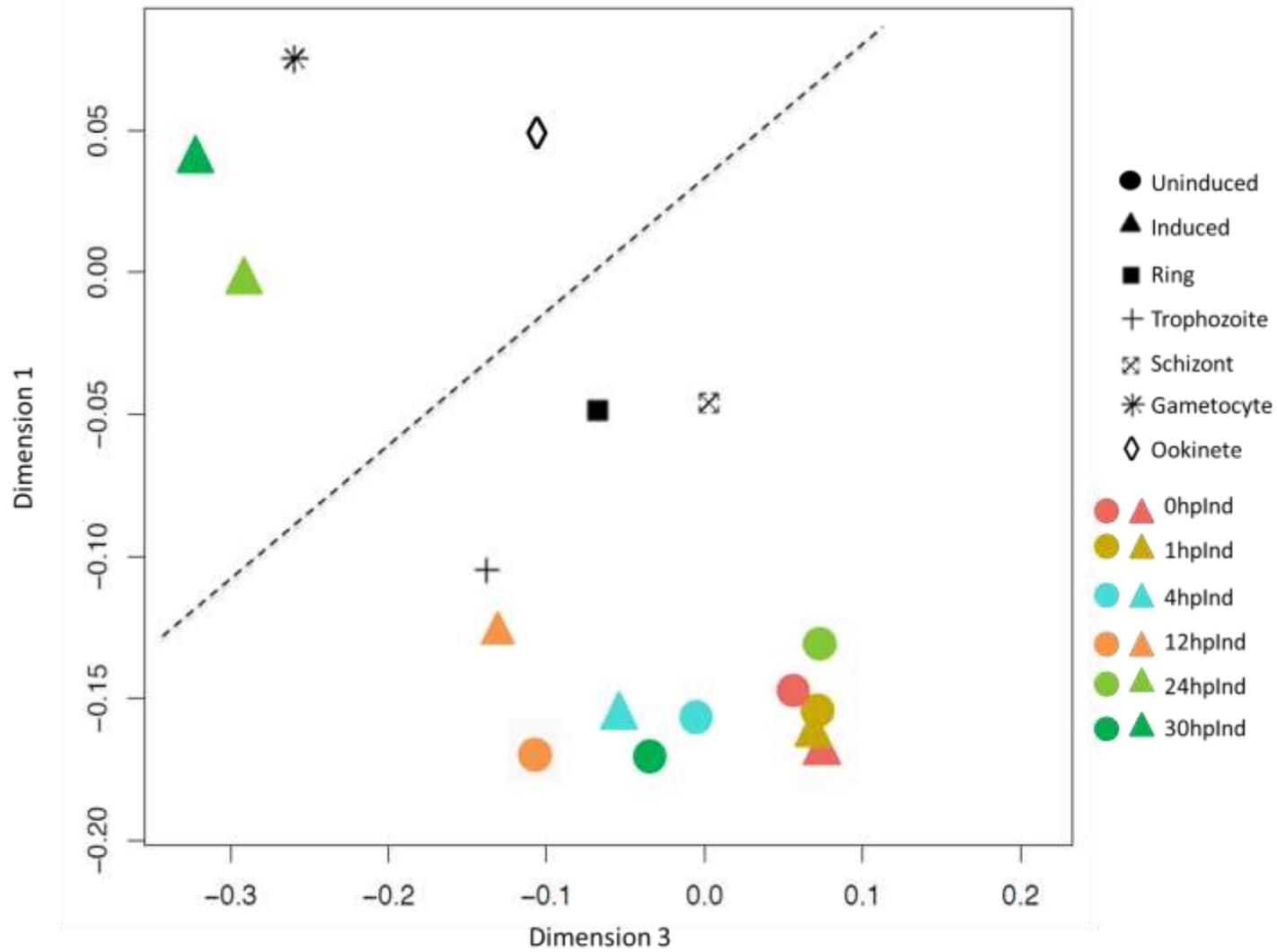
Supplementary Figure 2 PCA of top 20% differentially expressed genes in the pilot time course
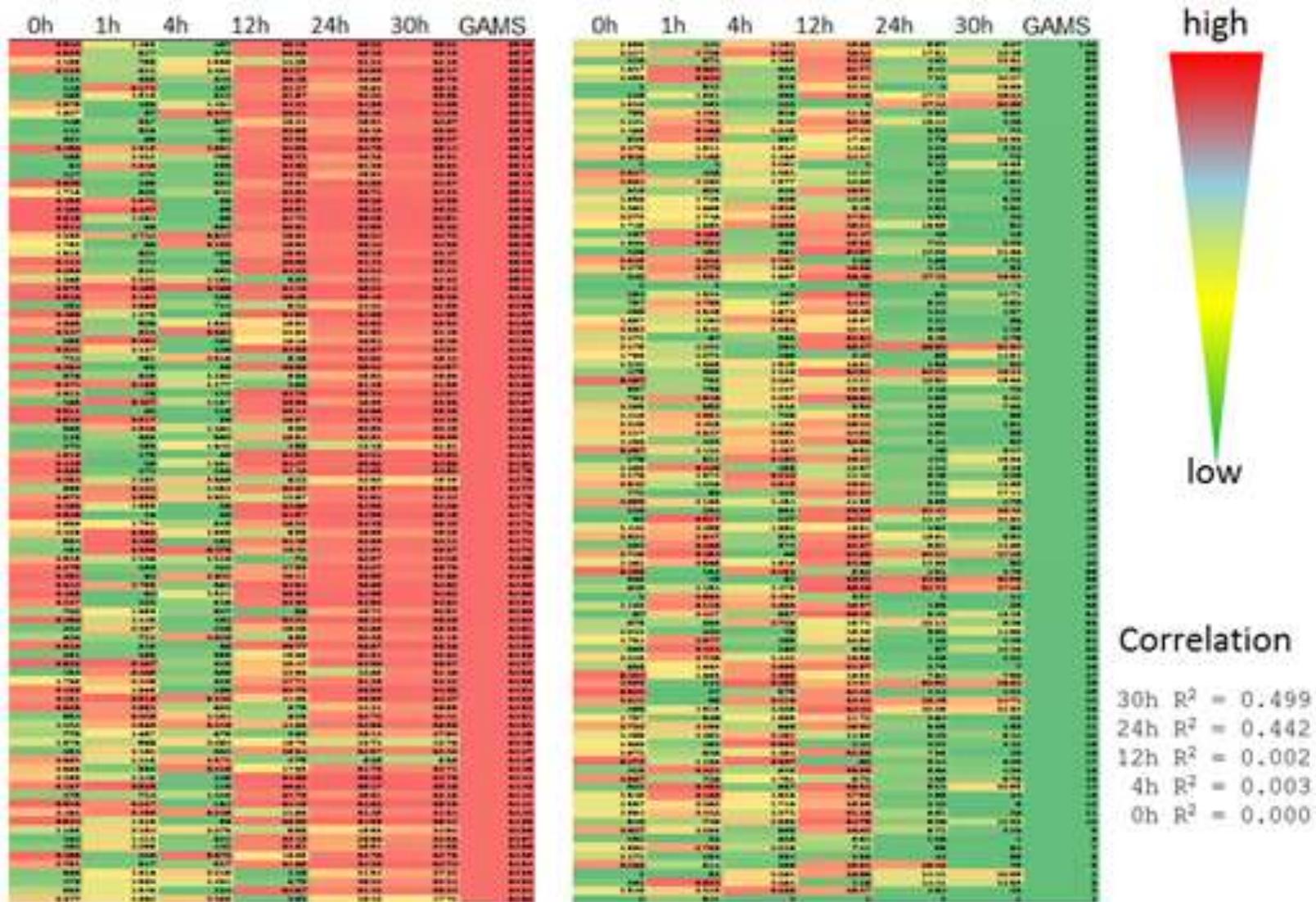
Supplementary Figure 3 Correlation of induced samples to purified gametocytes

Primer no.	Sequence	Used for
P1	GTGAAGTTCAAATATGTGAAAAAACAATAAATGAATTTAGC	5' P1 integration
P2	GTTCTTTATCTGTTCCTTC	3' P1 Integration
P3	TAGTTAGCTTAAATTGTCCAACTGG	5' P2 Integration
P4	CATAAACGGTTTATTTAAAGTCATTTTTGG	3' P2 Integration
P5	ACTTTGGTGACAGATACTACTG	GIMO 5' R
P6	GTAAACTTAAGCATAAAGAGCTCG	GIMO 3' F
P7	ccGCGGCCGCCTAGAATTATTAATATATATGAATATATATACATCGTTGTATG	DiCre 5' R
P8	CTAGTCGTGCAATACTG	DiCre 3' F
P9	CTGGGTATTCTGGCAGAAG	EG/CFP 5' R
P10	cgACTAGTCCCGGGCTTAACATTCACATATATTAATAATTTTAAT	EG/CFP 3' F
P11	ATGGTTGGTTCGCTAAACTG	Selectable marker F
P12	CTGGGTATTCTGGCAGAAG	Selectable marker R
P13	GCATTTTACACTATTTTGCCATAAG	qPCR reporter F
P14	CTA GCT TTT GCG TTA GCC ATA TAA C	qPCR reporter Unexcised
P15	CGT TCT TAG AAG ATC TAC TCA TAT AAC	qPCR reporter excised
P16	GCATACTATATTATGACTATCCTTGCC	Reporter qPCR control F
P17	CAT TTA TTA GCC CCT TAG TTA AAG CAG	Reporter qPCR control R
P18	GCAAATATTTTCTGACGTTTACC	ap2-g inducible 5' F
P19	CAT CAT CTG AGT GGT TTT GAA TAT CAG	ap2-g inducible 3' R
P20	Gtctcttcaatgattcataaatag	ap2-g inducible integration R
P21	ACAAATTCAATGAACTTCTAATATGACTC	ap2-g inducible integration F
P22	GTCAGTTTTAAACTCAAATGATTCAG	ap2-g aPCR flip control F
P23	CTT AAA AGT CTT TAA ATT GAT CTT TAT G	ap2-g aPCR flip control R
P24	CAT CAT CTG AGT GGT TTT GAA TAT CAG	ap2-g flipping R
P25	CGT ATA TTT TCC CTC AAA TTA ATT AAA TAA ATG	ap2-g unflipped F
P26	GCATTTTACACTATTTTGCCATAAG	ap2-g flipped F
GU3607	GTA AAC TTA AGC ATA AAG AGC TCG	hDHFR
GU3608	GTAG TAT CTG TCA CCA AAG TCA C	yfcu F
GU3609	G AGA CAC ACC TTA ACA TTT AAC TTG	PBANKA_060120
GU3610	CGA GTT GCT CAC ATA TAC AAT ATA TG	PBANKA_081180
GU3611	GAA GAA ATA TAT GAC AAC GGT GTT AAT G	PBANKA_090850
GU3612	CAT AAA GAG CTC GAA AAG AAT TAA GC	hDHFR promoter R
GU3613	AGC ATG CAC GAG GAA CAG C	PBANKA_093360
GU3614	GTG ACT TTG GTG ACA GAT ACT AC	yfcu F
GU3615	GAG TAC ATC AAG TGA TTT ATT GAT GC	PBANKA_021390

GU3616	GTG TGC ATA TAT GTA TAG TGA AAT TGG	PBANKA_031270
GU3617	AGA GCT CGA AAA GAA TTA AGC TG	yfcu F
GU3618	GAT ATG CTT CGA TTT ATC AAT ATT CC	PBANKA_111500
GU3619	CTGCTATGGCTAAAATCAAATTTAGT	PBANKA_101830
GU3620	CAA CAA ATT TTA TCA AAA GCC AAA AAT AC	PBANKA_112040
GU3621	GTA TAT AAT AAT ATA GGC TGT AGA CAT G	PBANKA_121830
GU3622	GGATATGCATGCGTATATATTTCTC	PBANKA_070470
GU3623	CCC TGT TTA TCT AGT CGA TAT TTT G	PBANKA_135250
GU3624	GTATTATTAGAGAAGCCATTACTATTC	PBANKA_142920
GU3625	GC AAA AAG TAT TCC ACA AAT TAA TGT TAG	PBANKA_133500
GU3626	GG TAA CTT ATT TTA CCA TCA AAG AAT AAC	PBANKA_133810
GU3627	CT TAA AAA TAT TAA TAT ATA TGT ATG CAC ATG	PBANKA_113860
GU3628	GGT CTG GAA TTT CTT CAA CAT TTG	PBANKA_050440
GU3629	CGA AGC TCA AGT TCG AGT TC	PBANKA_010510
GU3630	GGT ACA TGA CGA TGT TGT GTA TG	PBANKA_080720
GU3631	ATC ACT ATT TCA TCT TAA CGA TAA CG	PBANKA_122550
GU3632	GTG AAA CTC ATT ACC ATT TTA AAT GTG	PBANKA_122990
GU3633	GATATGATATATGTTTTGTCTTTTATCTAC	PBANKA_136040
GU3634	GGTA TAG ACT TAA ATG AAG CCA ATC	PBANKA_050830
GU3732	GTA GAT ATG CGC AAG GCA CCC	PBANKA_121260
GU3733	TTA GCG CGC ATA TAT GCG CTA A	PBANKA_071650
GU3734	GTCGTTCAAGGAACATTATTAGGAAGC	PBANKA_010240
GU3735	CTC ATC GTT CAT GGA TTT AAG CAA CTT C	PBANKA_082800
GU3736	CTT AGA TGG TAA TGT TCG TAG TAT TGT G	PBANKA_141810
GU3737	CAT GGT ATA GCA TAA TAA ACA TAT TGA ATG AG	PBANKA_143520
GU3738	GTT CTT AAT AAT AAC CAT TCT ATC AAT TCC AC	PBANKA_090240
GU3739	GTA CAG TAG GCC TAT ATT TAT AAA TAA TAT TAA TTT TG	PBANKA_041340
GU3740	TAT CCC TTA TCA CAA TCA TAC CCT TAT A	PBANKA_112510
GU3741	CAC ACT TAA GGA TAA ATA ATA TTA GGA ATA TAT AC	PBANKA_142930
GU3742	GAGATGTTCATGTATGTATGGGTTCAG	PBANKA_083040
GU3749	GAA AAC CAA CAT ATG ATG AGT TTC CAA G	PBANKA_140430
GU3750	GAG AGA TGA TGA TTC TGA AGA AAA TAT AGA C	PBANKA_050500
GU3751	TAA TAA CTA TAA GTA TAT ATC CTA TAA TTT TGC TTT CT	PBANKA_130810
GU3752	GTT TGT GAT AAG AAT GGG AAC TCT ATA TTT G	PBANKA_041720
GU3753	CACTTTGTAATAATGTGTACATATATGTATACAC	PBANKA_051060

Supplementary Table 1 Primers used in this work