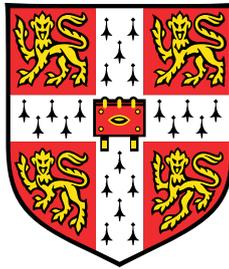


# Characterisation of *Plasmodium* parasite sexual commitment and development



**Juliana Catherine Maria Cudini**

Wellcome Sanger Institute  
University of Cambridge

This dissertation is submitted for the degree of  
*Doctor of Philosophy*

Gonville & Caius College

October 2021



For Mom, Dad, and Abby



## **Declaration**

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

Juliana Catherine Maria Cudini

October 2021



## Acknowledgements

First and foremost, I'd like to thank my primary supervisor Dr. Mara Lawniczak for her guidance, friendship, understanding, and most of all, patience over the last four years. I don't think either of us could have predicted the wayward directions my scientific whims headed in, but she was always there to help nudge the steering wheel back to (tractable) solid ground. I can't thank her enough and I certainly wouldn't be the scientist I am without her. In addition, I would like to thank my co-supervisor Prof. Dominic Kwiatkowski for many of the same grounding conversations, and to my thesis committee Dr. Martin Hemberg and Dr. Aylwyn Scally for their many useful comments.

I'd also like to thank the members of the Lawniczak lab, past and present, for their camaraderie and (sometimes bemused) assistance as I stumbled my way from a computer screen to a lab bench, in particular Dr. Ginny Howick, Dr. Arthur Talman and Dr. Sunil Kumar Dogga. Also, a resounding thank you to the newly minted *Dr.* Andy Russell for four years of collaboration and friendship.

A PhD is rarely done in isolation, and I'd like to thank so many of the friends I've made that have helped me to reach this point. To my fellow inmates at 164 and 207, thank you for keeping my spirits up whilst we were locked down, in particular Tom Albrow-Owen, Catherine Walker, Katherine Brightwell and Yasmine Shafiq. To Hannah Jagoe, Emma Carpenter, and Joa Hoshizaki, thank you for keeping me (and my cells) alive during many late nights in the malaria labs. I also wouldn't be here without the endless support, advice, and encouragement from Tapoka Mkandawire, Anna Appios, Brittany Howell, Duncan Berger, Rebecca Dell, Maria Stroyakovski, Cameron Smith, and many other members of the Sanger Institute and Gonville & Caius College communities that I have forgotten to mention. To my Canadian cheerleaders Alyx Riddell, Katie Pedersen and Stephanie Küntz, thank you for encouraging me to follow my passions overseas, and to my adopted Canadian

family on the Cambridge ice hockey team for keeping me grounded outside of my research. Additionally, thank you to Angela Harper for being my rock, my best friend, and my voice of reason throughout. I am truly privileged to have studied alongside such an incredible and accomplished group of people.

Finally, and perhaps most importantly, I'd like to thank my family for their endless encouragement, inspiration, and unwavering support. My sister Abby, wise beyond her years, has kept me level with her sometimes infuriatingly lucid life advice. And my parents, Mark and Irene, though they may regret the destruction I inflicted upon the local flora and fauna from an early age, never once discouraged my curiosity, even as it took me further and further afield. I owe everything I have to them.

## Abstract

Malaria is a devastating disease responsible for over 400,000 deaths each year. The disease is caused by a single-celled parasite of the genus *Plasmodium*, which establishes infection via a bite from an Anopheline mosquito. While the parasite progresses through a complex range of life stages, it is the blood stages, or the intraerythrocytic developmental cycle (IDC), that cause the large majority of harmful symptoms. During the course of the IDC, a parasite grows in size within a red blood cell until it is able to multiply itself asexually many times and burst from the cell as individual infectious units, each one then able to infect a new red blood cell and restart the cycle. This pattern of asexual reproduction and re-invasion of fresh cells allows the parasite population to swell to impressive sizes within a host.

While the IDC growth cycle can keep a parasite population happily established within the host, it is not able to allow passage between hosts. Thus, as the parasite progresses through the IDC, it must make a decision. Either it can continue into another cycle of asexual growth in that host, or sexually (and terminally) differentiate into gametocytes, the transmissible form of the parasite, and thus gain an opportunity to transfer to a new host. Gametocytogenesis, the formation of these sexual forms, is therefore essential for malaria transmission, and an attractive target for transmission blocking interventions. Despite its importance, we know little about sex-specific gene expression or how the decision to become male or female is made. Efforts to understand gametocytogenesis have been hampered by the fact that gametocytes often represent less than 1% of the total population of parasites circulating in a host, meaning any sexual transcriptional signal is lost amidst an abundance of asexuals. Single cell RNA-sequencing has revolutionised our ability to capture rare populations, providing an ideal window into heterogeneity between parasites and developmental processes at high resolution.

In this thesis, I use 10x Genomics single cell capture to sample the transcriptome of over 30,000 single cells from time points spanning the sexual developmental pathway of *P. falciparum*, from asexual growth, to sexual commitment, and into sexual maturity. I first use the data collected to generate a high quality reference atlas for gametocyte development. From this, I profile a number of global changes underlying sexual commitment, development, and maturity into males and females. By mixing two genetically distinct parasite strains (NF54 and 7G8), I place these findings in a larger context, describing differences in development that occur between strains of the same species. Lastly, I complete my profile of transcriptional changes underlying parasite development by exploring the localisation of the lesser profiled non-coding expression to specific regions of the life cycle, and how they may contribute to transmission.

# Table of contents

<b>List of figures</b>	<b>xv</b>
<b>List of tables</b>	<b>xix</b>
<b>Abbreviations</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Global Burden of Malaria . . . . .	1
1.2 <i>Plasmodium</i> parasites and their phylogeny . . . . .	2
1.2.1 Apicomplexans: a phylum of obligate parasites . . . . .	2
1.2.2 The <i>Plasmodium</i> genus . . . . .	3
1.3 The human malaria parasite <i>Plasmodium falciparum</i> . . . . .	3
1.3.1 Life cycle at a glance . . . . .	4
1.3.2 Global diversity of natural infections . . . . .	6
1.3.3 Strains used in laboratory practice . . . . .	8
1.3.4 The <i>Plasmodium falciparum</i> Genome . . . . .	10
1.4 Malaria parasites, in the age of transcriptomics . . . . .	10
1.4.1 Global patterns of expression . . . . .	11
1.4.2 Mechanisms for regulating transcription . . . . .	12
1.4.3 Non-coding transcription . . . . .	14
1.4.4 Transcription at a single-cell level . . . . .	18

---

1.5	Gametocytogenesis . . . . .	20
1.5.1	The stages of sexual development . . . . .	21
1.5.2	Sexual commitment: when and why? . . . . .	22
1.5.3	Regulating sexual commitment . . . . .	24
1.5.4	Sex determination and sexual maturity . . . . .	26
<b>2</b>	<b>Thesis Outline</b>	<b>29</b>
<b>3</b>	<b>An atlas of gametocyte development in <i>Plasmodium falciparum</i></b>	<b>31</b>
3.1	Introduction . . . . .	31
3.2	Results . . . . .	33
3.2.1	Constructing an atlas of gametocyte development . . . . .	33
3.2.2	scRNAseq provides resolution regarding the timing of key events during sexual development . . . . .	43
3.2.3	Profiling global changes in gene expression during sexual development . . . . .	49
3.2.4	Sexual differentiation precedes a shared sexual progenitor state	52
3.2.5	The gametocyte atlas as a tool for the malaria research community . . . . .	57
3.3	Discussion . . . . .	58
3.4	Materials & Methods . . . . .	59
3.4.1	Parasite culture and sample collection . . . . .	59
3.4.2	Library preparation, single cell sequencing and read mapping	60
3.4.3	SCT filtering, normalisation and annotation . . . . .	61
3.4.4	Single cell transcriptome analyses . . . . .	63
<b>4</b>	<b>Inter-strain differences in expression during sexual development</b>	<b>67</b>
4.1	Introduction . . . . .	67
4.2	Results . . . . .	68

---

4.2.1	Deconvolution of strains using scRNAseq data . . . . .	68
4.2.2	Differential expression in gametocyte-specific genes between strains . . . . .	70
4.3	Discussion . . . . .	74
4.4	Materials & Methods . . . . .	76
4.4.1	Parasite culture, sample preparation, and sequencing . . . . .	76
4.4.2	SCT Strain assignment and doublet removal . . . . .	77
4.4.3	Differential expression analysis . . . . .	78
<b>5</b>	<b>Investigating long non-coding RNA expression in <i>Plasmodium falciparum</i></b>	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Results . . . . .	81
5.2.1	Generation of a data set with improved depth for lncRNA detection . . . . .	81
5.2.2	Mapping and detection of lncRNA in scRNAseq data . . . . .	82
5.2.3	lncRNAs show stage-specific expression across the <i>Plasmodium</i> life cycle . . . . .	89
5.2.4	scRNAseq as a tool for selecting non-coding candidates of interest . . . . .	95
5.3	Discussion . . . . .	98
5.4	Materials & Methods . . . . .	99
5.4.1	Parasite culturing and sample collection . . . . .	99
5.4.2	Single cell sequencing, barcode assignment and read mapping . . . . .	99
5.4.3	SCT filtering, normalisation and annotation . . . . .	100
5.4.4	Sample merging and lncRNA annotation . . . . .	101
5.4.5	lncRNA filtering and further data analysis . . . . .	102
<b>6</b>	<b>Conclusions &amp; Future Directions</b>	<b>105</b>

6.1	Biological insights . . . . .	105
6.1.1	A progenitor stage of indeterminate sex in early gametocytes	105
6.1.2	Specificity of long non-coding RNAs . . . . .	106
6.2	Cell atlases: the first step . . . . .	107
6.2.1	Mapping across across experiments . . . . .	107
6.2.2	Towards a comprehensive picture of expression . . . . .	108
	<b>References</b>	<b>111</b>
	<b>Appendix A Supplementary Tables</b>	<b>137</b>
A.1	Supplementary material for Chapter 3 . . . . .	137

# List of figures

1.1	The life cycle of <i>Plasmodium</i> parasites . . . . .	5
1.2	Global diversity of 7000 <i>P. falciparum</i> samples . . . . .	8
1.3	Mechanisms of ncRNA-mediated <i>var</i> gene silencing . . . . .	17
1.4	Regulation of sexual commitment by antisense <i>gdv1</i> transcript . . . . .	18
1.5	Stages of sexual development in <i>P. falciparum</i> . . . . .	22
1.6	Pathways to sexual commitment . . . . .	26
3.1	Experimental design . . . . .	34
3.2	Cell Ranger barcode white-listing struggles with continuous cell types	35
3.3	Asexual life stages require custom quality control thresholds . . . . .	37
3.4	Cell quality thresholds do not remove putative gametocyte cells . . . . .	38
3.5	Integration of multiple data sets allows for full coverage of gametocyte development . . . . .	40
3.6	Batch correction improves data integration . . . . .	41
3.7	A single-cell atlas of <i>P. falciparum</i> gametocyte development . . . . .	42
3.8	Fitting pseudotime to sexual and asexual developmental trajectories	44
3.9	Combining scRNAseq with bulk RNAseq allows for timing of gametocyte departure . . . . .	45
3.10	Staging gametocyte transcriptomes . . . . .	47
3.11	Gene expression in early gametocytes . . . . .	48

---

3.12 Gene-based clustering reveals patterns of co-expression during sexual and development . . . . .	50
3.13 Average expression of clusters enriched for asexual-development genes . . . . .	51
3.14 Average expression of clusters enriched for sexual-development genes	52
3.15 Gene expression during sexual differentiation . . . . .	54
3.16 Gene expression during sexual differentiation . . . . .	56
3.17 Mapping to the gametocyte atlas allows for identification of outliers in single cell data . . . . .	58
4.1 Proportion of NF54 and 7G8 parasite strains captured in each sample	69
4.2 Assessing strain imbalance in each sample . . . . .	71
4.3 Pseudotime-associated differentially expressed genes between NF54 and 7G8 gametocytes . . . . .	73
4.4 Increased expression of genes contained in a segmental duplication in the 7G8 genome . . . . .	74
4.5 Differential expression of genes expressed during early gametocyte development . . . . .	75
5.1 lncRNA classification . . . . .	80
5.2 New scRNAseq chemistry improves mRNA detection . . . . .	83
5.3 lncRNA detection and expression . . . . .	84
5.4 Average lncRNA expression is largely independent of AT content, strand and length . . . . .	85
5.5 lncRNA expression is correlated with proximal mRNA expression . .	87
5.6 Schematic for lncRNA classification . . . . .	89
5.7 lncRNA show stage specific expression . . . . .	90
5.8 lncRNA exhibit cascade-like peak expression throughout asexual development . . . . .	92

---

5.9	lncRNAs are expressed during gametocyte development and sexual differentiation . . . . .	94
5.10	Predictions for lncRNA stage-specific expression match those estimated from bulk RNAseq . . . . .	96
5.11	lncRNAs are expressed in close proximity to coding regions implicated in gametocyte development . . . . .	97
5.12	Expression of anti- <i>gdv1</i> antisense transcript can be detected in scRNAseq data . . . . .	98



# List of tables

1.1	Origins of select <i>P. falciparum</i> clones/strains often used in laboratory culture . . . . .	9
3.1	Quality Statistics per Sample . . . . .	39
A.1	Cluster assignments for gene by cell gene UMAP . . . . .	137



# Abbreviations

asLncRNA	Antisense long non-coding RNA
GAM	Generalised additive model
IDC	Intraerythrocytic developmental cycle
IMC	Inner membrane complex
iRBC	Immature red blood cell
lincRNA	Long intergenic non-coding RNA
lncRNA	Long non-coding RNA
mRNA	Messenger RNA (protein coding)
NAT	Natural antisense transcript
NCC	Next cycle conversion
ncRNA	Non-coding RNA
Pol II	RNA Polymerase II
RBC	Red blood cell
SCC	Same cycle conversion
scRNAseq	Single-cell RNA sequencing
SNP	Single nucleotide polymorphism
TSS	Transcription start site
UMI	Unique molecular identifier
UTR	Untranslated region
VCF	Variant call file
WGS	Whole genome sequencing



# Chapter 1

## Introduction

### 1.1 The Global Burden of Malaria

Malaria is a devastating blood-borne disease caused by an infection with a single-celled parasite of the genus *Plasmodium*, often defined by a broad clinical presentation of periodic fevers and chills. The disease is both ancient and prolific, with historical accounts of its signature recurrent fevers reported in nearly every inhabited continent (Cox, 2010). Symptoms often begin as a general malaise, later progressing to fever, which can be accompanied by a range of other sequelae. While prognosis at this stage is generally good, the disease can quickly progress into more severe forms that can lead to convulsions, jaundice, respiratory distress, or even coma (World Health Organization, 2015). If the disease progresses to include neurological symptoms, it is classed as cerebral malaria, which, without treatment, is nearly always fatal (Bartoloni and Zammarchi, 2012).

Malaria has had a profound effect on the human population. Its prevalence and impact throughout human history is so widespread that it is often cited as the strongest known driver of human evolution, with a number of blood-related phenotypes tracing their origins to constitute some form of protection against the disease (Carter and Mendis, 2002; Kwiatkowski, 2005). Despite global efforts to eliminate malaria launched by the World Health Organization (WHO) in the mid 1900s, the disease still affected 229 million people in 2019, and led to an excess of 400,000 deaths (World Health Organization, 2020). While elimination efforts have limited the distribution of the disease to tropical regions, this still places over

40% of the global population at risk of infection, and the majority of deaths occur in children under the age of five living in regions facing extreme poverty (Hay et al., 2004).

The 2020 edition of the World Malaria Report, released annually by the WHO, reflects on the successes of the last 20 years of global malaria control efforts, and highlights a number of outstanding challenges for the future (World Health Organization, 2020). The report estimates that since the turn of the century, over 1.1 billion cases of malaria have been avoided, and mortality rates have fallen steadily (albeit decreasing at a slower rate in recent years). Projections for further reductions in morbidity and mortality are at risk, however. Growing resistance to drugs threatens both parasite and vector control strategies, and the loss of the *pfhrp2/3* locus commonly used for rapid diagnostic testing endangers global surveillance efforts. The recent success of the RTS,S vaccine remains a beacon of hope amongst these challenges, but continued research, both basic and applied, are essential to support vaccination efforts with further transmission prevention and drug treatment (Keating, 2020).

## **1.2 *Plasmodium* parasites and their phylogeny**

### **1.2.1 Apicomplexans: a phylum of obligate parasites**

*Plasmodium* parasites sit within a larger phylum of obligate intracellular parasites known as the Apicomplexa, named as such for the apical shape formed by organisms in this group to facilitate host cell invasion. Apicomplexans evolved from free-living dinoflagellate-like organisms, whose first instance of parasitism was thought to occur upon ingestion by marine worms (O'Donoghue, 2017). Due to this, most apicomplexan parasites have retained the use of the gut epithelium in some form over the course of their life-cycles, often for sexual reproduction (Gibson, 2021). The phylum contains a number of parasitic lineages relevant to human and animal health in addition to *Plasmodium*, including *Toxoplasma* and *Cryptosporidium*.

### 1.2.2 The *Plasmodium* genus

Within the Apicomplexa, the genus *Plasmodium* sits under the order of Haemosporidian parasites (literally meaning 'blood spores'), for which all known hosts are vertebrates. The genus, previously believed to contain only parasites infecting mammalian species, has been since proven to be polyphyletic, containing a number of sub-genera infecting a mixture of primate, mammalian, and avian hosts (Böhme et al., 2018; Galen et al., 2018). In total, there are over 150 known species of *Plasmodium* parasites that infect a wide range of vertebrate hosts, including lizards and birds, in addition to mammals (Garnham, 1966).

Currently, we are aware of six species of *Plasmodium* that can infect humans. Of these, *Plasmodium falciparum* is the most widespread and deadly, followed by *Plasmodium vivax*, which causes most cases of malaria outside of the African continent. Sister taxa to *P. vivax* is *Plasmodium knowlesi*, which is largely believed to primarily infect primates and humans only through zoonoses, but has been listed as a human malaria parasite due to the high number of cases reported, especially in Malaysia (Jeyaprakasam et al., 2020; Singh et al., 2004). Two closely related species of *Plasmodium ovale* (*P. ovale curtisi* and *P. ovale wallikeri*), as well as *Plasmodium malariae* cause disease globally, but to a lesser severity, and are often found in co-infections with *P. falciparum* (Autino et al., 2012; Sutherland et al., 2010).

In research, the rodent malaria *Plasmodium berghei*, first described in thicket rats in what is now the Democratic Republic of the Congo (Vincke et al., 1948), is often used to model *Plasmodium* infection in humans. While its relevancy in modelling human clinical phenotypes has been called into question (White et al., 2010), *P. berghei* is often regarded as an acceptable proxy for transcriptomic analyses as it shares a number of conserved core genes and transcriptional mechanisms with human malaria (Bushell et al., 2017; Howick et al., 2019; Orr et al., 2012; Yeoh et al., 2017)

## 1.3 The human malaria parasite *Plasmodium falciparum*

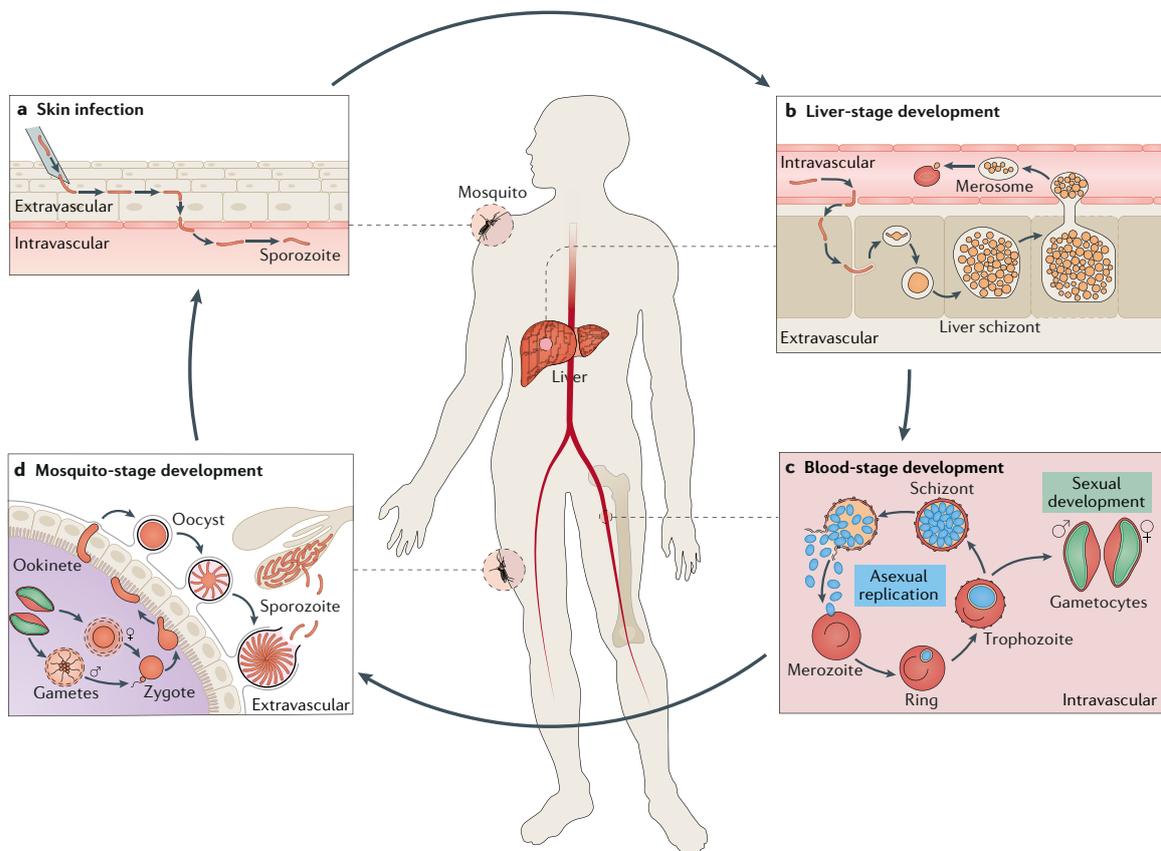
While *P. falciparum* was not the first species of human malaria to be described, it has certainly risen to prominence as the most significant. It is responsible for

the most malaria-related deaths each year, dominating 97% of the infections in the WHO African Region, which constitutes > 90% of total cases globally (World Health Organization, 2020). This is largely attributed to both the pathology of the parasite and the success of its most common vector, *Anopheles gambiae* (Sinka et al., 2020). Due to its resounding impact on global health, *P. falciparum* is the focus of most malaria-related research.

### 1.3.1 Life cycle at a glance

*Plasmodium falciparum* parasites progress through a number of complex stages across their life-cycle, each distinct in both their morphology and function. Despite differences in behaviour, physiology, and distribution of both their hosts and insect vectors, *Plasmodium* species infecting mammals have retained remarkable conservation with which they progress through their developmental cycle (Ngotho et al., 2019). The cycle, summarised in **Fig. 1.1** below, begins when a female mosquito of the genus *Anopheles* takes up a blood meal by injecting her proboscis into the vasculature of a mammalian host (**Fig. 1.1a**). In order to facilitate feeding, the female will inject a salivary cocktail of proteins into her host aimed at, amongst other functions, suppressing immune response to the intrusion and increasing blood flow to the area (Ribeiro and Francischetti, 2003; Vogt et al., 2018). If the female is carrying *Plasmodium*, the parasite will hijack this feeding event, taking advantage of this salivary ejection to send highly motile forms known as sporozoites into the vasculature of the host. The exchange is very quick, with imaging studies demonstrating salivary discharge speeds of 50-500  $\mu\text{L/s}$ , roughly translating to 1-2.5 sporozoites released with each second the mosquito feeds (Frischknecht et al., 2004; Jin et al., 2007).

Within minutes of a feed, hundreds of *Plasmodium* sporozoites flood the host vasculature, where they remain until they reach the host's liver (**Fig. 1.1b**) (Rosenberg et al., 1990). The invasion of host hepatocytes by the sporozoites in the liver precedes an initial round of asexual replication and cellular division, termed schizogony, to form a multi-nucleated schizont containing thousands of individual daughter parasites (Mota et al., 2001). Upon maturation and egress from the hepatocyte, each of these daughter parasites, or merozoites, are released back into the host's bloodstream, where they employ a suite of surface proteins to bind



**Figure 1.1** The life cycle of *Plasmodium* parasites

Reprinted with permission from Springer Nature: Nature Reviews Microbiology (Venugopal et al., 2020), copyright 2020.

to and invade circulating red blood cells (RBCs) (Cowman et al., 2017; Iyer et al., 2007).

This invasion marks the beginning of the intraerythrocytic development cycle (IDC), a sub-cycle within the larger developmental course of the parasite in which, in most *Plasmodium* species, the largest cell numbers are obtained and the majority of clinical malaria symptoms arise (**Fig. 1.1c**) (Venugopal et al., 2020). These cell numbers are achieved through successive rounds of invasion, schizogony, and release; exponentially increasing the number of parasites in the blood with each cycle. The cycle itself comprises of three distinct developmental forms. The formation of ring-stage parasites, named for their biconcave shape, directly follows RBC invasion. These parasites then begin a course of growth and maturation, filling out the ring shape into round trophozoites, feeding off of and modulating the host

cell as they develop (Bannister and Mitchell, 2003). The sub-cycle is concluded with a schizogony event in which DNA replication and nuclear division once again produce schizonts containing daughter merozoites capable of invading new RBCs, perpetuating asexual growth. The release of merozoites from late-stage schizonts is mediated by synchronised cell rupture across many infected RBCs — leading to the classical clinical presentation of cyclic fever in patients suffering from malaria (Hawking, 1970).

Eventual egress from the IDC mediates the continuation of the larger developmental programme of a *Plasmodium* parasite and dispersion into new hosts via another Anopheline vector. During this process, a small number of parasites within the IDC will diverge from the path of schizogony and instead commit to a process of maturation terminating in the production of sexually differentiated male and female forms, called gametocytes. Using a number of mechanisms, these gametocytes gather in vasculature near the surface of the skin, where they wait to be ingested as part of the blood meal of a vector mosquito (Venugopal et al., 2020).

Successful transmission of sexual forms relies on the uptake of both male and female gametocytes by a mosquito, which sets off a flurry of further developmental stages (**Fig. 1.1d**). Almost immediately upon entry, male and female gametocytes activate to form gametes. Highly motile male gametes traverse the blood meal present in the mosquito gut in order to find and mate with a female, forming zygotes. Burying itself into the wall of the mosquito's midgut membrane, these ookinetes then exit the gut epithelium and enter the basal lamina where they develop into oocysts. Oocysts perform many rounds of replication to eventually develop into sacs full of single-celled sporozoites. These sporozoites are released into the haemolymph of the mosquito and some reach the salivary glands, where they wait to be injected during the next time the mosquito takes a blood meal (Matuschewski 2006).

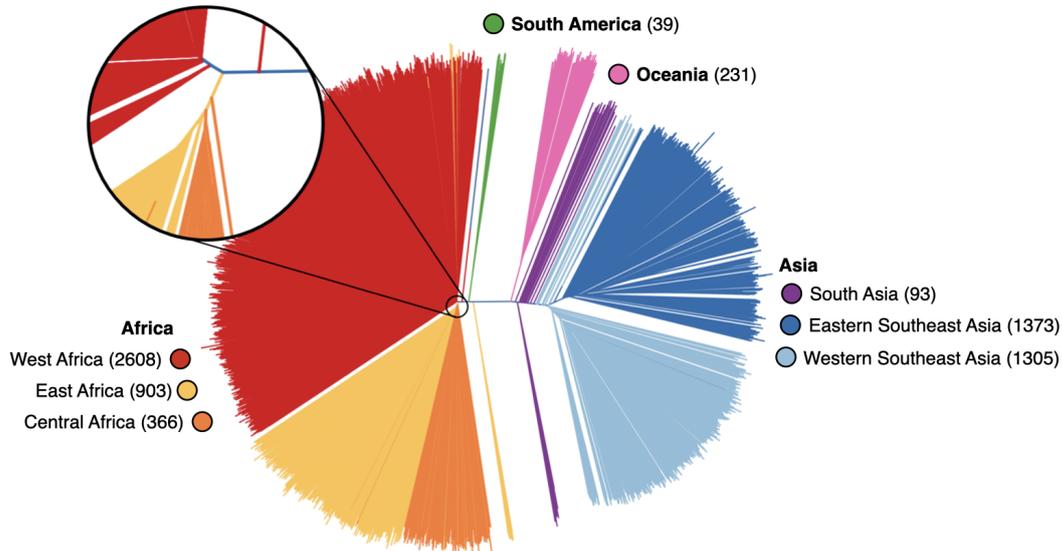
### 1.3.2 Global diversity of natural infections

The obligation for sexual reproduction in the mosquito bears the possibility for genetic 'reshuffling' each time a mosquito takes an infected blood meal. Any resultant heterogeneity can then act as a breeding ground for selection — generating sub-populations best suited to the local environmental context. Well before the

structure of DNA had even been discovered, researchers and clinicians saw this variability manifest in the clinical presentations of their patients, even those believed to be infected with the same species. This led French physician Émile Marchoux to declare the existence of distinct 'groups' within the same species (Marchoux, 1922), forming the basis for the strain theory of malaria a couple of decades later. The origins of strain classification in malaria are reviewed by McKenzie et al. (2008). Originally, strains were grouped on the basis of clinical phenotypes, named for the geographical region the strain was thought (often erroneously) to have originated. Further classifications were added as the theory developed, such as their response to antimalarial treatments or the degree to which they were 'tolerated' by the community in which they were found, later understood to be immunogenic tolerance. The use of strains to describe *P. falciparum* still remains today, but a formal definition of what constitutes a strain has never been properly formed, and the language has largely shifted from 'strain' to 'clone' or 'isolate'. This is likely because the release of the *P. falciparum* genome in 2002 by Gardner et al., and the flurry of population genomics studies it fostered, revealed a greater level of genetic diversity in parasite populations than could be reliably categorised into discrete strains.

Where parasite population structure is easier to define is on a geographical scale. A number of landmark studies, first using Sanger sequencing and later adopting whole genome sequencing (WGS), found that globally, *P. falciparum* forms geographical clusters by continent/region (**Fig. 1.2**), and that this structure can continue to scale locally in areas where transmission (and thus gene flow) are low or sporadic (Anderson et al., 2000; Kidgell et al., 2006; Manske et al., 2012; Neafsey et al., 2008; Volkman et al., 2007). These structures can then be used to follow malaria transmission patterns by grouping samples by phylogenetic relationships or shared variation. This approach forms the basis for genomic surveillance of *P. falciparum*, and has led to innumerable advancements (summarised by Neafsey et al. (2021)) in our understanding of parasite transmission, including the historical distribution of malaria across the globe (Yalcindag et al., 2012), the introduction and spread of drug resistance (Amato et al., 2018), and restraints to diversity (bottlenecks) imposed by malaria control strategies (Manske et al., 2012).

Genomic surveillance is also useful in vaccine development. Like most pathogens caught in a race for survival with their host, much of *P. falciparum*'s variation lies in



**Figure 1.2 Global diversity of 7000 *P. falciparum* samples**

Neighbour-joining tree of 7000 whole genome samples. Figure modified from Pearson et al. (2019). Sample sizes for each region are noted in parentheses.

surface antigens that are exported to the host RBC cell membrane (Volkman et al., 2007), complicating a 'one size fits all' approach to developing a global malaria vaccine (Ouattara et al., 2015). Further complicating these efforts is the ability of the parasite to switch the surface antigen PfEMP1 through the conditional expression of one of 60 known *var* genes (Scherf et al., 1998). Continued surveillance of *P. falciparum* diversity both regionally and globally can help to identify regions of the genome with immunogenic potential that are more conserved across populations.

### 1.3.3 Strains used in laboratory practice

The word 'strain' is still commonly used to refer to distinct, clonal (or presumed clonal) parasite lines used in laboratory settings. The development of a protocol to isolate parasites from natural infections and propagate them into continuous asexual *in vitro* cultures revolutionised malaria research (Trager and Jensen, 1976). Perhaps the most commonly used strains are NF54, which was isolated in the Netherlands in 1981 (Ponnudurai et al., 1981) but believed to have originated from Africa (Preston et al., 2014), and its clone, 3D7. Other popular strains, a number of which are summarised in **Table 1.1**, have been isolated from *P. falciparum* infections around the world, and serve as standardised proxies for global

parasite diversity outside of natural settings. What's more, many isolates and their clones were adapted to culture before the widespread use of modern antimalarial drugs, and thus serve as useful controls for assessing sensitivities and elucidating mechanisms of drug action (van Schalkwyk et al., 2013). The spatiotemporal separation represented by these strains has also led to molecular insights regarding the mechanisms of surface antigen switching and differential metabolic requirements between parasites, which can inform both drug and vaccine design (Awandare et al., 2018; Ke et al., 2011).

**Table 1.1 Origins of select *P. falciparum* clones/strains often used in laboratory culture**

Strain/Clone	Origin	Region	Year first reported	Reference
<b>NF54</b>	Netherlands/Africa	Europe/Africa	1981	Ponnudurai et al.
<b>3D7</b> (NF54)	Netherlands/Africa	Europe/Africa	1987	Walliker et al.
<b>FCR3</b>	The Gambia	Africa	1981	Jensen et al.
<b>K1</b>	Thailand	Asia	1981	Thaithong and Beale
<b>D10</b> (FC27)	Papua New Guinea	Oceania	1983	Anders et al.
<b>IT</b> (Ituxi084)	Brazil	South America	1983	Udeinya et al.
<b>HB3</b> (Honduras I/CDC)	Honduras	Central America	1984	Bhasin and Trager
<b>7G8</b> (IMTM22)	Brazil	South America	1984	Burkot et al.
<b>W2</b> (Indo III/CDC)	Indochina	Asia	1988	Oduola et al.
<b>DD2</b> (W2mef)	Indochina	Asia	1988	Wellems et al.
<b>GB4</b> (Ghana III/CDC)	Ghana	Africa	2003	Sullivan et al.
<b>MRA-1241</b>	Cambodia	Asia	2011	Witkowski et al.
<b>NF135.C10</b>	Cambodia	Asia	2013	Teirlinck et al.
<b>NF166.C8</b>	Guinea	Africa	2017	McCall et al.

Adapted from van Schalkwyk et al. (2013). Parent lines from which strains were cloned denoted in parentheses

Despite representing a narrow snapshot of evolutionary time, there is good evidence to suggest the use of these strains as models for natural infections is valid. Before Manske et al. (2012) performed the first global sample of natural diversity via WGS, many selected lab strains as geographical representatives (Neafsey et al., 2008; Volkman et al., 2007) for single nucleotide polymorphism (SNP) genotyping. Another experiment used strains NF54, MRA-1241, and 7G8 to model parasite populations from Africa, Asia, and the Americas to show that these geographically distinct populations likely co-evolved with the *Anopheles* species specific to their region, evading complement-mediated immunity in the mosquito (Molina-Cruz et al., 2015). Lab strains are even used to test vaccine efficacy in different genetic

backgrounds via controlled human malaria infections (CHMI) (Moser et al., 2020), after they were found by Jensen et al. (1981) to remain infectious to humans, and NF54 currently acts as the genetic background for the PfSPZ candidate vaccine (Walk et al., 2017).

The 'catch', when it comes to using lab strains, lies in culture adaptation. A study subjecting recently cloned clinical isolates from The Gambia to 48 days of continuous *in vitro* culture revealed that cultures gained a number of loss-of-function SNPs upon adapting to *in vitro* culture, and that these SNPs were also present in many of the strains listed in **Table 1.1** that have been in use since the 1980s (Claessens et al., 2017).

### 1.3.4 The *Plasmodium falciparum* Genome

The most complete, annotated, and curated whole genome assembly for *Plasmodium falciparum* is of the 3D7 laboratory clone, and its production would not have been possible without the quantities of DNA afforded by large-scale continuous *in vitro* culturing developed by Trager and Jensen in 1976. The full genome, released in 2002 by Gardner et al., contains 23.33Mbp distributed across 14 chromosomes, and includes an apicoplast and mitochondrial genome in addition to the nuclear one. The genome harbours a strong base composition bias, with A and T comprising 80.6% of bases in protein coding regions, and over 90% outside of these regions (Hamilton et al., 2017). Manually annotated and updated regularly via geneDB ([www.geneDB.org](http://www.geneDB.org)), the current 3D7 genome contains 5720 annotated genes, of which 5318 are protein coding. The full genome's release facilitated a boom of functional annotation over the next decade, bringing the proportion of genes with unknown functions from 67% to 33% (Böhme et al., 2019). Still, this leaves a third of the genes in the genome for which we lack any functional annotation, highlighting the need for further transcriptional and functional studies.

## 1.4 Malaria parasites, in the age of transcriptomics

The genomic repertoire of *P. falciparum*, along with the other members of the Apicomplexa, has been shaped by millions of years of evolution, perhaps the

most dramatic of which being the transition to obligate parasitism by the phylum's ancestor (White and Suvorova, 2018). It is estimated that this transition led to the shedding of over 4000 ancestral genes relating to free-living growth deemed unnecessary for a parasitic lifestyle (Woo et al., 2015). These losses were then offset by the expansion and adaptation of genes relating to host-parasitism, including an expansion in extracellular proteins in direct contact with the host, cytoskeletal proteins to facilitate host-cell invasion, and DNA/RNA binding proteins to regulate these processes (Balaji et al., 2005; Woo et al., 2015). The genes utilised by *P. falciparum* have thus been carefully curated over hundreds of thousands of years of successful parasitism. Understanding when and where these genes are used provides a road map researchers can use to reconstruct, and ultimately disrupt, the events leading to malaria infection.

### 1.4.1 Global patterns of expression

In the years that followed the release of the *P. falciparum* genome, numerous studies have facilitated the refinement of gene model annotation by profiling splice junctions, transcriptional start sites (TSS), expression outside coding regions, defining untranslated regions, and isoforms (Adjalley et al., 2016; Chappell et al., 2020; Kensche et al., 2016; Otto et al., 2010; Siegel et al., 2014; Sorber et al., 2011). The latest estimates predict that expression can be reliably detected in 89% of the genome, and that 78% originate from coding regions (Chappell et al., 2020). Improved accuracy in defining transcriptional units by Chappell et al. demonstrated that the genome of *P. falciparum* is remarkably compact, with very few regions that do not display some sort of function relating to transcription or transcriptional control.

The expression of coding genes has been best defined in the intra-erythrocytic development cycle (IDC). The earliest studies profiled RNA expression using cDNA microarrays, demonstrating a 'cascade' of expression over the course of the IDC (Bozdech et al., 2003; Le Roch et al., 2003), and that these cascades were largely conserved across species and strains (Bozdech et al., 2008; Llinás et al., 2006). In many cases, genes were found to be transcribed just before the protein they encoded was needed, leading to the description of the cascade as 'just-in-time' transcription (Bozdech et al., 2003; Le Roch et al., 2004). This pattern was found

to be relatively robust to environmental perturbations due to drug treatments (Ganesan et al., 2008; Natalang et al., 2008), leading to the belief that global transcription across the IDC is 'hard-wired'. While globally this may be the case, studies have detected 'wobble-room' in the tightly-regulated cascade (Hu et al., 2010), one of which being in response to febrile temperatures (Oakley et al., 2007; Rawat et al., 2021). Recently, this has led to the discovery of a transcription factor responsible for mediating the response (Tintó-Font et al., 2021), lending credence to the hypothesis that even 'hard-wired' transcriptional mechanisms can still sense, respond and adapt to environmental changes.

### 1.4.2 Mechanisms for regulating transcription

Transcription in malaria parasites is largely carried out by RNA polymerase II (Pol II), which, like in many species, facilitates transcription by recruiting a number of proteins required for transcript generation and processing via its C-terminal domain (CTD) (Chapman et al., 2008; Ukaegbu and Deitsch, 2015). For the most part, transcription follows the same known processes as in other eukaryotes: a protein coding gene is flanked by 3' and 5' untranslated regions (UTRs), often contains efficiently spliced introns, and is regulated via an upstream transcription start site (TSS) (Hughes et al., 2010). Most regions contain an upstream core promoter whose function can be modulated by enhancers (Crabb and Cowman, 1996). The exact mechanisms by which transcription is regulated in *Plasmodium* are not fully understood, but transcription factors and epigenetic control have been demonstrated to both play an important role.

#### Transcription factors

Transcription factors regulate gene expression through DNA binding in a sequence-specific manner, often enhancing expression by recruiting RNA polymerases to the site (or in rarer cases, blocking this recruitment to act as a repressor). In many cases, transcription factors can be combined to add complexity to a regulatory process. A stark lack of these factors in *P. falciparum* were first noted when the genome was sequenced in 2002 when compared to *S. cerevisiae*, an organism with a similar genome size and gene density (Gardner et al., 2002; Templeton et al.,

2004). Attempts to use sequence homology to other eukaryotic transcription factors discovered few candidates in the *P. falciparum* genome, until the discovery of a family of proteins containing apetala2 (AP2) integrase binding domains (Balaji et al., 2005). This family, homologous to apetala2/ethylene response factors in plants, represents a lineage-specific expansion in the Apicomplexans, leading to the name ApiAP2s. Currently there are 27 known ApiAP2s, and while they are thought to act as the primary class of regulatory proteins in *Plasmodium*, not all of them have been functionally annotated (Toenhake and Bártfai, 2019). Roughly half have been deemed essential for growth in the IDC in *Plasmodium* (Toenhake and Bártfai, 2019), and many show stage specific expression (Flueck et al., 2010; Iwanaga et al., 2012; Kafsack et al., 2014; Modrzynska et al., 2017; Sinha et al., 2014; Yuda et al., 2010). Other transcription factors have been found in addition to the ApiAP2s, with proteins containing HTH, KH-domain, ZnF-C2H2 and  $\beta$ -scaffold binding domains (Toenhake and Bártfai, 2019).

### Epigenetic control

Epigenetic modifications are another key aspect of gene expression regulation in *Plasmodium*, and likely act in concert with transcription factors to exert complex levels of control (Read et al., 2019). The use of Hi-C chromatin capture has revealed the nucleosome landscape across development to be dynamic and stage-specific (Bunnik et al., 2018; Hollin and Le Roch, 2020; Lemieux et al., 2013), for example compacted nuclei lead to lower transcription during the ring and late schizont stages of the IDC (Ay et al., 2014; Ponts et al., 2010), and more open conformation allowed for the active transcription and growth of trophozoites (Ay et al., 2014). The nucleosome of *P. falciparum* resembles that of eukaryotes, but the parasites also contain histone variants (H2A.Z, H2B.Z, H3.3, CenH3) that can be swapped with the canonical core histones (Hollin and Le Roch, 2020; Hughes et al., 2010; Kensche et al., 2016). These histones can then be modified via posttranslational modification (PTM) of their N-terminal tails. Trimethylation of lysines 9 or 36 on H3 (H3K9me3 or H3K36me3) are the most likely to be associated with heterochromatin and transcriptional repression, and that acetylation of lysine 9 or trimethylation of lysine 4 (H3K9ac or H3K4me3) are most commonly associated with euchromatin and active promoters (Hollin and Le Roch, 2020).

Histone-modifying proteins are responsible for the bulk of PTM of histones and chromatin remodelling, and a large number are predicted to be encoded in the *P. falciparum* genome (Doerig et al., 2015). Of these, PfHP1 (Heterochromatin protein 1) plays a critical role in silencing the expression of multi-gene families, such as *var* genes, and genes required for gametocytogenesis (Brancucci et al., 2014; Filarsky et al., 2018; Flueck et al., 2009; Pérez-Toledo et al., 2009). Many of these proteins form larger protein complexes with histones, and histone pull down experiments coupled with mass spectrometry have identified many contain transcription factors, highlighting the cooperativity involved in all aspects of transcriptional modification in *Plasmodium* (Hoeijmakers et al., 2019; Miao et al., 2021; von Grüning et al., 2021).

### 1.4.3 Non-coding transcription

Perhaps the most intriguing mechanism for regulating gene expression in *P. falciparum* lies in its non-coding transcriptome. By definition, a non-coding transcript is one of any size that lacks evidence for coding potential. They are broadly classed into small non-coding RNAs (sncRNA), containing transcripts smaller than 200bp, and long non-coding RNAs (lncRNA), with transcripts larger than 200bp. From there, they are often further divided by their genomic location, secondary structure, or function. Like mRNA transcripts, lncRNAs can be spliced, capped, polyadenylated, and are often transcribed by Pol II (Krishnan and Mishra, 2014; Militello et al., 2005; Statello et al., 2021). They tend to be subject to faster rates of transcript decay, greater tissue specificity, and lower expression as compared to coding transcripts (Ransohoff et al., 2018). In humans, where non-coding transcripts are best described, many lncRNAs are essential regulators of normal gene expression, with disruptions in these regions implicated in a number of disease states including cancer (Slack and Chinnaiyan, 2019), Alzheimers Disease (Li et al., 2021), and cardiovascular diseases (Lorenzen and Thum, 2016). The mechanisms by which lncRNA mediate gene expression in eukaryotes are vast — to list them would be to describe most known mechanisms of transcriptional regulation. Transcriptional repression through chromatin modifications is one of the most common mechanisms (Beck et al., 2016). Indeed, many lncRNA have been noted to act as repressors of gene expression rather than activators, but this is in no way exclusive (Krishnan and Mishra, 2014).

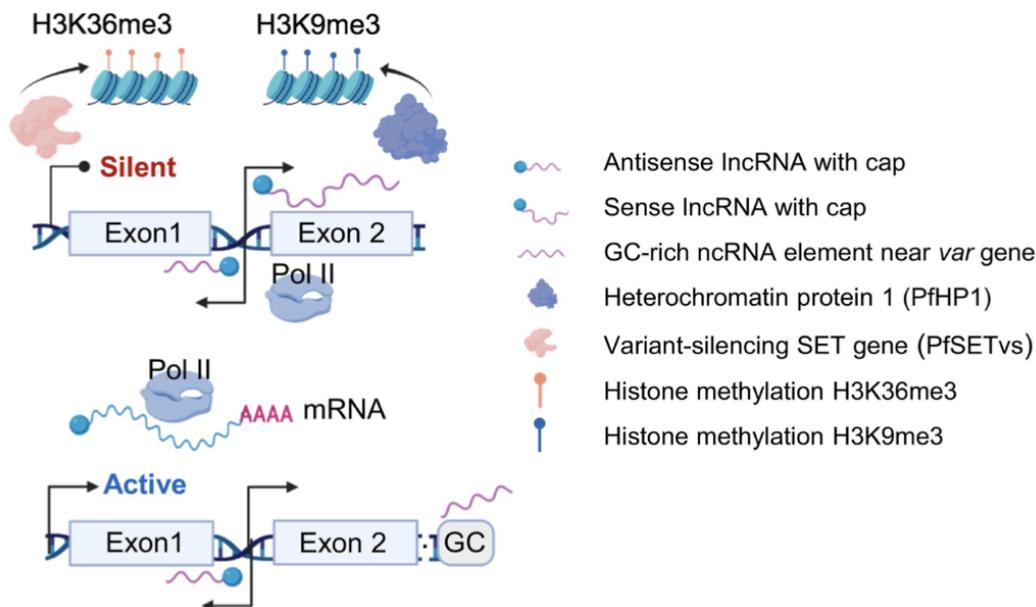
It has been postulated that non-coding RNAs could help explain the 'regulatory gap' between *P. falciparum*'s tightly controlled expression cascades and relatively low abundance of transcription factors, by regulating the transcription factors themselves or the sites they target (Gardner et al., 2002; Vembar et al., 2014). Early explorations using both tiling arrays and stranded RNAseq detected an abundance of natural antisense transcripts, or NATs, expressed in *P. falciparum*, often in a stage-specific manner, and that this occurred both in laboratory settings as well as natural infections (Broadbent et al., 2015, 2011; López-Barragán et al., 2011; Raabe et al., 2010; Siegel et al., 2014; Subudhi et al., 2014). NATs are lncRNAs defined as being antisense to a coding gene with either a complete or partial overlap, and often transcribed by Pol II. Their observed abundance in *P. falciparum* led Vembar et al. to wonder whether this represented an alternative form of gene regulation in the absence of RNAi machinery, as it has in other species, notably in budding yeast (*Saccharomyces cerevisiae*). While the evidence remains somewhat controversial, Alcid and Tsukiyama (2016) compared the non-coding transcriptome of six species of budding yeast and found those that lacked RNAi machinery had the greatest expansion in antisense lncRNA usage. Subsequent studies have suggested this expansion may be due to the absence of RNAi-mediated clearance of antisense transcripts by nucleases DICER and Xrn1, especially when these transcripts pair with mRNA to form dsRNA duplexes (Szachnowski et al., 2019; Yu et al., 2014).

Efforts to profile the non-coding transcriptome of malaria parasites have so far led to three well-characterised examples of non-coding regulation in *P. falciparum*: var gene silencing, regulation of sexual commitment via *gdv1*, and telomere maintenance; the mechanisms for the former two are summarised below. Beyond these, the remaining ncRNAs discovered have been recently called into question by Chappell et al. (2020). In this study, the authors develop an RNAseq protocol that circumvents the need for PCR amplification, due to the known under-representation of regions with extreme AT-enrichment in amplified libraries Kozarewa et al. (2009). This approach allowed for much more accurate clarification of the boundaries of the UTRs (untranslated regions) of coding genes, and found that most of the regions previously identified as putative ncRNA regions were simply yet-unannotated UTRs, highlighting the need for an updated refinement to what we know about the annotated ncRNAs remaining in *P. falciparum* and where they are expressed.

### ***var* gene regulation**

*var* gene expression and silencing plays an essential role in *P. falciparum*'s pathology and virulence. The family is composed of 60 *var* genes, only one of which is expressed, while the rest are transcriptionally silent. LncRNA have been implicated in both *var* gene silencing and active expression, but the mechanisms underlying these processes are complex and not yet clearly understood (Li et al., 2020). As shown in **Fig. 1.3**, each *var* gene is comprised of two exons, one which is conserved (Exon 2), and one which is variable (Exon 1), separated by a conserved intron. The region is thought to contain two promoters, one upstream of the first exon that regulates the *var* mRNA expression, and the other within the conserved intron which can act bidirectionally (Li et al., 2020). The intronic promoter is believed to produce two non-coding transcripts, one that lies antisense to Exon 1, and another sense transcript (often termed the 'sterile transcript') that contains only Exon 2 (Epp et al., 2009). In general, the sense transcript is thought to act to repress *var* gene expression, potentially through the recruitment of the methyltransferase PfSET2 (also called PfSETVs) via Pol II (Jiang et al., 2013; Ukaegbu et al., 2014). The antisense transcript is generally associated with an increase in *var* expression — Amit-Avraham et al. (2015) and Jiang et al. (2013) found active *var* gene expression to be correlated with that of the antisense transcript, and that its disruption led to decreased expression of this *var* gene and subsequent *var* gene switching, and Jing et al. (2018) demonstrated that by artificially expressing antisense transcripts, the mutually exclusive expression of *var* genes was broken down. It is believed this bidirectional promoter is on in both silent and active *var* loci, but its activity may be asynchronous, as conflicting studies have detected its antisense transcript in both states (Epp et al., 2009), or, more convincingly, just in the active state (Amit-Avraham et al., 2015; Ukaegbu et al., 2014), and the sense transcript has been found in both states but in a stage-specific manner (Ukaegbu et al., 2014). The role of intronic ncRNA transcripts in *var* gene regulation is likely part of a larger picture that has yet to be properly elucidated. Bryant et al. (2017) found that removal of the *var2csa* intron did not affect silencing of the gene in non-ring stages, but did increase expression in ring stages. Additional studies have implicated a whole new type of lncRNA, one that is expressed from GC-rich regions located near *var* gene regions, in *var* gene activation, and one of the first examples of a ncRNA

to act *in trans* to regulate expression (Barcons-Simon et al., 2020; Guizetti et al., 2016).



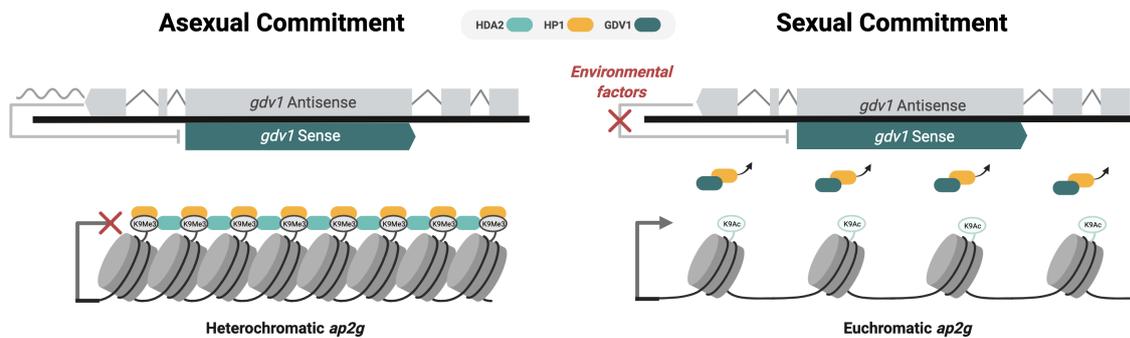
**Figure 1.3 Mechanisms of lncRNA-mediated *var* gene silencing**

The *var* gene intron, the lncRNAs transcribed from it, and histone modifications, have been associated with *var* gene silencing. The antisense lncRNA transcribed from the intron and ncRNA from a proximal GC-rich element of internally located *var* genes have been associated with *var* gene activation. Source: Li et al. (2020)

### GDV1 antagonism

GDV1, as will be discussed in a later section, is a critical modulator of sexual commitment and early gametocyte development in *Plasmodium* (Eksi et al., 2012). Follow up study implicated an antisense transcript, now named *asgdv1*, in the gene's mechanism of action (Filarsky et al., 2018). The transcript, originally detected by Broadbent et al. (2011), contains a number of putative introns, the largest of which completely overlaps the gene body of *gdv1*. GDV1 promotes sexual differentiation by evicting heterochromatin protein 1 (HP1) from H3K9me3 on heterochromatin, one of which being the *ap2g* locus, thereby allowing for *ap2g* to be expressed (Fig. 1.4). *ap2-g*, which will also be discussed in a later section, is the master regulator of sexual commitment in *Plasmodium* (Kafsack et al., 2014). Production of the antisense transcript *asgdv1* however, negatively inhibits *gdv1* expression and HP1

silencing prevents sexual commitment. Filarsky et al. also found this relationship to be choline-responsive — the addition of choline, a repressor of gametocyte commitment (Brancucci et al., 2017), showed a dose-dependent increase in GDV1 expression.



**Figure 1.4 Regulation of sexual commitment by antisense *gdv1* transcript**

GDV1 promotes *ap2g* acetylation and expression by evicting HP1 on heterochromatin, and this process is negatively regulated by expression of *asgdv1*, a lncRNA antisense to the *gdv1* locus. This regulation can be disrupted in the presence of choline, promoting sexual conversion. Adapted from Rea et al. (2018)

#### 1.4.4 Transcription at a single-cell level

Much of what is known regarding gene expression in malaria, as highlighted above, has been gathered by capturing, sequencing, and quantifying RNA from parasites. The process of reconstructing the transcriptional landscape present within a cell at the point of capture, otherwise known as the transcriptome, is referred to as 'transcriptomics'. Until recently, because the RNA content contained within a single cell is insufficient to meet minimum requirements for sequencing, 'representative' transcriptomes were built by pooling the RNA from a population of cells, garnering the name 'bulk RNAseq'. Single-cell RNA sequencing, as the name suggests, overcomes the limitations of bulk sequencing, and facilitates the generation of a single transcriptome to represent the activity of a single cell, multiplied over a population (Aldridge and Teichmann, 2020).

### **Capturing single cells**

Since the first application of single-cell RNAseq (scRNAseq) (Tang et al., 2009), a variety of protocols have been developed to isolate, capture, and sequence the RNA in individual cells (reviewed in Chappell et al. (2018)). Many protocols share an overarching workflow: isolate the contents of a cell, mark those contents as belonging to that cell, pool the contents from many cells, and sequence. Among the most common techniques are Smart-seq2 (Picelli et al., 2014), which isolates and tags cell contents within a single tube or well in a plate, and 10X Chromium (Zheng et al., 2017), which uses microfluidics to separate cells and isolate their contents within an enclosed droplet. The latter is the technology used in this thesis, thus an overview of the protocol is described below.

### **Overview of 10X Chromium cell capture**

Using 10X Chromium, single cell isolation is achieved by passing cells through an 8-well microfluidic chip (the protocol, as paraphrased below, is described in detail in Zheng et al. (2017), as well as on the 10X Genomics website). As cells are processed through the chip, they are encapsulated within an oil droplet, along with a reverse transcription (RT) master mix and a barcoded gel bead. Each bead is tagged with oligonucleotides that contain, amongst sequencing primers and adaptors, a unique 16bp cell barcode, a 10-12bp (depending on chemistry) unique molecular identifier (UMI), and an anchored 30bp oligo-dT. While the cell barcode sequence is common to all oligonucleotides tethered to a bead, the UMI is random, and unique to each oligonucleotide. A gel bead successfully captured within an oil droplet is called a Gel bead in EMulsion, or a GEM. For cells that get encapsulated in a GEM, lysis begins immediately. Within the safe confines of the oil droplet, gel beads disintegrate and release their oligonucleotides, which capture polyadenylated transcripts from the lysed cell via the oligo-dT primer. A subsequent RT reaction thus results in cDNA that can be traced back to a GEM (via the cell barcode; as long as the GEM contained a single cell, GEM and cell are synonymous), and a single RNA molecule (via the UMI). Once amplified, pooled and sequenced, a transcriptome can be reconstructed for each cell in the form of a cell x gene matrix; where each row contains only reads containing its unique cell barcode, and the columns populated

by the number of UMIs mapped to each gene. Thousands of cells can be profiled this way within a single run.

### scRNAseq of malaria parasites

Single-cell transcriptomics has enriched our understanding of gene expression in malaria parasites. The technology's benefit, amongst others, is in the heterogeneity it captures between cells that are otherwise lost by 'averaging' a population during bulk RNAseq (Nötzel et al., 2018). In malaria parasites this has allowed, for example, for the discovery of distinct developmental trajectories underlying sexual development (Bancells et al., 2019), as well as the detection of sub-populations that form under temperature stress (Rawat et al., 2021), during host-cell invasion (Hentzschel et al., 2021), or in response to differing mosquito environments (Witmer et al., 2021).

ScRNAseq has also been used to update and expand upon the profiles of global gene expression underlying developmental stages using bulk approaches. The Malaria Cell Atlas project now contains data from three published single cell studies, and profiles gene expression as parasites progress through nearly every stage of their development (Howick et al., 2019; Real et al., 2021; Reid et al., 2018). Missing from this project, however, is a comprehensive portrait of *P. falciparum*'s sexual development pathway.

## 1.5 Gametocytogenesis

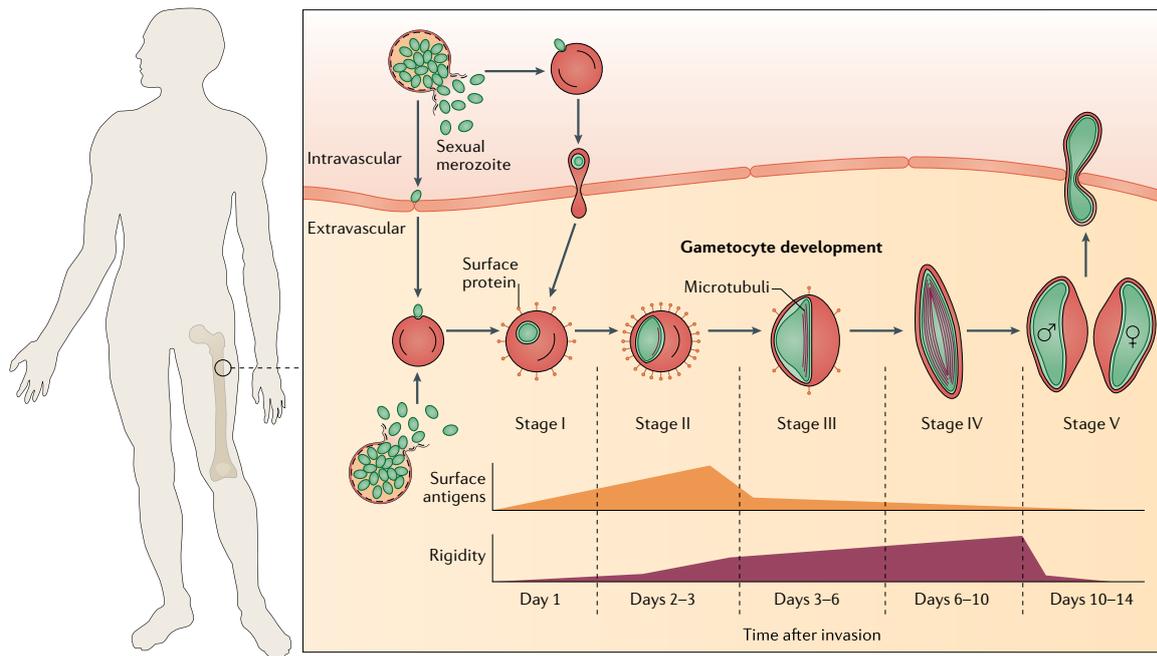
Whilst canonical life-cycle descriptions of *Plasmodium* parasites often start at the initial infectious bite and parasite inoculum of a new host, one could also make a case that the 'beginning' of the ongoing life cycle sits at the point of egress from asexual development into sexual development during the blood stages. The point at which these processes diverge demarcate two overarching goals underlying the parasite's biology; proliferation within a host vs. transmission between hosts. It also represents an important demarcation in our current strategies to combat malaria infection. Most commonly used antimalarials are designed to target the asexual blood stages, as they are the cause of much of the disease and morbidity associated

with malaria. Strategies aimed at instead disrupting the transmission of infection to a novel host are known as 'transmission-blocking' (Sinden, 2017). Transmission-blocking strategies have been identified as a crucial step in malaria eradication and further drug and vaccine interventions (Challenger et al., 2021; Delves et al., 2018; Rabinovich et al., 2017), and improved understanding of gametocyte biology can aid in their development.

### 1.5.1 The stages of sexual development

The sexual development of *Plasmodium falciparum*, which takes place over the course of 12-14 days, is often described in terms of its five distinct morphological stages (**Fig. 1.5**). The first stage is nearly morphologically indistinguishable from asexual trophozoites, identified originally only through the lack of knob formations in the RBC membrane that typify asexual trophozoites (Sinden, 1982), and the slightly darker collection of pigment (Baker, 2010). Later studies have gone on to identify a number of early markers of gametocyte differentiation, and it is only through the expression of these markers that stage I gametocytes can be reliably identified (Carter et al., 1989; Schneider et al., 2004). By stage II, the formation of the IMC, or inner membrane complex, is initiated, and microtubule formation begins to pinch the shape of the parasite into a distinctive lemon shape, allowing it to be distinguished from asexual parasites and stage I gametocytes (Sinden, 1982). IMC and spindle formation continues to flatten the shape of the parasites into stage III. The formation of a flat edge along one side of the parasite produces a distinctive 'D-shape' (Dixon et al., 2012). The final two stages of sexual development encompass the distinctive crescent or banana-shaped parasites; earning the 'falciform' namesake of *P. falciparum* (named after the latin falx, for 'sickle' or 'curved'). Microtubule elongation renders stage IV parasites completely tubular with pointed ends, filling up over 50% of the original RBC volume, leaving the only remaining unfilled RBC membrane space as the flattened Laveran's bib (Dixon et al., 2012; Ngotho et al., 2019; Sinden, 1982). Once fully mature, the final stage V parasites are identified by the rounding of the pointed edges and slight bend into the distinctive crescent shape.

Only mature (stage V) *P. falciparum* gametocytes can be detected in circulation. The early stages (stage I-IV) sequester out of circulation and are enriched in the



**Figure 1.5 Stages of sexual development in *P. falciparum***

Reprinted with permission from Springer Nature: Nature Reviews Microbiology (Venugopal et al., 2020), copyright 2020.

host's haematopoietic niche in the bone marrow (Aguilar et al., 2014; Farfour et al., 2012; Joice et al., 2014). Invasion of this niche is thought to be facilitated by the expression of proteins exported to the RBC surface, such as GEXP07 and GEXP10 (Dantzler et al., 2019; Silvestrini et al., 2010; Tibúrcio et al., 2012). Gametocytes then develop within this niche for ~2-10 days. Over this time, cell rigidity increases, and this is thought to prevent the premature clearance of the immature RBC (iRBC) in which the gametocyte is developing by the bone marrow (Dixon et al., 2012; Peatey et al., 2013; Venugopal et al., 2020). This rigidity is rapidly 'switched off' in mature parasites (stage V) to facilitate their re-entry into circulation (Dearnley et al., 2016; Tibúrcio et al., 2012).

### 1.5.2 Sexual commitment: when and why?

The move towards sexual differentiation brings about a permanency that is not present in the asexual cycle; whereas asexual parasites always have the potential to undergo gametocytogenesis in subsequent rounds of development, sexual differentiation represents a terminal and irreversible trajectory towards sexual maturity.

In all well-studied species of *Plasmodium* in mammals, this process takes longer than asexual development, and as such represents a relatively costly investment of resources, especially if not picked up by a mosquito (Gautret and Motard, 1999). With a 12-14 day developmental time for sexual development as compared to a 48hr IDC, this problem is especially pertinent in *P. falciparum* parasites. As a result, investment in sexual development is usually kept remarkably low, estimated to be ~1% in vivo (Cao et al., 2019; Eichner et al., 2001).

### **Environmental factors influencing commitment**

The commitment to transmission is thus a molecular decision that is not made lightly by a *Plasmodium* parasite, instead likely made dynamically amidst a careful balance of environmental stimuli (Carter et al., 2013; Reece et al., 2009). It is often hypothesised these stimuli signal stress or flag the potential for worsening conditions within a host, and prompt sexual conversion in order for the parasite to prioritise transmission (Greischar et al., 2016; Koella and Antia, 1995; Meibalan and Marti, 2017), however given the time delay between commitment and maturity in *P. falciparum*, this may not be the only motivation for adjusting commitment (Neveu et al., 2020). Environmental shifts may signal, on a more local scale, the presence or absence of conditions that would favour sexual development over asexual growth, or vice versa (Venugopal et al., 2020).

The parasite's ability to detect environmental cues has been shown to be mediated by the release of EVs, extracellular vesicles, that probe the environment and mediate communication between parasites upon encountering stress (Mantel et al., 2013; Regev-Rudzki et al., 2013). The most studied environmental effector of sexual commitment is the lipid lysophosphatidylcholine (LysoPC); a host-derived fat molecule involved in the synthesis of membrane component phosphatidylcholine. Low levels of LysoPC act as a signal of host nutrient-deficiency and have been shown to stimulate production of *P. falciparum* gametocytes *in vitro* (Brancucci et al., 2017). Other environmental effectors of sexual commitment include the host's genetic background (Robert et al., 1996), host pathology like fever (Usui et al., 2019), anaemia (Bruce et al., 1990) or stress hormones (Maswoswe et al., 1985), parasite density or local competition (Reece et al., 2005; Talman et al., 2004), and antimalarial drug interventions (Buckling et al., 1999; Ono et al., 1993).

### 1.5.3 Regulating sexual commitment

#### The sexual switch

Despite the identification of gametocytes by Laveran in the late 1800s, and the release of the genome sequence in 2002, the genetic mechanisms underlying early gametocyte commitment and regulation remained largely unknown until the early 2010's. Early identification of key regulators relied on the loss of gametocyte-forming capacity often following successive blood-passages *in vitro*, leading to the discovery of a 30kb deletion present in strains where gametocyte-forming capacity was lost. Reinstatement of the gene contained within this region in gametocyte-deficient parasites was sufficient to restore the formation of sexual stages, and the protein was subsequently named GDV-1, gametocyte development protein 1 (Eksi et al., 2012). Later, a similar approach was used to identify the master regulator of sexual development as the transcription factor AP2-G (Kafsack et al., 2014; Sinha et al., 2014). Recent evidence suggests the regulation of sexual development itself is epigenetically mediated. Throughout the asexual cycle, the *ap2-g* locus exists in an epigenetically repressed state via histone H3 lysine9 trimethylation (H3K9me3), heterochromatin protein 1 (HP1), and histone deacetylase (HDA2) (Brancucci et al., 2014) (**Fig. 1.4**). GDV-1 can remove HP1 from *ap2-g*, thus de-repressing (activating) sexual development (Filarsky et al., 2018). Interestingly, while HP1 silencing of *ap2-g* is highly conserved across *P. falciparum*, *P. berghei*, *P. yoelii*, *P. chabaudi*, *P. knowlesi*, and *P. vivax* (Fraschka et al., 2018), the *gdv1* locus has been lost in rodent *Plasmodium*, hinting at a secondary pathway to regulate *ap2-g* expression in this lineage (Josling et al., 2018).

The activation and expression of AP2-G activates a cascade of downstream events. Following an initial increase in *ap2-g* expression, the locus is re-repressed, potentially mediated by concurrent increases in the expression in ISWI, SNF2L, HP1, and HDA2 and other SWI/SNF helicases (Poran et al., 2017; van Biljon et al., 2019). Following this re-repression is a putative 'check-point', in which *ap2-g* expression is dependent on transcript stabilisation, potentially mediated by PF3D7\_1222400, an ApiAP2 transcription factor often lost during culture adaptation. Past the 'check-point', *ap2-g* expression once again rises in a self-mediated positive-feedback loop, initiating nearly all the mechanisms required for early sexual development and down-regulating those specific to asexual growth (Josling et al., 2020). These

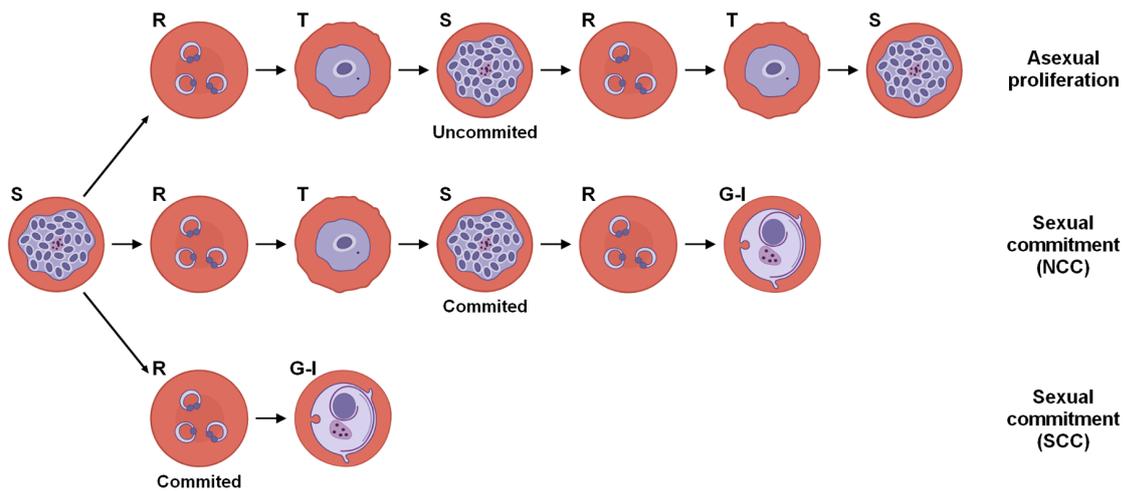
earliest genes include well-known early gametocyte markers *pfs16*, *pfg14.744*, *pfg27*, invasion markers such as SERA proteases, *msrp1*, *maf1*, exported proteins *gexp02* and *gexp05*, and histone modulators *lsd2* and *sir2a* (Josling et al., 2020, 2018; Poran et al., 2017; van Biljon et al., 2019).

### The timing of commitment

Observations from RBC monolayer plaque assays formed the basis for the long-held belief that sexual commitment occurred during the schizont stage, in which all progeny of a single schizont would develop sexually or asexually, depending on the commitment state of the parental schizont, and would never mix (Bruce et al., 1990; Inselburg, 1983). Similarly, all progeny of a sexually-committed schizont would develop into either males or females, aligning the time of sexual commitment with the timing of sex determination (Smith et al., 2000). Kafsack et al. also, upon the discovery of the *ap2-g* locus as a master sexual regulator, observed an ‘all-or-none’ pattern in *ap2-g*-tagged fluorescence upon release from *ap2-g+* or *ap2-g-* schizonts, respectively (Kafsack et al., 2014), lending further credence to this theory. A more recent application of parasite plaque assays have challenged this theory, observing a mixture of sexual and asexual progeny from the schizont at rates up to 40% (Bancells et al., 2019).

The advent of single-cell RNA sequencing (scRNAseq), however, demonstrated the existence of sexual commitment even earlier than the schizont stage. Using a *P. falciparum* line in which endogenous *ap2-g* was conditionally targeted for destruction with an FKBP destabilisation domain unless rescued with a Shield-1 ligand (Shld-1), Bancells et al. demonstrated that the timing of sexual commitment along the asexual cycle itself was variable, and depended on *ap2-g* transcript stabilisation. When stabilised with Shld in the early-ring stage, parasites would frequently undergo sexual commitment and conversion before schizogony, representing same cycle commitment (SCC), whereas when *ap2-g* was absent in ring-stages and Shld-1 was applied in the late stages of asexual development, conversion continued in the proceeding cycle, following schizogony (Bancells et al., 2019) (**Fig. 1.6**). By identifying parasites committed to each route, the authors identified only three (*pfg14.748*, PF3D7\_1476600, EPF1) proteins to be significantly enriched in NCC parasites, suggesting functional differences between commitment cycles may be

subtle to none. Josling et al., however, expanded upon this observation, using ChIP-Seq to identify differential occupancy of *ap2-g* over 391 targets between gametocytes resulting from NCC vs. SCC, and over 270 targets that were bound by *ap2-g* exclusively in one cycle (Josling et al., 2020).



**Figure 1.6 Pathways to sexual commitment**

Source: (Hollin and Le Roch, 2020)

### 1.5.4 Sex determination and sexual maturity

Parasites fated to develop into gametocytes have one more crossroads set in front of them — their sex. *Plasmodium* parasites lack sex chromosomes, meaning sex differences are mediated and identified through changes in gene expression and subsequent protein levels (Tadesse et al., 2019). Upon maturity and activation in the mosquito, each female gametocyte will produce one female macrogamete capable of mating, whereas a male gametocyte will further divide into up to eight motile microgametes, leading to a bias towards females in the gametocyte sex ratio of most *Plasmodium* species (Carter and Graves, 1988). This ratio is not static, however, and can vary over the course of infection (Paul et al., 2002), as well as in response to antimalarial drugs (Tadesse et al., 2019). Intriguingly, it can also vary in the presence of parasites of the same species from differing genetic backgrounds (Reece et al., 2005) and in mixed-species infections (Bousema et al., 2008; McKenzie et al., 2002).

While the mechanisms behind sexual commitment have been illuminated by the discovery of AP2-G the switch, if one exists, to regulate sex remains unknown. The timing of this switch is also elusive, it is not known if commitment to gametocytogenesis and a particular sex are concurrent or indeed independent. A recent study paired a screen of a barcoded pool of genetic mutants (Schwach et al., 2015) with scRNAseq in order to profile the expression of mutants that perturbed sex ratios in reporter lines of *P. berghei* (Russell et al., 2021). The authors profiled ten mutants capable of altering sex ratios and found that some were detected prior to a detectable sexual dimorphism. Whether or not these markers are expressed in the same way or have the same effect on *P. falciparum* gametocytes has yet to be determined.



# Chapter 2

## Thesis Outline

Despite substantial progress, many aspects of malaria gametocyte biology remain unclear. Bridging these knowledge gaps is important in the development and deployment of novel transmission-blocking strategies and vaccines against malaria. In this thesis, I aim to facilitate further exploration of gametocyte biology by using single-cell RNA sequencing to profile the gene expression changes underlying sexual development in *P. falciparum*.

In **Chapter 3**, I first describe the collection and curation of over 30,000 *P. falciparum* cells to generate the gametocyte development atlas, the fourth instalment to the Malaria Cell Atlas project. I then use the data set to explore early, mid, and mature stages of gametocyte development that can be difficult to access using bulk approaches. I also investigate a stage of immature gametocyte development that precedes sexual dimorphism on a transcriptional level.

Using the same data set in **Chapter 4**, I compare the developmental trajectories of two laboratory strains of *P. falciparum* as they progress through sexual development, underscoring the high degree of conservation in this process.

In **Chapter 5**, I follow asexual and sexual development through the lens of the long non-coding transcriptome, generating a new data set with improved transcript capture efficiency in order to do so.

Finally, I place my work in the greater context of the field of malaria research in **Chapter 6**, and highlight ways in which these data may be used to further elucidate gametocyte biology in the future.



# Chapter 3

## An atlas of gametocyte development in *Plasmodium falciparum*

### 3.1 Introduction

The formation of gametocytes from asexually proliferating cells represents an essential step in *Plasmodium spp.* transmission and serves as the boundary between inter- and intra- host proliferation. The tipping point is believed to rest in the parasite's ability to sense and respond to changes in its host environment, leading to its description as a molecular 'decision' to differentially invest in either developmental programme (Carter et al., 2013). The decision is laden with the finality and inefficiency of gametocytogenesis, which requires a substantial investment of resources to develop a single gametocyte, and the perpetuity of continued asexual proliferation, which can exponentially multiply a population in a shorter period of time. It is not surprising, then, that sexual stages are relatively rare within a parasite population, with generally only ~1% detected as sexual forms *in vivo* (Cao et al., 2019; Eichner et al., 2001).

The skewed ratio of asexual:sexual parasites in a culture or infection can make sexual stages difficult to isolate and study. Despite this, a plethora of '-omics' studies made great strides to signpost the major transcriptional changes underlying asexual and sexual development (Bozdech et al., 2003; Eksi et al., 2012; Le Roch et al., 2003; López-Barragán et al., 2011; Silvestrini et al., 2010). The discovery of the master regulator of sexual development, AP2-G, (Kafsack et al., 2014; Sinha et al.,

2014) helped to place a starting point for sexual development, and subsequent work identifying the targets of AP2-G further refined the early transcriptional events in gametocyte development (Kent et al., 2018; Poran et al., 2017).

To date, there exists a patchwork of transcriptomic data covering most events in gametocyte development, however only two complete time courses in *P. falciparum* (van Biljon et al., 2019; Young et al., 2005). Both of these studies sampled regularly from *in vitro* parasite cultures, induced experimentally to trigger gametocytogenesis, from the point of induction until the point of sexual maturity. While incredibly rich, samples from these studies represent mixtures of asexual populations as well as mixtures of both stages and sexes in later time points. In asexual parasites, single-cell RNA sequencing has been demonstrated as a useful tool in deconvoluting the signal present in mixed populations sequenced in bulk (Howick et al., 2019; Reid et al., 2018), as well as further resolving hitherto undetected transcriptional heterogeneity. These data sets have then gone on to serve as useful 'anchoring points' for further perturbation studies (Hentzschel et al., 2021; Rawat et al., 2021), and even comparisons between apicomplexan species (Xu et al., 2020). A similar atlas-style approach to characterising sexual development in *Plasmodium falciparum* would thus hopefully serve a similar purpose; stitching together existing transcriptomic data from bulk approaches and acting as a map to guide future studies.

In this chapter, I use single cell RNA sequencing to follow gametocytes as they differentiate, develop, and mature into late male and female stages. Combining this data into a complete atlas of sexual development, I then harness the temporal resolution afforded by this type of data to explore key events in the sexual development of *P. falciparum*.

**Personal contributions:** I maintained all parasite cultures and assisted with the generation of all single-cell preparations, and performed all analysis, unless otherwise listed.

**Additional contributions:** The project was conceived by Dr. Mara Lawniczak, Dr. Arthur Talman, Dr. Andrew Russell, Dr. Virginia Howick, and myself. Drs Andrew Russell and Virginia Howick assisted with single cell preparations. Library preparation and sequencing were performed by the Wellcome Sanger sequencing pipelines facilities.

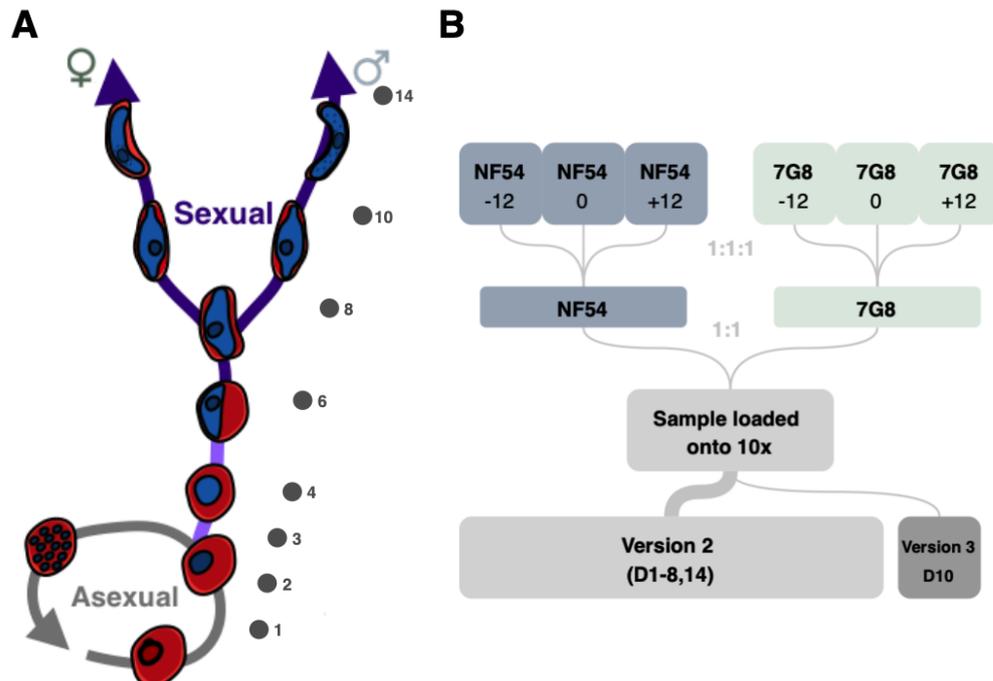
## 3.2 Results

### 3.2.1 Constructing an atlas of gametocyte development

To attempt to capture the full diversity of transcriptomic changes over the course of gametocyte development, I designed a multi-day, multi-sample single-cell RNAseq experiment using 10x Genomics 3' single-cell RNA capture. The design, detailed in **Fig. 3.1a**, involved 8 individual samples taken from *P. falciparum* parasite cultures over the course of gametocyte development (14 days). From days 1-4 post gametocyte induction, we took a sample daily in an attempt to ensure we captured the earliest stages of sexual commitment and development, as these stages can be the hardest to detect morphologically or through marker gene expression. Following this, we sampled more sparsely, aiming to capture the later stages of development through samples taken on days 6, 8, 10, and 14.

The aim of the sampling design was to capture the widest range of transcriptional events as they occurred throughout development. As such, I attempted to capture 24hr of developmental time within a sample taken on each day, even though the sample was taken at one point in time on that day. To do so, I split cultures into three flasks, synchronised at 12-hr intervals, and combined them to form the sample for each day. I repeated this procedure for two separate strains of *P. falciparum* parasites, NF54 and 7G8 (a comparison of these strains is discussed in Chapter 4). As shown in **Fig. 3.1b**, the sample taken on each targeted day was formed by mixing the three synchronised flasks for each strain at equal ratios, then combining the samples for each strain at equal ratios. Of note is the 10x Chromium version chemistry used to capture single-cell samples. For all samples except for that taken on day 10, cells were captured with Version 2 chemistry. On day 10, we took two identical samples, one with Version 2 chemistry, one with Version 3. Due to a reagent failure, the Version 2 sample for day 10 failed during library preparation, and thus the Version 3 sample had to be added to the Version 2 data. Samples from days 8 and 14 also failed at the library preparation stage.

Following library preparation and sequencing, we successfully captured single-cell transcriptomes from six of the original eight sampling dates. From each sample, I then sought to extract high-quality transcriptomes to combine into a comprehensive atlas. For each sample due to differences in library preparation and sequencing,

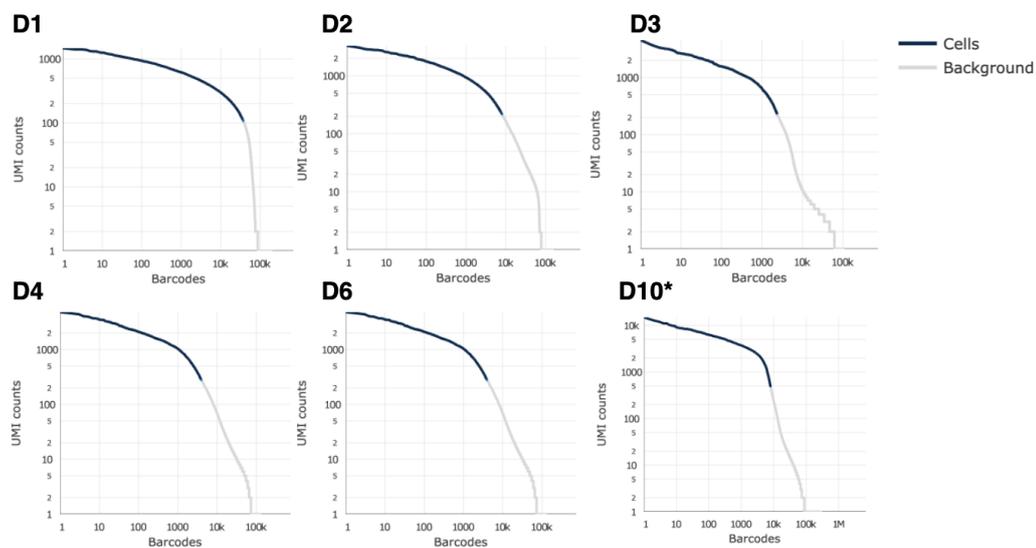


### Figure 3.1 Experimental design

Experimental design for capturing single-cell transcriptomes over the course of gametocyte development. **(A)** A schematic representing gametocyte development from asexual replication into sexual maturity. Grey dots represent the days (numbered post gametocyte induction) in which samples were taken from cultures and loaded onto a 10x Chromium controller for single-cell capture. **(B)** Representation of the makeup of a single sample. Parasites from two *P. falciparum* strains (NF54 and 7G8) were synchronised 24hrs apart in 12hr intervals, and combined in equal ratios on each sampling day. Samples from each strain were then combined and loaded onto the 10x controller. All samples but day 10 were processed using 10x Chromium 3' Version 2 chemistry, with day 10 successfully processed using Version 3 chemistry

I performed quality control and normalisation separately before later combining them. The first steps involved removing barcodes that represented empty droplets that did not capture a cell, those that failed to capture a cell but did encapsulate cell-free or damaged RNA that was released, or those that captured dead or degraded cells. I used 10x software Cell Ranger to map transcriptomes to the *P. falciparum* genome, which produces a white-list of barcodes that correspond to 'real' cells, based on the number of transcripts assigned to each cell barcode. The approach relies upon the assumption that the number of transcripts captured in 'real' cells will be substantially higher than for empty or degraded cells, and thus aims to algorithmically find the 'knee', or inflection point, between the distributions

of transcripts in 'real' vs. 'empty' barcodes. Where this approach struggles, which was the case with my data, is when heterogeneous cell populations differ in the number of transcripts they have available for capture, leading to an overlap in the distribution of transcripts captured per cell and obscuring the inflection point between barcodes that captured 'real' cells and cell-free debris. This led to an observed continuum of transcript capture in the samples I collected, and a likely underestimation of the true number of 'real' cell barcodes added to the Cell Ranger white-list (Fig. 3.2).



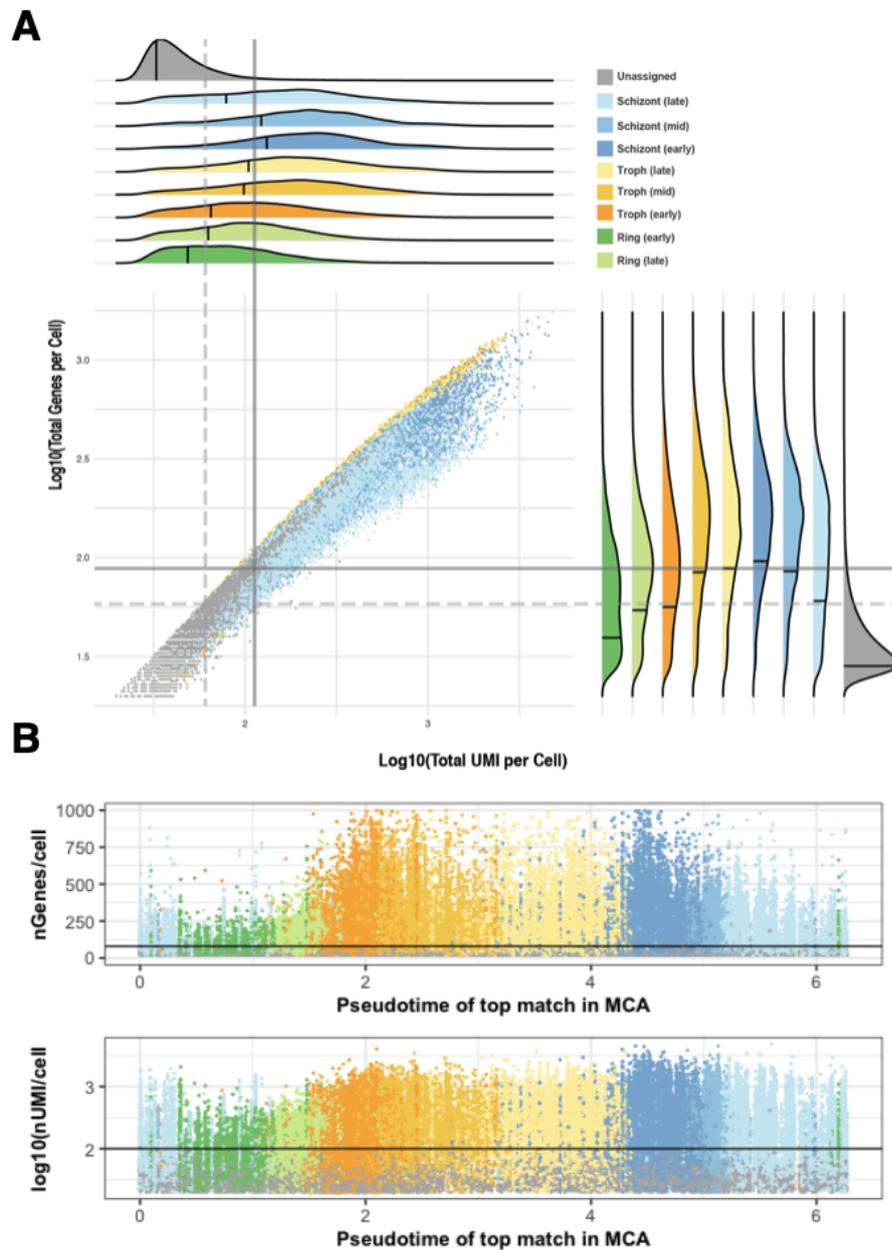
**Figure 3.2** Cell Ranger **barcode white-listing struggles with continuous cell types**

Number of unique molecular identifiers (UMIs) per cell barcode, split by sample. UMIs were mapped and assigned to barcodes using Cell Ranger. Barcodes coloured in blue represent those white-listed by Cell Ranger.

I thus carried out my own custom cell barcode white-listing by considering any cell barcode from the raw output matrices from Cell Ranger with  $> 10$  transcripts as a potential 'real' cell. I matched these putative cells to their closest correlate within the START-seq2 *P. berghei* atlas (Howick et al., 2019) using `scmap-cell` (v1.8.0) (Kiselev et al., 2018), which uses a  $k$ -nearest neighbour search to identify the closest matching cell in the atlas to the query. Barcodes that reached a cosine similarity of  $> 0.2$  to their  $k$ -nearest neighbour in the reference were considered a possible parasite cell and assigned the stage of the neighbour. As expected, I found the number of transcripts captured fluctuated by assigned stage, with ring and late

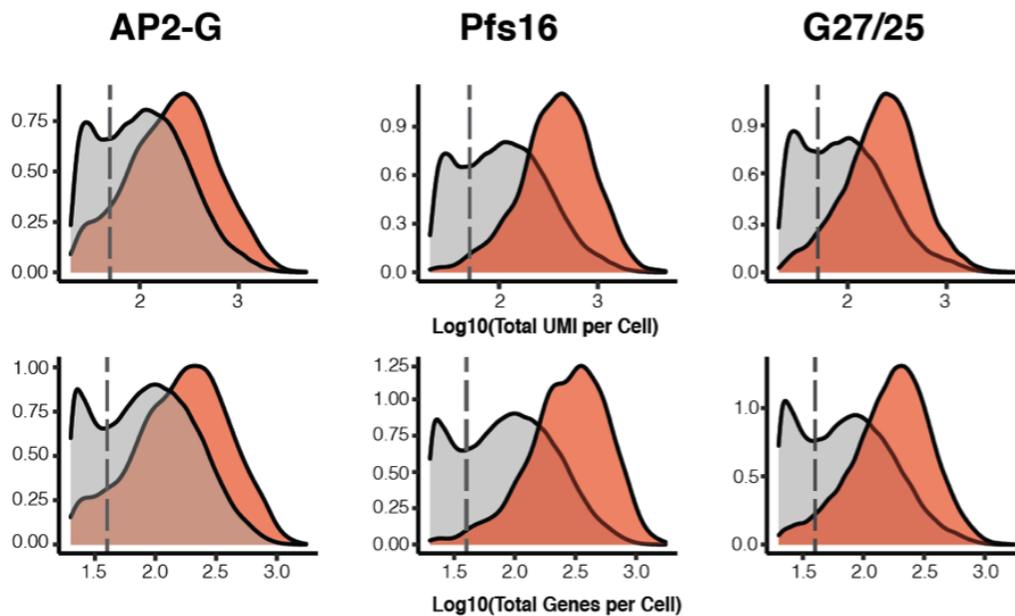
schizont stages displaying the lowest number of transcripts and genes captured per cell, which Howick et al. (2019) also observed for *P. berghei* and Poran et al. (2017) in *P. falciparum* (**Fig. 3.3**). I thus implemented a two-step quality control threshold that aimed to preserve capture of low mRNA stages without contaminating the data set with low-quality cells from other stages.

First, I white listed any cell barcode matching an early/late ring or a late schizont with > 50 UMI/cell or < 50 genes/cell. For all other barcodes, even those without an assigned stage, I then raised the quality threshold to discard those with < 100 UMIs or < 100 genes. These thresholds allowed me to discard the vast majority of low quality cells that did not match any stage in the Smart-seq2 *P. berghei* atlas, while 'rescuing' ring and schizont stages that were mislabelled by Cell Ranger as debris. Importantly, however, by including unassigned cells that still passed the quality threshold, I retained cells that may not have matched the Atlas due to differences in species or underrepresented cell types in the atlas, such as immature gametocytes. I found that the thresholds did not remove the majority of cells in the data sets expressing markers of sexual commitment or early gametocyte development (**Fig. 3.4**).



### Figure 3.3 Asexual life stages require custom quality control thresholds

**(A)** Log<sub>10</sub> UMIs captured per barcode plotted against log<sub>10</sub> genes per barcode for any barcode with > 10 UMIs from samples D1,2,3,4 and 6, coloured by closest stage assignment in the *P. berghei* atlas (Howick et al., 2019). Ridge plots above and to the right of the plot show the distributions of each value by stage. The dotted grey line marks the adjusted quality threshold for low-mRNA stages (early/late rings, late schizonts), as compared to the filter applied to the rest of the stages (solid grey line). **(B)** Genes/cell and UMI/cell plotted against the pseudotime value assigned to the top cell match in the *P. berghei* atlas. Black lines correspond to the 'upper' cell quality threshold applied to non ring/late schizont cells



**Figure 3.4 Cell quality thresholds do not remove putative gametocyte cells**

Log<sub>10</sub> UMI and genes/cell for cells expressing (> log(1)) three markers of gametocyte commitment (AP2-G) and early gametocyte development (Pfs16, G27/25), shown in red. Distributions of cells not expressing each marker are shown in grey. Dotted line represents the 'upper' quality thresholds applied to all non-ring/late schizont cells.

Lastly, I removed putative doublet cells from each sample. To do this, I leveraged the fact that the samples contained a mixture of two *P. falciparum* genotypes and used `souporcell` (Heaton et al., 2020) to identify cells containing single nucleotide polymorphisms (SNPs) from both NF54 and 7G8, thus representing inter-strain doublets. Identifying intra-strain doublets, which can occur if two different parasite cells of the same strain are encapsulated within one droplet, was more difficult due to the continuous nature of asexual development and a lack of discrete cell types. I thus sought to eliminate the most 'problematic' of doublets, i.e. those that contained cells of very disparate life stages ('heterotypic doublets') that could bias analyses more so than doublets containing cells of the same strain and stage ('homotypic doublets'). I used Louvain clustering with multilevel refinement as implemented in `Seurat` to first group cells into many highly-localised clusters, and then merged these into larger clusters that were transcriptomically similar using `DoubletDecon` (DePasquale et al., 2019). The tool then uses a weighted approach to assign a doublet likelihood score to cells most likely to be heterotypic doublets

between these broader clusters in order to preserve transitional or progenitor cells that exist 'between' continuous cell types.

Overall, I retained 35,978 total cells across the six successful samples (quality statistics summarised in **Table 3.1**). In each sample, I captured an increasing number of gametocyte cells, with multiple samples capturing the bifurcation point in which gametocyte departure from the asexual cycle is visible via UMAP reduction. The breadth of sampling meant that combined, I was able to follow the complete gametocyte development trajectory through to mature gametocytes (**Fig. 3.5**).

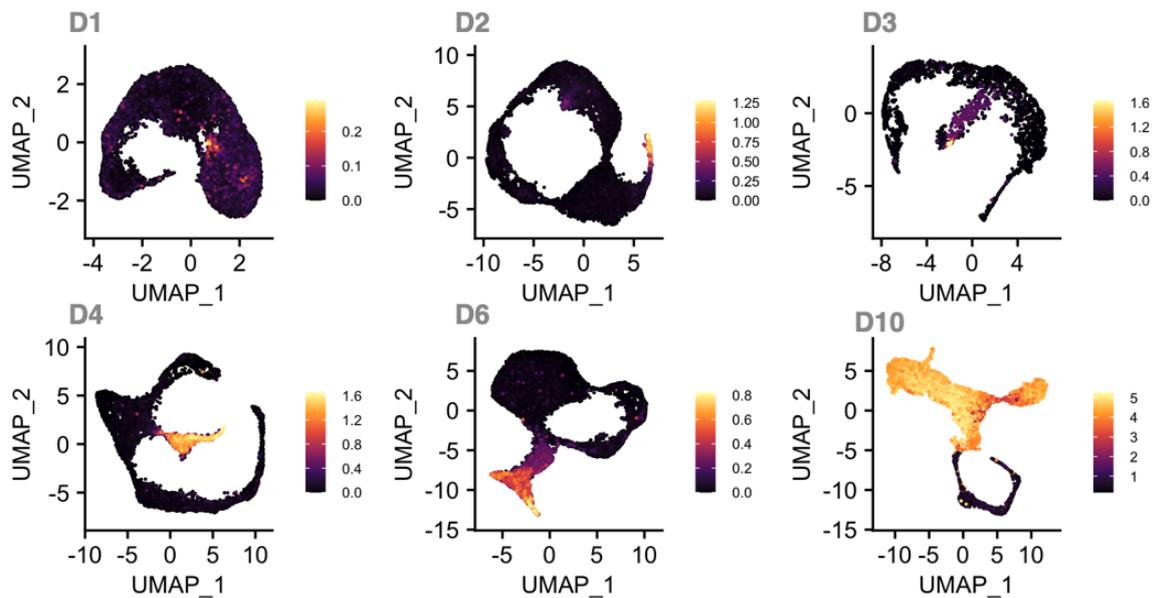
**Table 3.1 Quality Statistics per Sample**

Sample	nUMI/cell	nGene/cell	Total cells
<b>D1</b>	348.41	247.38	10,070
<b>D2</b>	536.55	308.28	6,231
<b>D3</b>	644.61	398.32	2,347
<b>D4</b>	751.82	411.68	3,653
<b>D6</b>	318.90	210.89	6,634
<b>D10*</b>	2082.32	1016.88	7,043

\* Version 3 Chemistry

To explore the complete trajectory of gametocyte development in this data, I then needed to combine the data from each experiment into a single object containing all single cell transcriptomes. I found that simply merging the transcriptomes (using the `merge` function) from each data set into one object without correction led to substantial batch effects that made the data incomparable (illustrated for days 4 & 6 in **Fig. 3.6**). This necessitated the use of *Seurat*'s data set integration tool, which works by using a set of genes that display reliable variability between data sets to 'anchor' cells from one data set to similar cells in another. The result is a batch-corrected matrix of similarity scores between cells that allow for cells of different batches to cluster together on a PCA or UMAP.

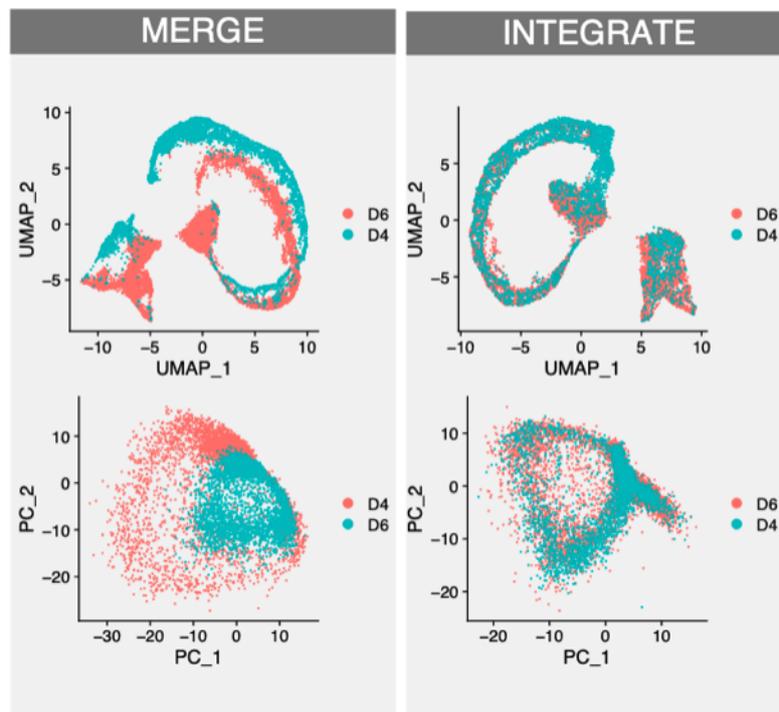
The resulting object contained all 36,000 cells from the six 10x samples. Dimensionality reduction via UMAP (McInnes et al., 2018) showed cell transcriptomes organising into the asexual cycle as a circular ring. Gametocytes emerged from this ring as a Y-shaped object, with more gametocytes captured as days progressed



**Figure 3.5 Integration of multiple data sets allows for full coverage of gametocyte development**

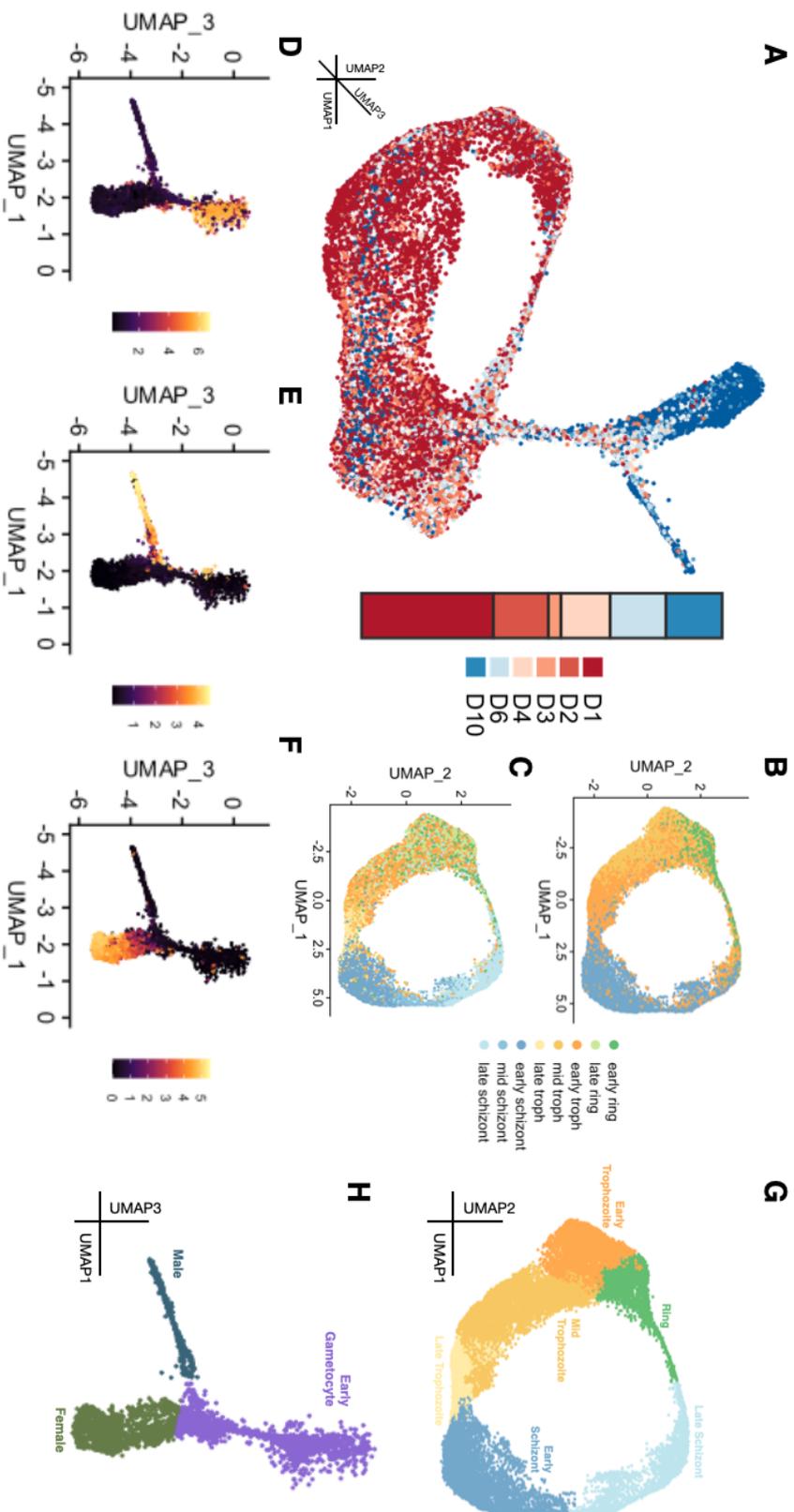
(A) UMAP projections of single-cell transcriptomes captured from each sampling day, coloured by *pfs16* expression (B) to mark gametocytes. Colour intensity is proportional to scaled, log-normalised expression values.

(Fig. 3.7a), reflective of the differentiation into male and female sexual forms. Correlation of each cell with both the *P. berghei* atlas and stage-specific bulk RNAseq (López-Barragán et al., 2011) allowed me to assign a life stage to Louvain clusters (Fig. 3.7b,c). For gametocytes, I used the expression of *pfgexp02*, one of the earliest known markers of gametocyte expression (Portugaliza et al., 2019), to identify early gametocyte clusters. In a similar manner, I used the expression of *pfs25* and PF3D7\_0905300 (a component of the dynein heavy chain) to identify female and male clusters, respectively.



**Figure 3.6 Batch correction improves data integration**

PCA and UMAP projections of cells from D4 and D6 samples when single-cell transcriptomes from each data set are merged using Seurat's `merge` command into a single object without correction (left), or integrated using Seurat's integration tool.



**Figure 3.7 A single-cell atlas of *P. falciparum* gametocyte development**

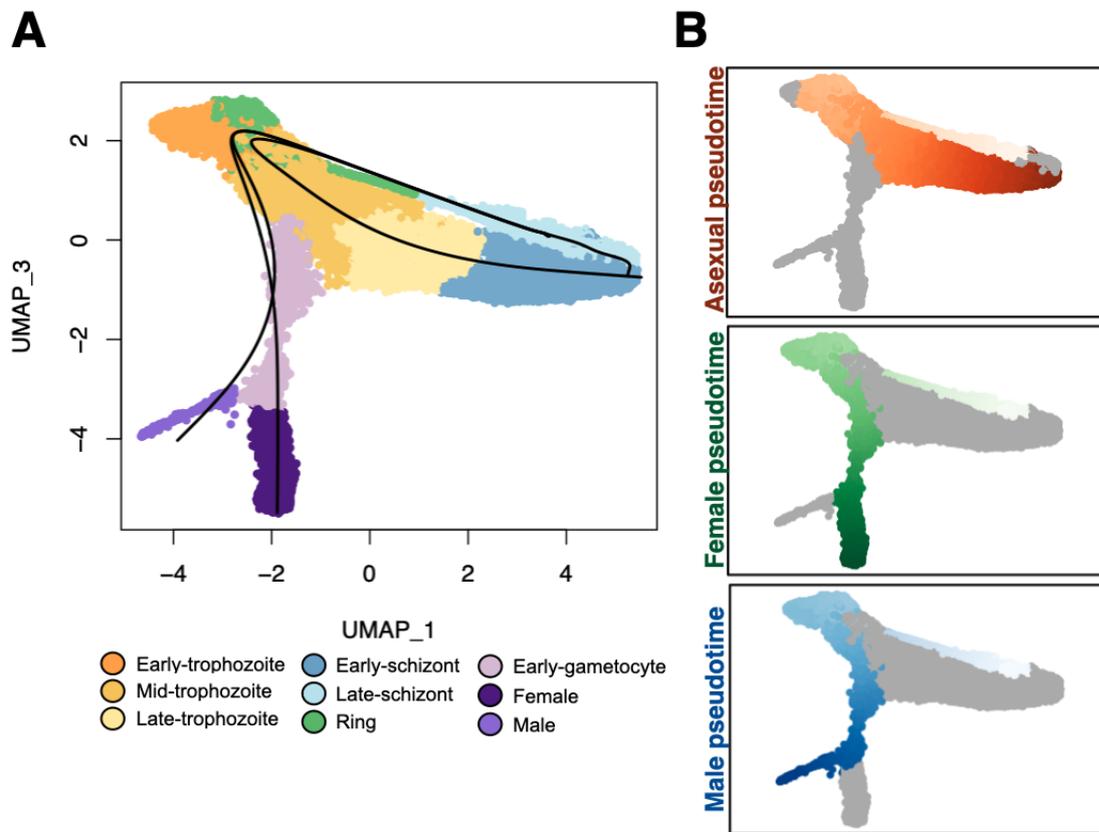
(A) 3-dimensional UMAP projection of the integrated gametocyte atlas object, coloured by the day in which each single cell transcriptome was collected. The proportion of total cells comprised by each sample are shown in the bar chart on the right. (B) UMAP of just asexual cells, coloured by the stage most highly correlated from bulk RNAseq (López-Barragán et al., 2011), or the (C) *P. berghei* atlas (Howick et al., 2019). (D-E) Expression of early gametocyte markers (D) *gexp02*, (E) PF3D7\_0905300, and (F) *pf525*. (G-H) UMAP projections of the gametocyte atlas with life stages annotated.

### 3.2.2 scRNAseq provides resolution regarding the timing of key events during sexual development

At the completion of a stage-resolved, merged and quality controlled gametocyte atlas, I then used the data to explore the transcriptional changes underlying gametocyte development. One of the greatest advantages of using single-cell technologies in favour of bulk data is the improved resolution with which transcriptional changes can be profiled. To harness this in my own data, I first needed to order cells according to their relative position along each developmental trajectory, or pseudotime. Although named as a type of 'time', pseudotime is not required to be linear, and isn't always a good proxy for clock time. Instead, pseudotime is a set of values assigned to cells to mark their relative order in a developmental process. It is better described as a proxy for a cell's position in transcriptional space, allowing observed transcriptional events to be ordered in developmental time, regardless of how long (in clock time) they take to occur.

I chose to explore three trajectories for cells in my data set: starting from ring stage cells, which were 1) those that leave the asexual cycle and develop into mature female gametocytes, 2) those that develop into mature male gametocytes, and 3) those that do not differentiate into sexual forms and continue to propagate asexually. I used the tool `slingshot` (Street et al., 2018) to fit a curve to each of these trajectories using the underlying UMAP embedding and annotated cell clusters, and assign to each cell a pseudotime value that described its position along that curve (**Fig. 3.8a**). To ensure shared cells (i.e. asexuals that may become sexual, or sexual cells that may become male or female) could be compared across trajectories, the three curves were fit simultaneously such that cells that were shared between trajectories would receive virtually the same pseudotime values, and that the relative values assigned to cells unique to one lineage could be compared to those with a similar value in another (**Fig. 3.8b**).

I first used the assigned pseudotime ordering to identify when, in 'time' & stage I could first see evidence for sexual departure from the asexual cycle. I defined observable sexual departure as the point in which the base of the sexual 'Y' shape left the circular asexual ring during the early-mid trophozoite stage in the UMAP embedding (**Fig. 3.9a**). To narrow down this window to align with real clock time, I correlated my single cell expression with that of bulk RNAseq experiments that

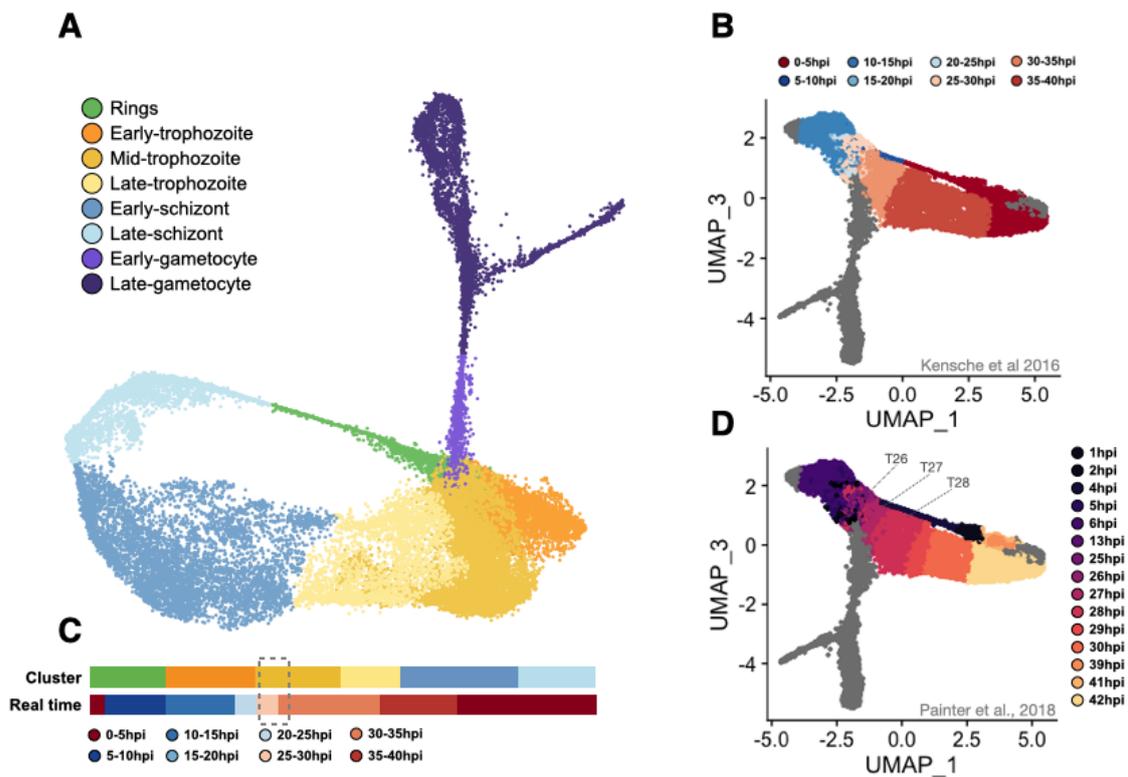


**Figure 3.8 Fitting pseudotime to sexual and asexual developmental trajectories**

(A) 2-dimensional UMAP plot, coloured by stage. Three estimated pseudotime trajectories (black lines, calculated using `slingshot` (Street et al., 2018)), were seeded at ring stages, and terminated in early schizonts (asexual trajectory), males (sexual trajectory 1), or females (sexual trajectory 2). (B) Cells are coloured by their relative position (light to dark) along each of the three pseudotime trajectories.

were synchronised and captured at regular intervals following merozoite invasion. I found bulk correlations at a single-cell level to be difficult to interpret given the level of noise present and the incomplete transcript capture in each cell, and thus opted to instead create 'pseudobulk' clusters of cells with similar pseudotimes, thus correlating the average expression of a 'slice' of pseudotime with real clock time based on bulk RNAseq. The log-normalised expression values were averaged across all cells in 146 clusters, which were defined in 150-cell intervals along the asexual trajectory. The resulting 146 pseudobulk RNAseq 'samples' were then matched to their highest correlate (using a Pearson's correlation test) in the bulk data set. I first correlated expression with that of Kensche et al. (2016), and using

this narrowed the window of sexual departure within trophozoites to roughly 20-30 hours post merozoite invasion (**Fig. 3.9b,c**). The same technique with the hourly resolution of the Painter et al. (2017) data set allowed me to further narrow this window to 26-28hpi (**Fig. 3.9d**). These estimates confirm what was seen previously (Kent et al., 2018; Pelle et al., 2015), and corresponds to the estimated time in which trophozoites are sequestered in tissue vasculature (Pelle et al., 2015), as well as the timing of *pfs16* expression (Bruce et al., 1994).



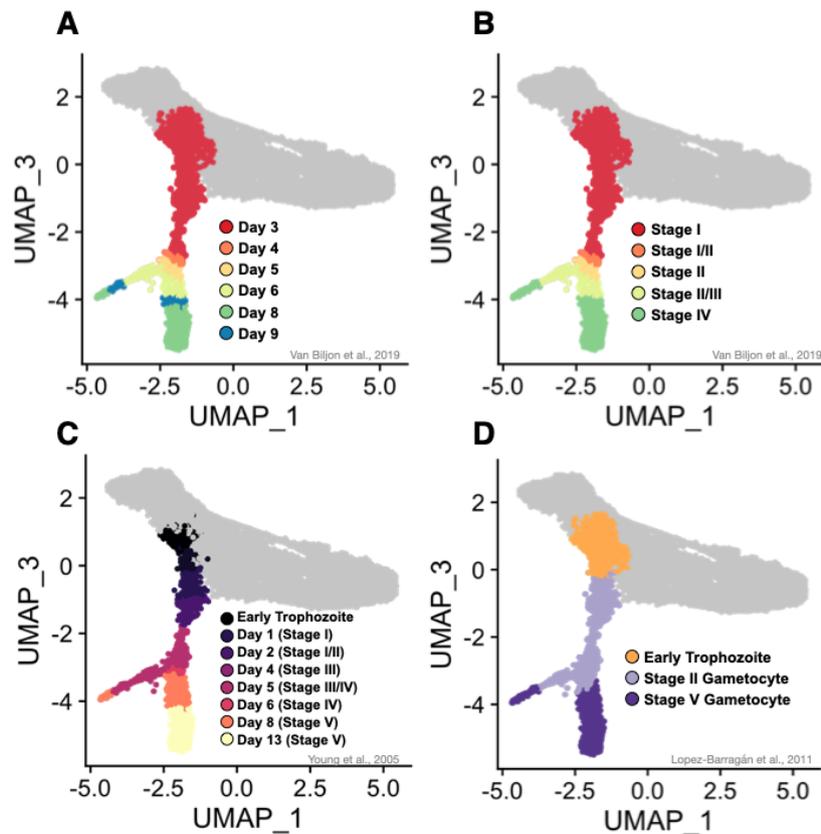
**Figure 3.9 Combining scRNAseq with bulk RNAseq allows for timing of gametocyte departure**

(**A**) 3-Dimensional UMAP, coloured by stage. (**B**) 2-dimensional UMAP, coloured by the sample from Kensche et al. (2016) that best correlates to each pseudotime cluster along the asexual trajectory. These correlates are summarised in a bar plot in (**C**) to demonstrate the point of sexual departure. (**D**) Top correlate of each pseudotime cluster to RNAseq samples from Painter et al. (2017). Clusters correlating to time points near the point of sexual departure are labelled above. All correlation scores are based on Pearson's correlation.

I took the same approach to assign stages (I-V) to the gametocyte clusters in my data, as the asynchronous nature of the experimental design meant I collected a range of gametocyte stages in each sample. Using the same pseudotime window

approach, I grouped cells into 150-cell groups along female and male pseudotime and correlated their average expression to the recent micro-array time course of gametocyte development by van Biljon et al. (2019). I found the top correlate for each pseudotime cluster followed a similar temporal progression to the samples from those of the bulk data set, beginning with day three (note, these labels refer to those used in the van Biljon et al. time course rather than the sample labels from this study), which is the first day gametocytes were observed in the study, and ending with day 8/9 post gametocyte induction, which approximately matches the same sampling days in which gametocytes were collected in this study (**Fig. 3.10a**). I was then able to roughly sketch out the relative stage (I-V) for each pseudotime cluster based on the most highly correlated sampling day in the bulk data (**Fig. 3.10b**). It is important to note that these samples were collected in bulk and thus do not represent the transcriptome of a pure stage, but rather a mixture of gametocyte stages, and therefore the stages assigned in **Fig. 3.10b** to pseudotime clusters likely represent the most abundant stage from each sample day.

I found the initial point of sexual departure correlated most highly with the transcriptomes of early stage I gametocytes from van Biljon et al., and continued to do so throughout the stalk of the Y shape. Directly preceding sexual differentiation, clusters begin to correlate to later stage II and stage III stages, with stage III dominating the clusters located at the earliest observed bifurcation point of male and female gametocytes in the UMAP structure. The appearance of transcriptomic bifurcation during stage III is in keeping with the fact that it is stage III in which the earliest markers of sexual dimorphism are visible via microscopy (Sinden, 1983). I found a similar pattern when I correlated expression to an earlier micro array time course (Young et al., 2005), with the bifurcation point matching sampling days from this micro array that were dominated by Stage III and IV gametocytes (**Fig. 3.10c**). The distribution of early stages in the stalk and mature stages in the branches of the sexual trajectories was further confirmed by correlation with a more granular collection of bulk RNAseq from stage II and stage V gametocytes by López-Barragán et al. (2011) (**Fig. 3.10d**).



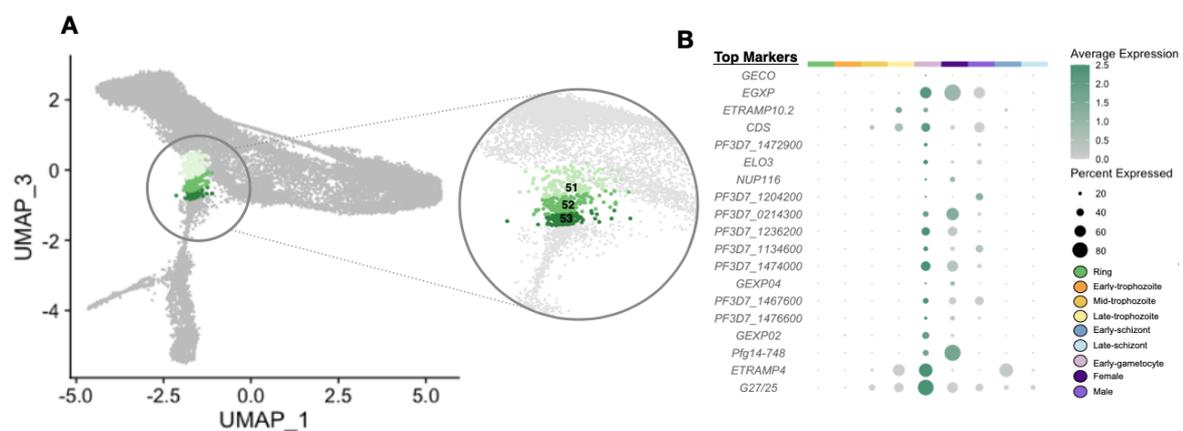
**Figure 3.10 Staging gametocyte transcriptomes**

2-dimensional UMAPs highlighting the pseudotime clusters that follow sexual differentiation, coloured by the top correlate to previous bulk data sets. **(A)** Correlation to microarray samples collected by van Biljon et al. (2019), or **(B)** the most abundant gametocyte stage in each of these samples. **(C)** Micro array data from Young et al. (2005). The most abundant stage from each sample is listed. **(D)** RNAseq data from López-Barragán et al. (2011). All correlations were done using Pearson's correlation tests.

Differential expression analysis of the cells directly preceding sexual departure further confirmed their correlation with early stage I gametocytes. Due to the similarities in morphology to trophozoite stages (Baker, 2010), as well as the difficulty in removing asexual stages in the earliest stages of gametocyte growth, bulk transcriptomic samples of early gametocytes often contain a mixture of sexual and asexual parasites. To better understand the transcriptional changes underlying the bifurcation of sexual and asexual parasites, I combined (for greater power) the first three clusters in pseudotime following sexual departure (**Fig. 3.11a**) and used MAST to identify 19 genes that showed a  $> 1$  log-fold difference in expression

as compared to the rest of the cells in the life cycle (**Fig. 3.11b**). I found that a number of these markers have been previously described as well-known markers of early sexual stage parasites, including *pfg27/25* and *pfg14-748*, and *nup116*, a downstream target of master regulator PfAP2-g (Bancells et al., 2019; Josling et al., 2020). These genes are involved in protein export and membrane remodelling, some of the earliest known functions in a developing gametocyte (Silvestrini et al., 2010).

These well-known markers of gametocytogenesis lend confidence to the use of these cells as the first observable forms of developing gametocytes. In addition to these well known markers, I also detected 8 genes that were significantly differentially expressed in these cells for which function is still listed as unknown. This includes PF3D7\_1134600, a putative Zn-finger transcription factor unique to *Plasmodium spp.* that infect primates (Ukaegbu et al., 2014). While the function of these genes cannot be ascertained from transcriptomic data alone, their placement at the earliest stages of gametocyte differentiation makes them interesting targets for further study into their use as gametocyte-blocking strategies or novel bio-markers for gametocyte detection.



### Figure 3.11 Gene expression in early gametocytes

**(A)** Location of cells, clustered according to pseudotemporal progression, used to identify markers of very early gametocyte differentiation. Cells from clusters 51-53 were combined to improve statistical power. These cells were then subject to differential expression analysis via MAST to identify genes with a  $> 1$  log-fold change in expression compared to the rest of the cells (grey). **(B)** Dotplot of 19 marker genes identified highly expressed in the cells contained in the clusters in (A). Dot size is proportional to the proportion of cells in each stage expressing the gene, and colour intensity is proportional to log-normalised average expression level.

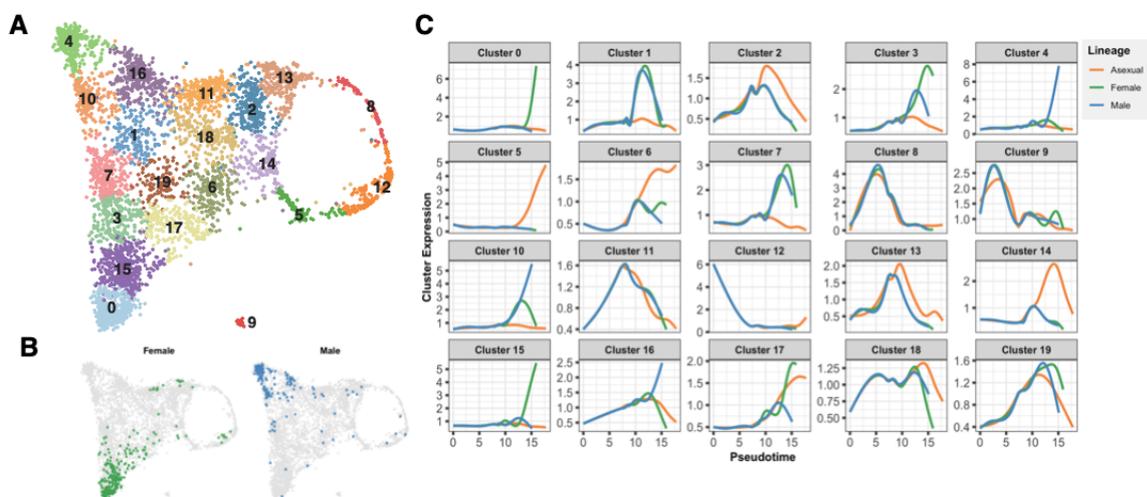
### 3.2.3 Profiling global changes in gene expression during sexual development

Leveraging the number and breadth of cells collected in the data set, I aimed to gain a high-level picture of gene expression across the course of *P. falciparum* asexual and sexual development. In much the same way that information regarding global patterns in gene expression can be gained by clustering cells whose transcriptomes resemble one another in reduced dimensional space, one can cluster genes based on the cells that co-express them. In doing so, one can speculate upon possible mechanisms of action to genes whose functions are yet unknown based on a 'guilt-by-association' approach.

I thus transposed my cell x gene expression matrix to create a correlation (similarity) matrix of genes based on their log-normalised expression pattern in each cell (gene x cell). I found the signal to be more sensitive to noise than the cell x gene matrix, and thus filtered out genes that were detected in fewer than 100 cells (for a total of 4806 genes). From there, I used UMAP to collapse the matrix into 2-dimensional space, and used kMeans clustering as implemented in `scikit-learn` to assign each gene to one of 20 clusters. The cluster assignments for all genes can be found in **Table A.1**. I chose the UMAP algorithm due to its recent demonstration as a robust means of collapsing high-dimension transcriptomic data while preserving biological interpretability (Dorrity et al., 2020). Each of the 20 kMeans clusters contained between 50-300 genes, with most containing 100-200. The distribution of these clusters and their underlying UMAP shape resembled that of the cell x gene graph (called the 'cell UMAP', for simplicity), with a larger Y or triangle-like shape attached to a circle (**Fig. 3.12a**). To investigate whether the protrusion corresponded to sexual development like it does in the cell UMAP, I mapped sex-specific markers to those identified by Lasonder et al. (2016), who used fluorescent markers to separate the transcriptomes of late stage males and females onto the gene x cell ('gene UMAP') graph. The placement of these markers confirmed that the clusters at the tips of the Y shape on the gene graph (0 and 4) were indeed enriched for genes highly expressed in male or female gametocytes (**Fig. 3.12c**).

Plotting the distribution of each cluster over pseudotime (as estimated by a generalised additive model (GAM) fit to each lineage using `geom_smooth`) revealed

that many of these clusters were pseudotime-specific (**Fig. 3.12c**). Clusters found in the circular portion of the gene UMAP, like 8,13,14,and 5, peak at the beginning, middle, and end of asexual pseudotime, respectively. Other clusters peak after the male and female lineages have diverged down their path to sexual development. At this point (approximately pseudotime value 10), clusters enriched in sexual stages increase, and some begin to show clear patterns of sexual dimorphism. Cluster 9 contained genes with the lowest total expression out of the genes selected and thus was believed to contain noisy genes that were not removed during initial filtering.

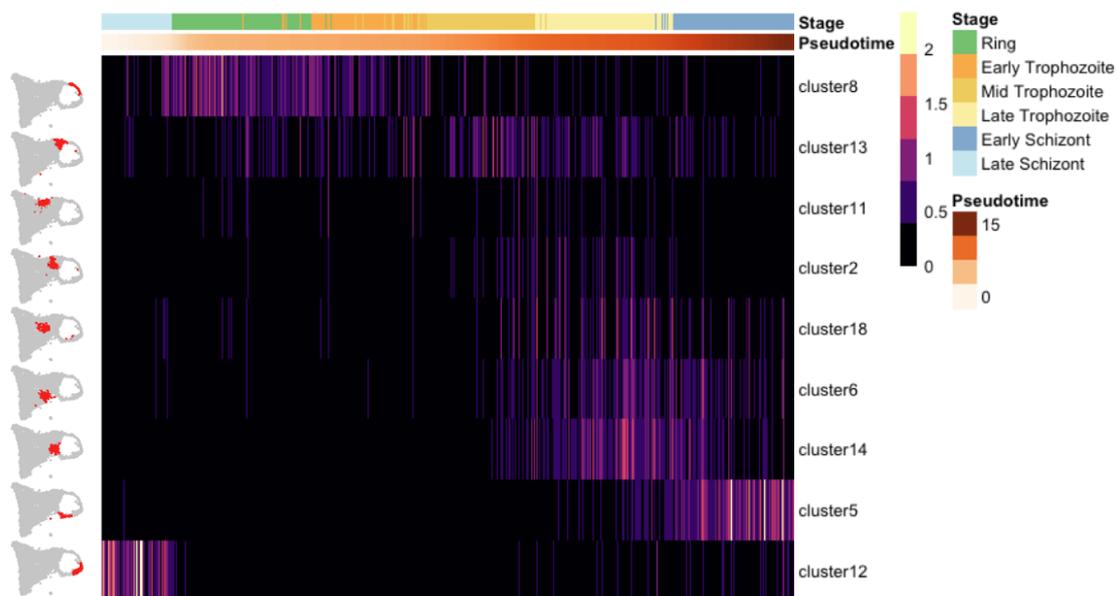


### Figure 3.12 Gene-based clustering reveals patterns of co-expression during sexual and development

(A) 2-dimensional UMAP of 4806 genes, based on their expression in each cell, where each dot represents a single gene. Genes are coloured by cluster assignment, which was carried out using kMeans clustering on the predicted UMAP dimensions, with  $k=20$ , and implemented by the `scikit-learn` package in python. (B) Location of genes identified as differentially expressed in male (blue) or female (green) gametocytes by Lasonder et al. (2016). (C) Average expression of genes in each kMeans cluster by pseudotime, coloured by each of the three trajectories. Expression over time was smoothed by fitting a generalised additive model (GAM) with a span of 0.7.

Narrowing in on the average expression of clusters surrounding the 'ring' in the gene UMAP revealed that these clusters indeed corresponded to genes implicated in tightly-time controlled cascades along the IDC, with enriched gene ontology (GO) processes indicative of those involved in asexual proliferation, such as host cell entry, growth and metabolism, protein export, and DNA replication (**Fig. 3.13**). These clusters organised around the gene UMAP ring in order of developmental time,

echoing what is seen in the cell UMAP shape. The abrupt switch-like pattern to genes co-expressed over the IDC mirrors that seen previously in single-cell sequencing of *Plasmodium* parasites (Howick et al., 2019; Real et al., 2021; Reid et al., 2018), and is in keeping with the notion of a 'just in time' pattern of translational regulation, first hypothesised with the use of bulk transcriptomic approaches (Bozdech et al., 2008; Le Roch et al., 2004).

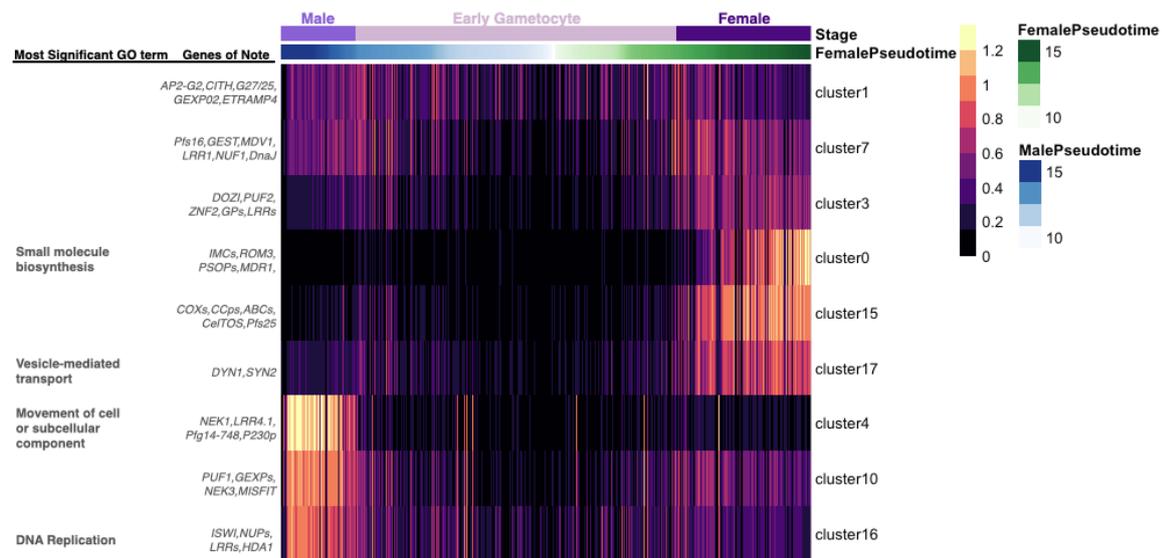


**Figure 3.13 Average expression of clusters enriched for asexual-development genes**

Heatmap of scaled gene expression along asexual pseudotime, averaged for selected gene clusters, selected from kMeans clustering of a gene x cell UMAP object. Columns represent individual cells, rows represent gene clusters. The location of each cluster on the gene graph is displayed to the left of each row, coloured in red.

I took the same approach to examine the rest of the gene clusters, those that showed evidence for expression in the male and female lineages following their departure from the asexual cycle (**Fig. 3.14**). As I confirmed in **Fig. 3.12c**, I found clusters 0, 15, and 4 to be the most sexually dimorphic in their expression, and they contained a number of well-studied genes often used as markers for the sexes, including *pfs25* in females and *Nek3*, and *P230* in males. In both sexes, clusters 1 and 7 were expressed at the point of sexual departure, which contained early gametocyte marker *pfs16*, the transcription factor *ap2-g2*, and a number of other

genes involved in RNA metabolism. Development in females is then marked by increasing expression of genes involved in the formation of the inner membrane complex, vesicle transport, and the expression of ookinete-specific genes such as CelTOS and secreted ookinete proteins (PSOPs).



**Figure 3.14 Average expression of clusters enriched for sexual-development genes**

Heatmap of scaled gene expression along sexual pseudotime, averaged for selected gene clusters, selected from kMeans clustering of a gene x cell UMAP object. Columns represent individual cells, rows represent gene clusters. Selected genes, as well as significant gene ontology (GO) terms for each cluster are listed.

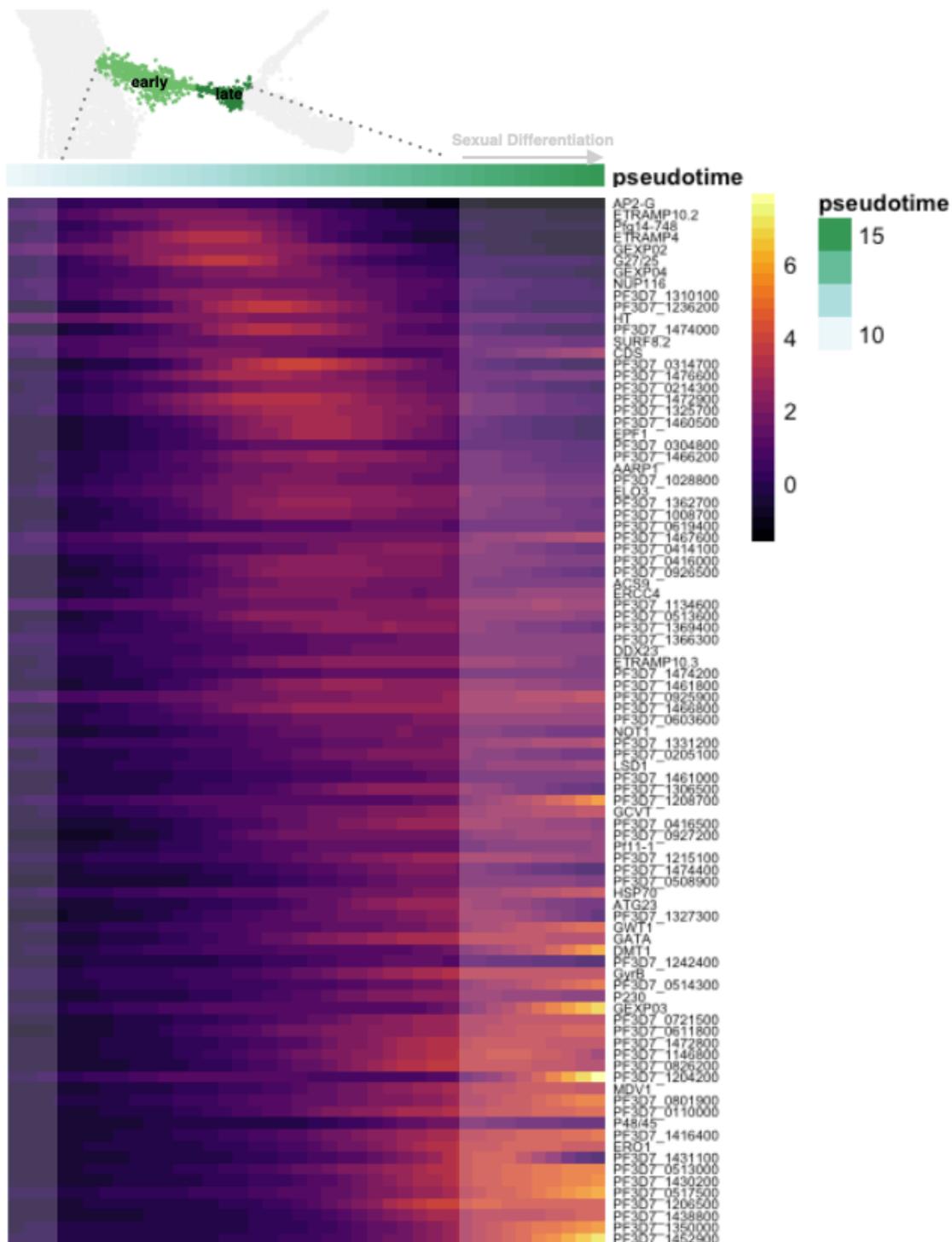
### 3.2.4 Sexual differentiation precedes a shared sexual progenitor state

While the systems underlying the 'switch' from asexual parasites to committed and eventually sexual parasites are relatively well characterised, the mechanics behind sex-determination in *Plasmodium* are less well known. This includes the timing of sex-determination — which may occur at the point of sexual commitment or after a period of generic 'sexual growth', in which sex remains undetermined until a later point. To explore these possibilities, I focused my analysis to one of the most intriguing aspects of the UMAP structure of the data — the 'stalk' of the Y, where

parasites are developing sexually but do not yet show distinct branching patterns evident of sexual dimorphism. Before exploring the latter, I wanted to expand upon the markers that differentiated early gametocytes I found in **Fig. 3.11** to profile what defines this potentially early stage of 'sexually ambiguous' growth, or if I could detect signs of sexual dimorphism. As I described in section 3.2.2, I found the majority of pseudotime clusters throughout the stalk to map to samples containing stage I gametocytes, only appearing to transition to match the transcriptome of more mature stages directly before sexual differentiation. This may have occurred for a number of reasons, not limited to sampling bias, 'snapping' of cells to the nearest similar cluster with a high density of related cells (i.e., cells being 'pulled' towards trophozoites or later stages), or may represent a dramatic 'switch' in transcription that distances (in transcriptional space) the transcriptome of budding gametocytes from the trophozoites they developed from.

I thus combined pseudotime clusters (in order to get better power in my differential expression analysis) based on their bulk correlations into two larger 'early' and 'late' clusters, that represented roughly the potential checkpoint from stage I/II gametocytes to more mature forms. I again used MAST to identify genes that displayed differential expression between the two clusters. Instead of implementing a fold-change cut-off, which is well-suited for finding marker genes but can miss weaker signals, I used the `tradeSeq` (Van den Berge et al., 2020) tool to fit a generalised additive model (GAM) to describe the expression of each gene identified by MAST over the pseudotime trajectories I fit previously in section 3.2.2. Using the `associationTest` function, I whittled down my list of MAST candidate genes to only those whose average expression (as modelled by the GAM) demonstrated significant association with pseudotime. 91 genes in total met these criteria.

I then used rank-two ellipse seriation (Chen, 2002) to rank genes based on a correlation matrix of the fitted values for each of the 91 genes significantly up-regulated in the stalk over pseudotime. I found that, while not expressed to a high degree, *pfap2-g* was ranked 'earliest' in a cascade-like progression throughout the stalk (**Fig. 3.15**). Following *pfap2-g*, the majority of genes with 'narrow' expression exclusive to the stalk were those identified by my initial screen of early gametocytes, including *nup116*, *Pf14.748*, and *gexp02*, as well as a number of putative downstream targets of AP2-g as identified by Josling et al. (2020).



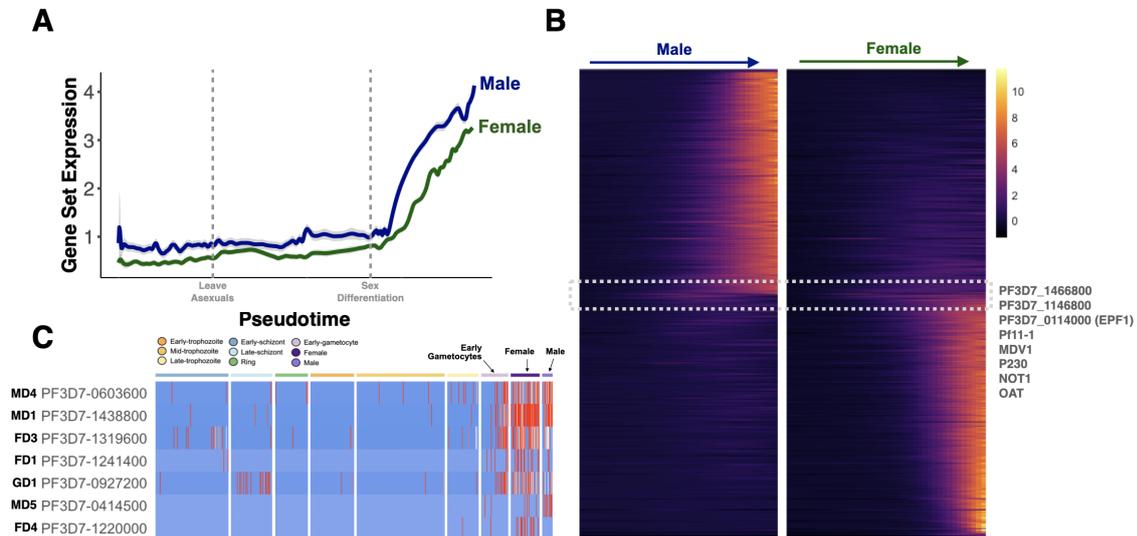
### Figure 3.15 Gene expression during sexual differentiation

Heatmap of 91 genes differentially expressed in the stalk. Genes were selected based on marker gene detection by MAST (Finak et al., 2015) for pseudotime clusters containing 150 cells each. Female pseudotime was selected as the values for pseudotime cluster groupings, but contain the same cells/values as male pseudotime due to simultaneous fitting. Gene order was determined via rank-two ellipse seriation (Chen, 2002) using the `seriate` package in R.

Finally, I explored the end of the stalk — that is, the branching point in the UMAP in which sexual differentiation into males and females is visible. Using genes previously identified as enriched in males or females by Lasonder et al. (2016), I observed a general 'switch-like' increase in expression of male and female specific transcript sets from only after the branching point between males and females on the UMAP (**Fig. 3.16a**). This is in contrast to what van Biljon et al. (2019) observed in a micro-array time course of gametocyte development. Using the same transcript sets (from Lasonder et al. (2016)), they observed an increase in male-specific transcripts before that of the female-specific. Given the ability to differentiate sexual populations in single-cell data, this discrepancy may reflect an over-representation of early male gametocytes in early samples of the micro-array time course, thus leading to an increase in male transcripts before the female. It could, however, also result from the regression of early signals towards the mean when hundreds of genes expressed in > 7000 cells are averaged together. Thus, I instead defined my own set of male/female specific genes using MAST (with a log-threshold change > 1), and plotted them over pseudotime to see if any showed signs of expression prior to the branching point in the UMAP. In total, I identified 617 sex-specific transcripts, and used `tradeSeq` to plot their expression over pseudotime (**Fig. 3.16b**). As before, the majority of sex-specific expression was switch-like, with genes rapidly increasing in expression at the same time, following the bifurcation of the male and female lineages on the UMAP. A small subset of genes, however, seemed to increase in expression before the rest. These included PF3D7\_1466800 and PF3D7\_1146800, two transcripts found to enriched in early females by van Biljon et al. (2019), potentially before the observation of morphological differences between the sexes. They also included three important early gametocyte genes (*pf11-1*, *mdv1* and *epf1*(PF3D7\_0114000)) recently found by Singh et al. (2021) to decrease in expression in an *pfap2-g2* knockout line, and by Xu et al. (2020) as a target of AP2-G2.

This global switch-like increase in sex-specific transcripts was also observed in a recent single-cell study of gametocytogenesis by Russell et al. (2021). This study however also identified a number of potential regulators of this global switch using sex-specific knockouts in a *P. berghei* mutational screen, a number of which are expressed in early gametocytes before sex differences are observed. Of the 7 mutants in that study with syntenic orthologs in *P. falciparum*, I also observed

expression of these genes in the progenitors, or early gametocyte stalk, of the UMAP preceding sex-specific expression (**Fig. 3.16c**).



### Figure 3.16 Gene expression during sexual differentiation

**(A)** Average expression of male or female specific genes over pseudotime. Genes were those selected by van Biljon et al. (2019) from the list identified by Lasonder et al. (2016) as enriched in late stage male or female gametocytes. **(B)** Heatmap of expression of male or female specific genes enriched in males or females in this data set, as determined by performing MAST (Finak et al., 2015) on cells labelled 'Male' or 'Female', with a log-fold change threshold  $> 1$ . Genes determined to be expressed 'earlier' than the cascade of male/female specificity, via visual inspection, are labelled to the right. **(C)** Expression of putative sex-determination genes identified in Russell et al. (2021)

Correlation analysis of these genes with others found in the area failed to detect any reliable signals of co-expression with sex-specific markers. Further narrowing in and sub-clustering of cells in any point of the asexual branch preceding sexual differentiation, or early gametocytes preceding the bifurcation of the sexes, in any of the samples, failed to demonstrate evidence for sub-structure indicative of a male/female fate. These findings are likely influenced by the inherent variability present in single cell expression data, especially given the less-efficient capture of 10x version 2 chemistry. In spite of this, the clustering of cells in reduced dimensional space (UMAP), the concerted surge of sex-specific expression into males and females in cells most closely correlated to stages III/IV, and the lack of observable sex-specific sub-clusters in the stalk all suggest the existence of a gametocyte progenitor stage after sexual cells have departed the asexual cycle, but before transcriptionally (or morphologically) identifiable as a male or female

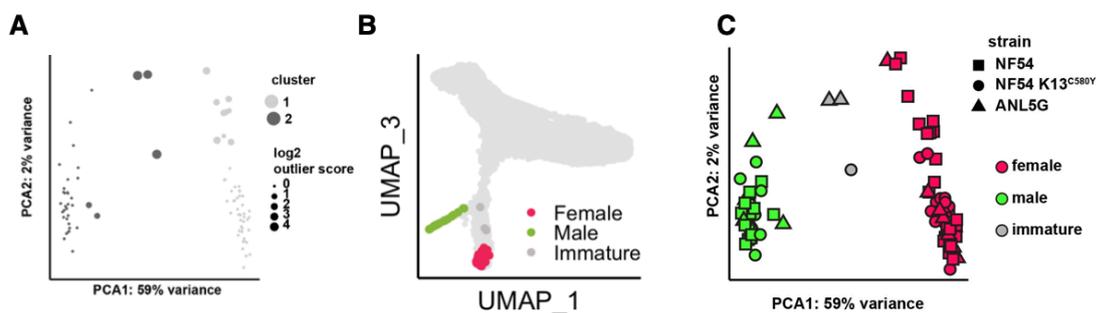
gametocyte. This early, clearly sexual but not yet clearly sex-specific structure has been observed in other single cell experiments of *Plasmodium* parasites (Hentzschel et al., 2021; Poran et al., 2017; Rawat et al., 2021) in addition to Russell et al. (2021). It is possible that this progenitor stage exists due to the shared biology of stage Ib - stage IV gametocytes as they sequester in tissue and bone marrow and out of circulation (Venugopal et al., 2020). This is similar to what is observed in sexually committed schizont cells, which share a great deal of similarity with their non-committed counterparts, and this is hypothesised to be due to the requirement of a sexually committed schizont to perform the same schizont-specific function until it begins differentiation Josling et al. (2018).

### 3.2.5 The gametocyte atlas as a tool for the malaria research community

Single-cell RNAseq allows for the generation of high resolution data sets in which both rare and common cell types and processes can be captured. This is especially useful when researching *Plasmodium* parasites; the continuous nature of the IDC and the sexual departure from it make it difficult for bulk approaches to examine a slice of this time without the possibility of contaminating stages. To remedy this, this data set contains, in the form of pseudotime, cells ranked according to their position along these continuous processes. By mapping RNAseq data from bulk or single-cell samples, the malaria research community can gain a better understanding of the composition of their samples and the range of transcriptional space they capture. This is especially useful for studies involving gametocytes, where early stages difficult to discern from asexual stages, and males and females can be difficult to separate.

As an example of the data set's utility, I mapped a single-cell RNAseq data set of mature males and females collected for our recent addition to the Malaria Cell Atlas (Real et al., 2021). Part of the study included a novel fluorescent assisted cell sorting (FACS) gating strategy using a DNA-RNA double stain in order to separate mature male and female gametocytes. These cells were then single-cell sequenced using Smart-seq2 and examined to determine the accuracy of the gating strategy. Cluster analysis of the data identified three outlier cells that did not show evidence of expression for male or female markers (**Fig. 3.17a**). To understand if these cells

were perhaps errant asexual stage parasites, abnormal male/female stages, or simply contamination, I mapped them to their closest correlate in my data set using `scmap-cell`. I found that while most cells clustered to the tips of the male or female branches, the three outlier cells fell earlier in the trajectory of sexual differentiation, and thus represented immature gametocytes that had been collected along with the more mature stages (**Fig. 3.17b**). Thus, we were able to more accurately label the cells collected as 'male', 'female', or 'immature' (**Fig. 3.17c**) and better assess the accuracy of the sorting strategy.



**Figure 3.17 Mapping to the gametocyte atlas allows for identification of outliers in single cell data**

(A) Cluster analysis of sorted single-cell 15-day old gametocyte populations processed using Smart-seq2. Gametocyte populations were double stained with Hoechst (DNA) and Pylonin Y (RNA) to sort gametocytes by sex. Each dot represents one gametocyte. Dot sizes represent the outlier score of the cluster assignments. (B) Position of sorted gametocyte cells when mapped to its closest neighbour in the gametocyte atlas using `scmap-cell`. Cells are coloured by assigned sex. (C) PCA showing the assigned sex for each cell. The different shapes represent the three different parasite lines sorted, two deriving from the canonical lab strain NF54 background, with a third deriving from a recently culture-adapted Cambodian field isolate, APL5G. Figures A and C were made by Dr. Kathrin Witmer, and were published in Real et al. (2021), on which I am a co-author. Parasite culture, sorting, sequencing and analysis were performed by Drs. Virginia Howick and Kathrin Witmer. Data mapping and sex assignment in figure B were performed by myself.

### 3.3 Discussion

The gametocyte development atlas is an important addition to the pre-existing Malaria Cell Atlas project, comprising 35,978 single cell transcriptomes spanning the trajectory of sexual development and differentiation in *P. falciparum*. Such

resolution makes the study of sub-populations and rare cell types possible without mixing from other stages, and for the deconvolution of bulk samples containing mixed populations (Briggs et al., 2021). This is particularly useful in the evaluation of sex ratios in transcriptomic data, which can be inconsistent if determined using microscopy alone (Tadesse et al., 2019). The assignment of a pseudotime value as a proxy for developmental progression can act as an anchor for future single-cell data sets wishing to contextualise cells based on their relative positions along this continuum. Finally, as highlighted by Nötzel and Kafsack (2021), by making my quality controlled, annotated transcriptomes available as part of an interactive web application ([www.malariacellatlas.org](http://www.malariacellatlas.org)), I hope to make these data accessible and interpretable even by those unfamiliar with single cell RNAseq analysis.

It is important to note, however, that this collection is just the first step towards building a comprehensive tool for the malaria community. Whilst I prioritised completeness in coverage by sampling asynchronised parasites, the addition of synchronised single cell transcriptomes paired with microscopic images would allow a coupling of transcriptomic changes with morphological ones, which would be especially beneficial during sexual differentiation, and given that the five stages of gametocyte development have been assigned as such based on morphological observations. What's more, the addition of samples from later time points including mature and activated gametocytes could aid in bridging the gap between the present atlas and pre-existing single cell data during the mosquito stages (Real et al., 2021), creating a more 'complete' picture of gametocyte development.

## 3.4 Materials & Methods

### 3.4.1 Parasite culture and sample collection

Two *P. falciparum* strains (NF54, 7G8) were cultured separately in O+ blood in complete RPMI 1640 culture medium at 37°C in a gas mixture of 5% O<sub>2</sub>/5% CO<sub>2</sub>/90% N<sub>2</sub>, as described in (Howick et al., 2019). To ensure transcriptional heterogeneity due to slightly different developmental stages encompassed with a 24 hour window was represented by a single "day" sampling point each strain was split into three flasks synchronised at 12-hour intervals using two rounds of 5% sorbitol (Lambros and Vanderberg, 1979). Thus the sample taken on each targeted day was

actually represented by 6 flasks (-12, 0, +12 hours and two genotypes). As such, for each strain, each single cell sampling point represents a combination of three replicates grown in different flasks. Due to the overlap in the stages captured across days, each sample then constitutes an additional technical replicate (although not complete, as there are bound to be a few stages which by nature were only captured on individual days). These replicates are important to include, as differential gene expression can vary across the same cell types between biological replicates (Squair et al., 2021). Whether these samples represent biological replicates, however, are unclear, as they were all seeded from the same original flask and are thus non-independent. A more careful consideration of the design may have been to seed each flask from a separate thawed parasite population and proceeded to combine them for each sample.

Sexual commitment was induced at 1% parasitemia and 3% hematocrit through four consecutive days of growth and incomplete media change until gametocytes were observed. Culture media during sexual growth was supplemented with 10% human serum (obtained locally in accordance with ethically approved protocols). Samples for single-cell sequencing were taken 2,3,4,6,10 and 14 days after initial seeding. After six days, 10% heparin (Sigma, prepared at 1 $\mu$ L/mL) was added to cultures to prevent further asexual growth. For each sample, cultures from both strains across all three synchronisation flasks were combined such that equal numbers of parasites were added from each, using estimated parasitemia and RBC density. The parasitemia and RBC density of this new sample was then re-calculated before loading.

### **3.4.2 Library preparation, single cell sequencing and read mapping**

Samples from each day were loaded onto a single inlet on a 10x Genomics Chromium Controller according to the manufacturer's instructions with a target cell recovery of 10,000 cells. The reagents required for each run were mixed between two Chromium Single Cell 3' GEM, Library & Gel Bead Kits, which due to quality control between batches by 10x Genomics is normally not a problem, however the RT enzyme from the second kit, which was obtained from another lab, was believed to have expired or deactivated due to repeat freeze-thawing. Due to this, the samples

collected on days 8 and 14 failed to generate sufficient libraries for sequencing. I attempted to pool what libraries we did manage to recover with one from a previous day, however did not recover a meaningful number of cells from either of the runs that passed later quality control thresholds. All single-cell emulsions were prepared using a 10x Chromium Single Cell 3' Library v2 chemistry, with the exception of Day 10, where v3 chemistry was used. The initial intention was for the 'new' v3 chemistry to serve as a comparison with the older v2 kits, however the sample taken for D10 with the v2 kit suffered from the same reagent failure as the day 8 and 14 samples.

Successful 10x libraries were sequenced on separate run lanes of a HiSeq 4000 (Illumina) using 150-bp paired-end sequencing. Raw reads were assigned barcodes and UMIs using `Cell Ranger` single-cell software (version 3.0.0) using default parameters (Zheng et al., 2017). Processed reads (cDNA inserts) were aligned to the *P. falciparum* 3D7 reference genome ([www.sanger.ac.uk/resources/downloads/protozoa/](http://www.sanger.ac.uk/resources/downloads/protozoa/)). UMIs were quantified and single-cell expression matrices generated using `Cell Ranger` (version 3.0.0).

### 3.4.3 SCT filtering, normalisation and annotation

#### Cell quality thresholds

Single-cell expression matrices were processed using the `Seurat` (v3.2.2) R package (Stuart et al., 2019). Poor-quality cells were removed from the data based on life stage, as the average total number of genes expressed change over the course of the *P. falciparum* IDC (Howick et al., 2019), and stages with lower expression are often disproportionately discarded by the `Cell Ranger` default cell whitelisting process. I thus considered any barcode from the raw output matrices associated with > 10 UMIs as a putative cell and conducted a custom whitelisting approach. This approach was applied to all samples containing asexual parasites (days 1-6), and identified putative parasite cells by projecting the raw expression matrices on to the *P. berghei* cell atlas (Howick et al., 2019) using `scmap-cell` (v1.8.0) (Kiselev et al., 2018), which uses a k-nearest neighbour search to identify the closest matching cell in the atlas to the query. To accommodate for the difference in species, a table of 1-to-1 orthologs was used that was compiled by Howick et al. using OrthoMCL

(Li et al., 2003). Barcodes that reached a cosine similarity of  $> 0.2$  to their k-nearest neighbour in the reference were considered a possible parasite cell and assigned the stage of the neighbour. Cells identified as an early/late ring or late schizont containing  $< 50$  UMI/cell and  $< 50$  genes/cell were removed. Barcodes mapped to late stages, or barcodes not assigned to a stage in the reference (likely including gametocytes), were removed if they contained  $< 100$  UMIs/cell and  $< 80$  genes/cell.

### Doublet detection

Cells were demultiplexed into their strain of origin (NF54 or 7G8) via the `souporcell` pipeline (v2.0) (Heaton et al., 2020). Cells classified as containing variants from both NF54 and 7G8 by the pipeline were considered to be inter-strain doublets and removed. Remaining intra-strain doublets were identified and removed using `DoubletDecon` (v1.1.4) with NCBI taxonomy ID “5833” (for *Plasmodium falciparum*) and a rho-prime value of 1.3 (DePasquale et al., 2019).

### Normalisation

Resulting single-cell transcriptomes (SCTs) were then log-normalised, scaled, and further processed using `Seurat` (v.3.2.2). A variance stabilising transformation (vst) was performed on each sample to identify the 750 most highly variable genes used to perform a principal component analysis. A jackstraw plot was generated per sample to determine the number of significant principal components used to perform dimensionality reduction via three-dimensional Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) and were visualised using the `ggplot2` (Wickham, 2016) and `rgl` packages in R version 3.6.0 (R Core Team, 2019). Cells were clustered via the `FindClusters` command in `Seurat`, using a Louvain clustering algorithm with multilevel refinement.

### Data integration

Cells from all successful single cell samples were first combined using the `merge` command in `Seurat`, which simply combines cells from each experiment into the

same object, without performing batch correction. To perform batch correction, samples from each day were integrated using Seurat's integration tool (Stuart et al., 2019). First, 2000 anchoring genes were identified using `FindIntegrationAnchors` command, and cells were integrated using the `IntegrateData` command, using the top 10 principal components.

### 3.4.4 Single cell transcriptome analyses

#### Pseudotime

Each cell was ordered along pseudotime using `slingshot` version 1.4.0 (Street et al., 2018). UMAP embeddings were used as input. Cells assigned as 'rings' were chosen as the starting cluster with `start.clus` and 'late schizonts', 'males', or 'females' were chosen as the terminating clusters with `end.clus`. I tried to replace the default smoother implemented in `slingshot` (smooth spline) with a periodic loess smoother, as it is more appropriate for systems with closed loops (like the IDC), however this functionality was not yet accommodated for in the software package, and thus the default smoother was used instead.

#### Correlation to bulk data

All Pearson's correlations were performed using the `cor` command in R. The stage/sample in bulk data that had the highest  $R^2$  value was chosen as the best match for each cell/pseudobulk average.

#### Marker gene detection

Marker genes for clusters were identified using `MAST` (Finak et al., 2015) with a log-fold change threshold  $> 1$  on the log-normalised count data. Importantly, this was done using the non-integrated counts. Integrated counts have been batch corrected to allow for cell-cell comparisons with regards to clustering and dimensionality reductions, but the counts produced are not suitable for comparisons of gene expression differences (information found on the Seurat package's github issues page). To ensure counts were not being influenced by the large difference in

capture efficiency in the day 10 sample (which was processed using v3 chemistry), I performed marker gene detection with and without day 10 cells and found no difference in the enriched marker genes detected.

### Gene clustering

Genes were selected for clustering by selecting genes that were detected ( $> 0$ ) in more than 100 cells, resulting in a total of 4806 genes. A gene x cell matrix was created by transposing the cell x gene matrix. In order to correct for variability in total expression across genes, counts were then normalised by dividing by the average expression for each gene, and log-normalising. I then created two objects using the `umap.UMAP` function in python's `scikit-learn` package (Pedregosa, 2011), one with two dimensions (for visualisation) and one with 10 dimensions (for higher resolution in clustering), both with the same parameters (`n_neighbors=20`, `min_dist=0.01`, `metric='correlation'`, `spread=1`). Genes were clustered based on the 10-dimension embedding using `k=20`. Average expression for each cluster over pseudotime was then estimated by generating an average expression value for all genes of a cluster for each cell, using a modified version of the code presented in (Poran et al., 2017), the code for which was also used in (Real et al., 2021). A generalised additive model (GAM) smoother was then fit to represent the expression over pseudotime using `geom_smooth` with a 0.5 span, and `n=100`.

### GO term enrichment

GO term enrichment (Ashburner et al., 2000) was performed using the `clusterProfiler` package in R with GO annotations retrieved via the `org.Pf.plasmo.db` (Carlson, 2019). GO terms specifying 'biological processes' were selected.

### Modelling gene expression over pseudotime

Gene expression over pseudotime was estimated using `tradeSeq` (Van den Berge et al., 2020). I chose to use this package for a few reasons. One, quantifying and comparing gene expression changes over pseudotime using cell by cell values

is liable to noise and differences in total gene expression between genes. Many packages, the most common of which being `Monocle` (Trapnell et al., 2014), reduce this uncertainty by fitting a model/curve to describe this relationship in order to more robustly cluster genes with similar curves. I chose to use `tradeSeq` to fit these curves as it uses a negative binomial generalised additive model (NB-GAM) to model expression values, as opposed to a generalised linear model (GLM) fit by `Monocle` and other packages, and I believe the additive model to be more sensitive to subtle changes in gene expression (also described in (Song and Li, 2021)). Importantly, `tradeSeq` also allows for the fitting of expression values across genes in multiple lineages (`Monocle` currently only allows for one). Count matrices for `tradeSeq` were prepared based on the 'RNA' matrix rather than the 'integrated' one for reasons mentioned previously. Raw (non-normalised, non-logged) counts were used as `tradeSeq` implements its own internal normalisation. Pseudotime values and cell weights were supplied from `slingshot`. The `fitGAM` command was run on a randomly down-sampled (to  $n=10,000$  cells) subset of the entire atlas object to reduce computational time, however since I was interested largely in gametocyte cells, I only randomly sampled from asexual cells. The `evaluateK` function was used on a sub-sample of 200 genes to estimate the optimal number of knots to input to the `gitGAM` function (which was 8).



# Chapter 4

## Inter-strain differences in expression during sexual development

### 4.1 Introduction

Gene expression in *P. falciparum* is a tightly controlled process that results in a highly synchronised wave of expression throughout development. It has been suggested, however, that parasites can deviate from these patterns in order to 'bet-hedge' and respond to changes in their environment (Brancucci et al., 2014; Filarsky et al., 2018; Fraschka et al., 2018; Rovira-Graells et al., 2012). Strains of *P. falciparum* have been shown to differ in their antigenic potential (Awandare et al., 2018), their metabolism (Ke et al., 2011), their sporozoite infectivity (McCall et al., 2017), and importantly, in their sexual commitment rates (Gebru et al., 2017). Recent studies have also demonstrated the ability of malaria parasites to detect and respond to changes in mosquito environments (Witmer et al., 2021) and temperature stress (Rawat et al., 2021; Tintó-Font et al., 2021).

The transcriptional differences that have been found to exist between strains are usually subtle (Bozdech et al., 2003; Kamaliddin et al., 2021; Llinás et al., 2006; Tarr et al., 2018). Single-cell RNA sequencing has allowed for greater resolution in detecting strain-specific differences in *Toxoplasma gondii* (Xue et al., 2020), revealing a potential shift in developmental progression between strains. In this chapter, I use single-cell RNAseq in a similar fashion to profile the development of two strains, 7G8 and NF54. Understanding the differences that may exist in

gametocyte development between strains is important as they are often used to test transmission-blocking drugs (Delves et al., 2012).

**Personal contributions:** I maintained all parasite cultures and assisted with the generation of all single-cell preparations, and performed all analysis, unless otherwise listed.

**Additional contributions:** The project was conceived by Dr. Mara Lawniczak, Dr. Andrew Russell, Dr. Virginia Howick, and myself. Drs Andrew Russell and Virginia Howick assisted with single cell preparations, as mentioned in the previous chapter.

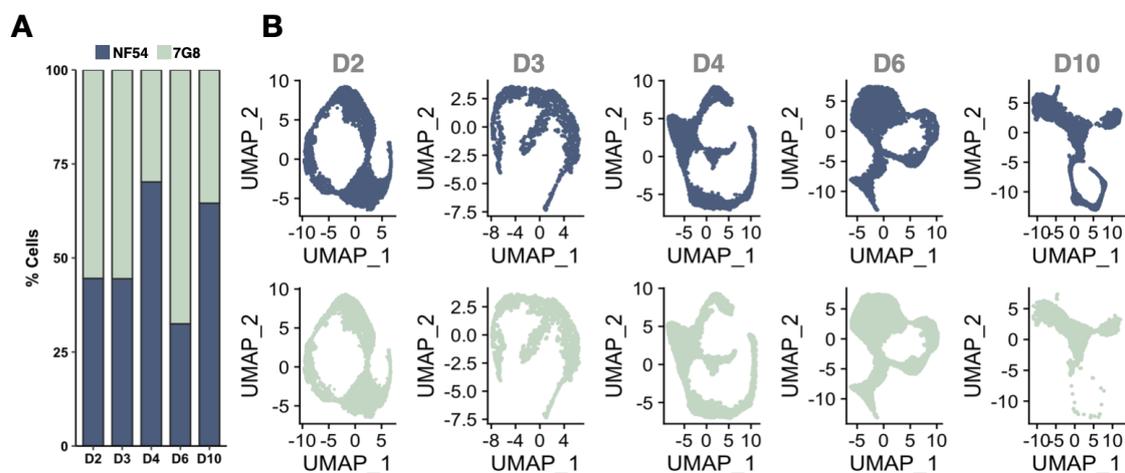
## 4.2 Results

### 4.2.1 Deconvolution of strains using scRNAseq data

The sampling strategy for the gametocyte atlas generated in Chapter 3 contained the inclusion of two genetically distinct *P. falciparum* strains. The aim of this approach was to investigate whether I could detect variation in gene expression during gametocyte development across the two strains. Parasites from NF54 and 7G8 were thus cultured separately and mixed together in a (targeted) 1:1 ratio based on parasitemia and RBC density estimations on each day of sampling. As these cells were not barcoded or tagged according to which strain they originated from, I used `souporcell` (Heaton et al., 2020) to assign a strain to each cell based on single nucleotide polymorphisms (SNPs) present in the read data, matching variants to those present in variant call files (VCFs) for 3D7 (a clone of NF54) and 7G8 from the latest release of the Pf6K data set (Pearson et al., 2019).

As I was interested in specifically examining strain differences as they pertain to gametocytogenesis, I began by narrowing my analysis to just include samples from the time course experiment that contained a detectable/substantial population of gametocytes (assessed by visually examining the expression values for early gametocyte marker *pfs16* (PF3D7\_0406200), as can be seen in Fig. 3.5 in Chapter 3), and thus excluded the D1 sample. For the remainder of the samples, I observed a relatively equal proportion of cells assigned to NF54 or 7G8 after sequencing and quality control (save for days 4 and 10, where NF54 comprised just under 75% of the sample) (**Fig. 4.1a**). Splitting the UMAP generated for each sample by

strain revealed that by and large, I had generated good coverage of gametocyte development in both strains (**Fig. 4.1b**), however there were samples for which strain capture was biased. The sample taken on D10, for example, contained a small number of asexual parasites along with mature gametocytes for the NF54 cells, but almost exclusively mature gametocytes for 7G8. Due to this, and given the abundance of asexual cells present in the other data sets, I included only sexual cells (as assessed by clusters expressing *pfs16* and *pfgexp02*) in the analysis for this sample going forward.



**Figure 4.1 Proportion of NF54 and 7G8 parasite strains captured in each sample**

**(A)** Proportion of two *P. falciparum* strains (7G8 and NF54) captured in each gametocyte time course sampling date that contained gametocytes (i.e., D1 was excluded from this analysis). For details on experimental design, sequencing, and quality control of each sample, see Fig. 3.1. For each sample, 7G8 and NF54 parasites were cultured separately and mixed 1:1 at the time of sample loading onto a 10X Chromium controller. **(B)** UMAP plots for each sample, split by strain. Each dot represents a cell assigned as NF54 or 7G8. Strain assignment was achieved by variant mapping and clustering based using `souporcell` (Heaton et al., 2020), using variant call files (VCF) for 3D7 (a clone of NF54) and 7G8 from the latest release of the Pf6K data set (Pearson et al., 2019).

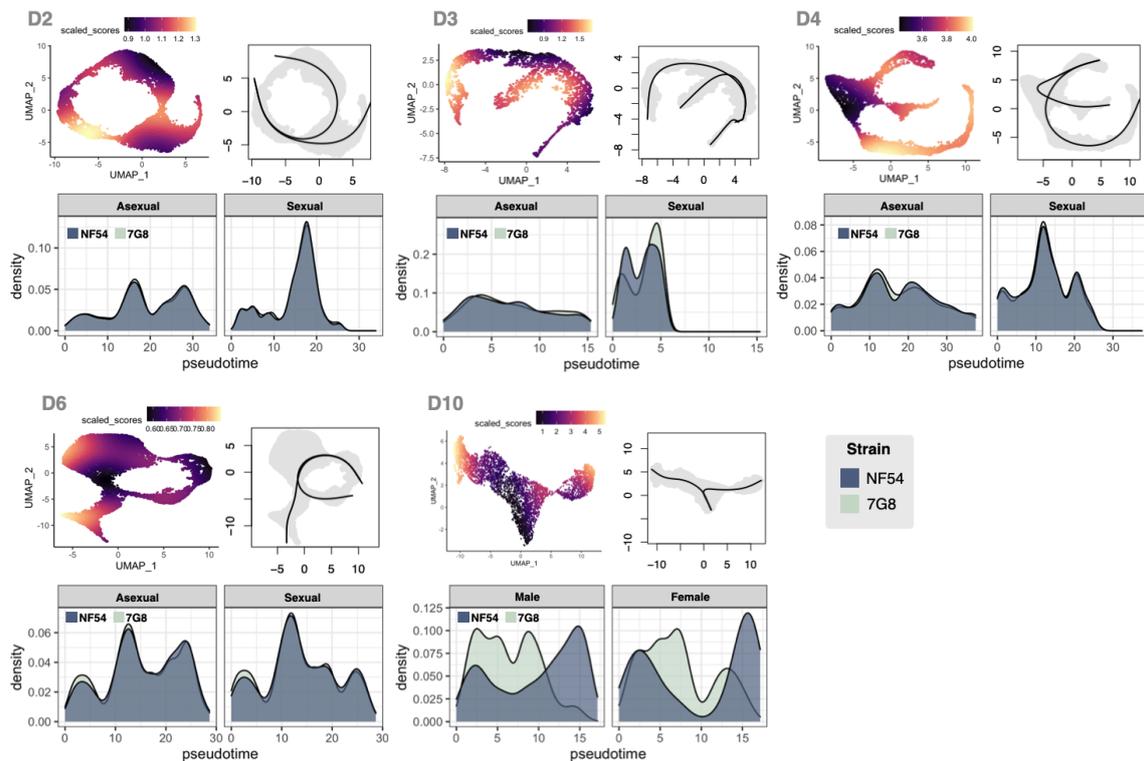
Before I began to investigate transcriptional differences between the two strains, I first wanted to ensure their distributions were not unbalanced in a way that would lead to a significant difference in expression between strains based solely on proportional differences. I used the tool `condiments` (de Bézieux et al., 2021) to assign an imbalance score, which considers both the local and global cell neighbourhoods between conditions to score each cell by how 'out of balance' it is as compared to the cells around it. This score is then smoothed across all cells in order to identify regions in which the imbalance in conditions (in this case, strains) is greatest. I

found that while all data sets contained some level of observed imbalance for the most part imbalance scores remained low, with the exception of days 4 and 10, where the total proportion of strains captured overall differed the most (**Fig. 4.2**). I further tested these imbalances by fitting pseudotime trajectories to the asexual/sexual (or male/female, in the case of D10) development paths present in each sample using `slingshot` (Street et al., 2018). From there, I was able to visualise the abundance of each strain as they progressed along these trajectories (**Fig. 4.2**). All days passed ( $p > 0.05$ ) a `topologyTest` as implemented in `condiments`, which aims to assess whether conditions differ significantly in how they progress along pseudotime to determine if separate pseudotime trajectories would be more appropriate to fit. The days 2,3,4, and 6 samples also showed no evidence for significant differences in strain abundance over pseudotime. The day 10 sample failed ( $p < 0.05$ ) the latter of these tests, as it displayed a large imbalance in stages at the very ends of the male/female trajectories. To remedy this, I removed mature male and female gametocyte populations from D10 for which the imbalance score was greater than 4, containing  $> 90\%$  NF54 cells.

#### 4.2.2 Differential expression in gametocyte-specific genes between strains

Differential expression analysis of all cells (assigned to a sexual trajectory in each sample) using `MAST` (Finak et al., 2015) revealed very few differences in gene expression between the two strains, implying a high degree of conservation in the global processes underlying gametocytogenesis, which is similar to what has been observed in asexual parasites (Llinás et al., 2006; Tarr et al., 2018). Owing to both differences in overall transcript capture efficiency and relative strain imbalance, signals of differential expression between strains were difficult to detect in samples collected with 10X version 2 chemistry (D1,D2,D3,D6) when combined with D10, which was collected with version 3 chemistry, and thus their analyses were separated.

In total, only four genes reached significant ( $p < 0.05$ ) differential expression between strains in days 2,3,4, and 6. 7G8 cells (largely in asexual cells preceding sexual differentiation) were enriched for expression of *kahrp*, knob-associated histidine rich protein, essential for knob-formation and cytoadherence (Crabb et al.,



**Figure 4.2 Assessing strain imbalance in each sample**

Assessment of strain imbalance across 5 sampling days. For each day, *condiments* (de Bézieux et al., 2021) was used to calculate an imbalance score between the distribution of NF54 and 7G8 cells across a global UMAP (top left, per day). *slingshot* was then used to fit pseudotime trajectories to cells from both strains (top right, per day) such that the distribution of each strain could be visualised over pseudotime (bottom, per day).

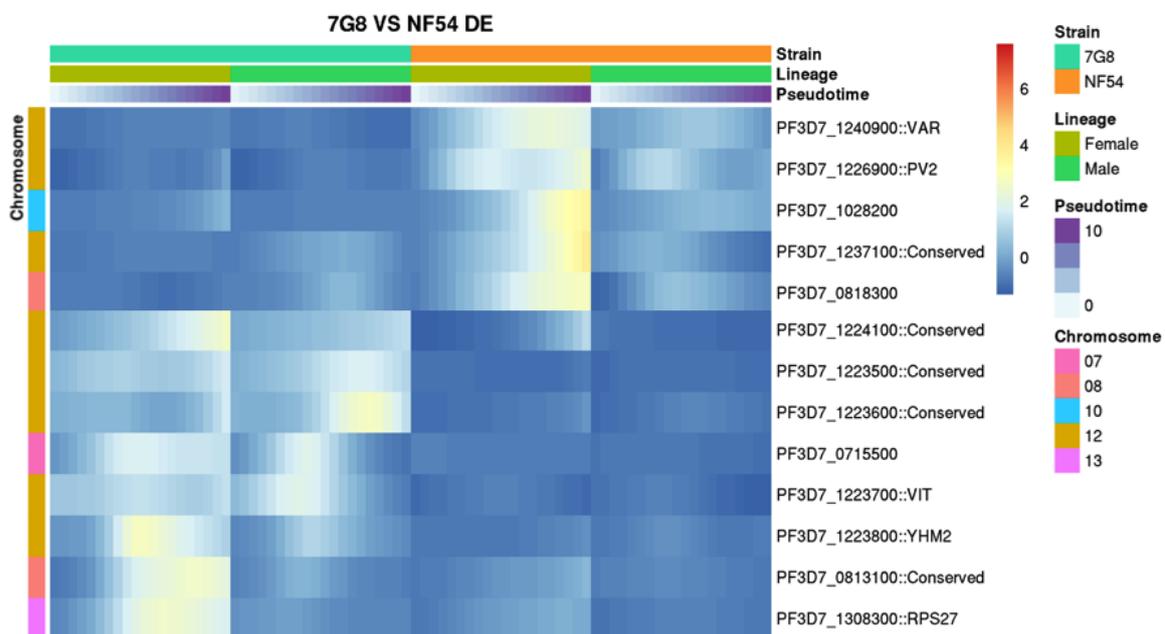
1997), but the average log-fold change was low ( $< 0.5$ ). The promoter for this gene has been shown to vary by copy number depending on strain, with 7G8 containing the most copies (6, as compared to 2 in 3D7) of the strains examined (Otto et al., 2018). In addition, PF3D7\_1327300 showed enriched expression in 7G8 gametocytes. This gene has been previously demonstrated as being upregulated in gametocytes (Lu et al., 2017), and a study comparing clinical and lab isolates of *P. falciparum* found it to have higher expression in laboratory-adapted strains as compared to recent field isolates (Tarr et al., 2018). In NF54 gametocytes, I observed higher expression of PF3D7\_0417300, a putative LETM1-like protein, and *pfg27/25*, an aforementioned marker of early gametocyte development and also shown by Tarr et al. to vary in expression by strain. While significant, the log-fold change in expression between strains for these two genes were very low ( $< 0.2$ ).

The improved depth of capture in the D10 data processed with 10X version 3 chemistry allowed for greater power in differential expression analysis between strains. As mentioned previously, the D10 sample was first filtered to contain only gametocyte stages, and then again to remove the 'tips' of the male and female branches where the imbalance between NF54 and 7G8 cells were the greatest. An initial MAST run with the same detection thresholds as were used with the D2-4,6 samples (which were relatively generous, requiring only a  $> 0.1$  log-fold change in expression) returned  $> 1700$  differentially expressed genes between conditions, with  $> 1100$  of them significant. To narrow this list down, generalised additive models (GAMs) were fit to describe the expression of each of these significant genes in each strain over pseudotime using *tradeSeq* (Van den Berge et al., 2020) in order to detect genes that differed the most between strains but also over developmental time. An association test between gene expression, strain, and pseudotime identified 12 significantly differentially expressed genes in total (**Fig. 4.3**).

Five genes were enriched in NF54 cells (**Fig. 4.3**), and were all associated with mid-late stage female gametocyte development. These included a component of the parasitophorous vesicle membrane (PVM), *pfpv2*, which facilitates PVM rupture during parasite activation and egress (Sologub et al., 2011), a gene encoding a putative Zn-finger protein PF3D7\_1028200 also shown to vary in expression levels between strains 3D7 and HB3 (like 7G8, also originating from South America) by Kamaliddin et al. (2021), and two genes with unknown functions (PF3D7\_1237100 and PF3D7\_0818300) shown previously to be enriched (and expressing protein) in late females and ookinetes (López-Barragán et al., 2011; Silvestrini et al., 2010). A member of the VAR gene family PF3D7\_1240900 encoding the highly variable surface protein PfEMP1 was also found to be enriched in NF54 female cells.

In 7G8 cells, 8 genes were significantly enriched over pseudotime (**Fig. 4.3**). Enriched in 7G8 females were a ribosomal subunit *rps27* and a gene with an unknown function (PF3D7\_0813100), the latter appearing in a ChIP-seq pulldown experiment targeting binding to the transcription factor AP2-I (Santos et al., 2017) and confirmed via ATACseq (Toenhake et al., 2018). A putative epsilon subunit of ATP-synthase PF3D7\_0715500 was also expressed transiently in 7G8 early gametocyte cells. The most significant of these findings, however, was PF3D7\_1223500, a gene with an unknown function that showed  $> 2$  fold enrichment in 7G8 cells

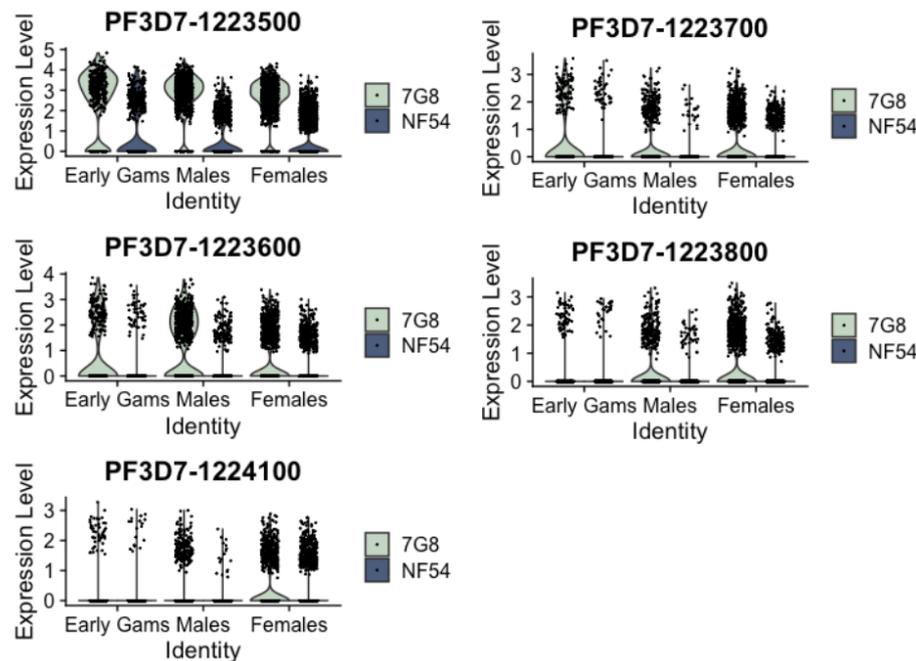
as compared to NF54 in this sample. Notably, this gene is located in a known segmental duplication in chromosome 12 of the 7G8 genome (Moser et al., 2020; Turkiewicz et al., 2020), and 4 other genes from this duplication were also significantly enriched in 7G8 in this sample, including vacuolar iron transporter *VIT* and citrate/oxoglutarate transporter *YHM2* (Fig. 4.4). This duplication also contains the *pfgch1* gene, which has been associated with sulfadoxine-pyrimethamine resistance (Turkiewicz et al., 2020), but this gene was not significantly enriched in 7G8 cells in this sample.



**Figure 4.3 Pseudotime-associated differentially expressed genes between NF54 and 7G8 gametocytes**

Heatmap displaying gene expression over pseudotime for male and female trajectories in the D10 sample, split by strain. Genes displayed are those that showed both a significant association with pseudotime and differed between conditions, as calculated by *tradeSeq* (Van den Berge et al., 2020). Trajectories were truncated in order to remove regions where NF54/7G8 imbalance was high ( $> 4$  imbalance score, as calculated by *condiments*). Genes (rows) are clustered hierarchically based on expression, and labelled by chromosome. This figure was made by Jesse Rop. Analysis was performed collaboratively.

Genes identified using *tradeSeq* captured nearly all of those identified by MAST (with an increased log-fold threshold of 0.5), save for four genes that were enriched in NF54 cells (Fig. 4.5). Interestingly, these contained two known markers of early gametocyte development, *pfs16* and *etramp4*. In the absence of genetic



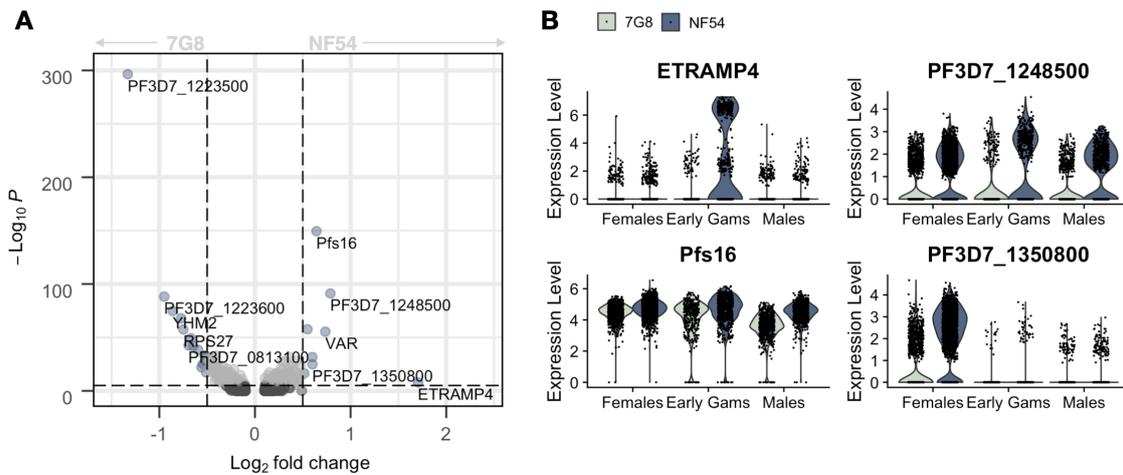
**Figure 4.4 Increased expression of genes contained in a segmental duplication in the 7G8 genome**

Violin plots demonstrating the distribution of log-expression values of five genes upregulated in 7G8 parasites and part of a segmental duplication on chromosome 12, split by strain and assigned stage. Width is proportional to the number of observations

perturbations or associated proteomics, it is difficult to ascribe an underlying mechanism as to why these genes were enriched in NF54 populations, and could be reflective of differences strain sexual conversion rates, as NF54 has been shown to display sexual conversion rates up to 4-fold higher than 7G8 lines *in vitro* (Reader et al., 2015). The remaining two genes identified by MAST as enriched in NF54 cells, PF3D7\_1248500, a putative Bax-1 inhibitor, and PF3D7\_1350800, which lacks a functional annotation, were both found to be enriched in gametocytes (Lasonder et al., 2016; López-Barragán et al., 2011; Silvestrini et al., 2010).

### 4.3 Discussion

Commitment to gametocytogenesis can vary widely between infections (Gebru et al., 2017; Usui et al., 2019), as well as in their observed sex ratios (Tadesse et al., 2019). In this chapter, I analyse single-cell RNAseq data for two strains,



### Figure 4.5 Differential expression of genes expressed during early gametocyte development

(A) All differentially expressed genes between 7G8 and NF54 as identified by MAST (Finak et al., 2015). Genes passing both the p-value ( $p < 0.05$ ) and log-fold change ( $> 0.5$ ) thresholds are coloured in blue and labelled. (B) Violin plots demonstrating the distribution of log-expression values of the four significantly differentially expressed genes identified by MAST that were not already detected by *tradeSeq*, split by strain.

representing geographically distinct populations (see 1.1 for locations), to explore how they differ in gene expression during sexual development. Similar to what has been observed previously (Chappell et al., 2020; Llinás et al., 2006; Tarr et al., 2018; Usui et al., 2019), the two strains exhibited a high degree of conservation in gene expression, and differences between strains were subtle. This demonstrates promise in using the NF54 gametocyte atlas described in Chapter 3 as a scaffold for mapping RNAseq from natural infections, as the trajectories in sexual development should be similar enough to allow for comparisons between gametocytes acquired from clinical samples. The variations that were observed included genes encoding transport proteins found to be duplicated in 7G8 along with others associated with drug resistance (Turkiewicz et al., 2020), as well as a number of early gametocyte markers.

As demonstrated in this chapter, improved capture efficiency can vastly improve sensitivity to these changes, and it will be important to ensure sufficient capture in the future in order to further explore these changes in expression level and how they may affect strain-specific sexual development. This is especially important

given the lack of biological replication present in the experimental design, where biological variation may present as inter-strain variation.

## 4.4 Materials & Methods

### 4.4.1 Parasite culture, sample preparation, and sequencing

Details regarding the culturing conditions and sampling strategies for these samples are described in Chapter 3. NF54 and 7G8 parasites were cultured separately for the entirety of the sampling period and mixed 1:1 at the time of sampling. Parasites were mixed according to parasitemia estimates calculated from manual counting of Giemsa-stained methanol-fixed smears and RBC densities estimated using a haemocytometer (Neubauer C-Chip, NanoEnTek). These strains were chosen to represent geographically distinct variation in gametocyte development (with NF54 likely originating in Africa, and 7G8 originating from South America). Both have been adapted to culture for many years, and thus while perhaps less clinically relevant, were less likely to simply pick up signals of lab adaptation rather than strain-specific variation in sexual development.

Details regarding sample preparation and single cell sequencing via 10X Chromium are the same as can be found in Chapter 3. All barcoded reads were mapped to the *Plasmodium falciparum* 3D7 reference genome v3 (January 2016)(<ftp://ftp.sanger.ac.uk/pub/genedb/releases/latest/Pfalciparum/>) using STAR as implemented in the 10X software Cell Ranger version 3.0.0, and reads assigned to barcodes using default parameters (`Cell Ranger count -id=.. -transcriptome=.. -fastq=..`). It is of course important to note that reference genome assemblies and associated annotation files exist for both NF54 and 7G8 and may have served as more appropriate references for read mapping and transcript quantification. My reason for choosing the 3D7 reference genome largely came down to the level of curation for each reference, as 3D7 remains the only reference genome consistently curated and updated to GeneDB and eventually PlasmoDB (Aurrecochea et al., 2009). In addition to this, annotations of both 7G8 and NF54 have been generated (manually, in the case of 7G8, and automatically, in the case of NF54) using the 3D7 genome as a reference, and neither have been updated to reflect subsequent updates to this reference. Personal correspondence with Dr. Ulrike

Böhme (who is involved in *Plasmodium* genome reference annotation for GeneDB) confirmed the v3 3D7 reference genome as the likely best choice for read mapping for reads originating from both strains. For reference, the 7G8 genome considered can be found at [https://plasmodb.org/plasmo/app/downloads/Current\\_Release/Pfalciparum7G8/](https://plasmodb.org/plasmo/app/downloads/Current_Release/Pfalciparum7G8/), and the same for NF54 at [https://www.ebi.ac.uk/ena/data/view/GCA\\_000401695.2](https://www.ebi.ac.uk/ena/data/view/GCA_000401695.2). All instances of Cell Ranger were run by Dr. Yong Gu as part of Sanger Sequencing Pipelines.

#### 4.4.2 SCT Strain assignment and doublet removal

Single cell transcriptome (SCT) filtering, quality control, library-size normalisation and inter-species doublet removal were performed on all SCTs, agnostic of strain. The rationale behind this decision was to prioritise comparability between strains, and because I did not believe, given the level of global conservation in gene expression between strains found previously (Chappell et al., 2020; Llinás et al., 2006; Tarr et al., 2018) that gene expression in these two strains, which were sampled and sequenced together in the same run, would differ so much that they would require separate quality control. As a sanity check, however, post strain assignment, I separated cells assigned to either strain and performed library size normalisation and highly variable gene analysis separately, and found no effect on expression matrices or selected variable genes.

At the point of SCT filtering but before doublet removal, barcodes were assigned strains (or identified as a mixture of the two strains) using `souporcell` (Heaton et al., 2020), which assigns cell to  $k$  clusters based on the variants detected in the cDNA reads for each cell. The tool was run with additional of `-k 2 -p 1 -known_genotypes` to specify 2 desired clusters (corresponding to each strain), a haploid genome, and the inclusion of known genotypes (without this parameter, the tool will predict clusters *de novo* from the observed variants). Variant call files (VCFs) supplied were obtained from the Pf6k data set (Pearson et al., 2019) (<https://www.malariagen.net/data/catalogue-genetic-variation-p-falciparum-v6.0>). Sample IDs '7G8' and 'PG0051-C'(3D7) were extracted from the data set using `bcftools view -S` (version 1.3).

As mentioned in Chapter 3, any cell identified by `souporcell` as containing variants corresponding to both strains were removed as doublets. The identification

of inter-strain doublets, variable gene selection, and dimensionality reduction were then performed on cells from both strains.

### 4.4.3 Differential expression analysis

#### Strain imbalance and estimating pseudotime

Imbalance scores were calculated for each sample separately using `condiments v1.0.0` (de Bézieux et al., 2021). Cell clusters for each day, obtained previously in Chapter 3 using Louvain clustering with multilevel refinement based on the calculated UMAP embeddings, as implemented in `Seurat` version 3.2.2. These clusters, as well as the UMAP embeddings, were then used as input to `slingshot` version 1.4.0 (Street et al., 2018) to fit asexual and sexual trajectories for each day (or male/female trajectories for D10) and assign each cell a relative pseudotime value. 'Start' and 'end' clusters (due to the nature of the asexual ring) were manually annotated using `start.clus` and `end.clus`.

#### Differential expression over pseudotime

Generalised additive models (GAMs) were fit to the expression of each gene over pseudotime, split by lineage (asexual vs sexual, or male vs female) as well as condition (strain) using `tradeSeq` (Van den Berge et al., 2020) version 1.5.10. The number of knots required for fitting in each sample (in other words, the limits of non-overlapping 'windows' of pseudotime in which GAMs are estimated) were estimated using the `evaluateK` function estimated on 500 genes. Differentially expression genes between conditions were identified using an `associationTest`, implemented in `tradeseq`.

#### MAST

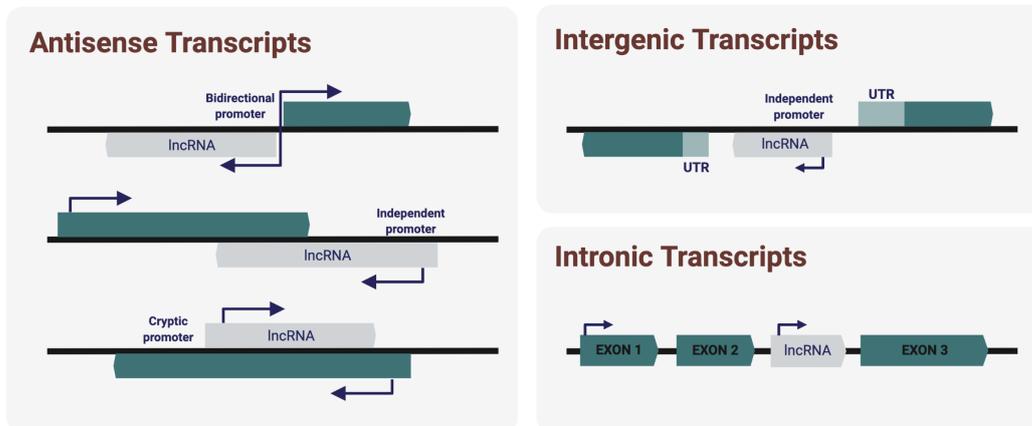
Genes expressed differentially between strains, independent of pseudotime, were calculated using `MAST` (Finak et al., 2015) as implemented in the `FindMarkers` function in `Seurat`, with a minimal log-fold threshold change of 0.1 (for 10X version 2 samples) or 0.5 (for the 10X version 3 sample, D10).

# Chapter 5

## Investigating long non-coding RNA expression in *Plasmodium falciparum*

### 5.1 Introduction

Non-coding transcripts are transcripts that show little to no evidence for coding potential, but instead may play a role in transcriptional regulation. Originally thought to be transcriptional noise, evidence over the past decade has demonstrated the significance of these transcripts in nearly all aspects of transcriptional regulation (Statello et al., 2021). They can be long (> 200bp) or short (< 200bp), and employ a great deal of diversity in the mechanisms by which they affect transcription. Among these, long non-coding RNAs (lncRNAs) are some of the best described. These transcripts can be polyadenylated and spliced like mRNA transcripts, but usually show greater specificity and lower average expression than mRNA (Azzalin et al., 2007; Krishnan and Mishra, 2014; Militello et al., 2005; Statello et al., 2021). lncRNAs can be classified based on their genomic location, secondary structure, size, cellular location, or function, which often lead to somewhat confusing overlapping/redundant annotations in the literature. Perhaps the most simplistic view places lncRNA into one of three classifications, described in (**Fig. 5.1**). Antisense lncRNAs (aslncRNAs), also called natural antisense transcripts (NATs), are those that sit antisense to a coding region. Intergenic long non-coding RNAs (lincRNAs) are those that do not overlap any known coding regions. Intronic lncRNAs sit within the bounds of a coding region and are expressed from the intron of these genes.



**Figure 5.1 IncRNA classification**

Classification of lncRNA. Antisense transcripts are those that overlap a gene on the anti-sense strand. They are often generated via a bidirectional promoter controlling the sense mRNA, via a cryptic promoter that sits within the bounds of the sense RNA gene body, or via their own independent promoter. In this chapter, I simply refer to these as 'asLncRNA'. Intergenic transcripts are those that do not overlap any known coding regions, including extended UTRs. In this chapter I refer to these as 'lincRNA'. Intronic transcripts are those that are expressed in sense to a coding region, and sit within an exon. Due to the difficult nature in differentiating these transcripts from sense transcription, they are not separately distinguished in the Hoshizaki et al. (in prep) data set.

A number of lncRNA-mediated regulatory mechanisms have been detected in *Plasmodium falciparum* and other apicomplexan parasites (Li et al., 2020). The best characterised examples in *P. falciparum* implicate lncRNA transcripts in *var* gene silencing and activation, as well as in the regulation of sexual conversion to gametocytes, two of the most important processes in *Plasmodium* virulence and transmission. Annotating the putative locations of lncRNA regions in the genome were largely carried out by RNA tiling arrays or strand-specific RNAseq (Broadbent et al., 2015, 2011; Kensche et al., 2016; López-Barragán et al., 2011; Siegel et al., 2014). A recent RNAseq protocol was developed to access the underrepresented AT-rich areas of the *P. falciparum* genome, and found a great deal of these annotations to be previously undiscovered UTRs, rather than non-coding transcripts (Chappell et al., 2020). In light of this, the location of putative lncRNAs have been re-annotated using the updated UTR and TSS (transcription start site) models from Chappell et al. (2020) and (Adjalley et al., 2016) in combination with RNAseq evidence from

two full-length RNAseq datasets generated via Oxford Nanopore (ONT) long-read sequencing (Hoshizaki et al., *in prep*).

Previous studies found lncRNA expression in *P. falciparum* to be highly stage specific, implying they may have a role in regulating the 'cascade-like' expression also observed in mRNA (Broadbent et al., 2015; Siegel et al., 2014). Given many of these relationships may have instead been simply detection of UTR expression, confirmation of the stage-specificity of non-coding transcription in *P. falciparum* is needed. Single-cell RNAseq has been previously shown to be useful in profiling the expression of lncRNAs in a highly time-resolved manner, and has been used to implicate lncRNAs in T-cell development in cancer (Luo et al., 2020), hematopoietic stem cell differentiation (Zhou et al., 2019), and stage-specific expression during embryonic development in mice (Zhang et al., 2014).

In this chapter, I profile the expression of these updated lncRNA annotations across the genome as they are expressed in my single-cell data set of *P. falciparum* asexual and sexual development in order to better understand their stage-specificity, how they may be regulating development, and to identify targets for further functional investigation.

**Personal contributions:** I assisted with the maintenance of all parasite cultures and single-cell preparations. I performed all analysis, unless otherwise listed.

**Additional contributions:** The project was conceived by Dr. Mara Lawniczak, Dr. Sunil Kumar Dogga, and myself. Dr. Sunil Kumar Dogga assisted with single cell preparations. Library preparations and sequencing were performed by the Wellcome Sanger sequencing pipelines facilities.

## 5.2 Results

### 5.2.1 Generation of a data set with improved depth for lncRNA detection

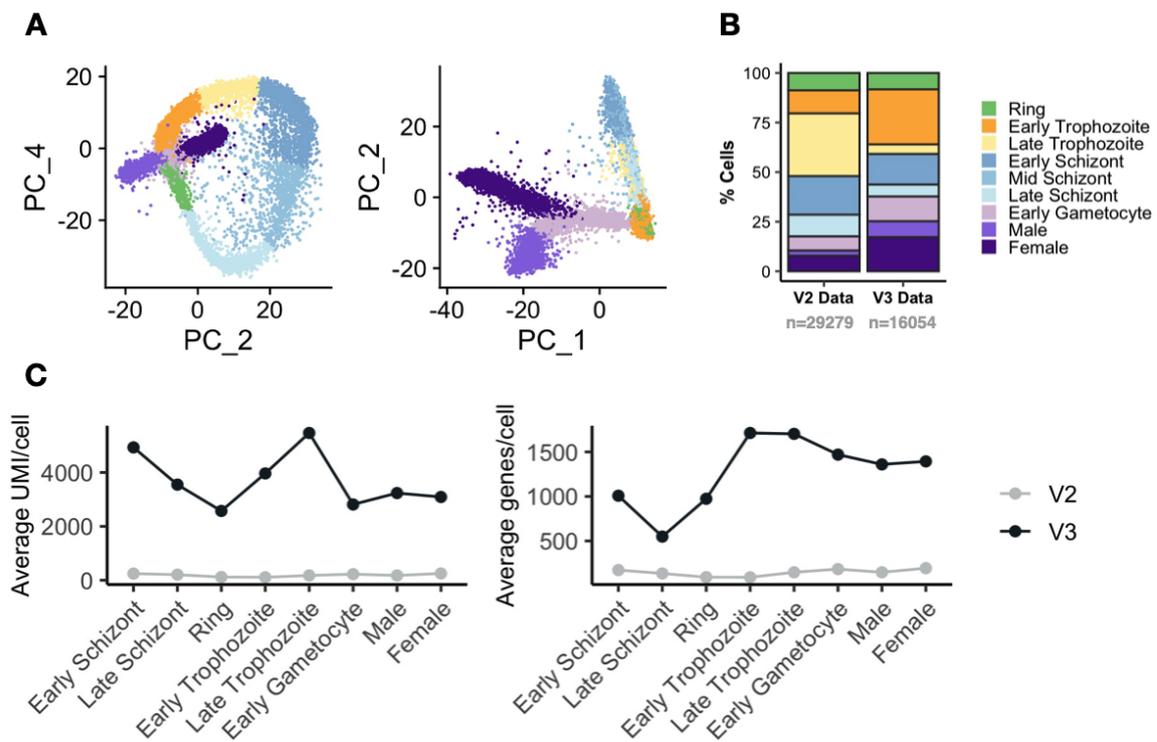
Given the relatively low abundance of lncRNA in comparison to mRNA, the difficulties plaguing cell and transcript capture in scRNAseq of malaria parasites are only heightened when exploring the non-coding transcriptome. To improve my chances

of capturing tangible lncRNA expression, I generated a second, smaller version of the gametocyte atlas using the newer 10x Genomics Chromium 3' V3 chemistry, which promises more efficient cell capture and library efficiency than the previous chemistry. This chemistry has also been shown to detect a greater proportion of lncRNA as compared to the plate-based Smart-seq2 in two human cancer cell lines (Wang et al., 2019).

My experimental design for this second data set (V3 data) was a pared down version of the first (V2 data), containing 3 samples taken on days 1, 5, and 10 post-induction, in order to capture lncRNA expression throughout asexual and sexual development, and only included one strain (NF54). I processed these data in the same manner as the original atlas; first performing quality control, normalisation, and doublet removal on each sample individually, and then merging them into one object. Stages were assigned as previously by mapping to *P. berghei* cells collected using Smart-seq2 from the Malaria Cell Atlas (Howick et al., 2019) (**Fig. 5.2a**). In the case of gametocytes (that are less well represented in the Atlas), I used the marker genes *pfs16* (PF3D7\_0406200) to mark early gametocytes, *pfs25* (PF3D7\_1031000) to mark females, and PF3D7\_0905300, a component of the dynein heavy chain, to mark males. The total distribution of stages were similar across the two data sets, albeit with a slightly increased proportion of gametocytes in the V3 data (**Fig. 5.2b**). As expected, I observed substantially better sensitivity in V3 samples. Regardless of stage, which can vary in mRNA abundance, the average number of both transcripts and genes captured per cell were considerably higher using V3 chemistry, with an average gain of approximately 2000 UMIs/cell and 800 genes/cell across stages (**Fig. 5.2c**). The improved sensitivity allowed me to capture a greater number of lncRNA transcripts per cell and I continued with this data set as the focus of my analysis.

### 5.2.2 Mapping and detection of lncRNA in scRNAseq data

All 2223 putative lncRNA regions were mapped to the *P. falciparum* genome using STAR as part of the 10x Genomics Cell Ranger pipeline, in a gtf file containing a combination of both lncRNA annotations and coding sequence annotations. Of the annotated regions, I did not find them to be biased by strand but indeed by location,



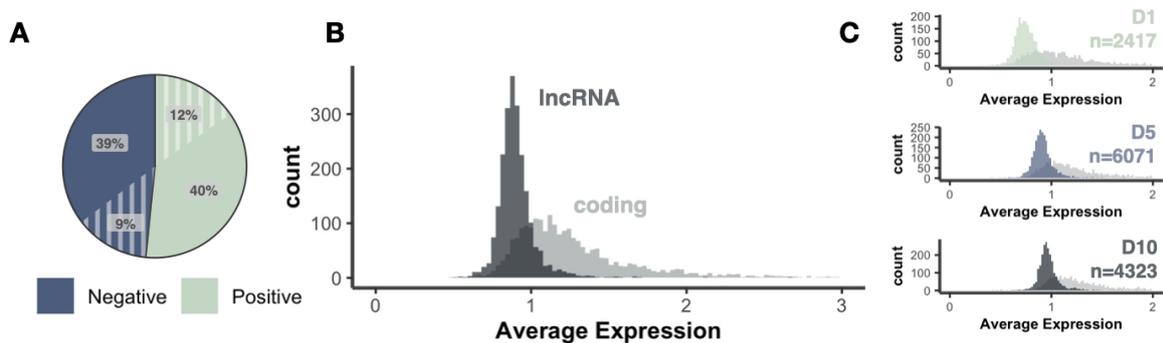
**Figure 5.2 New scRNAseq chemistry improves mRNA detection**

(A) Principal component projection of all cells collected as part of the V3 experiment, coloured by life stage as assigned by mapping to the *P. berghei* cells from the Malaria Cell Atlas (Howick et al., 2019) via `scmap-cell`. Stage colours correspond to those outlined in (B). (B) Composition of V2 and V3 data sets by life stage. (C) Average UMI captured per cell (left) and genes detected per cell (right) across life stage for both V2 and V3 data sets.

with most lncRNAs detected lying anti-sense to a gene, and fewer to intergenic regions (Fig. 5.3a).

I first checked whether the expression detected in lncRNA regions were biased by AT content (which may arise due to PCR bias), length (which normally isn't considered a substantial problem in UMI-based protocols, but has not been well investigated in non-coding transcripts), or strand. Because total expression can be affected by the proportion of cells expressing the transcript (i.e., heavily specialised transcripts expressed to a high degree in a low number of cells will have a low total expression count), I instead calculated the average expression for a lncRNA in all cells for which the log-normalised expression value in that cell was greater than 0. Thus, it is simply a measure of the average expression in cells expressing

that lncRNA, and is less biased by cell proportions. In doing this, I also removed 38 annotated lncRNA for which no expression greater than 0 was detected in any cell. To compare, I calculated the same values for a random sample of coding genes that was the same size as the number of lncRNAs. As expected, the average expression of all lncRNA investigated was lower than that of coding regions (**Fig. 5.3b**), a phenomenon which has also been detected in mammals (Cabili et al., 2011). These values did not differ substantially between batches (**Fig. 5.3c**).

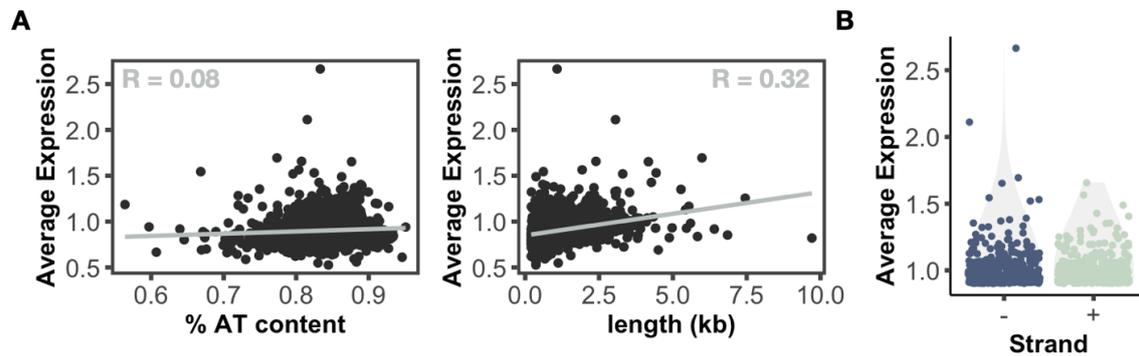


**Figure 5.3 lncRNA detection and expression**

(A) Proportion of total lncRNA annotations from the positive and negative strands. Solid quadrants denote lncRNAs that lie antisense to a gene, and hashed quadrants denote lncRNAs that lie in intergenic regions (B) Average expression of all mapped coding and non-coding regions (C) Expression values from (B), split by data set.

Expression remained largely independent of lncRNA AT content but showed a slight positive correlation with lncRNA length (**Fig. 5.4a**). This may be due to a gene-length bias, which is thought to be more of a problem in full-length RNAseq protocols rather than UMI-based protocols, but could have also occurred in this instance if there were an abundance of internal priming events that took place during reverse transcription within the highly AT-rich areas that exist outside coding regions where lncRNAs are found. The correlation ( $R=0.32$ ) was low, however, and not statistically significant ( $p > 0.05$ ). In addition, the highest average expression ( $> 2.5$ ) existed in regions less than 5kb in length, leading me to carry on with my analysis without correcting for lncRNA length. I also found no evidence for bias in lncRNA expression by strand (**Fig. 5.4b**).

Once I was more confident that detected lncRNA levels were not an artefact of an obvious technical fault, I interrogated their expression levels and how they related to their overlapping (in the case of anti-sense lncRNAs), or nearest (in the

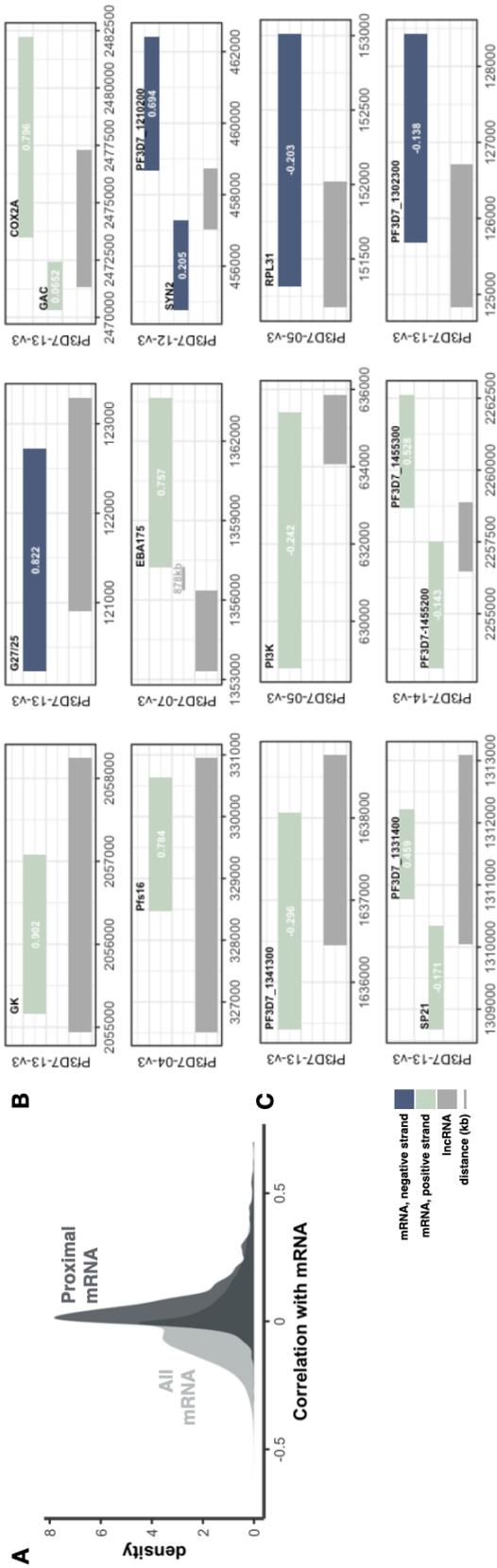


**Figure 5.4 Average lncRNA expression is largely independent of AT content, strand and length**

(A) Association of AT richness and lncRNA length on average expression levels. R values were calculated using a two-sided Pearson's correlation test (B) Effect of DNA strand on average expression levels.

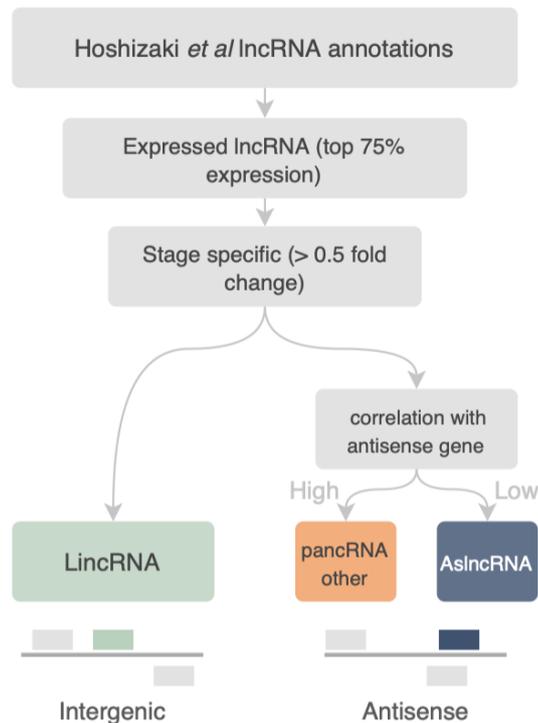
case of intergenic lncRNAs) gene. To do this, I randomly sampled 50 (this number was chosen to economise computational time) lncRNAs from both anti-sense and intergenic locations and correlated their expression (using a Pearson's correlation test) with their most proximal coding gene, and compared these to their correlation scores with all other genes. These two distributions are not totally independent due to a number of reasons including shared function/distribution, *cis*-acting effects with other proximal genes, or *trans*-acting effects on distal genes. Despite this, it was clear that proximal (< 500bp apart) lncRNA:mRNA pairs, on average, displayed more highly correlated expression than more distal (> 500bp apart) lncRNA:mRNA pairs (Fig. 5.5a). This trend, as well as the relative distribution of correlation values, is consistent with other investigations of lncRNA regions in single cell data in mammalian systems (Zhou et al., 2019). Of the highest lncRNA:mRNA correlates, most were in lncRNA that sat directly anti-sense to one or two genes (Fig. 5.5b). Where there were two proximal genes (like in the case of an aslncRNA in between *cox2a* and *gac*), it was often the case that the lncRNA would correlate with only one of the genes, and often the gene upstream on a different strand to the lncRNA, suggesting that the lncRNA expression may be associated with a bidirectional promoter. It wasn't always the case, however, that lncRNA expression was correlated with that of the nearest gene, as seen in Fig. 5.5c. Thus, while some of the lncRNA expression I detected may result from a potentially artefactual or non-biologically relevant mismatching of reads from the opposite

strand, or as the result of a leaky or bidirectional promoter, this wasn't exclusively the case, and the majority of lncRNA:mRNA pairs showed correlation scores > 20%. I therefore decided to keep lncRNA that displayed a very high correlation with their neighbouring gene in my analysis, and instead re-classify them as pancRNAs, or potentially promoter-associated lncRNAs.



**Figure 5.5 IncRNA expression is correlated with proximal mRNA expression**  
**(A)** Distribution of two-sided Pearson's correlation scores of lncRNA expression with all mRNA (light grey) or mRNA that are directly anti-sense (in the case of asLncRNA) or most proximal (in the case of lincRNA) to the lncRNA. **(B)** Genomic location of the top 6 most highly correlated lncRNA:proximal mRNA pairs, with the lncRNA on the bottom of each plot in grey, and proximal gene(s) above, coloured by strand (light blue = positive strand, dark blue = negative strand). Pearson's correlation with lncRNA expression are shown in each gene block. Where a lncRNA is intergenic, the distance between the lncRNA and gene body is shown by a grey line with the distance in kb above it. The chromosome for which each lncRNA:proximal mRNA pair is located is listed on the left of the plot. **(C)** same as (B), with lowest 6 correlated lncRNA:proximal mRNA pairs.

I began to narrow my analysis on lncRNA that may be biologically relevant before classifying them based on their genomic context and correlated expression. As described in (Fig. 5.6), I first removed any lncRNA that sat in the bottom 25% of average expression values across all lncRNAs. I assumed that while these regions could still contain biologically important features, their expression was too low to perform meaningful analysis and were difficult to distinguish from background noise. Of the remaining 1669 regions, I then used MAST (Finak et al., 2015) to identify lncRNAs that were significantly (adjusted p-value < 0.05) differentially expressed in one stage and displayed greater than a 0.5 log fold-change in their expression, compared to the rest of the stages. 194 lncRNA showed stage-specific expression in this way, and nearly 70% of them (135) were also detected as highly variable genes (HVG) (where the HVG analysis was performed using Seurat's vst method, on the combined V3 object using all lncRNA and mRNA counts, with n=2000). I then classified these regions into one of three options: 1) lincRNA (long intergenic non-coding RNA), or those for which I did not detect another coding gene for > 500bp on either side, on either strand, 2) asLncRNA (anti-sense long non-coding RNA), those that overlapped a coding gene on the opposite strand or sat < 500bp away from one on the opposite strand and displayed a low (< 0.6) correlation with their expression, or 3) putative pancRNA (promoter-associated non-coding RNA), for those that sat anti-sense to, and displayed highly (> 0.6) correlated expression with a coding gene. The annotated boundaries used for coding regions included the extended UTR models from Chappell et al. (2020). This, along with the extended 500bp window, caused some previously annotated intergenic lncRNAs to be re-classified as asLncRNAs. In total, I detected expression for 139 asLncRNAs, 32 LincRNAs, and 23 pancRNAs. It is important to note that classification as a pancRNA does not imply that the observed correlated expression is definitively the result of a leaky, cryptic, or bidirectional promoter, but rather a label to assign a certain level of caution when interpreting the expression levels of these regions, as they may be more likely to represent technical error, cDNA priming from sense transcripts during reverse transcription, a yet undiscovered UTR, non-functional expression, etc. Even in the face of these, the correlation with a neighbouring gene could still be explained by an underlying biological relationship between the lncRNA and the gene, and thus I chose not to remove them completely from the analysis.



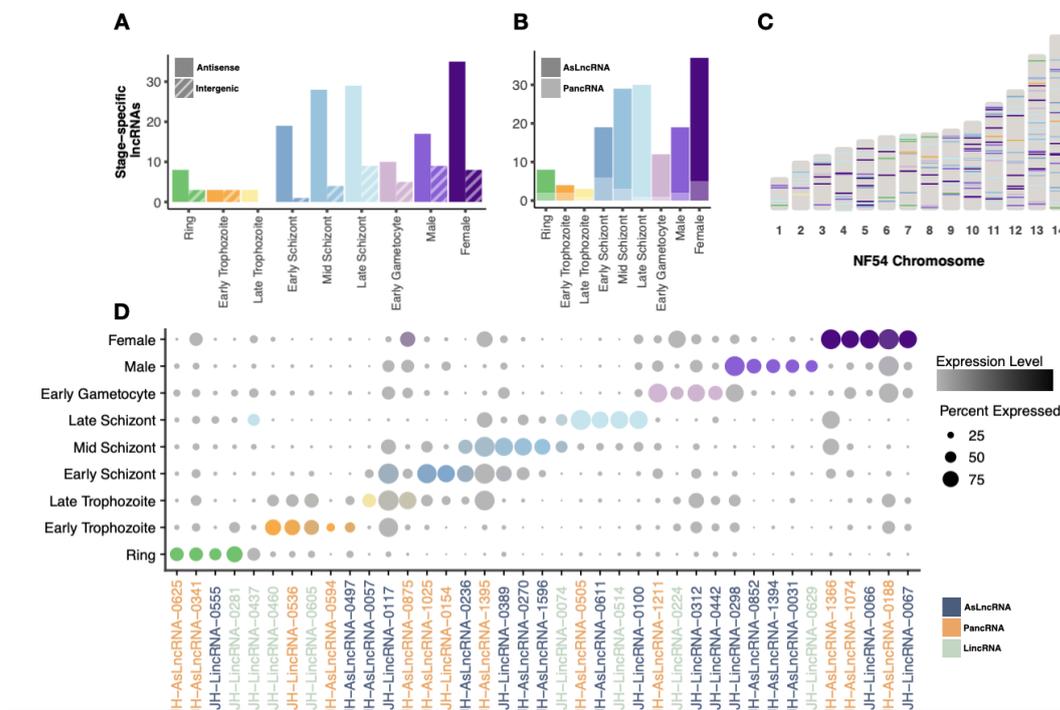
**Figure 5.6 Schematic for lncRNA classification**

Workflow diagram for lncRNA classification. lncRNAs are first filtered by average expression (top 75%) and stage-specificity ( $> 0.5$  fold change in expression as measured using MAST). lncRNAs sitting anti-sense to a gene are then further subdivided into those highly correlated ( $> 0.6$  Pearson's correlation score) with adjacent mRNA expression, labelled putative promoter-associated lncRNAs (pancRNA).

### 5.2.3 lncRNAs show stage-specific expression across the *Plasmodium* life cycle

Once filtered and re-classified, I then began exploring the regions in which lncRNAs displayed localised expression. Of the original 1669 annotations with sufficient average expression, 194 showed significant stage-specific expression. Stage specificity was largely localised to the schizont and gametocyte stages, in both anti-sense and intergenic regions (**Fig. 5.7a**). Reassuringly, potential pancRNAs did not make up the majority of stage-specific anti-sense lncRNAs, nor did I observe an enrichment in any particular stage (**Fig. 5.7b**). It is not possible to determine if this represents a greater propensity for lncRNA usage and expression in these stages or instead an increased likelihood of capture. Interestingly, however, trophozoites, which

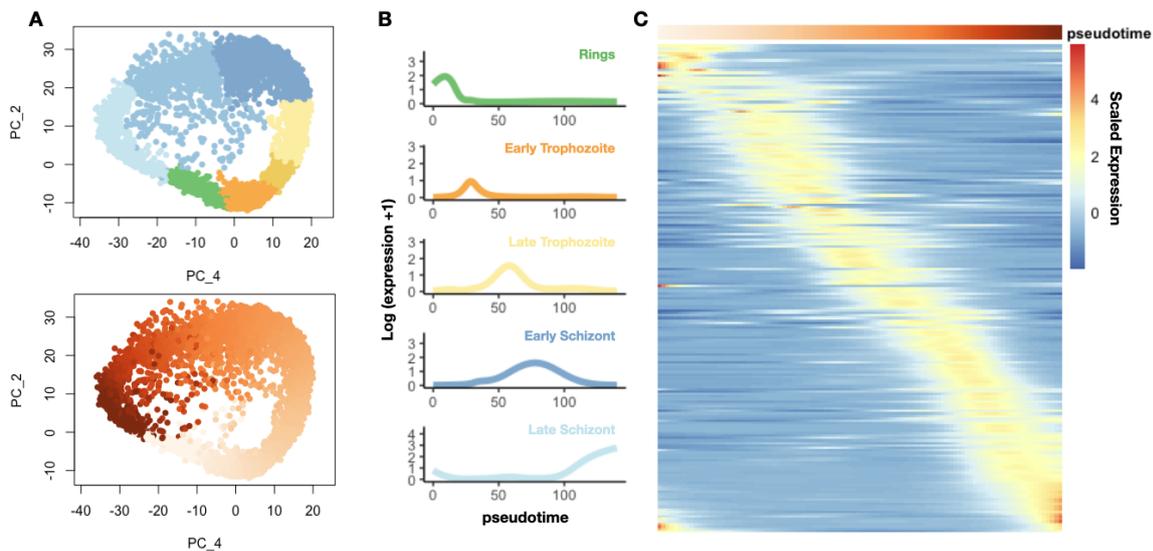
make up the largest proportion of stages in the V3 data and displayed the highest average nUMI and nGenes/cell rates across stages, had some of the lowest number of stage-specific lncRNAs. I did not observe any substantial bias towards chromosomal location in these lncRNAs, save for a small clustering of schizont-specific lncRNAs in the middle of chromosome 13. Of the stage-specific lncRNA markers with the highest fold-change in expression, only about a third were classified as putative pancRNAs, confirming that the observed stage-specific expression is not dominated by potential artefacts of stage-specific mRNA expression (**Fig. 5.7d**).



**Figure 5.7** IncRNA show stage specific expression

(A) Number of lncRNA detected with greater than 0.5 log fold-change (calculated using MAST) in expression in each life stage, split by where the lncRNA was detected (anti-sense to a gene, or in an intergenic region) (B) Same as in (A), but subsetted to just include anti-sense lncRNA, split by classification type. Lighter colours were re-classified as putative pancRNAs, darker remained asLncRNA (C) Location of lncRNA across the *P. falciparum* NF54 chromosomes. Lines represent lncRNA, coloured by stage as seen in (A) and (B) (D) lncRNA expression of the top 5 lncRNA per stage with the greatest fold change in expression. Dot size is proportional to percentage of cells in each stage expressing the lncRNA, and colour intensity is proportional to scaled expression values, and coloured by life stage. lncRNA labels are coloured by reclassification type (Note: these may differ from the initial naming convention).

It has been previously demonstrated that coding gene expression throughout asexual development in the IDC is tightly regulated in *P. falciparum*, such that gene expression cascades are switched 'on' and 'off' as a parasite develops from ring to schizont (Reid et al., 2018). This cascade approach has also been demonstrated to be likely true of lncRNAs in bulk RNAseq (Broadbent et al., 2015). To leverage the resolution of single cell data in order to more accurately order lncRNA within their respective stages, I used `slingshot` (Street et al., 2018) to order cells according to their relative position along asexual development, and then `tradeSeq` (Van den Berge et al., 2020) to fit a generalised additive model (GAM) to represent the expression of each lncRNA along this pseudotemporal trajectory (**Fig. 5.8a**). This allowed me to order all asexual lncRNAs based on the pseudotime value corresponding to the peak density of the GAM, or in other words, order the lncRNAs by the timing of their peak expression, even within those that had been assigned the same stage (**Fig. 5.8b**). I observed remarkable specificity in the timing of peak expression of each lncRNA, demonstrating the same cascading effect seen in mRNA transcription, confirming, and adding greater temporal resolution to, the findings of Broadbent et al. (2015).

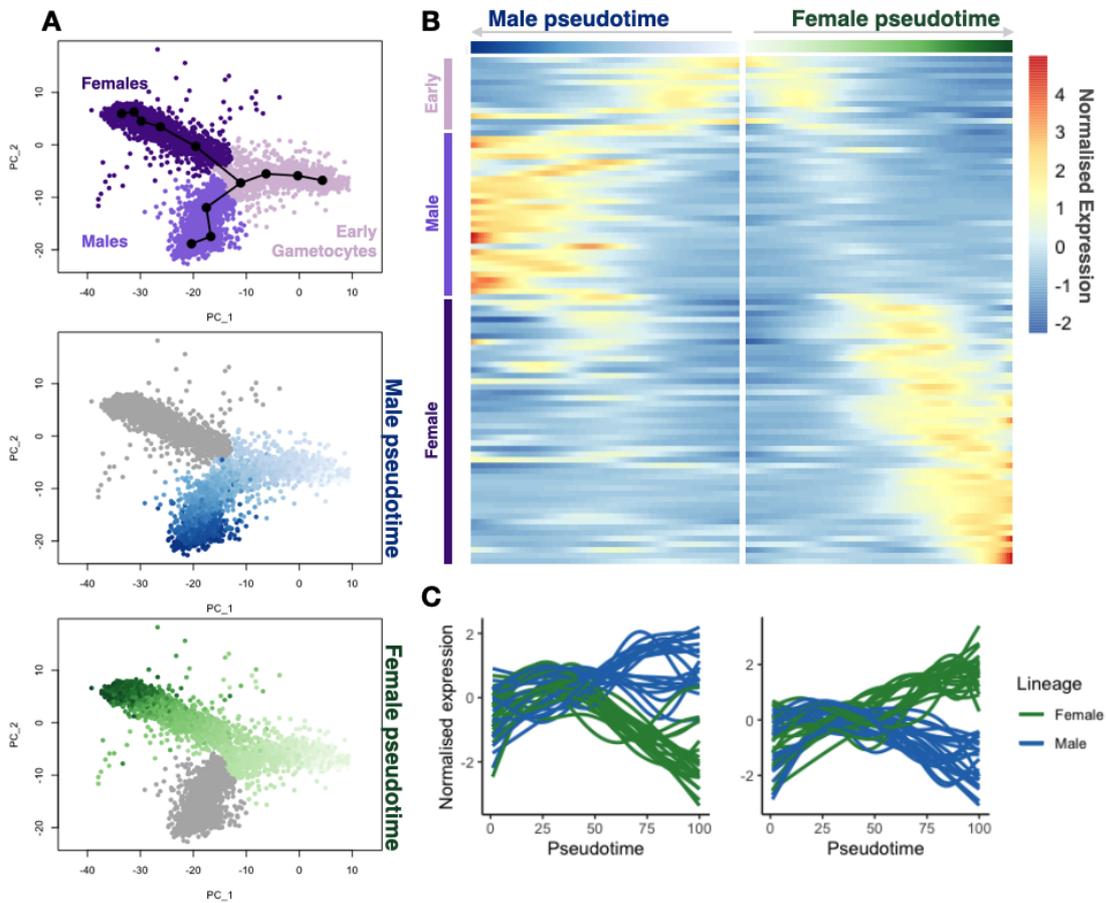


**Figure 5.8 IncRNA exhibit cascade-like peak expression throughout asexual development**

(A) Principal component projections of cells assigned to asexual stages in V3 data, coloured by stage (top) and by pseudotime (bottom). Stages colours correspond to the colours used in (B). Pseudotime was calculated using `slingshot`. Pseudotime colours correspond to those seen in (C) (B) GAM fit to the expression of the top IncRNA marker per stage, by fold-change in expression, over pseudotime. GAMs fit using the package `tradeSeq` (C) Heatmap of GAM models representing IncRNA expression (rows) over pseudotime windows (columns), beginning at rings on the left, and ending at late schizonts on the right.

I then carried out the same analysis on the sexual stages in the V3 data, this time using `slingshot` to order cells along two pseudotemporal trajectories, from early gametocytes into late males or late females (Fig. 5.9a). As done previously, the trajectories were simultaneously fit such that shared cells that were common to both the male and female trajectories (i.e., cells labelled as early gametocytes) would receive the same pseudotime values and thus could be compared on a similar scale. I again fitted models to each sexual stage-specific IncRNA along both the male and female trajectories using `tradeSeq`. Much like I observed in asexual stages, I found the peak densities of IncRNA expression were restricted to relatively narrow windows of pseudotime within sexual development (Fig. 5.9b). I also found many of these IncRNAs to be differentially expressed between male and female development. I used RSEC hierarchical clustering as part of the `clusterExperiment` package in R

to cluster lncRNA expression models that were differentially expressed between males and females in order to observe their expression dynamics over time. While I found that most sex-specific lncRNA, as in mRNA, increased in expression around the same time (corresponding to the bifurcation point in the PCA projection), there were subtle differences in when lncRNA began to increase, and to what degree they decreased in the opposite sex (**Fig. 5.9c**), which could represent clusters of lncRNAs that are used to regulate expression at different times during sexual development.



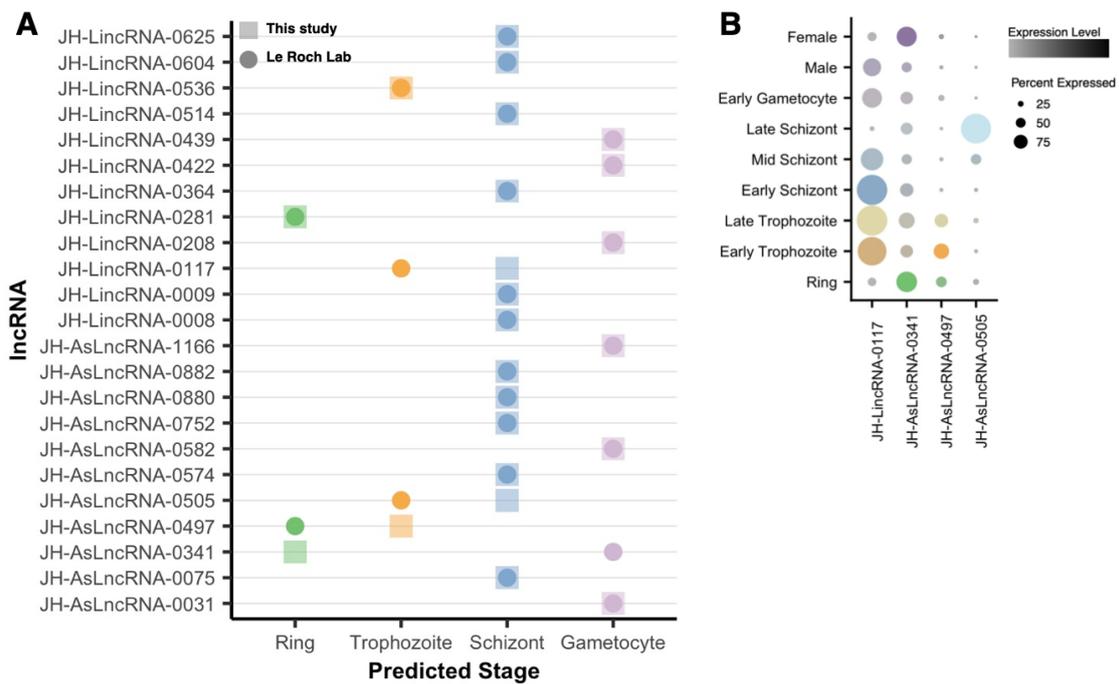
**Figure 5.9** IncRNAs are expressed during gametocyte development and sexual differentiation

(A) Principal component projections of cells assigned to asexual stages in V3 data, coloured by stage (top panel), and by pseudotime along a male (middle panel) or female (bottom panel) developmental trajectory. Stages colours correspond to the colours used in (B). Pseudotime was calculated using `slingshot` (B) Heatmap of GAM models representing lncRNA expression (rows) over pseudotime windows (columns), across male and female development. (C) GAM model fits for genes expressed differentially across male or female pseudotime as assessed by a `differentialEndTest` in `tradeSeq`. Genes were clustered using RSEC hierarchical clustering.

### 5.2.4 scRNAseq as a tool for selecting non-coding candidates of interest

The increased temporal resolution assigned to the expression of putative lncRNAs in *Plasmodium spp.* via scRNAseq can serve as an ideal initial screen for candidates for genetic modification or further study, especially as meaningful data can be obtained from re-mapping pre-existing single-cell transcriptomic data sets. To demonstrate this, I assessed the accuracy of the stage-assignments I predicted using my single cell data to predictions produced by collaborators in the Le Roch lab at UC Riverside, who performed bulk RNA sequencing on highly synchronised parasite cultures to determine the peak expression of a number of lncRNA regions. Of those that overlapped with the annotations used in this study and had stages assigned, I found that predictions from the scRNAseq data matched those of the Le Roch lab in over 80% of cases (19 agreed, of 23 total) (**Fig. 5.10a**). Of those for which there was a discrepancy in predicted stage, all but one (JH-AsLncRNA-505) showed evidence of co-expression in the stage assigned by the Le Roch lab (**Fig. 5.10b**), highlighting the difficulty often faced in assigning discrete stages to a continuous life cycle. Overall, however, these findings are encouraging, and suggest that the expression patterns of non-coding regions in single cell data can not only re-capitulate, but expand upon those obtained from bulk RNA sequencing.

I then created a short-list of locations of interest for further investigation into lncRNA function in gametocyte development. I found a number of reliably expressed, stage-specific lncRNA regions that lay proximal or anti-sense to a number of transcription factors or genes that have been shown to be important in gametocyte development (**Fig. 5.11**), including the master regulator of sexual development, *pfap2-g*, histone-modifying enzyme *pfsir2a*, early gametocyte markers (*pfs16*, *pfg27/25*, and *pfgexp22*), as well as sex-specific genes such as *dozi* in females and kinesin-8b (PF3D7\_0111000) in males. Some of these lncRNAs showed high correlation with the expression of the most proximal gene (such as *gk*), while others showed more dissimilar expression from their proximal mRNA pair. Correlated expression, or lack thereof, is not always reliable evidence for functional association. The lincRNA JH-LincRNA-0309, for example, sits about 2kb upstream of the *gdv1* gene on chromosome 9 and displays little correlation with detected *gdv1* mRNA expression levels (**Fig. 5.11**). In a recent ChIRP-seq experiment, however, collaborators in the Le Roch lab (manuscript in prep) found

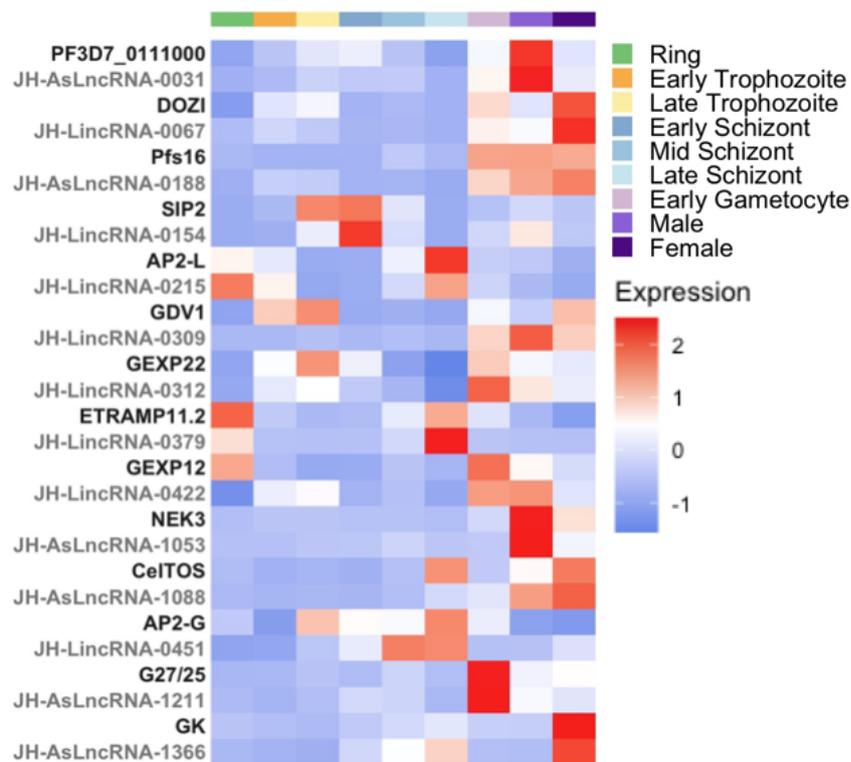


**Figure 5.10 Predictions for lncRNA stage-specific expression match those estimated from bulk RNAseq**

**(A)** Comparison of predictions for peak expression of lncRNA regions between this study (scRNAseq, squares) and a collaborative study with the Le Roch lab (bulk RNAseq, circles). **(B)** lncRNA expression of 4 lncRNA regions for which predictions disagree. Dot size is proportional to percentage of cells in each stage expressing the lncRNA, and colour intensity (coloured by stage) is proportional to scaled expression values.

two peaks associated with JH-LincRNA-0309, one of which is at the *gdv1* locus, and another at gametocytogenesis-implicated gene PF3D7\_035600, suggesting a possible role of this lncRNA in GDV1 modification, despite the lack of observed overlapping expression.

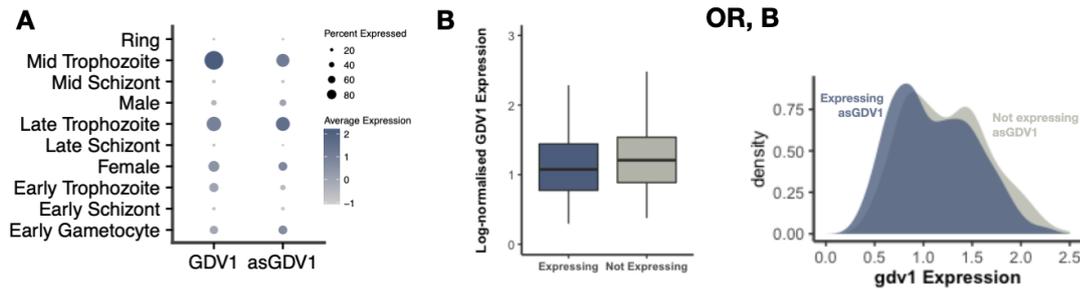
JH-LincRNA-0309 is of course not the only lncRNA associated with GDV1. As mentioned earlier, *gdv1* expression has been shown to be regulated by a lncRNA directly anti-sense to the *gdv1* locus (Filarsky et al., 2018). This transcript, termed here as '*asgdv1*' is already annotated in the current version of the *P. falciparum* 3D7 genome on PlasmoDB as PF3D7\_0935390, and thus was included in the gtf I used in order to explore its expression in my single cell data. The expression of *asgdv1* was localised, for the most part, to the same stages as sense *gdv1*, except for a reduction in expression of *asgdv1* in early to mid trophozoites as compared to



**Figure 5.11 IncRNAs are expressed in close proximity to coding regions implicated in gametocyte development**

Heatmap of scaled, log-normalised expression averaged by stage for selected mRNA:IncRNA pairs. Coding regions are coloured in black, with their paired proximal IncRNA region coloured below in grey

*gdv1* (Fig. 5.12a). To explore whether my expression data was sensitive enough to detect a shift in *gdv1* expression in the presence of *asgdv1*, I examined the *gdv1* expression levels in any cell in which *gdv1* was detected (i.e. cells where *gdv1* > 0 after log normalisation and correction). I then split these cells based on the presence (> 0 counts) or absence (= 0 counts) of *asgdv1* expression. I observed a slight decrease in overall *gdv1* expression in cells not co-expressing *asgdv1* as compared to those with detectable *asgdv1* expression, but the shift was relatively subtle (Fig. 5.12b). Given the technical noise and dropout rate in single cell RNA sequencing, it is difficult to ascribe this shift a biological cause, however, the mere detection of *asgdv1* in this data creates promise for use of scRNAseq for tracing the effects of future modifications of *asgdv1*.



**Figure 5.12 Expression of anti-*gdv1* antisense transcript can be detected in scRNAseq data**

(A) lncRNA expression of sense *gdv1* (PF3D7\_0935400) and anti-sense transcript *as-gdv1* (PF3D7\_0935390) across sexual and asexual life stages. Dot size is proportional to percentage of cells in each stage expressing the lncRNA, and colour intensity is proportional to average log-normalised, scaled expression values (B) *gdv1* expression of all *gdv1*-expressing cells (i.e., cells with > 0 detected *gdv1* transcripts) that do (blue), or do not (grey) co-express detectable (> 0 log-normalised counts) levels of *asgdv1*

### 5.3 Discussion

Overall, the results from this chapter demonstrate the promise in using scRNAseq data to explore the non-coding transcriptome of *Plasmodium falciparum* parasites. While this approach has already been exploited in single cell data sets from mice, humans, and other model systems (Liu et al., 2016; Luo et al., 2020; Shaath et al., 2021; Xie et al., 2018), it has yet to be explored in *Plasmodium spp.*, likely due to the small (but growing) supply of available data sets, the difficulty in efficient cell/transcript capture, and a relative lack of robust non-coding annotations (as compared to humans or mice). I propose that the approach I take here may serve as an example of a non-coding 'atlassing' step that can be added to analysis pipelines when generating or revisiting single cell RNAseq data in *P. falciparum*, or in the event of improved lncRNA annotations. The cell numbers that can be achieved with a droplet-based cell capture approach like 10x Chromium allow for this sort of screen to be done on a large scale, accessing both common and rare population subtypes or conditions with the numbers needed to robustly assess expression level perturbations. The advent of new approaches for sequencing full-length transcripts from single cell data (like single-cell IsoSeq from Pacific Biosciences, or Direct RNAseq from Oxford Nanopore) make this possibility even more powerful, potentially

overcoming the short-comings that can befall 10x Chromium's 3' amplification approach.

## 5.4 Materials & Methods

### 5.4.1 Parasite culturing and sample collection

*Plasmodium falciparum* parasites of the NF54 strain were maintained and gametocytes induced in the same manner as was described in Chapter 3. As was done previously in Chapter 3, to maintain heterogeneity flasks were synchronised in 12hr intervals (T-12, T0, T+12) and combined on the date of collection. Two sets of these three synchronisations were performed, such that each set was induced to initiate gametocytogenesis 72hr apart. Samples for single cell sequencing were taken from Set #1 on day 1, 5, and 10 post gametocyte induction, with Set #2 existing as a back-up in case of contamination. Four days following initial gametocyte induction, 10% heparin (Sigma, prepared at 1  $\mu$ L/mL) was added (Sigma) to flasks to prevent further asexual growth. For each sample, cultures from both strains across all three synchronisation flasks were combined such that equal numbers of parasites were added from each, using estimated parasitemia and RBC density as calculated via manual counting following staining with a 10% Giemsa stain on methanol-fixed slides. RBC densities were estimated using a haemocytometer C-chip (NanoEnTek). The parasitemia and RBC density of this new sample was then re-calculated before loading. All cultures for this experiment were maintained by Dr. Sunil Kumar Dogga, and samples were collected by myself and Dr. Sunil Kumar Dogga.

### 5.4.2 Single cell sequencing, barcode assignment and read mapping

Samples were loaded onto a 10x Genomics Chromium according to manufacturer's instructions, each in a separate inlet, with a 10,000-cell target recovery. Single-cell emulsions were prepared using a 10x Chromium Single Cell 3' Library v3 chemistry. Individual 10x libraries were then pooled and sequenced on a single lane of a NovaSeq 6000 (Illumina) using 150-bp paired-end sequencing. Library

preparation and sequencing were performed by the Sanger Institute sequencing facilities staff. Raw reads were assigned barcodes and UMIs using the `Cell Ranger` single-cell software (version 6.1) using default parameters.

Reads were then mapped first to the updated *Plasmodium falciparum* 3D7 reference genome, (PlasmoDB, release 53) (Aurrecochea et al., 2009), which incorporated the extended UTR models and annotations from Chappell et al. (2020) using the 10x Genomics `Cell Ranger` pipeline. This first gtf file contained only the protein-coding regions and additional annotations available already on PlasmoDB, termed for simplicity as the 'mRNA only' mapping. We then performed a second, separate mapping step for each sample, in which the same reads were mapped to a second gtf file that contained the original annotations from the PlasmoDB reference genome, but with lncRNA annotations provided by Johanna Hoshiaki as part of her ongoing PhD project in collaboration with Drs. Adam Reid and Lia Chappell (termed the 'mRNA + lncRNA' mapping). The reads for each sample were mapped using the same version of `Cell Ranger`, with the same parameters, save for the additional parameter `-attribute=gene_biotype:lncRNA` input to `Cell Ranger mkref` in order to allow for regions labelled as lncRNAs to be included in mapping. Transcriptome and genome annotation files for mapping were downloaded and prepared collaboratively by myself and Dr. Sunil Kumar Dogga, and all instances of `Cell Ranger` were run by Dr. Yong Gu as part of the Sanger Institute sequencing facilities.

### 5.4.3 SCT filtering, normalisation and annotation

For consistency with procedures carried out on the V2 gametocyte atlas in Chapter 3, I performed all quality control, filtering, and initial normalisation on the 'mRNA only' count matrices for each sample, generating a custom list of quality controlled and annotated cell barcodes to subset the 'mRNA + lncRNA' mapped files with. I chose to include the original protein-coding annotations in the second gtf with the lncRNA as it has been shown previously that lncRNA counts can be overestimated in their absence, due to the misassignment of reads that would have otherwise mapped to protein-coding regions to lncRNA regions (Zheng et al., 2019). I performed cell barcode filtering, doublet detection, normalisation and dimensionality reduction on

the data in the same way as was described in Chapter 3 using R v3.6.0 (R Core Team, 2019) and Seurat v3.2.2 (Stuart et al., 2019).

Due to differences in version chemistry, the following were performed differently as to what was described in Chapter 3. First, the improved capture efficiency of 10x Version 3 chemistry necessitated new cell quality filtering thresholds, and thus, I discarded any cell barcode which contained fewer than 1000 UMI/cell or fewer than 300 genes/cell. Unlike in the Version 2 data of Chapter 3, I observed a substantial proportion of cells matching the transcriptome of late schizont and ring stage parasites in *P. berghei* cells collected for the Malaria Cell Atlas (which I had previously had to 'rescue' from Cell Ranger's built-in cell white-listing feature due to lower UMI capture when compared to the rest of the stages). Due to this, I performed all quality control and filtering on any cell barcode that was initially white-listed by Cell Ranger.

Filtered transcriptomes (SCTs) for the 'mRNA only' mappings were log-normalised and corrected for discrepancies in library size using Seurat v3.2.2 individually by sample. A variance stabilising transformation (vst) was performed on each sample to identify the 750 most highly variable genes used to perform a principal component analysis. Cells were clustered according to the most significant PCs (as determined by a jackstraw analysis) using Louvain clustering with multilevel refinement via Seurat's FindClusters command with a resolution parameter of 0.4. Clusters were then manually assigned a life cycle stage based on the stage of the best matched cell from the Malaria Cell Atlas assigned to the majority of cells in the cluster and confirmed by collapsing the counts from each cluster into a pseudo-bulk sample and correlating this with stage-specific bulk RNAseq data from López-Barragán et al. (2011).

#### 5.4.4 Sample merging and lncRNA annotation

Individual V3 samples were first merged without correction into a single object using the merge command in Seurat. In the same way I did in Chapter 3, I also integrated the data using Seurat's IntegrateData function (described in Stuart et al.) which attempts to merge individual single cell data sets in a manner that corrects for batch effects whilst preserving biological variation. In comparing the two methods (merging the SCTs from each sample without correction, vs. integrating samples

and correcting their expression values for batch effects), I noticed that unlike in the Version 2 data, data that were simply merged showed very little evidence for clustering by batch. Because methods available to integrate single cell data are still imperfect and can regress out more subtle biological signals (Luecken and Theis, 2019), and seeing as my data sets were collected from the same parasite cultures and processed and sequenced by the same operators on the same machines, that I merged the data from each sample, rather than batch-correcting and integrating the data like I did with the original gametocyte atlas in Chapter 3.

Once the data sets were merged into a single object (which I call the 'mRNA V3 data'), I then processed the count matrices generated from the 'mRNA + lncRNA' mapping into *Seurat* and filtered and merged these objects to contain the same cell barcodes that were retained in the merged, filtered mRNA V3 data, and imported the same principal component embeddings. My rationale for doing so was so that I could assess the stage-specificity of the observed lncRNA values based on stages assigned from the protein-coding transcriptomes alone, without potential bias from lncRNA expression values. The count matrices for the mRNA + lncRNA mappings were, however, still log-normalised and corrected for library size separately to those of the mRNA only mapping, including both mRNA and lncRNA counts in this process. I then used this merged, filtered, and corrected data set in all downstream analyses.

#### 5.4.5 lncRNA filtering and further data analysis

All data were visualised using *ggplot2* v3.3.5 (Wickham, 2016) in R. The 'average expression' metric for lncRNA expression was calculated by averaging all cells with non-zero counts from the 'data' slot (i.e., log-normalised and depth-corrected expression values) for a given lncRNA. All Pearson correlation tests performed using the two-sided `cor.test` function in R. Differential expression analysis performed using *MAST* (Finak et al., 2015) as implemented in *Seurat* v3.2.2 `FindAllMarkers`, returning only positive markers with an average log fold-change threshold of 0.5. Pseudotime ordering was performed separately on sexual and asexual cells by first subsetting the original V3 objects to contain only sexual or asexual cells, and then using the original PCA embeddings of these cells and stage clusters as input into *slingshot* (Street et al., 2018), with 'rings' and 'early gametocytes' used as starting clusters for their respective data sets. Generalised additive models to

---

describe lncRNA expression over time were fit using `tradeSeq` (Van den Berge et al., 2020).



# Chapter 6

## Conclusions & Future Directions

In this thesis, I present a single-cell atlas of gametocyte development in *Plasmodium falciparum* that I have designed, built, curated and integrated. I demonstrate how single-cell RNA sequencing can be used to access the transcriptomes of 'uncontaminated' cell populations that are difficult to purify in bulk transcriptomics, such as very early gametocytes or pure male or female populations, and how this approach can be expanded to quantify transcriptional variation between parasites of differing genetic backgrounds. Furthermore, I broaden the scope of what has been explored in the single-cell transcriptomes of malaria parasites to include long non-coding RNA expression for the first time and demonstrate a workflow through which pre-existing data sets may be analysed in a similar manner. Ultimately these chapters, and the data generated within them, demonstrate how a cell atlas can act as a research tool — not only to uncover novel biological insights, but also as a spring board from which a multitude of further analyses can be launched.

### 6.1 Biological insights

#### 6.1.1 A progenitor stage of indeterminate sex in early gametocytes

The processes underpinning sex determination are still not well understood in *Plasmodium*. In my first chapter, I demonstrate the existence of a period of sexual

development in which parasites are clearly sexual but do not yet display signs of sexual dimorphism. It does not imply that these cells are not pre-committed to becoming male or female (as has been seen in plaque assays by Smith et al. (2000) and Silvestrini et al. (2000)), but rather that in the current data, a cell's sex cannot be determined based on expression alone. This finding mirrors that which is found in *P. berghei* by Russell et al. (2021) and while not specifically analysed, is also present in *P. berghei* cells from Hentzschel et al. (2021). As was discussed in Chapter 3, these sexually indeterminate transcriptomes likely reflect the shared biology of these stages — as far as we know, all early gametocyte cells sequester into the hematopoietic niche, regardless of their eventual sex (Venugopal et al., 2020). Indeed, van Biljon et al. (2019) found clusters expressed during 'transitional' stages of gametocyte development to be enriched for genes involved in maintaining the cell rigidity that defines these stages and maintains their retention in the bone marrow (Dixon et al., 2012; Peatey et al., 2013). The fact that this shared transcriptome is followed by a sharp and sudden increase in sex-specific gene expression likely implies that the factors regulating sex-specific biology are more complex than what can be observed in an unperturbed transcriptome alone, or may not be detectable in transcriptomic data at all. In support of this, the observed expression of potential sex-specific regulators in the bipotential cluster by Russell et al. was subtle, and only revealed following a targeted screen identifying genes for which deletions affected overall sex ratio. Of genes that had a syntenic ortholog in *P. falciparum*, I observed similar expression patterns in putative gametocyte progenitor cells to that of Russell et al. It would be interesting to investigate whether these genes display the same sex-ratio alterations in *P. falciparum* as they do in *P. berghei*, especially given the disparity in the duration of sexual maturity between the species. These findings highlight the importance of 'layering' context on top of the biological insights gained from the gametocyte atlas in order to better understand function.

### 6.1.2 Specificity of long non-coding RNAs

In Chapter 5, I demonstrated the utility of 10X Chromium data in detecting and profiling the expression of long non-coding RNAs during the *P. falciparum* sexual and asexual blood stages. The observed stage specificity of these transcripts and increased resolution over that which was detected by the Le Roch lab in bulk

RNAseq imply that single-cell data can be used to screen for lncRNAs important for development, especially given the increased specificity often displayed by ncRNAs (Li et al., 2020). To this end, myself and Joa Hoshizaki are currently generating CRISPR/Cas9 mutants for lncRNAs I detected as highly expressed during gametocyte stages to investigate their potential role in regulating sexual development or maturity. Previous studies have demonstrated the use of this technique in investigating lncRNA activity in *Plasmodium* (Bryant et al., 2017; Guizetti et al., 2016). This work is particularly exciting in light of the recent discovery of a novel aslncRNA that mediates a switch between meiosis and mitosis in yeast (Andric et al., 2021), where meiosis is believed to be controlled through active repression, and only switched on in response to worsening environmental conditions (Harigaya et al., 2006).

## 6.2 Cell atlases: the first step

The re-purposing of a single-cell RNAseq data, intended to profile mRNA expression, to instead screen the non-coding transcriptome is a good example of how the greatest value in a single-cell atlas may lie beyond mRNA transcription, and instead in its combinatorial power with additional paired data.

### 6.2.1 Mapping across across experiments

Already, the *P. berghei* atlas generated by (Howick et al., 2019) has been used to provide context for subsequent single-cell data sets of *Plasmodium* parasites (Hentzschel et al., 2021; Rawat et al., 2021; Real et al., 2021; Ruberto et al., 2021), allowing for a level of standardisation and comparability between the studies (Briggs et al., 2021). Standardisation could furthermore allow for the combination of multiple single-cell data sets in order to explore species-specific cell types (Tarashansky et al., 2021). As I demonstrate in Chapter 5, 10X Chromium's updated chemistry vastly improves capture efficiency, and will hopefully allow for future combinatorial approaches to be less liable to collection protocol (Nötzel et al., 2018). The recent release of *P. berghei* data collected with 10X Chromium version 3 chemistry by Hentzschel et al. provides the perfect test case for this approach

— mapping this data to the atlas produced in this thesis is a logical next step in exploring species-specific differences in sexual development.

In addition to wild-type atlases, I demonstrate in Chapter 4 that strain-specific variation in expression can be quantified using my atlas, but that strains do not vary so much in their expression that it precludes comparison. As such, my atlas may be used as a 'scaffold' of sorts to facilitate the comparison of gametocytes collected from clinical samples and sequenced using bulk or single-cell approaches, as is demonstrated in Howick et al. (2019). In a similar sense, mapping data collected following genetic perturbations (such as the knock-out mutants in Russell et al. (2021)) to my atlas could help to identify cases in which perturbations led to stunted or alternative developmental pathways during sexual development. In fact, this approach is what led to the recent discovery of a master regulator of bradyzoite differentiation in *Toxoplasma gondii* (Waldman et al., 2020). Ultimately, the 'holy grail' of perturbation-based sequencing along with single-cell RNAseq would be perturbations on a single cell level using CRISPR (Replogle et al., 2020). While this approach is likely beyond the reaches of possibility in *P. falciparum* given the difficulty in applying the technology to organisms in bulk, it would be extremely valuable in elucidating the function of coding and non-coding transcripts.

## 6.2.2 Towards a comprehensive picture of expression

There are many approaches that have single-cell applications beyond capturing mRNA, but not all of them have been used to explore *Plasmodium* biology. Combining modalities can often provide the context required to understand transcriptional processes that cannot be ascertained with one modality alone.

### CITE-seq

One such example of this already in the literature is the use of CITE-seq (Cellular Indexing of Transcriptomes and Epitopes by Sequencing) in combination with scRNAseq by Hentzschel et al. (2021). CITE-seq allows for the capture of RNA from a cell whilst also obtaining a read out of the expression of a tagged protein on the cell's surface. Using this approach, Hentzschel et al. were able to tag each

transcriptome by the host cell it had invaded and thus elucidate transcriptomic differences underlying host cell preference.

### **ATACseq**

Perhaps the most useful addition to the gametocyte atlas as it stands would be associated single-cell ATAC (Assay for Transposase-Accessible Chromatin using sequencing) sequencing data. This approach uses a mutated Tn5 transposase to tag accessible chromatin, thereby allowing for the identification of promoters and sites with active transcription (Buenrostro et al., 2013). It has been used successfully to identify important regulatory regions in *P. falciparum* in bulk (Ruiz-Orera et al., 2014; Toenhake et al., 2018), however the AT-richness of the genome still precludes its use at a single cell level (unpublished work not detailed in this thesis). If adjustments to the current scATACseq protocol can facilitate single-cell ATAC data, a profile of accessibility could be drawn for cell population subsets as they commit to gametocytogenesis and differentiate into males and females, and could provide further resolution into the gametocyte progenitor phase observed in the expression data.

### **Long-read transcriptomes**

As I describe in Chapter 5, single-cell RNAseq can be used to detect and profile the expression of non-coding genome. While this approach can be used to mine pre-existing data for non-coding expression, ideally a complete profile would employ a more targeted approach to explore non-coding expression. Long-read technologies such as those available from Pacific Biosciences (PacBio) or Oxford Nanopore Technologies (ONT) have adapted their protocols in order to allow for long-read, full-length transcripts to be captured, which can allow for improved profiling of non-coding transcription, as well as differential isoform usage. In bulk, long-read sequencing of mRNA transcripts have been used to profile the total splicing landscape of *P. falciparum* and *Toxoplasma gondii*. We have since applied PacBio Iso-seq to the single-cell transcriptomes presented in Chapter 5 and intend to further resolve variations in non-coding expression and splicing in each stage of sexual development.

In summary, single-cell technologies can provide an important level of clarity to complex transcriptional and post-transcriptional mechanisms in *Plasmodium* parasites. The work presented in this thesis will hopefully provide the malaria research community with a scaffold from which to build a better picture of sexual development in the parasite, with the ultimate aim of reducing the transmission and human impact of the deadly disease.

# References

- Adjalley, S. H., Chabbert, C. D., Klaus, B., Pelechano, V., and Steinmetz, L. M. (2016). Landscape and dynamics of transcription initiation in the malaria parasite *Plasmodium falciparum*. *Cell Rep.*, 14(10):2463–2475.
- Aguilar, R., Magallon-Tejada, A., Achtman, A. H., Moraleda, C., Joice, R., Cisteró, P., Li Wai Suen, C. S. N., Nhabomba, A., Macete, E., Mueller, I., Marti, M., Alonso, P. L., Menéndez, C., Schofield, L., and Mayor, A. (2014). Molecular evidence for the localization of *Plasmodium falciparum* immature gametocytes in bone marrow. *Blood*, 123(7):959–966.
- Alcid, E. A. and Tsukiyama, T. (2016). Expansion of antisense lncRNA transcriptomes in budding yeast species since the loss of RNAi. *Nat. Struct. Mol. Biol.*, 23(5):450–455.
- Aldridge, S. and Teichmann, S. A. (2020). Single cell transcriptomics comes of age. *Nat. Commun.*, 11(1):4307.
- Amato, R., Pearson, R. D., Almagro-Garcia, J., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Drury, E., Stalker, J., Miotto, O., Fairhurst, R. M., and Kwiatkowski, D. P. (2018). Origins of the current outbreak of multidrug-resistant malaria in southeast asia: a retrospective genetic study. *Lancet Infect. Dis.*, 18(3):337–345.
- Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E., and Dzikowski, R. (2015). Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.*, 112(9):E982–91.
- Anders, R. F., Brown, G. V., and Edwards, A. (1983). Characterization of an S antigen synthesized by several isolates of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.*, 80(21):6652–6656.
- Anderson, T. J., Paul, R. E., Donnelly, C. A., and Day, K. P. (2000). Do malaria parasites mate non-randomly in the mosquito midgut? *Genet. Res.*, 75(3):285–296.
- Andric, V., Nevers, A., Hazra, D., Auxilien, S., Menant, A., Graille, M., Palancade, B., and Rougemaille, M. (2021). A scaffold lncRNA shapes the mitosis to meiosis switch. *Nat. Commun.*, 12(1):770.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver,

- L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.*, 25(1):25–29.
- Aurrecochea, 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, Jr, C. J., Treatman, C., and Wang, H. (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.*, 37(Database issue):D539–43.
- Autino, B., Noris, A., Russo, R., and Castelli, F. (2012). Epidemiology of malaria in endemic areas. *Mediterr. J. Hematol. Infect. Dis.*, 4(1):e2012060.
- Awandare, G. A., Nyarko, P. B., Aniweh, Y., Ayivor-Djanie, R., and Stoute, J. A. (2018). Plasmodium falciparum strains spontaneously switch invasion phenotype in suspension culture. *Sci. Rep.*, 8(1):5782.
- Ay, F., Bunnik, E. M., Varoquaux, N., Bol, S. M., Prudhomme, J., Vert, J.-P., Noble, W. S., and Le Roch, K. G. (2014). Three-dimensional modeling of the p. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.*, 24(6):974–988.
- Azzalin, C. M., Reichenbach, P., Khorauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, 318(5851):798–801.
- Baker, D. A. (2010). Malaria gametocytogenesis. *Mol. Biochem. Parasitol.*, 172(2):57–65.
- Balaji, S., Babu, M. M., Iyer, L. M., and Aravind, L. (2005). Discovery of the principal specific transcription factors of apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.*, 33(13):3994–4006.
- Bancells, C., Llorà-Batlle, O., Poran, A., Nötzel, C., Rovira-Graells, N., Elemento, O., Kafsack, B. F. C., and Cortés, A. (2019). Revisiting the initial steps of sexual development in the malaria parasite Plasmodium falciparum. *Nat Microbiol.*, 4(1):144–154.
- Bannister, L. and Mitchell, G. (2003). The ins, outs and roundabouts of malaria. *Trends Parasitol.*, 19(5):209–213.
- Barcons-Simon, A., Cordon-Obras, C., Guizetti, J., Bryant, J. M., and Scherf, A. (2020). CRISPR interference of a clonally variant GC-Rich noncoding RNA family leads to general repression of var genes in Plasmodium falciparum. *MBio*, 11(1).
- Bartoloni, A. and Zammarchi, L. (2012). Clinical aspects of uncomplicated and severe malaria. *Mediterr. J. Hematol. Infect. Dis.*, 4(1):e2012026.
- Beck, Z. T., Xing, Z., and Tran, E. J. (2016). LncRNAs: Bridging environmental sensing and gene expression. *RNA Biol.*, 13(12):1189–1196.

- Bhasin, V. K. and Trager, W. (1984). Gametocyte-forming and non-gametocyte-forming clones of *Plasmodium falciparum*. *Am. J. Trop. Med. Hyg.*, 33(4):534–537.
- Böhme, U., Otto, T. D., Cotton, J. A., Steinbiss, S., Sanders, M., Oyola, S. O., Nicot, A., Gandon, S., Patra, K. P., Herd, C., Bushell, E., Modrzynska, K. K., Billker, O., Vinetz, J. M., Rivero, A., Newbold, C. I., and Berriman, M. (2018). Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res.*, 28(4):547–560.
- Böhme, U., Otto, T. D., Sanders, M., Newbold, C. I., and Berriman, M. (2019). Progression of the canonical reference malaria parasite genome from 2002–2019. *Wellcome Open Res*, 4:58.
- Bousema, J. T., Drakeley, C. J., Mens, P. F., Arens, T., Houben, R., Omar, S. A., Gouagna, L. C., Schallig, H., and Sauerwein, R. W. (2008). Increased *Plasmodium falciparum* gametocyte production in mixed infections with *p. malariae*. *Am. J. Trop. Med. Hyg.*, 78(3):442–448.
- Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.*, 1(1):E5.
- Bozdech, Z., Mok, S., Hu, G., Imwong, M., Jaidee, A., Russell, B., Ginsburg, H., Nosten, F., Day, N. P. J., White, N. J., Carlton, J. M., and Preiser, P. R. (2008). The transcriptome of *Plasmodium vivax* reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proc. Natl. Acad. Sci. U. S. A.*, 105(42):16290–16295.
- Brancucci, N. M. B., 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(2):165–176.
- Brancucci, N. M. B., Gerdt, J. P., Wang, C., De Niz, M., Philip, N., Adapa, S. R., Zhang, M., Hitz, E., Niederwieser, I., Boltryk, S. D., Laffitte, M.-C., Clark, M. A., Grüning, C., Ravel, D., Blancke Soares, A., Demas, A., Bopp, S., Rubio-Ruiz, B., Conejo-Garcia, A., Wirth, D. F., Gendaszewska-Darmach, E., Duraisingh, M. T., Adams, J. H., Voss, T. S., Waters, A. P., Jiang, R. H. Y., Clardy, J., and Marti, M. (2017). Lysophosphatidylcholine regulates sexual stage differentiation in the human malaria parasite *Plasmodium falciparum*. *Cell*, 171(7):1532–1544.e15.
- Briggs, E. M., Warren, F. S. L., Matthews, K. R., McCulloch, R., and Otto, T. D. (2021). Application of single-cell transcriptomics to kinetoplastid research. *Parasitology*, 148(10):1223–1236.
- Broadbent, K. M., Broadbent, J. C., Ribacke, U., Wirth, D., Rinn, J. L., and Sabeti, P. C. (2015). Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC Genomics*, 16:454.

- Broadbent, K. M., Park, D., Wolf, A. R., Van Tyne, D., Sims, J. S., Ribacke, U., Volkman, S., Duraisingh, M., Wirth, D., Sabeti, P. C., and Rinn, J. L. (2011). A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol.*, 12(6):R56.
- 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.
- Bruce, M. C., Carter, R. N., Nakamura, K., Aikawa, M., and Carter, R. (1994). Cellular location and temporal expression of the *Plasmodium falciparum* sexual stage antigen pfs16. *Mol. Biochem. Parasitol.*, 65(1):11–22.
- Bryant, J. M., Regnault, C., Scheidig-Benatar, C., Baumgarten, S., Guizetti, J., and Scherf, A. (2017). CRISPR/Cas9 genome editing reveals that the intron is not essential for var2csa gene activation or silencing in *Plasmodium falciparum*. *MBio*, 8(4).
- Buckling, A., Ranford-Cartwright, L. C., Miles, A., and Read, A. F. (1999). Chloroquine increases *Plasmodium falciparum* gametocytogenesis in vitro. *Parasitology*, 118 ( Pt 4):339–346.
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, 10(12):1213–1218.
- Bunnik, E. M., Cook, K. B., Varoquaux, N., Batugedara, G., Prudhomme, J., Cort, A., Shi, L., Andolina, C., Ross, L. S., Brady, D., Fidock, D. A., Nosten, F., Tewari, R., Sinnis, P., Ay, F., Vert, J.-P., Noble, W. S., and Le Roch, K. G. (2018). Changes in genome organization of parasite-specific gene families during the *Plasmodium* transmission stages. *Nat. Commun.*, 9(1):1910.
- Burkot, T. R., Williams, J. L., and Schneider, I. (1984). Infectivity to mosquitoes of *Plasmodium falciparum* clones grown in vitro from the same isolate. *Trans. R. Soc. Trop. Med. Hyg.*, 78(3):339–341.
- Bushell, E., Gomes, A. R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modrzynska, K., Schwach, F., Martin, R. E., Mather, M. W., McFadden, G. I., Parts, L., Rutledge, G. G., Vaidya, A. B., Wengelnik, K., Rayner, J. C., and Billker, O. (2017). Functional profiling of a *Plasmodium* genome reveals an abundance of essential genes. *Cell*, 170(2):260–272.e8.
- Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, 25(18):1915–1927.
- Cao, P., Collins, K. A., Zaloumis, S., Wattanakul, T., Tarning, J., Simpson, J. A., McCarthy, J., and McCaw, J. M. (2019). Modeling the dynamics of *Plasmodium falciparum* gametocytes in humans during malaria infection. *Elife*, 8.

- Carlson, M. (2019). org.pf.plasmo.db: Genome wide annotation for malaria.
- Carter, L. M., Kafsack, B. F. C., Llinás, 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(1):135–147.
- Carter, R. and Graves, P. M. (1988). Gametocytes. In Wernsdorfer, W. H. and McGregor, I., editors, *Malaria : principles and practice of malariology*, pages 233–305. Churchill Livingstone.
- Carter, R., Graves, P. M., Creasey, A., Byrne, K., Read, D., Alano, P., and Fenton, B. (1989). Plasmodium falciparum: an abundant stage-specific protein expressed during early gametocyte development. *Exp. Parasitol.*, 69(2):140–149.
- Carter, R. and Mendis, K. N. (2002). Evolutionary and historical aspects of the burden of malaria. *Clin. Microbiol. Rev.*, 15(4):564–594.
- Challenger, J. D., Olivera Mesa, D., Da, D. F., Yerbanga, R. S., Lefèvre, T., Cohuet, A., and Churcher, T. S. (2021). Predicting the public health impact of a malaria transmission-blocking vaccine. *Nat. Commun.*, 12(1):1494.
- Chapman, R. D., Heidemann, M., Hintermair, C., and Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD. *Trends Genet.*, 24(6):289–296.
- Chappell, L., Ross, P., Orchard, L., Russell, T. J., Otto, T. D., Berriman, M., Rayner, J. C., and Llinás, M. (2020). Refining the transcriptome of the human malaria parasite Plasmodium falciparum using amplification-free RNA-seq. *BMC Genomics*, 21(1):395.
- Chappell, L., Russell, A. J. C., and Voet, T. (2018). Single-Cell (multi)omics technologies. *Annu. Rev. Genomics Hum. Genet.*, 19:15–41.
- Chen, C.-H. (2002). Generalized association plots: Information visualization via iteratively generated correlation matrices. *Stat. Sin.*, (12):7–29.
- Claessens, A., Affara, M., Assefa, S. A., Kwiatkowski, D. P., and Conway, D. J. (2017). Culture adaptation of malaria parasites selects for convergent loss-of-function mutants. *Sci. Rep.*, 7:41303.
- Cowman, A. F., Tonkin, C. J., Tham, W.-H., and Duraisingh, M. T. (2017). The molecular basis of erythrocyte invasion by malaria parasites. *Cell Host Microbe*, 22(2):232–245.
- Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasit. Vectors*, 3(1):5.
- Crabb, B. S., Cooke, B. M., Reeder, J. C., Waller, R. F., Caruana, S. R., Davern, K. M., Wickham, M. E., Brown, G. V., Coppel, R. L., and Cowman, A. F. (1997). Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell*, 89(2):287–296.

- Crabb, B. S. and Cowman, A. F. (1996). Characterization of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.*, 93(14):7289–7294.
- Dantzler, K. W., Ma, S., Ngotho, P., Stone, W. J. R., Tao, D., Rijpma, S., De Niz, M., Nilsson Bark, S. K., Jore, M. M., Raaijmakers, T. K., Early, A. M., Ubaida-Mohien, C., Lemgruber, L., Campo, J. J., Teng, A. A., Le, T. Q., Walker, C. L., Hermand, P., Deterre, P., Davies, D. H., Felgner, P., Morlais, I., Wirth, D. F., Neafsey, D. E., Dinglasan, R. R., Laufer, M., Huttenhower, C., Seydel, K., Taylor, T., Bousema, T., and Marti, M. (2019). Naturally acquired immunity against immature *Plasmodium falciparum* gametocytes. *Sci. Transl. Med.*, 11(495).
- de Bézieux, H. R., Van den Berge, K., Street, K., and Dudoit, S. (2021). Trajectory inference across multiple conditions with condiments: differential topology, progression, differentiation, and expression.
- Dearnley, M., Chu, T., Zhang, Y., Looker, O., Huang, C., Klonis, N., Yeoman, J., Kenny, S., Arora, M., Osborne, J. M., Chandramohanadas, R., Zhang, S., Dixon, M. W. A., and Tilley, L. (2016). Reversible host cell remodeling underpins deformability changes in malaria parasite sexual blood stages. *Proc. Natl. Acad. Sci. U. S. A.*, 113(17):4800–4805.
- Delves, M., Plouffe, D., Scheurer, C., Meister, S., Wittlin, S., Winzeler, E. A., Sinden, R. E., and Leroy, D. (2012). The activities of current antimalarial drugs on the life cycle stages of *Plasmodium*: a comparative study with human and rodent parasites. *PLoS Med.*, 9(2):e1001169.
- Delves, M. J., Angrisano, F., and Blagborough, A. M. (2018). Antimalarial Transmission-Blocking interventions: Past, present, and future. *Trends Parasitol.*, 34(9):735–746.
- DePasquale, E. A. K., Schnell, D. J., Van Camp, P.-J., Valiente-Alandí, Í., Blaxall, B. C., Grimes, H. L., Singh, H., and Salomonis, N. (2019). DoubletDecon: Deconvoluting doublets from Single-Cell RNA-Sequencing data. *Cell Rep.*, 29(6):1718–1727.e8.
- Dixon, M. W. A., Dearnley, M. K., Hanssen, E., Gilberger, T., and Tilley, L. (2012). Shape-shifting gametocytes: how and why does *p. falciparum* go banana-shaped? *Trends Parasitol.*, 28(11):471–478.
- Doerig, C., Rayner, J. C., Scherf, A., and Tobin, A. B. (2015). Post-translational protein modifications in malaria parasites. *Nat. Rev. Microbiol.*, 13(3):160–172.
- Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S., and Trapnell, C. (2020). Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat. Commun.*, 11(1):1537.
- Eichner, M., Diebner, H. H., Molineaux, L., Collins, W. E., Jeffery, G. M., and Dietz, K. (2001). Genesis, sequestration and survival of *Plasmodium falciparum* gametocytes: parameter estimates from fitting a model to malariatherapy data. *Trans. R. Soc. Trop. Med. Hyg.*, 95(5):497–501.

- 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(10):e1002964.
- Epp, C., Li, F., Howitt, C. A., Chookajorn, T., and Deitsch, K. W. (2009). Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite Plasmodium falciparum. *RNA*, 15(1):116–127.
- Farfour, E., Charlotte, F., Settegrana, C., Miyara, M., and Buffet, P. (2012). The extravascular compartment of the bone marrow: a niche for Plasmodium falciparum gametocyte maturation? *Malar. J.*, 11:285.
- Filarsky, M., Fraschka, S. A., Niederwieser, I., Brancucci, N. M. B., Carrington, E., Carrió, E., Moes, S., Jenoe, P., Bártfai, R., and Voss, T. S. (2018). GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science*, 359(6381):1259–1263.
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., Linsley, P. S., and Gottardo, R. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.*, 16:278.
- Flueck, C., Bartfai, R., Niederwieser, I., Witmer, K., Alako, B. T. F., Moes, S., Bozdech, Z., Jenoe, P., Stunnenberg, H. G., and Voss, T. S. (2010). A major role for the Plasmodium falciparum ApiAP2 protein PfsIP2 in chromosome end biology. *PLoS Pathog.*, 6(2):e1000784.
- Flueck, C., Bartfai, R., Volz, J., Niederwieser, I., Salcedo-Amaya, A. M., Alako, B. T. F., Ehlgén, F., Ralph, S. A., Cowman, A. F., Bozdech, Z., Stunnenberg, H. G., and Voss, T. S. (2009). Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog.*, 5(9):e1000569.
- Fraschka, S. A., Filarsky, M., Hoo, R., Niederwieser, I., Yam, X. Y., Brancucci, N. M. B., Moring, F., Mushunje, A. T., Huang, X., Christensen, P. R., Nosten, F., Bozdech, Z., Russell, B., Moon, R. W., Marti, M., Preiser, P. R., Bártfai, R., and Voss, T. S. (2018). Comparative heterochromatin profiling reveals conserved and unique epigenome signatures linked to adaptation and development of malaria parasites. *Cell Host Microbe*, 23(3):407–420.e8.
- Frischknecht, F., Baldacci, P., Martin, B., Zimmer, C., Thiberge, S., Olivo-Marin, J.-C., Shorte, S. L., and Ménard, R. (2004). Imaging movement of malaria parasites during transmission by anopheles mosquitoes. *Cell. Microbiol.*, 6(7):687–694.
- Galen, S. C., Borner, J., Martinsen, E. S., Schaer, J., Austin, C. C., West, C. J., and Perkins, S. L. (2018). The polyphyly of Plasmodium: comprehensive phylogenetic analyses of the malaria parasites (order haemosporida) reveal widespread taxonomic conflict. *R Soc Open Sci*, 5(5):171780.

- Ganesan, K., Ponmee, N., Jiang, L., Fowble, J. W., White, J., Kamchonwongpaisan, S., Yuthavong, Y., Wilairat, P., and Rathod, P. K. (2008). A genetically hard-wired metabolic transcriptome in *Plasmodium falciparum* fails to mount protective responses to lethal antifolates. *PLoS Pathog.*, 4(11):e1000214.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906):498–511.
- Garnham, P. C. C. (1966). Malaria parasites and other haemosporidia. *Malaria Parasites and other Haemosporidia*.
- Gautret, P. and Motard, A. (1999). Periodic infectivity of *Plasmodium* gametocytes to the vector. a review. *Parasite*, 6(2):103–111.
- Gebru, T., Lalremruata, A., Kreamsner, P. G., Mordmüller, B., and Held, J. (2017). Life-span of in vitro differentiated *Plasmodium falciparum* gametocytes. *Malar. J.*, 16(1):330.
- Gibson, W. (2021). The sexual side of parasitic protists. *Mol. Biochem. Parasitol.*, 243:111371.
- Greischar, M. A., Mideo, N., Read, A. F., and Bjørnstad, O. N. (2016). Predicting optimal transmission investment in malaria parasites. *Evolution*, 70(7):1542–1558.
- Guizetti, J., Barcons-Simon, A., and Scherf, A. (2016). Trans-acting GC-rich non-coding RNA at *var* expression site modulates gene counting in malaria parasite. *Nucleic Acids Res.*, 44(20):9710–9718.
- Hamilton, W. L., Claessens, A., Otto, T. D., Kekre, M., Fairhurst, R. M., Rayner, J. C., and Kwiatkowski, D. (2017). Extreme mutation bias and high AT content in *Plasmodium falciparum*. *Nucleic Acids Res.*, 45(4):1889–1901.
- Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., Chikashige, Y., Hiraoka, Y., Yamashita, A., and Yamamoto, M. (2006). Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature*, 442(7098):45–50.
- Hawking, F. (1970). The clock of the malaria parasite. *Sci. Am.*, 222(6):123–131.
- Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., and Snow, R. W. (2004). The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect. Dis.*, 4(6):327–336.

- Heaton, H., Talman, A. M., Knights, A., Imaz, M., Gaffney, D. J., Durbin, R., Hemberg, M., and Lawniczak, M. K. N. (2020). Souporecell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods*, 17(6):615–620.
- Hentzschel, F., Gibbins, M. P., Attipa, C., Beraldi, D., Moxon, C. A., Otto, T. D., and Marti, M. (2021). Host cell maturation modulates parasite invasion and sexual differentiation in Plasmodium.
- Hoeijmakers, W. A. M., Miao, J., Schmidt, S., Toenhake, C. G., Shrestha, S., Venhuizen, J., Henderson, R., Birnbaum, J., Ghidelli-Disse, S., Drewes, G., Cui, L., Stunnenberg, H. G., Spielmann, T., and Bártfai, R. (2019). Epigenetic reader complexes of the human malaria parasite, Plasmodium falciparum. *Nucleic Acids Res.*, 47(22):11574–11588.
- Hollin, T. and Le Roch, K. G. (2020). From genes to transcripts, a tightly regulated journey in Plasmodium. *Front. Cell. Infect. Microbiol.*, 10:618454.
- Howick, V. M., Russell, A. J. C., Andrews, T., Heaton, H., Reid, A. J., Natarajan, K., Butungi, H., Metcalf, T., Verzier, L. H., Rayner, J. C., Berriman, M., Herren, J. K., Billker, O., Hemberg, M., Talman, A. M., and Lawniczak, M. K. N. (2019). The malaria cell atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. *Science*, 365(6455).
- Hu, G., Cabrera, A., Kono, M., Mok, S., Chahal, B. K., Haase, S., Engelberg, K., Cheemadan, S., Spielmann, T., Preiser, P. R., Gilberger, T.-W., and Bozdech, Z. (2010). Transcriptional profiling of growth perturbations of the human malaria parasite Plasmodium falciparum. *Nat. Biotechnol.*, 28(1):91–98.
- Hughes, K. R., Philip, N., Starnes, G. L., Taylor, S., and Waters, A. P. (2010). From cradle to grave: RNA biology in malaria parasites. *Wiley Interdiscip. Rev. RNA*, 1(2):287–303.
- Inselburg, J. (1983). Gametocyte formation by the progeny of single Plasmodium falciparum schizonts. *J. Parasitol.*, 69(3):584–591.
- Iwanaga, S., Kaneko, I., Kato, T., and Yuda, M. (2012). Identification of an AP2-family protein that is critical for malaria liver stage development. *PLoS One*, 7(11):e47557.
- Iyer, J., Grüner, A. C., Rénia, L., Snounou, G., and Preiser, P. R. (2007). Invasion of host cells by malaria parasites: a tale of two protein families. *Mol. Microbiol.*, 65(2):231–249.
- Jensen, J. B., Capps, T. C., and Carlin, J. M. (1981). Clinical drug-resistant falciparum malaria acquired from cultured parasites. *Am. J. Trop. Med. Hyg.*, 30(3):523–525.
- Jeyaprakasam, N. K., Liew, J. W. K., Low, V. L., Wan-Sulaiman, W.-Y., and Vythilingam, I. (2020). Plasmodium knowlesi infecting humans in southeast asia: What's next? *PLoS Negl. Trop. Dis.*, 14(12):e0008900.

- Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., Yang, W., Turner, L., Lavstsen, T., Theander, T. G., Peng, W., Wei, G., Jing, Q., Wakabayashi, Y., Bansal, A., Luo, Y., Ribeiro, J. M. C., Scherf, A., Aravind, L., Zhu, J., Zhao, K., and Miller, L. H. (2013). PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature*, 499(7457):223–227.
- Jin, Y., Kebaier, C., and Vanderberg, J. (2007). Direct microscopic quantification of dynamics of *Plasmodium berghei* sporozoite transmission from mosquitoes to mice. *Infect. Immun.*, 75(11):5532–5539.
- Jing, Q., Cao, L., Zhang, L., Cheng, X., Gilbert, N., Dai, X., Sun, M., Liang, S., and Jiang, L. (2018). *Plasmodium falciparum* var gene is activated by its antisense long noncoding RNA. *Front. Microbiol.*, 9:3117.
- Joice, R., Nilsson, S. K., Montgomery, J., Dankwa, S., Egan, E., Morahan, B., Seydel, K. B., Bertuccini, L., Alano, P., Williamson, K. C., Duraisingh, M. T., Taylor, T. E., Milner, D. A., and Marti, M. (2014). *Plasmodium falciparum* transmission stages accumulate in the human bone marrow. *Sci. Transl. Med.*, 6(244):244re5.
- Josling, G. A., Russell, T. J., Venezia, J., Orchard, L., van Biljon, R., Painter, H. J., and Llinás, M. (2020). Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nat. Commun.*, 11(1):1503.
- Josling, G. A., Williamson, K. C., and Llinás, M. (2018). Regulation of sexual commitment and gametocytogenesis in malaria parasites. *Annu. Rev. Microbiol.*, 72:501–519.
- Kafsack, B. F. C., Rovira-Graells, N., Clark, T. G., Bancells, C., Crowley, V. M., Campino, S. G., Williams, A. E., Drought, L. G., Kwiatkowski, D. P., Baker, D. A., Cortés, A., and Llinás, M. (2014). A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, 507(7491):248–252.
- Kamaliddin, C., Guillochon, E., Salnot, V., Rombaut, D., Huguet, S., Guillonneau, F., Houzé, S., Cot, M., Deloron, P., Argy, N., and Bertin, G. I. (2021). Comprehensive analysis of transcript and protein relative abundance during blood stages of *Plasmodium falciparum* infection. *J. Proteome Res.*, 20(2):1206–1216.
- Ke, H., Morrissey, J. M., Ganesan, S. M., Painter, H. J., Mather, M. W., and Vaidya, A. B. (2011). Variation among *Plasmodium falciparum* strains in their reliance on mitochondrial electron transport chain function. *Eukaryot. Cell*, 10(8):1053–1061.
- Keating, C. (2020). The history of the RTS,S/AS01 malaria vaccine trial. *Lancet*, 395(10233):1336–1337.
- Kensche, P. R., Hoeijmakers, W. A. M., Toenhake, C. G., Bras, M., Chappell, L., Berriman, M., and Bártfai, R. (2016). The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Res.*, 44(5):2110–2124.

- Kent, R. S., Modrzynska, K. K., Cameron, R., Philip, N., Billker, O., and Waters, A. P. (2018). Inducible developmental reprogramming redefines commitment to sexual development in the malaria parasite *Plasmodium berghei*. *Nature Microbiology*, 3(11):1206–1213.
- Kidgell, C., Volkman, S. K., Daily, J., Borevitz, J. O., Plouffe, D., Zhou, Y., Johnson, J. R., Le Roch, K., Sarr, O., Ndir, O., Mboup, S., Batalov, S., Wirth, D. F., and Winzeler, E. A. (2006). A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog.*, 2(6):e57.
- Kiselev, V. Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods*, 15(5):359–362.
- Koella, J. C. and Antia, R. (1995). Optimal pattern of replication and transmission for parasites with two stages in their life cycle. *Theor. Popul. Biol.*, 47(3):277–291.
- 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(4):291–295.
- Krishnan, J. and Mishra, R. K. (2014). Emerging trends of long non-coding RNAs in gene activation. *FEBS J.*, 281(1):34–45.
- Kwiatkowski, D. P. (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.*, 77(2):171–192.
- Lambros, C. and Vanderberg, J. P. (1979). Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J. Parasitol.*, 65(3):418–420.
- Lasonder, E., Rijpma, S. R., van Schaijk, B. C. L., Hoeijmakers, W. A. M., Kensche, P. R., Gresnigt, M. S., Italiaander, A., Vos, M. W., Woestenenk, R., Bousema, T., Mair, G. R., Khan, S. M., Janse, C. J., Bártfai, R., and Sauerwein, R. W. (2016). Integrated transcriptomic and proteomic analyses of *p. falciparum* gametocytes: molecular insight into sex-specific processes and translational repression. *Nucleic Acids Res.*, 44(13):6087–6101.
- Le Roch, K. G., Johnson, J. R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S. F., Williamson, K. C., Holder, A. A., Carucci, D. J., Yates, 3rd, J. R., and Winzeler, E. A. (2004). Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res.*, 14(11):2308–2318.
- Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., and Winzeler, E. A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, 301(5639):1503–1508.
- Lemieux, J. E., Kyes, S. A., Otto, T. D., Feller, A. I., Eastman, R. T., Pinches, R. A., Berriman, M., Su, X.-Z., and Newbold, C. I. (2013). Genome-wide profiling of chromosome interactions in *Plasmodium falciparum* characterizes nuclear architecture and reconfigurations associated with antigenic variation. *Mol. Microbiol.*, 90(3):519–537.

- Li, D., Zhang, J., Li, X., Chen, Y., Yu, F., and Liu, Q. (2021). Insights into lncRNAs in Alzheimer's disease mechanisms. *RNA Biol.*, 18(7):1037–1047.
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.*, 13(9):2178–2189.
- Li, Y., Baptista, R. P., and Kissinger, J. C. (2020). Noncoding RNAs in apicomplexan parasites: An update. *Trends Parasitol.*, 36(10):835–849.
- Liu, S. J., Nowakowski, T. J., Pollen, A. A., Lui, J. H., Horlbeck, M. A., Attenello, F. J., He, D., Weissman, J. S., Kriegstein, A. R., Diaz, A. A., and Lim, D. A. (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.*, 17:67.
- Llinás, M., Bozdech, Z., Wong, E. D., Adai, A. T., and DeRisi, J. L. (2006). Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res.*, 34(4):1166–1173.
- 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*, 12:587.
- Lorenzen, J. M. and Thum, T. (2016). Long noncoding RNAs in kidney and cardiovascular diseases. *Nat. Rev. Nephrol.*, 12(6):360–373.
- Lu, X. M., Batugedara, G., Lee, M., Prudhomme, J., Bunnik, E. M., and Le Roch, K. G. (2017). Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite *Plasmodium falciparum*. *Nucleic Acids Res.*, 45(13):7825–7840.
- Luecken, M. D. and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.*, 15(6).
- Luo, H., Bu, D., Shao, L., Li, Y., Sun, L., Wang, C., Wang, J., Yang, W., Yang, X., Dong, J., Zhao, Y., and Li, F. (2020). Single-cell long non-coding RNA landscape of T cells in human 2 cancer immunity. *bioRxiv*.
- Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., O'Brien, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J.-B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S. M., Marsh, K., Jiang, H., Su, X.-Z., Amaratunga, C., Fairhurst, R., Socheat, D., Nosten, F., Imwong, M., White, N. J., Sanders, M., Anastasi, E., Alcock, D., Drury, E., Oyola, S., Quail, M. A., Turner, D. J., Ruano-Rubio, V., Jyothi, D., Amenga-Etego, L., Hubbart, C., Jeffreys, A., Rowlands, K., Sutherland, C., Roper, C., Mangano, V., Modiano, D., Tan, J. C., Ferdig, M. T., Amambua-Ngwa, A., Conway, D. J., Takala-Harrison, S., Plowe, C. V., Rayner, J. C., Rockett, K. A., Clark, T. G., Newbold, C. I., Berriman, M., MacInnis, B., and Kwiatkowski, D. P. (2012). Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*, 487(7407):375–379.

- 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(5):521–534.
- Marchoux, E. (1922). Multiplicite des races dans les trois formes de parasites du paludisme. *Bulletin de la Société de Pathologie Exotique*, 15:108–109.
- Maswoswe, S. M., Peters, W., and Warhurst, D. C. (1985). Corticosteroid stimulation of the growth of *Plasmodium falciparum* gametocytes in vitro. *Ann. Trop. Med. Parasitol.*, 79(6):607–616.
- McCall, M. B. B., Wammes, L. J., Langenberg, M. C. C., van Gemert, G.-J., Walk, J., Hermsen, C. C., Graumans, W., Koelewijn, R., Franetich, J.-F., Chishimba, S., Gerdson, M., Lorthiois, A., van de Vegte, M., Mazier, D., Bijker, E. M., van Hellemond, J. J., van Genderen, P. J. J., and Sauerwein, R. W. (2017). Infectivity of *Plasmodium falciparum* sporozoites determines emerging parasitemia in infected volunteers. *Sci. Transl. Med.*, 9(395).
- McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction.
- McKenzie, F. E., Jeffery, G. M., and Collins, W. E. (2002). *Plasmodium malariae* infection boosts *Plasmodium falciparum* gametocyte production. *Am. J. Trop. Med. Hyg.*, 67(4):411–414.
- McKenzie, F. E., Smith, D. L., O'Meara, W. P., and Riley, E. M. (2008). Strain theory of malaria: The first 50 years. *Advances in Parasitology*, 66:1–46.
- Meibalan, E. and Marti, M. (2017). Biology of malaria transmission. *Cold Spring Harb. Perspect. Med.*, 7(3).
- Miao, J., Wang, C., Lucky, A., Liang, X., Min, H., Adapa, S. R., Jiang, R., Kim, K., and Cui, L. (2021). A unique GCN5 histone acetyltransferase complex controls erythrocyte invasion and virulence in the malaria parasite *Plasmodium falciparum*.
- Militello, K. T., Patel, V., Chessler, A.-D., Fisher, J. K., Kasper, J. M., Gunasekera, A., and Wirth, D. F. (2005). RNA polymerase II synthesizes antisense RNA in *Plasmodium falciparum*. *RNA*, 11(4):365–370.
- Modrzynska, K., Pfander, C., Chappell, L., Yu, L., Suarez, C., Dundas, K., Gomes, A. R., Goulding, D., Rayner, J. C., Choudhary, J., and Billker, O. (2017). A knockout screen of *ApiAP2* genes reveals networks of interacting transcriptional regulators controlling the *Plasmodium* life cycle. *Cell Host Microbe*, 21(1):11–22.
- Molina-Cruz, A., Canepa, G. E., Kamath, N., Pavlovic, N. V., Mu, J., Ramphul, U. N., Ramirez, J. L., and Barillas-Mury, C. (2015). *Plasmodium* evasion of mosquito immunity and global malaria transmission: The lock-and-key theory. *Proc. Natl. Acad. Sci. U. S. A.*, 112(49):15178–15183.

- Moser, K. A., Drábek, E. F., Dwivedi, A., Stucke, E. M., Crabtree, J., Dara, A., Shah, Z., Adams, M., Li, T., Rodrigues, P. T., Koren, S., Phillippy, A. M., Munro, J. B., Ouattara, A., Sparklin, B. C., Dunning Hotopp, J. C., Lyke, K. E., Sadzewicz, L., Tallon, L. J., Spring, M. D., Jongsakul, K., Lon, C., Saunders, D. L., Ferreira, M. U., Nyunt, M. M., Laufer, M. K., Travassos, M. A., Sauerwein, R. W., Takala-Harrison, S., Fraser, C. M., Sim, B. K. L., Hoffman, S. L., Plowe, C. V., and Silva, J. C. (2020). Strains used in whole organism *Plasmodium falciparum* vaccine trials differ in genome structure, sequence, and immunogenic potential. *Genome Med.*, 12(1):6.
- Mota, M. M., Pradel, G., Vanderberg, J. P., Hafalla, J. C., Frevert, U., Nussenzweig, R. S., Nussenzweig, V., and Rodríguez, A. (2001). Migration of *Plasmodium* sporozoites through cells before infection. *Science*, 291(5501):141–144.
- Natalang, O., Bischoff, E., Deplaine, G., Proux, C., Dillies, M.-A., Sismeiro, O., Guigon, G., Bonnefoy, S., Patarapotikul, J., Mercereau-Puijalon, O., Coppée, J.-Y., and David, P. H. (2008). Dynamic RNA profiling in *Plasmodium falciparum* synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics*, 9:388.
- Neafsey, D. E., Schaffner, S. F., Volkman, S. K., Park, D., Montgomery, P., Milner, Jr, D. A., Lukens, A., Rosen, D., Daniels, R., Houde, N., Cortese, J. F., Tyndall, E., Gates, C., Stange-Thomann, N., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Ferreira, M. U., Moraes, S. d. L., Dash, A. P., Chitnis, C. E., Wiegand, R. C., Hartl, D. L., Birren, B. W., Lander, E. S., Sabeti, P. C., and Wirth, D. F. (2008). Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol.*, 9(12):R171.
- Neafsey, D. E., Taylor, A. R., and MacInnis, B. L. (2021). Advances and opportunities in malaria population genomics. *Nat. Rev. Genet.*
- Neveu, G., Beri, D., and Kafsack, B. F. (2020). Metabolic regulation of sexual commitment in *Plasmodium falciparum*. *Curr. Opin. Microbiol.*, 58:93–98.
- Ngotho, P., Soares, A. B., Hentzschel, F., Achcar, F., Bertuccini, L., and Marti, M. (2019). Revisiting gametocyte biology in malaria parasites. *FEMS Microbiol. Rev.*, 43(4):401–414.
- Nötzel, C. and Kafsack, B. F. C. (2021). There and back again: malaria parasite single-cell transcriptomics comes full circle. *Trends Parasitol.*, 37(10):850–852.
- Nötzel, C., Poran, A., and Kafsack, B. F. C. (2018). Single-Cell transcriptome profiling of protozoan and metazoan parasites. *Trends Parasitol.*, 34(9):731–734.
- Oakley, M. S. M., Kumar, S., Anantharaman, V., Zheng, H., Mahajan, B., Haynes, J. D., Moch, J. K., Fairhurst, R., McCutchan, T. F., and Aravind, L. (2007). Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic *Plasmodium falciparum* parasites. *Infect. Immun.*, 75(4):2012–2025.
- O'Donoghue, P. (2017). Haemoprotozoa: Making biological sense of molecular phylogenies. *Int. J. Parasitol. Parasites Wildl.*, 6(3):241–256.

- Oduola, A. M., Weatherly, N. F., Bowdre, J. H., and Desjardins, R. E. (1988). Plasmodium falciparum: cloning by single-erythrocyte micromanipulation and heterogeneity in vitro. *Exp. Parasitol.*, 66(1):86–95.
- Ono, T., Ohnishi, Y., Nagamune, K., and Kano, M. (1993). Gametocytogenesis induction by berenil in cultured Plasmodium falciparum. *Exp. Parasitol.*, 77(1):74–78.
- 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.
- Otto, T. D., Böhme, U., Sanders, M., Reid, A., Bruske, E. I., Duffy, C. W., Bull, P. C., Pearson, R. D., Abdi, A., Dimonte, S., Stewart, L. B., Campino, S., Kekre, M., Hamilton, W. L., Claessens, A., Volkman, S. K., Ndiaye, D., Amambua-Ngwa, A., Diakite, M., Fairhurst, R. M., Conway, D. J., Franck, M., Newbold, C. I., and Berri-man, M. (2018). Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. *Wellcome Open Res*, 3:52.
- Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., Newbold, C., and Llinás, M. (2010). New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol. Microbiol.*, 76(1):12–24.
- Ouattara, A., Barry, A. E., Dutta, S., Remarque, E. J., Beeson, J. G., and Plowe, C. V. (2015). Designing malaria vaccines to circumvent antigen variability. *Vaccine*, 33(52):7506–7512.
- Painter, H. J., Carrasquilla, M., and Llinás, M. (2017). Capturing in vivo RNA transcriptional dynamics from the malaria parasite Plasmodium falciparum. *Genome Res.*, 27(6):1074–1086.
- Paul, R. E. L., Brey, P. T., and Robert, V. (2002). Plasmodium sex determination and transmission to mosquitoes. *Trends Parasitol.*, 18(1):32–38.
- Pearson, R. D., Amato, R., Kwiatkowski, D. P., and MalariaGEN Plasmodium falciparum Community Project (2019). An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples.
- Peatey, C. L., Dixon, M. W. A., Gardiner, D. L., and Trenholme, K. R. (2013). Temporal evaluation of commitment to sexual development in Plasmodium falciparum. *Malar. J.*, 12:134.
- Pedregosa, F. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830.
- Pelle, K. G., Oh, K., Buchholz, K., Narasimhan, V., Joice, R., Milner, D. A., Brancucci, N. M., Ma, S., Voss, T. S., Ketman, K., Seydel, K. B., Taylor, T. E., Barteneva, N. S., Huttenhower, C., and Marti, M. (2015). Transcriptional profiling defines dynamics of parasite tissue sequestration during malaria infection. *Genome Med.*, 7(1):19.

- Pérez-Toledo, K., Rojas-Meza, A. P., Mancio-Silva, L., Hernández-Cuevas, N. A., Delgadillo, D. M., Vargas, M., Martínez-Calvillo, S., Scherf, A., and Hernandez-Rivas, R. (2009). Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res.*, 37(8):2596–2606.
- Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using smart-seq2. *Nat. Protoc.*, 9(1):171–181.
- Ponnudurai, T., Leeuwenberg, A. D., and Meuwissen, J. H. (1981). Chloroquine sensitivity of isolates of Plasmodium falciparum adapted to in vitro culture. *Trop. Geogr. Med.*, 33(1):50–54.
- Ponts, N., Harris, E. Y., Prudhomme, J., Wick, I., Eckhardt-Ludka, C., Hicks, G. R., Hardiman, G., Lonardi, S., and Le Roch, K. G. (2010). Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res.*, 20(2):228–238.
- Poran, A., Nötzel, C., Aly, O., Mencia-Trinchant, N., Harris, C. T., Guzman, M. L., Hassane, D. C., Elemento, O., and Kafsack, B. F. C. (2017). Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature*, 551(7678):95–99.
- Portugaliza, H. P., Llorà-Batlle, O., Rosanas-Urgell, A., and Cortés, A. (2019). Reporter lines based on the gexp02 promoter enable early quantification of sexual conversion rates in the malaria parasite Plasmodium falciparum. *Sci. Rep.*, 9(1):14595.
- Preston, M. D., Campino, S., Assefa, S. A., Echeverry, D. F., Ocholla, H., Amambua-Ngwa, A., Stewart, L. B., Conway, D. J., Borrmann, S., Michon, P., Zongo, I., Ouédraogo, J.-B., Djimde, A. A., Doumbo, O. K., Nosten, F., Pain, A., Bousema, T., Drakeley, C. J., Fairhurst, R. M., Sutherland, C. J., Roper, C., and Clark, T. G. (2014). A barcode of organellar genome polymorphisms identifies the geographic origin of Plasmodium falciparum strains. *Nat. Commun.*, 5:4052.
- R Core Team (2019). R: A language and environment for statistical computing.
- Raabe, C. A., Sanchez, C. P., Randau, G., Robeck, T., Skryabin, B. V., Chinni, S. V., Kube, M., Reinhardt, R., Ng, G. H., Manickam, R., Kuryshev, V. Y., Lanzer, M., Brosius, J., Tang, T. H., and Rozhdestvensky, T. S. (2010). A global view of the nonprotein-coding transcriptome in Plasmodium falciparum. *Nucleic Acids Res.*, 38(2):608–617.
- Rabinovich, R. N., Drakeley, C., Djimde, A. A., Hall, B. F., Hay, S. I., Hemingway, J., Kaslow, D. C., Noor, A., Okumu, F., Steketee, R., Tanner, M., Wells, T. N. C., Whittaker, M. A., Winzeler, E. A., Wirth, D. F., Whitfield, K., and Alonso, P. L. (2017). malERA: An updated research agenda for malaria elimination and eradication. *PLoS Med.*, 14(11):e1002456.

- Ransohoff, J. D., Wei, Y., and Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.*, 19(3):143–157.
- Rawat, M., Srivastava, A., Johri, S., Gupta, I., and Karmodiya, K. (2021). Single-Cell RNA sequencing reveals cellular heterogeneity and stage transition under temperature stress in synchronized *Plasmodium falciparum* cells. *Microbiol Spectr*, 9(1):e0000821.
- Rea, E., Le Roch, K. G., and Tewari, R. (2018). Sex in *Plasmodium falciparum*: Silence play between GDV1 and HP1. *Trends Parasitol.*, 34(6):450–452.
- Read, D. F., Cook, K., Lu, Y. Y., Le Roch, K. G., and Noble, W. S. (2019). Predicting gene expression in the human malaria parasite *Plasmodium falciparum* using histone modification, nucleosome positioning, and 3D localization features. *PLoS Comput. Biol.*, 15(9):e1007329.
- Reader, J., Botha, M., Theron, A., Lauterbach, S. B., Rossouw, C., Engelbrecht, D., Wepener, M., Smit, A., Leroy, D., Mancama, D., Coetzer, T. L., and Birkholtz, L.-M. (2015). Nowhere to hide: interrogating different metabolic parameters of *Plasmodium falciparum* gametocytes in a transmission blocking drug discovery pipeline towards malaria elimination. *Malar. J.*, 14:213.
- Real, E., Howick, V. M., Dahalan, F. A., Witmer, K., Cudini, J., Andradi-Brown, C., Blight, J., Davidson, M. S., Dogga, S. K., Reid, A. J., Baum, J., and Lawniczak, M. K. N. (2021). A single-cell atlas of *Plasmodium falciparum* transmission through the mosquito. *Nat. Commun.*, 12(1):3196.
- Reece, S. E., Duncan, A. B., West, S. A., and Read, A. F. (2005). Host cell preference and variable transmission strategies in malaria parasites. *Proc. Biol. Sci.*, 272(1562):511–517.
- Reece, S. E., Ramiro, R. S., and Nussey, D. H. (2009). Plastic parasites: sophisticated strategies for survival and reproduction?
- 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(5):1120–1133.
- Reid, A. J., Talman, A. M., Bennett, H. M., Gomes, A. R., Sanders, M. J., Illingworth, C. J. R., Billker, O., Berriman, M., and Lawniczak, M. K. (2018). Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *Elife*, 7.
- Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, E. J., Terry, J. M., Riordan, D. P., Srinivas, N., Fiddes, I. T., Arthur, J. G., Alvarado, L. J., Pfeiffer, K. A., Mikkelsen, T. S., Weissman, J. S., and Adamson, B. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.*, 38(8):954–961.
- Ribeiro, J. M. C. and Francischetti, I. M. B. (2003). Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu. Rev. Entomol.*, 48:73–88.

- Robert, V., Tchuinkam, T., Mulder, B., Bodo, J. M., Verhave, J. P., Carnevale, P., and Nagel, R. L. (1996). Effect of the sickle cell trait status of gametocyte carriers of *Plasmodium falciparum* on infectivity to anophelines. *Am. J. Trop. Med. Hyg.*, 54(2):111–113.
- Rosenberg, R., Wirtz, R. A., Schneider, I., and Burge, R. (1990). An estimation of the number of malaria sporozoites ejected by a feeding mosquito. *Trans. R. Soc. Trop. Med. Hyg.*, 84(2):209–212.
- Rovira-Graells, N., Gupta, A. P., Planet, E., Crowley, V. M., Mok, S., Ribas de Pouplana, L., Preiser, P. R., Bozdech, Z., and Cortés, A. (2012). Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res.*, 22(5):925–938.
- Ruberto, A. A., Bourke, C., Merienne, N., Obadia, T., Amino, R., and Mueller, I. (2021). Single-cell RNA sequencing reveals developmental heterogeneity among *Plasmodium berghei* sporozoites. *Sci. Rep.*, 11(1):4127.
- Ruiz-Orera, J., Messeguer, X., Subirana, J. A., and Alba, M. M. (2014). Long non-coding RNAs as a source of new peptides. *Elife*, 3:e03523.
- Russell, A. J. C., Sanderson, T., Bushell, E., Talman, A. M., Anar, B., Girling, G., Hunziker, M., Kent, R. S., Metcalf, T., Montandon, R., Pandey, V., Brett Roberts, A., Sayers, C., Schwach, F., Rayner, J. C., Voet, T., Modrzynska, K. K., Waters, A. P., Lawniczak, M. K. N., and Billker, O. (2021). Regulators of male and female sexual development critical for transmission of a malaria parasite.
- Santos, J. M., Josling, G., Ross, P., Joshi, P., Orchard, L., Campbell, T., Schieler, A., Cristea, I. M., and Llinás, M. (2017). Red blood cell invasion by the malaria parasite is coordinated by the PfAP2-I transcription factor. *Cell Host Microbe*, 21(6):731–741.e10.
- Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., and Lanzer, M. (1998). Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J.*, 17(18):5418–5426.
- Schneider, P., Schoone, G., Schallig, H., Verhage, D., Telgt, D., Eling, W., and Sauerwein, R. (2004). Quantification of *Plasmodium falciparum* gametocytes in differential stages of development by quantitative nucleic acid sequence-based amplification. *Mol. Biochem. Parasitol.*, 137(1):35–41.
- 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 Res.*, 43(Database issue):D1176–82.
- Shaath, H., Vishnubalaji, R., Elango, R., Khattak, S., and Alajez, N. M. (2021). Single-cell long noncoding RNA (lncRNA) transcriptome implicates MALAT1 in triple-negative breast cancer (TNBC) resistance to neoadjuvant chemotherapy. *Cell Death Discov*, 7(1):23.

- Siegel, T. N., Hon, C.-C., Zhang, Q., Lopez-Rubio, J.-J., Scheidig-Benatar, C., Martins, R. M., Sismeiro, O., Coppée, J.-Y., and Scherf, A. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. *BMC Genomics*, 15:150.
- 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(5):465–471.
- 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(7):1437–1448.
- Sinden, R. E. (1982). Gametocytogenesis of *Plasmodium falciparum* in vitro: an electron microscopic study. *Parasitology*, 84(1):1–11.
- Sinden, R. E. (1983). Sexual development of malarial parasites. *Adv. Parasitol.*, 22:153–216.
- Sinden, R. E. (2017). Developing transmission-blocking strategies for malaria control. *PLoS Pathog.*, 13(7):e1006336.
- Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S. S. G., Cox-Singh, J., Thomas, A., and Conway, D. J. (2004). A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet*, 363(9414):1017–1024.
- Singh, S., Santos, J. M., Orchard, L. M., Yamada, N., van Biljon, R., Painter, H. J., Mahony, S., and Llinás, M. (2021). The PfAP2-G2 transcription factor is a critical regulator of gametocyte maturation. *Mol. Microbiol.*, 115(5):1005–1024.
- Sinha, A., Hughes, K. R., Modrzynska, K. K., Otto, T. D., Pfander, C., Dickens, N. J., Religa, A. A., Bushell, E., Graham, A. L., Cameron, R., Kafsack, B. F. C., Williams, A. E., Llinas, M., Berriman, M., Billker, O., and Waters, A. P. (2014). A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature*, 507(7491):253–257.
- Sinka, M. E., Pironon, S., Massey, N. C., Longbottom, J., Hemingway, J., Moyes, C. L., and Willis, K. J. (2020). A new malaria vector in africa: Predicting the expansion range of *Anopheles stephensi* and identifying the urban populations at risk. *Proc. Natl. Acad. Sci. U. S. A.*, 117(40):24900–24908.
- Slack, F. J. and Chinnaiyan, A. M. (2019). The role of non-coding RNAs in oncology. *Cell*, 179(5):1033–1055.
- Smith, T. G., Lourenço, P., Carter, R., Walliker, D., and Ranford-Cartwright, L. C. (2000). Commitment to sexual differentiation in the human malaria parasite, *Plasmodium falciparum*. *Parasitology*, 121 ( Pt 2):127–133.

- Sologub, L., Kuehn, A., Kern, S., Przyborski, J., Schillig, R., and Pradel, G. (2011). Malaria proteases mediate inside-out egress of gametocytes from red blood cells following parasite transmission to the mosquito. *Cell. Microbiol.*, 13(6):897–912.
- Song, D. and Li, J. J. (2021). PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data. *Genome Biol.*, 22(1):124.
- Sorber, K., Dimon, M. T., and DeRisi, J. L. (2011). RNA-Seq analysis of splicing in *Plasmodium falciparum* uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic Acids Res.*, 39(9):3820–3835.
- Squair, J. W., Gautier, M., Kathe, C., Anderson, M. A., James, N. D., Hutson, T. H., Hudelle, R., Qaiser, T., Matson, K. J. E., Barraud, Q., Levine, A. J., La Manno, G., Skinnider, M. A., and Courtine, G. (2021). Confronting false discoveries in single-cell differential expression. *Nat. Commun.*, 12(1):5692.
- Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.*, 22(2):96–118.
- Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1):477.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, 3rd, W. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of Single-Cell data. *Cell*, 177(7):1888–1902.e21.
- Subudhi, A. K., Boopathi, P. A., Garg, S., Middha, S., Acharya, J., Pakalapati, D., Saxena, V., Aiyaz, M., Orekondy, H. B., Mugasimangalam, R. C., Sirohi, P., Kochar, S. K., Kochar, D. K., and Das, A. (2014). Natural antisense transcripts in *Plasmodium falciparum* isolates from patients with complicated malaria. *Exp. Parasitol.*, 141:39–54.
- Sullivan, J. S., Sullivan, J. J., Williams, A., Grady, K. K., Bounngaseng, A., Huber, C. S., Nace, D., Williams, T., Galland, G. G., Barnwell, J. W., and Collins, W. E. (2003). Adaptation of a strain of *Plasmodium falciparum* from Ghana to *aotus lemurinus griseimembra*, *a. nancymai*, and *a. vociferans* monkeys. *Am. J. Trop. Med. Hyg.*, 69(6):593–600.
- Sutherland, C. J., Tanomsing, N., Nolder, D., Oguike, M., Jennison, C., Pukritayakamee, S., Dolecek, C., Hien, T. T., do Rosário, V. E., Arez, A. P., Pinto, J., Michon, P., Escalante, A. A., Nosten, F., Burke, M., Lee, R., Blaze, M., Otto, T. D., Barnwell, J. W., Pain, A., Williams, J., White, N. J., Day, N. P. J., Snounou, G., Lockhart, P. J., Chiodini, P. L., Imwong, M., and Polley, S. D. (2010). Two non-recombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. *J. Infect. Dis.*, 201(10):1544–1550.
- Szachnowski, U., Andjus, S., Foretek, D., Morillon, A., and Wery, M. (2019). Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome. *Life Science Alliance*, 2(5):e201900407.

- Tadesse, F. G., Meerstein-Kessel, L., Gonçalves, B. P., Drakeley, C., Ranford-Cartwright, L., and Bousema, T. (2019). Gametocyte sex ratio: The key to understanding *Plasmodium falciparum* transmission? *Trends Parasitol.*, 35(3):226–238.
- Talman, A. M., Domarle, O., McKenzie, F. E., Ariey, F., and Robert, V. (2004). Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malar. J.*, 3:24.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., and Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*, 6(5):377–382.
- Tarashansky, A. J., Musser, J. M., Khariton, M., Li, P., Arendt, D., Quake, S. R., and Wang, B. (2021). Mapping single-cell atlases throughout metazoa unravels cell type evolution. *Elife*, 10.
- Tarr, S. J., Díaz-Ingelmo, O., Stewart, L. B., Hocking, S. E., Murray, L., Duffy, C. W., Otto, T. D., Chappell, L., Rayner, J. C., Awandare, G. A., and Conway, D. J. (2018). Schizont transcriptome variation among clinical isolates and laboratory-adapted clones of the malaria parasite *Plasmodium falciparum*. *BMC Genomics*, 19(1):894.
- Teirlinck, A. C., Roestenberg, M., van de Vegte-Bolmer, M., Scholzen, A., Heinrichs, M. J. L., Siebelink-Stoter, R., Graumans, W., van Gemert, G.-J., Teelen, K., Vos, M. W., Nganou-Makamdop, K., Borrmann, S., Rozier, Y. P. A., Erkens, M. A. A., Luty, A. J. F., Hermsen, C. C., Sim, B. K. L., van Lieshout, L., Hoffman, S. L., Visser, L. G., and Sauerwein, R. W. (2013). NF135.C10: a new *Plasmodium falciparum* clone for controlled human malaria infections. *J. Infect. Dis.*, 207(4):656–660.
- Templeton, T. J., Iyer, L. M., Anantharaman, V., Enomoto, S., Abrahante, J. E., Subramanian, G. M., Hoffman, S. L., Abrahamsen, M. S., and Aravind, L. (2004). Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res.*, 14(9):1686–1695.
- Thaithong, S. and Beale, G. H. (1981). Resistance of ten thai isolates of *Plasmodium falciparum* to chloroquine and pyrimethamine by in vitro tests. *Trans. R. Soc. Trop. Med. Hyg.*, 75(2):271–273.
- Tibúrcio, M., Niang, M., Deplaine, G., Perrot, S., Bischoff, E., Ndour, P. A., Silvestrini, F., Khattab, A., Milon, G., David, P. H., Hardeman, M., Vernick, K. D., Sauerwein, R. W., Preiser, P. R., Mercereau-Puijalon, O., Buffet, P., Alano, P., and Lavazec, C. (2012). A switch in infected erythrocyte deformability at the maturation and blood circulation of *Plasmodium falciparum* transmission stages. *Blood*, 119(24):e172–80.
- Tintó-Font, E., Michel-Todó, L., Russell, T. J., Casas-Vila, N., Conway, D. J., Bozdech, Z., Llinás, M., and Cortés, A. (2021). A heat-shock response regulated by the PfAP2-HS transcription factor protects human malaria parasites from febrile temperatures. *Nat Microbiol*, 6(9):1163–1174.
- Toenhake, C. G. and Bártfai, R. (2019). What functional genomics has taught us about transcriptional regulation in malaria parasites. *Brief. Funct. Genomics*, 18(5):290–301.

- Toenhake, C. G., Fraschka, S. A.-K., Vijayabaskar, M. S., Westhead, D. R., van Heerlingen, S. J., and Bártfai, R. (2018). Chromatin Accessibility-Based characterization of the gene regulatory network underlying *Plasmodium falciparum* Blood-Stage development. *Cell Host Microbe*, 23(4):557–569.e9.
- Trager, W. and Jensen, J. B. (1976). Human malaria parasites in continuous culture. *Science*, 193(4254):673–675.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.*, 32(4):381–386.
- Turkiewicz, A., Manko, E., Sutherland, C. J., Diez Benavente, E., Campino, S., and Clark, T. G. (2020). Genetic diversity of the *Plasmodium falciparum* GTP-cyclohydrolase 1, dihydrofolate reductase and dihydropteroate synthetase genes reveals new insights into sulfadoxine-pyrimethamine antimalarial drug resistance. *PLoS Genet.*, 16(12):e1009268.
- Udeinya, I. J., Graves, P. M., Carter, R., Aikawa, M., and Miller, L. H. (1983). *Plasmodium falciparum*: effect of time in continuous culture on binding to human endothelial cells and amelanotic melanoma cells. *Exp. Parasitol.*, 56(2):207–214.
- Ukaegbu, U. E. and Deitsch, K. W. (2015). The emerging role for RNA polymerase II in regulating virulence gene expression in malaria parasites. *PLoS Pathog.*, 11(7):e1004926.
- Ukaegbu, U. E., Kishore, S. P., Kwiatkowski, D. L., Pandarinath, C., Dahan-Pasternak, N., Dzikowski, R., and Deitsch, K. W. (2014). Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in *p. falciparum*. *PLoS Pathog.*, 10(1):e1003854.
- Usui, M., Prajapati, S. K., Ayanful-Torgby, R., Acquah, F. K., Cudjoe, E., Kakaney, C., Amponsah, J. A., Obboh, E. K., Reddy, D. K., Barbeau, M. C., Simons, L. M., Czesny, B., Raiciulescu, S., Olsen, C., Abuaku, B. K., Amoah, L. E., and Williamson, K. C. (2019). *Plasmodium falciparum* sexual differentiation in malaria patients is associated with host factors and GDV1-dependent genes. *Nat. Commun.*, 10(1):2140.
- van Biljon, R., van Wyk, R., Painter, H. J., Orchard, L., Reader, J., Niemand, J., Llinás, M., and Birkholtz, L.-M. (2019). Hierarchical transcriptional control regulates *Plasmodium falciparum* sexual differentiation. *BMC Genomics*, 20(1):920.
- Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., and Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.*, 11(1):1201.
- van Schalkwyk, D. A., Burrow, R., Henriques, G., Gadalla, N. B., Beshir, K. B., Hasford, C., Wright, S. G., Ding, X. C., Chiodini, P. L., and Sutherland, C. J. (2013). Culture-adapted *Plasmodium falciparum* isolates from UK travellers: in vitro drug sensitivity, clonality and drug resistance markers. *Malar. J.*, 12:320.

- Vembar, S. S., Scherf, A., and Siegel, T. N. (2014). Noncoding RNAs as emerging regulators of *Plasmodium falciparum* virulence gene expression. *Curr. Opin. Microbiol.*, 20:153–161.
- Venugopal, K., Hentzschel, F., Valkiūnas, G., and Marti, M. (2020). *Plasmodium* asexual growth and sexual development in the haematopoietic niche of the host. *Nat. Rev. Microbiol.*, 18(3):177–189.
- Vincke, I. H., Lips, M., and Others (1948). *Plasmodium berghei*, a new *Plasmodium* of a wild rodent in the belgian congo. *Ann. Soc. Belg. Med. Trop.*, 28(1):97–104.
- Vogt, M. B., Lahon, A., Arya, R. P., Kneubehl, A. R., Spencer Clinton, J. L., Paust, S., and Rico-Hesse, R. (2018). Mosquito saliva alone has profound effects on the human immune system. *PLoS Negl. Trop. Dis.*, 12(5):e0006439.
- Volkman, S. K., Sabeti, P. C., DeCaprio, D., Neafsey, D. E., Schaffner, S. F., Milner, Jr, D. A., Daily, J. P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M. T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D. B., Zainoun, J., Wiegand, R. C., Birren, B. W., Hartl, D. L., Galagan, J. E., Lander, E. S., and Wirth, D. F. (2007). A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.*, 39(1):113–119.
- von Grüning, H., Coradin, M., Mendoza, M. R., Reader, J., Sidoli, S., Garcia, B. A., and Birkholtz, L.-M. (2021). A dynamic and combinatorial histone code drives malaria parasite asexual and sexual development.
- Waldman, B. S., Schwarz, D., Wadsworth, 2nd, M. H., Saeij, J. P., Shalek, A. K., and Lourido, S. (2020). Identification of a master regulator of differentiation in toxoplasma. *Cell*, 180(2):359–372.e16.
- Walk, J., Reuling, I. J., Behet, M. C., Meerstein-Kessel, L., Graumans, W., van Gemert, G.-J., Siebelink-Stoter, R., van de Vegte-Bolmer, M., Janssen, T., Teelen, K., de Wilt, J. H. W., de Mast, Q., van der Ven, A. J., Diez Benavente, E., Campino, S., Clark, T. G., Huynen, M. A., Hermsen, C. C., Bijker, E. M., Scholzen, A., and Sauerwein, R. W. (2017). Modest heterologous protection after *Plasmodium falciparum* sporozoite immunization: a double-blind randomized controlled clinical trial. *BMC Med.*, 15(1):168.
- Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T., Corcoran, L. M., Burkot, T. R., and Carter, R. (1987). Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science*, 236(4809):1661–1666.
- Wang, X., He, Y., Zhang, Q., Ren, X., and Zhang, Z. (2019). Direct comparative analysis of 10X genomics chromium and smart-seq2.
- Wellems, T. E., Oduola, A. M. J., Fenton, B., Desjardins, R., Panton, L. J., and Rosario, V. E. d. (1988). Chromosome size variation occurs in cloned *Plasmodium falciparum* on in vitro cultivation. *Rev. Bras. Genet.*, pages 813–825.
- White, M. W. and Suvorova, E. S. (2018). Apicomplexa cell cycles: Something old, borrowed, lost, and new. *Trends Parasitol.*, 34(9):759–771.

- 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 Parasitol.*, 26(1):11–15.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Witkowski, B., Khim, N., Chim, P., Kim, S., Ke, S., Kloeung, N., Chy, S., Duong, S., Leang, R., Ringwald, P., Dondorp, A. M., Tripura, R., Benoit-Vical, F., Berry, A., Gorgette, O., Arieu, F., Barale, J.-C., Mercereau-Puijalon, O., and Menard, D. (2013). Reduced artemisinin susceptibility of *Plasmodium falciparum* ring stages in western cambodia. *Antimicrob. Agents Chemother.*, 57(2):914–923.
- Witmer, K., Dahalan, F. A., Metcalf, T., Talman, A. M., Howick, V. M., and Lawniczak, M. K. N. (2021). Using scRNA-seq to identify transcriptional variation in the malaria parasite ookinete stage. *Front. Cell. Infect. Microbiol.*, 11:604129.
- Woo, Y. H., Ansari, H., Otto, T. D., Klinger, C. M., Kolisko, M., Michálek, J., Saxena, A., Shanmugam, D., Tayyrov, A., Veluchamy, A., Ali, S., Bernal, A., del Campo, J., Cihlář, J., Flegontov, P., Gornik, S. G., Hajdušková, E., Horák, A., Janouškovec, J., Katris, N. J., Mast, F. D., Miranda-Saavedra, D., Mourier, T., Naeem, R., Nair, M., Panigrahi, A. K., Rawlings, N. D., Padron-Regalado, E., Ramaprasad, A., Samad, N., Tomčala, A., Wilkes, J., Neafsey, D. E., Doerig, C., Bowler, C., Keeling, P. J., Roos, D. S., Dacks, J. B., Templeton, T. J., Waller, R. F., Lukeš, J., Oborník, M., and Pain, A. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife*, 4:e06974.
- World Health Organization (2015). *Guidelines for the Treatment of Malaria*. World Health Organization, Geneva.
- World Health Organization (2020). *World malaria report 2020: 20 years of global progress and challenges*.
- Xie, T., Wang, Y., Deng, N., Huang, G., Taghavifar, F., Geng, Y., Liu, N., Kulur, V., Yao, C., Chen, P., Liu, Z., Stripp, B., Tang, J., Liang, J., Noble, P. W., and Jiang, D. (2018). Single-Cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Rep.*, 22(13):3625–3640.
- Xu, Y., Qiao, D., Wen, Y., Bi, Y., Chen, Y., Huang, Z., Cui, L., Guo, J., and Cao, Y. (2020). PfAP2-G2 is associated to production and maturation of gametocytes in *Plasmodium falciparum* via regulating the expression of PfMDV-1. *Front. Microbiol.*, 11:631444.
- Xue, Y., Theisen, T. C., Rastogi, S., Ferrel, A., Quake, S. R., and Boothroyd, J. C. (2020). A single-parasite transcriptional atlas of *Toxoplasma gondii* reveals novel control of antigen expression. *Elife*, 9.
- Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T. J., Aubouy, A., Balloux, F., Besnard, P., Bogreau, H., Carnevale, P., D'Alessandro, U., Fontenille, D., Gamboa, D., Jombart, T., Le Mire, J., Leroy, E., Maestre, A., Mayxay, M., Ménard, D., Musset, L., Newton, P. N., Nkoghé, D., Noya, O., Ollomo, B., Rogier, C., Veron, V., Wide, A., Zakeri, S., Carme, B., Legrand, E., Chevillon, C., Ayala, F. J., Renaud,

- F., and Prugnolle, F. (2012). Multiple independent introductions of *Plasmodium falciparum* in south america. *Proc. Natl. Acad. Sci. U. S. A.*, 109(2):511–516.
- Yeoh, L. M., Goodman, C. D., Mollard, V., McFadden, G. I., and Ralph, S. A. (2017). Comparative transcriptomics of female and male gametocytes in *Plasmodium berghei* and the evolution of sex in alveolates. *BMC Genomics*, 18(1):734.
- Young, J. A., Fivelman, Q. L., Blair, P. L., de la Vega, P., Le Roch, K. G., Zhou, Y., Carucci, D. J., Baker, D. A., and Winzeler, E. A. (2005). The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol. Biochem. Parasitol.*, 143(1):67–79.
- Yu, R., Jih, G., Iglesias, N., and Moazed, D. (2014). Determinants of heterochromatic siRNA biogenesis and function. *Mol. Cell*, 53(2):262–276.
- Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T., and Kaneko, I. (2010). Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol. Microbiol.*, 75(4):854–863.
- Zhang, K., Huang, K., Luo, Y., and Li, S. (2014). Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. *BMC Genomics*, 15:845.
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J., and Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, 8:14049.
- Zheng, H., Brennan, K., Hernaez, M., and Gevaert, O. (2019). Benchmark of long non-coding RNA quantification for RNA sequencing of cancer samples. *Gigascience*, 8(12).
- Zhou, J., Xu, J., Zhang, L., Liu, S., Ma, Y., Wen, X., Hao, J., Li, Z., Ni, Y., Li, X., Zhou, F., Li, Q., Wang, F., Wang, X., Si, Y., Zhang, P., Liu, C., Bartolomei, M., Tang, F., Liu, B., Yu, J., and Lan, Y. (2019). Combined Single-Cell profiling of lncRNAs and functional screening reveals that H19 is pivotal for embryonic hematopoietic stem cell development. *Cell Stem Cell*, 24(2):285–298.e5.



# Appendix A

## Supplementary Tables

### A.1 Supplementary material for Chapter 3

Table A.1 Cluster assignments for gene by cell gene UMAP

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0114000	EPF1	14	PF3D7_0825900		5	PF3D7_1230000		14
PF3D7_0113800		9	PF3D7_0825800		4	PF3D7_1229800	MyoD	3
PF3D7_0113000	GARP	6	PF3D7_0825700		4	PF3D7_1229700		4
PF3D7_0112200	MRP1	6	PF3D7_0825500	KRI1	7	PF3D7_1229600		4
PF3D7_0111600		13	PF3D7_0825300		4	PF3D7_1229500	CCT3	6
PF3D7_0111500		9	PF3D7_0825000		1	PF3D7_1229400	MIF	6
PF3D7_0111400	GEXP19	1	PF3D7_0824900		5	PF3D7_1229300		9
PF3D7_0111300		8	PF3D7_0824800		11	PF3D7_1229100	MRP2	11
PF3D7_0111000		0	PF3D7_0824500		5	PF3D7_1229000		13
PF3D7_0110800		7	PF3D7_0824300		1	PF3D7_1228900		5
PF3D7_0110700		0	PF3D7_0824200		8	PF3D7_1228800		8
PF3D7_0110600	PIP5K	1	PF3D7_0824100		3	PF3D7_1228700		8
PF3D7_0110500		8	PF3D7_0823300	GCN5	2	PF3D7_1228600	MSP9	9
PF3D7_0110400	RPB9	5	PF3D7_0823200		11	PF3D7_1228400		8
PF3D7_0110200	ERV1	5	PF3D7_0823000	VPS15	13	PF3D7_1228300	NEK1	0
PF3D7_0110100		7	PF3D7_0822900		3	PF3D7_1228100	LRR13	0
PF3D7_0110000		13	PF3D7_0822400		13	PF3D7_1227900		11
PF3D7_0109850		9	PF3D7_0822200		7	PF3D7_1227800		6
PF3D7_0109800	cPheRS	11	PF3D7_0822100	MED7	13	PF3D7_1227700		10
PF3D7_0109700	RRP36	5	PF3D7_0821800	SEC61B	4	PF3D7_1227600		8
PF3D7_0109600		5	PF3D7_0821700		11	PF3D7_1227500	CYC2	8
PF3D7_0109500		4	PF3D7_0821600	CLP1	13	PF3D7_1227400		11
PF3D7_0109400		13	PF3D7_0821300	DHX36	5	PF3D7_1227300		7
PF3D7_0109300		5	PF3D7_0821100	PK1	2	PF3D7_1227200	K1	2
PF3D7_0109100	CCp5	5	PF3D7_0821000		9	PF3D7_1227000		9
PF3D7_0109000	PHIL1	3	PF3D7_0820900		2	PF3D7_1226900		10
PF3D7_0108800		2	PF3D7_0820700	KDH	10	PF3D7_1226800	ATX3	1
PF3D7_0108700	PSOP24	10	PF3D7_0820500		13	PF3D7_1226700	RRP9	11
PF3D7_0108600		4	PF3D7_0820300		10	PF3D7_1226400		5
PF3D7_0108500		12	PF3D7_0820200	PGPS	2	PF3D7_1225900		8
PF3D7_0108300	ARP	8	PF3D7_0820100		13	PF3D7_1225800	UBA1	2
PF3D7_0108100		5	PF3D7_0820000	SRCAP	11	PF3D7_1225700		6
PF3D7_0107900		5	PF3D7_0819800		14	PF3D7_1225600		7
PF3D7_0107800	MRE11	2	PF3D7_0819700		9	PF3D7_1225500		13
PF3D7_0107700		5	PF3D7_0819600		5	PF3D7_1225400		4
PF3D7_0107600		14	PF3D7_0819000		8	PF3D7_1225200		8
PF3D7_0107500		3	PF3D7_0818900	HSP70	11	PF3D7_1225100		6
PF3D7_0107400		13	PF3D7_0818800	IMP4	5	PF3D7_1224900	SF3B6	5

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0107000.1		14	PF3D7_0818700		8	PF3D7_1224700		9
PF3D7_0106900	IspD	2	PF3D7_0818500		14	PF3D7_1224300	PABP	2
PF3D7_0106800	RAB5c	1	PF3D7_0818300		5	PF3D7_1224200		4
PF3D7_0106700	AARP2	13	PF3D7_0818200	14-3-3I	11	PF3D7_1224100		9
PF3D7_0106500		0	PF3D7_0818100		3	PF3D7_1224000	GCH1	6
PF3D7_0106400		5	PF3D7_0818000		1	PF3D7_1223800	YHM2	6
PF3D7_0106300	ATP6	2	PF3D7_0817900	HMGB2	1	PF3D7_1223700	VIT	2
PF3D7_0106200		1	PF3D7_0817800		1	PF3D7_1223600		11
PF3D7_0106100		7	PF3D7_0817700	RON5	9	PF3D7_1223500		1
PF3D7_0106000		1	PF3D7_0817600		1	PF3D7_1223400		3
PF3D7_0105900		4	PF3D7_0817500		1	PF3D7_1223300	GyrA	9
PF3D7_0105800		1	PF3D7_0817400		8	PF3D7_1223200		9
PF3D7_0105700		7	PF3D7_0817300		2	PF3D7_1223100	PKAr	3
PF3D7_0105600		8	PF3D7_0817200		10	PF3D7_1222900		5
PF3D7_0105200		7	PF3D7_0817100		13	PF3D7_1222800		4
PF3D7_0104500		6	PF3D7_0816800	DMC1	5	PF3D7_1222700	GAP45	3
PF3D7_0104400	LytB	5	PF3D7_0816600	ClpB1	11	PF3D7_1222400	ApiAP2	7
PF3D7_0104300	UBP1	3	PF3D7_0816400		13	PF3D7_1222300	GRP94	1
PF3D7_0104200		3	PF3D7_0816300		0	PF3D7_1222100		2
PF3D7_0103700	PSTK	13	PF3D7_0815800	VPS9	11	PF3D7_1221800		2
PF3D7_0103600		14	PF3D7_0815600	EIF3G	5	PF3D7_1221400	IMC1h	5
PF3D7_0103500		8	PF3D7_0815500		9	PF3D7_1221300		1
PF3D7_0103400		3	PF3D7_0815100		4	PF3D7_1221100.1		4
PF3D7_0103300		8	PF3D7_0815000	Sel3	4	PF3D7_1221000	SET10	2
PF3D7_0103200	NT4	11	PF3D7_0814400		1	PF3D7_1220900	HP1	9
PF3D7_0103100	VPS51	1	PF3D7_0814200	ALBA1	11	PF3D7_1220700		5
PF3D7_0102500	EBA181	3	PF3D7_0814000		6	PF3D7_1220400	DRN1	13
PF3D7_0102200	RESA	12	PF3D7_0813900		6	PF3D7_1220300		10
PF3D7_0220800	CLAG2	9	PF3D7_0813800	GMD	4	PF3D7_1220000		5
PF3D7_0220700		6	PF3D7_0813700	ABCF1	13	PF3D7_1219600	ATPase2	9
PF3D7_0220600		12	PF3D7_0813600		6	PF3D7_1219100		10
PF3D7_0220200		6	PF3D7_0813500		13	PF3D7_1219000		10
PF3D7_0220100		8	PF3D7_0813400		4	PF3D7_1218900		7
PF3D7_0220000	LSA3	12	PF3D7_0813300		6	PF3D7_1218800	PSOP17	5
PF3D7_0219800		10	PF3D7_0813200		10	PF3D7_1218400		7
PF3D7_0219600	RFC1	8	PF3D7_0813100		1	PF3D7_1218200		0
PF3D7_0219500		5	PF3D7_0813000		7	PF3D7_1218000	TRAMP	3
PF3D7_0218700	PRP45	7	PF3D7_0812600		5	PF3D7_1217900		1
PF3D7_0218600		8	PF3D7_0812500		8	PF3D7_1217700		1
PF3D7_0218400	DDX47	6	PF3D7_0812400	KARalpha	11	PF3D7_1217500		0
PF3D7_0218300		10	PF3D7_0812100		8	PF3D7_1217400		1
PF3D7_0218200		7	PF3D7_0812000		8	PF3D7_1217300		8
PF3D7_0218100		9	PF3D7_0811900		9	PF3D7_1217200	MRD1	11
PF3D7_0218000	RFC2	8	PF3D7_0811800		6	PF3D7_1216900		6
PF3D7_0217800	RPS26	6	PF3D7_0811700		9	PF3D7_1216700	PLP2	0
PF3D7_0217600		7	PF3D7_0811600		9	PF3D7_1216600	CelTOS	5
PF3D7_0217500	CDPK1	3	PF3D7_0811400		6	PF3D7_1216500	MDV1	13
PF3D7_0217200		14	PF3D7_0811300	CAF1	2	PF3D7_1216400		10
PF3D7_0217100		10	PF3D7_0811200	EMC1	11	PF3D7_1216300	SRP19	1
PF3D7_0217000		14	PF3D7_0811000	CUL1	8	PF3D7_1216000		4
PF3D7_0216900		0	PF3D7_0810900.1		1	PF3D7_1215700		0
PF3D7_0216600		1	PF3D7_0810600	DBP1	14	PF3D7_1215600		5
PF3D7_0216500		13	PF3D7_0810500	PPM7	4	PF3D7_1215500		0
PF3D7_0216400	VPS45	1	PF3D7_0810400	AQP2	4	PF3D7_1215400		7
PF3D7_0216300		11	PF3D7_0810300	PPM5	3	PF3D7_1215100		7
PF3D7_0216200		6	PF3D7_0810200	ABCK1	5	PF3D7_1214800	OBC13	5
PF3D7_0216100		0	PF3D7_0810000		5	PF3D7_1214500		4
PF3D7_0216000		13	PF3D7_0809900	JmjC1	11	PF3D7_1214100	PIGO	2
PF3D7_0215800	ORC5	0	PF3D7_0809800		0	PF3D7_1213900		5
PF3D7_0215600		13	PF3D7_0809700	RUVB1	0	PF3D7_1213500		4
PF3D7_0215500		0	PF3D7_0809600		8	PF3D7_1213400		5
PF3D7_0215400		10	PF3D7_0809400		13	PF3D7_1212900	BDP2	8
PF3D7_0215000	ACS9	14	PF3D7_0809200	pfa55-14	4	PF3D7_1212800		13
PF3D7_0214900	RON6	9	PF3D7_0808400	SEC28	10	PF3D7_1212700	EIF3A	11
PF3D7_0214800		3	PF3D7_0808300		8	PF3D7_1212600.1		4
PF3D7_0214700		3	PF3D7_0808100		2	PF3D7_1212300		0
PF3D7_0214600		3	PF3D7_0808000		6	PF3D7_1212200		0
PF3D7_0214400		13	PF3D7_0807900	TyrRS	10	PF3D7_1212100	PPP1	7
PF3D7_0214300		14	PF3D7_0807800	RPN10	1	PF3D7_1212000	TPx(GI)	2
PF3D7_0214100	SEC31	2	PF3D7_0807700	DegP	9	PF3D7_1211900	ATP4	3
PF3D7_0214000	CCT8	6	PF3D7_0807600		2	PF3D7_1211800	PfpUB	10

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0213900		7	PF3D7_0807500		2	PF3D7_1211600	LSD1	11
PF3D7_0213800		5	PF3D7_0807400	COQ10	4	PF3D7_1211400	PFJ4	11
PF3D7_0213600		14	PF3D7_0807300	RAB18	2	PF3D7_1211300	MCM8	2
PF3D7_0213500		0	PF3D7_0807100	DHH1	8	PF3D7_1211200		0
PF3D7_0213200		14	PF3D7_0806800		11	PF3D7_1211000		4
PF3D7_0213100	SIS1	13	PF3D7_0806700		14	PF3D7_1210600		14
PF3D7_0212800	MATE	7	PF3D7_0806600		0	PF3D7_1210200		5
PF3D7_0212600	SPATR	3	PF3D7_0806500		3	PF3D7_1210100	SYN2	1
PF3D7_0212500		11	PF3D7_0806300		2	PF3D7_1210000	RPL1	1
PF3D7_0212400		8	PF3D7_0806200		10	PF3D7_1209700		5
PF3D7_0212300		4	PF3D7_0806100		11	PF3D7_1209300	KROX1	2
PF3D7_0212100		2	PF3D7_0806000		13	PF3D7_1208900		8
PF3D7_0212000		7	PF3D7_0805800		4	PF3D7_1208800		13
PF3D7_0211900		8	PF3D7_0805700	FIKK8	9	PF3D7_1208700		4
PF3D7_0211800	AsnRS	11	PF3D7_0805500		11	PF3D7_1208100		11
PF3D7_0211700	TKL1	8	PF3D7_0805300		4	PF3D7_1207800		0
PF3D7_0211500		5	PF3D7_0805100		5	PF3D7_1207700		4
PF3D7_0211400	KASIII	5	PF3D7_0804900		3	PF3D7_1207300		4
PF3D7_0210900		11	PF3D7_0804800	CYP24	6	PF3D7_1207200		2
PF3D7_0210600		3	PF3D7_0804700		13	PF3D7_1207100		11
PF3D7_0210500		3	PF3D7_0804500		8	PF3D7_1207000		8
PF3D7_0210200		8	PF3D7_0804300		5	PF3D7_1206900		0
PF3D7_0210100.1		11	PF3D7_0804000		8	PF3D7_1206800		11
PF3D7_0210000	Sec61 $\gamma$	4	PF3D7_0803800		9	PF3D7_1206700	EIF5	8
PF3D7_0209800	UAP56	11	PF3D7_0803600.1		2	PF3D7_1206500		13
PF3D7_0209700		13	PF3D7_0803500		3	PF3D7_1206300		1
PF3D7_0209500		5	PF3D7_0803400	RAD54	5	PF3D7_1206200	EIF3C	11
PF3D7_0209300	IspF	6	PF3D7_0803200		1	PF3D7_1205900		10
PF3D7_0209000	P230	14	PF3D7_0803100	UTP14	7	PF3D7_1205800	HMGB3	2
PF3D7_0208900	P230p	0	PF3D7_0803000	CYP81	7	PF3D7_1205700	tGloII	0
PF3D7_0208800		5	PF3D7_0802800	CNA	9	PF3D7_1205600		0
PF3D7_0208600	RRF1	2	PF3D7_0802600	AC $\beta$	9	PF3D7_1205500		8
PF3D7_0208400		8	PF3D7_0802500		2	PF3D7_1205400		11
PF3D7_0208200	KRR1	11	PF3D7_0802400		11	PF3D7_1205000		4
PF3D7_0208100.1		10	PF3D7_0802200		6	PF3D7_1204400		4
PF3D7_0208000	SERA1	9	PF3D7_0802100	ApiAP2	10	PF3D7_1204300	EIF5A	11
PF3D7_0207800	SERA3	9	PF3D7_0802000	GDH3	6	PF3D7_1204200		4
PF3D7_0207700	SERA4	9	PF3D7_0801900		7	PF3D7_1204100		5
PF3D7_0207600	SERA5	9	PF3D7_0801800		11	PF3D7_1203700	NAPL	11
PF3D7_0207500	SERA6	9	PF3D7_0801700	SEN2	11	PF3D7_1203500	SUA5	5
PF3D7_0207400	SERA7	9	PF3D7_0801000		11	PF3D7_1203400	MFS4	2
PF3D7_0207100		3	PF3D7_0936800		12	PF3D7_1203300		14
PF3D7_0207000	MSP4	3	PF3D7_0936300	REX3	12	PF3D7_1203200	SRP14	4
PF3D7_0206900.1		3	PF3D7_0935900	REX1	12	PF3D7_1203100		4
PF3D7_0206800	MSP2	3	PF3D7_0935800	CLAG9	9	PF3D7_1203000	ORC1	8
PF3D7_0206500		1	PF3D7_0935500	GEXP22	2	PF3D7_1202900	HMGB1	2
PF3D7_0206200	PAT	3	PF3D7_0935200	VP533	7	PF3D7_1202600		10
PF3D7_0206000	RAD2	3	PF3D7_0935100	KSH1	6	PF3D7_1202300		14
PF3D7_0205900	RPN1	2	PF3D7_0934800	PKAc	9	PF3D7_1202000		14
PF3D7_0205800	PH1	4	PF3D7_0934700		7	PF3D7_1201800	COX19	10
PF3D7_0205700.1		4	PF3D7_0934400	ApiAP2	5	PF3D7_1201700		13
PF3D7_0205300		0	PF3D7_0934000		2	PF3D7_1201600	NEK3	14
PF3D7_0205100		14	PF3D7_0933800		0	PF3D7_1201500		7
PF3D7_0205000		0	PF3D7_0933500		8	PF3D7_1201000		6
PF3D7_0204900		1	PF3D7_0933400		4	PF3D7_1200800	FIKK12	12
PF3D7_0204800		5	PF3D7_0933300		11	PF3D7_1200700	ACS7	12
PF3D7_0204700	HT	2	PF3D7_0933200		1	PF3D7_1200600	VAR2CSA	12
PF3D7_0204500	AspAT	6	PF3D7_0933100		0	PF3D7_1371800		6
PF3D7_0204300		0	PF3D7_0933000		8	PF3D7_1370300	MAHRP1	12
PF3D7_0204200		3	PF3D7_0932800	CSE1	11	PF3D7_1369600		2
PF3D7_0204100		10	PF3D7_0932600	RPS6	10	PF3D7_1369400		7
PF3D7_0204000		0	PF3D7_0932500	DHHC6	1	PF3D7_1369300		13
PF3D7_0203900		13	PF3D7_0932300	M18AAP	4	PF3D7_1369200		1
PF3D7_0203800		4	PF3D7_0932200	PFN	3	PF3D7_1369100		5
PF3D7_0203700	MAK16	5	PF3D7_0932100		10	PF3D7_1368800	ERCC4	7
PF3D7_0203400		14	PF3D7_0932000		5	PF3D7_1368700		2
PF3D7_0203300	ERCC1	5	PF3D7_0931900	AKLP2	0	PF3D7_1368400		4
PF3D7_0203200		5	PF3D7_0931700		0	PF3D7_1368100	RPN11	10
PF3D7_0203100		1	PF3D7_0931500		4	PF3D7_1367700	AlaRS	6
PF3D7_0203000	ROPE	3	PF3D7_0931400		2	PF3D7_1367500		4
PF3D7_0202500	ETRAPMP2	12	PF3D7_0931100		11	PF3D7_1367200		0

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0202400		6	PF3D7_0931000		2	PF3D7_1367100		7
PF3D7_0202100	LSAP2	3	PF3D7_0930800		8	PF3D7_1366900		11
PF3D7_0202000	KAHRP	6	PF3D7_0930700		0	PF3D7_1366600	SRPR $\alpha$	13
PF3D7_0201900	EMP3	12	PF3D7_0930600.1		13	PF3D7_1366400	RHOP148	1
PF3D7_0201800	KAHsp40	12	PF3D7_0930500	DGK1	9	PF3D7_1366300		8
PF3D7_0323800		5	PF3D7_0930300	MSP1	3	PF3D7_1366200		4
PF3D7_0323700	SART1	2	PF3D7_0930200	LRR8	0	PF3D7_1365900		13
PF3D7_0323500	SMN	3	PF3D7_0930100		2	PF3D7_1365800		0
PF3D7_0323400	RIPR	3	PF3D7_0930000		4	PF3D7_1365600		9
PF3D7_0323200		10	PF3D7_0929600		5	PF3D7_1365500	GCVT	6
PF3D7_0323100		0	PF3D7_0929400	RhopH2	9	PF3D7_1365000		13
PF3D7_0322900		6	PF3D7_0929300		13	PF3D7_1364600		2
PF3D7_0322800		0	PF3D7_0929200		7	PF3D7_1364400		8
PF3D7_0322700		8	PF3D7_0928800		13	PF3D7_1364300	PRP16	10
PF3D7_0322600		0	PF3D7_0928700		7	PF3D7_1364200		8
PF3D7_0322500	IscA2	1	PF3D7_0928500		4	PF3D7_1364100	P92	9
PF3D7_0322400		11	PF3D7_0928200		11	PF3D7_1364000		10
PF3D7_0322300	DGAT	2	PF3D7_0928100		14	PF3D7_1363800	GEX1	5
PF3D7_0322000	CYP19A	6	PF3D7_0928000.1		1	PF3D7_1363500		1
PF3D7_0321900	CARL	8	PF3D7_0927800	COX5B	5	PF3D7_1363400	DOA1	8
PF3D7_0321800		0	PF3D7_0927600		1	PF3D7_1363200	PNKP	0
PF3D7_0321700		8	PF3D7_0927400		0	PF3D7_1363100		8
PF3D7_0321600	DDX42	13	PF3D7_0927300	FH	9	PF3D7_1363000		7
PF3D7_0321500		2	PF3D7_0927200		13	PF3D7_1362900		5
PF3D7_0321300		5	PF3D7_0927100		13	PF3D7_1362800		10
PF3D7_0321100		9	PF3D7_0926800		0	PF3D7_1362700		7
PF3D7_0321000		2	PF3D7_0926600		14	PF3D7_1362600		5
PF3D7_0320900	H2A.Z	2	PF3D7_0926500		7	PF3D7_1362400	Pcalp	1
PF3D7_0320800	DOZI	13	PF3D7_0926400		5	PF3D7_1362100		5
PF3D7_0320700	SPC2	10	PF3D7_0926300		14	PF3D7_1361900	PCNA1	8
PF3D7_0320600		13	PF3D7_0926200		3	PF3D7_1361800		3
PF3D7_0320500	Nico	3	PF3D7_0926100		8	PF3D7_1361300		4
PF3D7_0320200		4	PF3D7_0926000		0	PF3D7_1361200		8
PF3D7_0320100	SEC22	1	PF3D7_0925900		11	PF3D7_1361100	SEC24A	9
PF3D7_0320000	I2	1	PF3D7_0925800		8	PF3D7_1360800	FLN	2
PF3D7_0319700	ABCI3	14	PF3D7_0925300	aPRS	9	PF3D7_1360700	PIAS	8
PF3D7_0319500		13	PF3D7_0925200	RRP8	13	PF3D7_1360500	GC $\beta$	0
PF3D7_0319400		8	PF3D7_0925100		4	PF3D7_1360400		7
PF3D7_0319200		11	PF3D7_0925000		0	PF3D7_1360300		5
PF3D7_0319000	ATPase7	2	PF3D7_0924800	TRF1	4	PF3D7_1359900		1
PF3D7_0318900		2	PF3D7_0924700	SF3A3	7	PF3D7_1359700		10
PF3D7_0318700		3	PF3D7_0924600		4	PF3D7_1359600		2
PF3D7_0318500		14	PF3D7_0924500		10	PF3D7_1359500		13
PF3D7_0318400		0	PF3D7_0924400		14	PF3D7_1359300	DIS3	6
PF3D7_0318300		8	PF3D7_0924300	TPK	9	PF3D7_1359200	HMGB4	7
PF3D7_0318200	RPB1	2	PF3D7_0924100		11	PF3D7_1359000		14
PF3D7_0318000		4	PF3D7_0924000		3	PF3D7_1358800	RPS15	6
PF3D7_0317800		6	PF3D7_0923900		13	PF3D7_1358700	YOP1L	9
PF3D7_0317700		7	PF3D7_0923800.1		2	PF3D7_1358600		13
PF3D7_0317500	EG5	8	PF3D7_0923400		11	PF3D7_1358400		5
PF3D7_0317300		11	PF3D7_0922800		8	PF3D7_1358300	ROM7	14
PF3D7_0317200	CRK4	9	PF3D7_0922700	PRP18	13	PF3D7_1358200		11
PF3D7_0316900		3	PF3D7_0922600		5	PF3D7_1358100		13
PF3D7_0316800		6	PF3D7_0922500	PGK	6	PF3D7_1358000		2
PF3D7_0316500	NUF2	13	PF3D7_0922300		4	PF3D7_1357800	CCT4	6
PF3D7_0316400		8	PF3D7_0922200	SAMS	2	PF3D7_1357700	UTP21	7
PF3D7_0316300.2		9	PF3D7_0922100		11	PF3D7_1357500		0
PF3D7_0316200		8	PF3D7_0922000		0	PF3D7_1357400		14
PF3D7_0316100		5	PF3D7_0921900		11	PF3D7_1357000		6
PF3D7_0316000	MA	3	PF3D7_0921700		9	PF3D7_1356800	ARK3	1
PF3D7_0315700		5	PF3D7_0921600		8	PF3D7_1356100		1
PF3D7_0315400		2	PF3D7_0921200		11	PF3D7_1356000		5
PF3D7_0315300		0	PF3D7_0920900	PRP24	11	PF3D7_1355700	NIF3	11
PF3D7_0315000		5	PF3D7_0920700		3	PF3D7_1355500	PP5	14
PF3D7_0314900		13	PF3D7_0920600		8	PF3D7_1355100	MCM6	0
PF3D7_0314800		9	PF3D7_0920400		2	PF3D7_1354900		8
PF3D7_0314300	Der1-2	8	PF3D7_0920300		4	PF3D7_1354500	ADSS	6
PF3D7_0314200		8	PF3D7_0920200		2	PF3D7_1354300		1
PF3D7_0314000	P23	4	PF3D7_0920000	ELO3	13	PF3D7_1354200	IP5P	5
PF3D7_0313900		13	PF3D7_0919900		10	PF3D7_1354000		14
PF3D7_0313800		0	PF3D7_0919800		8	PF3D7_1353300		1

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0313600		14	PF3D7_0919600		14	PF3D7_1353100		12
PF3D7_0313400		5	PF3D7_0919400	PDI9	0	PF3D7_1352800	MON1	11
PF3D7_0313100	HRD3	6	PF3D7_0919200		3	PF3D7_1352700		8
PF3D7_0313000		5	PF3D7_0919100		10	PF3D7_1352500		9
PF3D7_0312800		6	PF3D7_0919000	NAPS	11	PF3D7_1352400		9
PF3D7_0312400	GSK3	1	PF3D7_0918900	gammaGCS	8	PF3D7_1352300		14
PF3D7_0312300	RPN12	1	PF3D7_0918700		3	PF3D7_1352100	ABCB6	1
PF3D7_0312200		0	PF3D7_0918400		0	PF3D7_1352000		2
PF3D7_0312100		1	PF3D7_0918100		4	PF3D7_1351900		7
PF3D7_0311800		10	PF3D7_0918000	GAP50	9	PF3D7_1351700	IMC1f	3
PF3D7_0311700		5	PF3D7_0917900	HSP70	2	PF3D7_1351600	GK	5
PF3D7_0311400		14	PF3D7_0917500		2	PF3D7_1351400		6
PF3D7_0311300		8	PF3D7_0917400		0	PF3D7_1351200		8
PF3D7_0311200		8	PF3D7_0917100	OGG1	8	PF3D7_1351000		8
PF3D7_0311100		8	PF3D7_0917000	MOP	8	PF3D7_1350800		4
PF3D7_0310500	DHX57	8	PF3D7_0916900		2	PF3D7_1350500		9
PF3D7_0310400	PIESP1	9	PF3D7_0916700	HoMu	3	PF3D7_1350300		7
PF3D7_0310300		13	PF3D7_0916400		7	PF3D7_1350200	SNRPE	5
PF3D7_0310200		2	PF3D7_0916200		0	PF3D7_1350100	KRS1	6
PF3D7_0310000		4	PF3D7_0916000	MFS2	5	PF3D7_1350000		5
PF3D7_0309900		7	PF3D7_0915800	GLTP	4	PF3D7_1349600		8
PF3D7_0309700	SBP2	14	PF3D7_0915700		13	PF3D7_1349500		10
PF3D7_0309600	PfP2	6	PF3D7_0915400	PFK9	6	PF3D7_1349300	TKL3	10
PF3D7_0309500	AS	11	PF3D7_0915200		0	PF3D7_1349200		6
PF3D7_0309200	ARK2	8	PF3D7_0915100	UBC9	14	PF3D7_1348800		0
PF3D7_0308900	SF3B1	8	PF3D7_0915000	NDH2	5	PF3D7_1348700	WDR16	0
PF3D7_0308700		3	PF3D7_0914900		1	PF3D7_1348600		9
PF3D7_0308600	PRPF19	11	PF3D7_0914800		5	PF3D7_1348400		13
PF3D7_0308500		6	PF3D7_0914600	ELF1	7	PF3D7_1348200		13
PF3D7_0308200	CCT7	7	PF3D7_0914500		8	PF3D7_1348000		13
PF3D7_0308100		9	PF3D7_0914400		4	PF3D7_1347900		8
PF3D7_0308000		0	PF3D7_0914100		1	PF3D7_1347700	ECT	7
PF3D7_0307900		0	PF3D7_0914000		5	PF3D7_1347600		4
PF3D7_0307700		8	PF3D7_0913900	RSSapi	4	PF3D7_1347500	ALBA4	6
PF3D7_0307600		1	PF3D7_0913800		3	PF3D7_1347400		13
PF3D7_0307500		1	PF3D7_0913700		1	PF3D7_1347200	NT1	11
PF3D7_0307400	ClpP	9	PF3D7_0913600		1	PF3D7_1347100	TOP3	5
PF3D7_0307300		14	PF3D7_0913500		10	PF3D7_1346800	P47	5
PF3D7_0307200		6	PF3D7_0913200	EF-1 $\beta$	6	PF3D7_1346700	P48/45	14
PF3D7_0306900		6	PF3D7_0913000		6	PF3D7_1346600		10
PF3D7_0306800	CCT2	6	PF3D7_0912900	RPN8	11	PF3D7_1346400		10
PF3D7_0306700		4	PF3D7_0912700		5	PF3D7_1346300	ALBA2	6
PF3D7_0306300	GRX1	5	PF3D7_0912600		0	PF3D7_1346100	SEC61	1
PF3D7_0306200	AHA1	5	PF3D7_0912500		10	PF3D7_1346000		5
PF3D7_0306100		5	PF3D7_0912400		2	PF3D7_1345900	SPC25	4
PF3D7_0305800		0	PF3D7_0912200		4	PF3D7_1345800		0
PF3D7_0305700		10	PF3D7_0912000		8	PF3D7_1345400		0
PF3D7_0305600		5	PF3D7_0911900	ICP	11	PF3D7_1345300		5
PF3D7_0305500		2	PF3D7_0911800		9	PF3D7_1345200	ROM6	2
PF3D7_0305300		11	PF3D7_0911700		2	PF3D7_1345000		4
PF3D7_0305200		1	PF3D7_0911600		10	PF3D7_1344900		0
PF3D7_0305100		8	PF3D7_0911500.1		4	PF3D7_1344700		8
PF3D7_0305000	EF-Ts	9	PF3D7_0911400		8	PF3D7_1344500	USB1	1
PF3D7_0304800		2	PF3D7_0911100		3	PF3D7_1344400		4
PF3D7_0304400.1		6	PF3D7_0911000		4	PF3D7_1344300		3
PF3D7_0304200	EHD	3	PF3D7_0910500		3	PF3D7_1344200	HSP110	9
PF3D7_0304100	IMC1e	5	PF3D7_0910300		11	PF3D7_1344100		7
PF3D7_0303900		4	PF3D7_0910200		2	PF3D7_1343800		10
PF3D7_0303700	BCKDH-E2	1	PF3D7_0909900		8	PF3D7_1343700	K13	10
PF3D7_0303500		8	PF3D7_0909700		14	PF3D7_1343600		0
PF3D7_0303200		9	PF3D7_0909600		9	PF3D7_1343400	RAD5	0
PF3D7_0303100		10	PF3D7_0909400		4	PF3D7_1343300		8
PF3D7_0302900		2	PF3D7_0909000		5	PF3D7_1343200		10
PF3D7_0302600	ABCB4	6	PF3D7_0908900		1	PF3D7_1343100		9
PF3D7_0302500	CLAG3.1	9	PF3D7_0908500		9	PF3D7_1343000	PMT	2
PF3D7_0302200	CLAG3.2	9	PF3D7_0908400		1	PF3D7_1342900	ApiAP2	11
PF3D7_0302100	SRPK1	0	PF3D7_0908300.1		4	PF3D7_1342800	PEPCK	4
PF3D7_0302000	PRP46	8	PF3D7_0908200		10	PF3D7_1342600	MyoA	3
PF3D7_0301700		12	PF3D7_0908100		1	PF3D7_1342400	CK2 $\beta$ 2	11
PF3D7_0301400		6	PF3D7_0907900	PDF	10	PF3D7_1342300		8
PF3D7_0424700	FIKK4.2	12	PF3D7_0907800		4	PF3D7_1342200		0

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0424600		12	PF3D7_0907600		7	PF3D7_1342000		6
PF3D7_0424500	FIKK4.1	12	PF3D7_0907500		13	PF3D7_1341900		5
PF3D7_0424200	RH4	3	PF3D7_0907400	ClpY	9	PF3D7_1341800	IMC1k	5
PF3D7_0424100	RH5	3	PF3D7_0907200		10	PF3D7_1341700		5
PF3D7_0423700	ETRAMP4	2	PF3D7_0907100		7	PF3D7_1341300		6
PF3D7_0423600		13	PF3D7_0906910		13	PF3D7_1341200		6
PF3D7_0423500	GAPM2	9	PF3D7_0906800		7	PF3D7_1340900	PiT	3
PF3D7_0423300		1	PF3D7_0906700	LRR9	6	PF3D7_1340600	DBR1	6
PF3D7_0423100		13	PF3D7_0906600		11	PF3D7_1340500		5
PF3D7_0422900		5	PF3D7_0906500		5	PF3D7_1340300		7
PF3D7_0422500	BRR2	8	PF3D7_0906400		13	PF3D7_1339700		9
PF3D7_0422400	RPS19	6	PF3D7_0906200		8	PF3D7_1339600		6
PF3D7_0422300		14	PF3D7_0906000	RNaseII	0	PF3D7_1339300		3
PF3D7_0422200		14	PF3D7_0905800		8	PF3D7_1338900		14
PF3D7_0422100		6	PF3D7_0905600	WDR66	0	PF3D7_1338800		5
PF3D7_0421900		4	PF3D7_0905500		1	PF3D7_1338700		10
PF3D7_0421700		0	PF3D7_0905400	RhopH3	9	PF3D7_1338600		9
PF3D7_0420600		10	PF3D7_0905300		0	PF3D7_1338500		0
PF3D7_0420500		14	PF3D7_0905100	NUP100	8	PF3D7_1338400		14
PF3D7_0420400	RRF2	4	PF3D7_0904900	CuTP	2	PF3D7_1338300		11
PF3D7_0420300	ApiAP2	3	PF3D7_0904800	RPA1	8	PF3D7_1338200		11
PF3D7_0420100	RIO2	11	PF3D7_0904700	HU	9	PF3D7_1338100	RPN3	10
PF3D7_0420000		11	PF3D7_0904600		8	PF3D7_1338000		9
PF3D7_0419900		2	PF3D7_0904300		0	PF3D7_1337800	CDPK5	3
PF3D7_0419800		13	PF3D7_0904200	PH	5	PF3D7_1337700		9
PF3D7_0419700	PF34	9	PF3D7_0904100		8	PF3D7_1337500		7
PF3D7_0419600	RANBP1	11	PF3D7_0904000		13	PF3D7_1337400		2
PF3D7_0419400		9	PF3D7_0903900	RPL32	11	PF3D7_1337300		13
PF3D7_0419300		9	PF3D7_0903800	CCp4	4	PF3D7_1336900	cTrpRS	6
PF3D7_0419000		0	PF3D7_0903700		1	PF3D7_1336800	NUDC	11
PF3D7_0418900		13	PF3D7_0903600.1		3	PF3D7_1336700		11
PF3D7_0418800		4	PF3D7_0903500		8	PF3D7_1336400		13
PF3D7_0418700		3	PF3D7_0903400	DDX60	6	PF3D7_1336200		1
PF3D7_0418600		3	PF3D7_0903300		2	PF3D7_1336000		10
PF3D7_0418300		2	PF3D7_0903200	RAB7	9	PF3D7_1335800		10
PF3D7_0418100		0	PF3D7_0903000		5	PF3D7_1335700		9
PF3D7_0418000		14	PF3D7_0902800	SERA9	6	PF3D7_1335100	MSP7	3
PF3D7_0417900		0	PF3D7_0901700		6	PF3D7_1334600	MSRP3	9
PF3D7_0417800	CRK1	8	PF3D7_1038400	Pf11-1	14	PF3D7_1334300	MSRP5	6
PF3D7_0417700		14	PF3D7_1038300		0	PF3D7_1334200		11
PF3D7_0417600		0	PF3D7_1038100		11	PF3D7_1334100		8
PF3D7_0417400		11	PF3D7_1038000.1		2	PF3D7_1334000		13
PF3D7_0417300		10	PF3D7_1038000.2		2	PF3D7_1333900		2
PF3D7_0417200	DHFR-TS	14	PF3D7_1037900		10	PF3D7_1333700	CenH3	4
PF3D7_0417000		5	PF3D7_1037600		5	PF3D7_1333600	UTP4	7
PF3D7_0416900		8	PF3D7_1037500	DYN2	5	PF3D7_1333500		4
PF3D7_0416800	SAR1	2	PF3D7_1037400		14	PF3D7_1333400		4
PF3D7_0416600		7	PF3D7_1037300	ADT	11	PF3D7_1333200	UBA1	2
PF3D7_0416500		7	PF3D7_1037000		2	PF3D7_1332900		11
PF3D7_0416400	HAT1	2	PF3D7_1036900		8	PF3D7_1332600	APN1	9
PF3D7_0416300	MCM9	10	PF3D7_1036800	ACT	9	PF3D7_1332400		7
PF3D7_0416200		3	PF3D7_1036500		3	PF3D7_1332200		3
PF3D7_0416100	GATA	1	PF3D7_1036000	MSP11	3	PF3D7_1331800		11
PF3D7_0416000		10	PF3D7_1035900	M566	3	PF3D7_1331700		1
PF3D7_0415900		6	PF3D7_1035700	DBLMSP	3	PF3D7_1331600	PTPLA	4
PF3D7_0415800		3	PF3D7_1035500	MSP6	3	PF3D7_1331400		4
PF3D7_0415700		3	PF3D7_1035400	MSP3	3	PF3D7_1331200		11
PF3D7_0415400		1	PF3D7_1035300	GLURP	3	PF3D7_1331100	POLQ	1
PF3D7_0415300	CRK3	14	PF3D7_1035200		9	PF3D7_1331000		3
PF3D7_0415200		0	PF3D7_1035100		9	PF3D7_1330800		5
PF3D7_0415100		4	PF3D7_1035000		14	PF3D7_1330700		1
PF3D7_0414700		3	PF3D7_1034900	MRScyt	6	PF3D7_1330500		13
PF3D7_0414600	APH	5	PF3D7_1034600	IF3b	9	PF3D7_1330300		4
PF3D7_0414500		0	PF3D7_1034500		9	PF3D7_1330100		9
PF3D7_0414300		6	PF3D7_1034200		4	PF3D7_1329900		5
PF3D7_0414100		14	PF3D7_1034000		1	PF3D7_1329700	ACBP1	13
PF3D7_0414000	SMC3	13	PF3D7_1033900		1	PF3D7_1329600		7
PF3D7_0413900	USP13	9	PF3D7_1033800		4	PF3D7_1329500		11
PF3D7_0413800		4	PF3D7_1033700	BDP1	4	PF3D7_1329300		8
PF3D7_0413700		2	PF3D7_1033600	Myb2	11	PF3D7_1329100	MyoC	10
PF3D7_0413600	RPT3	9	PF3D7_1033500	WDR70	5	PF3D7_1329000		11

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0413500	PGM2	10	PF3D7_1033200	ETRAMP10.2	2	PF3D7_1328900		0
PF3D7_0411900		8	PF3D7_1033100	AdoMetDC/OD	2	PF3D7_1328800	SIR2A	7
PF3D7_0411800		8	PF3D7_1033000		13	PF3D7_1328600		14
PF3D7_0411700		4	PF3D7_1032900		7	PF3D7_1328500		3
PF3D7_0411200		13	PF3D7_1032800	LRR1	13	PF3D7_1328300		3
PF3D7_0411000		5	PF3D7_1032700		7	PF3D7_1328200		2
PF3D7_0410900		14	PF3D7_1032300		4	PF3D7_1328100		10
PF3D7_0410800		8	PF3D7_1032200		4	PF3D7_1328000		0
PF3D7_0410700	RbgA	0	PF3D7_1032100	DCP1	7	PF3D7_1327700	UPF3B	14
PF3D7_0410300	PPM1	8	PF3D7_1032000	RimM	5	PF3D7_1327500		9
PF3D7_0410000	EVP1	3	PF3D7_1031800		9	PF3D7_1327400		9
PF3D7_0409800		10	PF3D7_1031600	GEXP15	11	PF3D7_1327300		7
PF3D7_0409700	RF2	5	PF3D7_1031500		13	PF3D7_1327100		4
PF3D7_0409600	RPA1	10	PF3D7_1031300		11	PF3D7_1327000		7
PF3D7_0409500		1	PF3D7_1031200	MORN1	4	PF3D7_1326900		13
PF3D7_0409400	DnaJ	7	PF3D7_1031100		5	PF3D7_1326700		5
PF3D7_0409300		11	PF3D7_1031000	Pfs25	5	PF3D7_1326600		8
PF3D7_0409200	SOF1	5	PF3D7_1030800		1	PF3D7_1326500		14
PF3D7_0409100	PRPF31	11	PF3D7_1030500	RPN9	2	PF3D7_1326100		10
PF3D7_0409000		2	PF3D7_1030300		1	PF3D7_1325900		8
PF3D7_0408500	FEN1	0	PF3D7_1030100	PRP22	1	PF3D7_1325800		14
PF3D7_0408300	ZRANB2	13	PF3D7_1029900		8	PF3D7_1325700		7
PF3D7_0408200		0	PF3D7_1029400		10	PF3D7_1325400		10
PF3D7_0408100		14	PF3D7_1029200		5	PF3D7_1325300		14
PF3D7_0408000		3	PF3D7_1028900	IMC1m	5	PF3D7_1325200		0
PF3D7_0407900		3	PF3D7_1028800		9	PF3D7_1325100		6
PF3D7_0407800		10	PF3D7_1028700	MTRAP	3	PF3D7_1324900	LDH	6
PF3D7_0407700		9	PF3D7_1028600		1	PF3D7_1324800	DHFS-FPGS	9
PF3D7_0407600		13	PF3D7_1028500		4	PF3D7_1324600		7
PF3D7_0407200	PTH2	5	PF3D7_1028300	EBP2	5	PF3D7_1324500		1
PF3D7_0406900		5	PF3D7_1028000		5	PF3D7_1324400		1
PF3D7_0406700		2	PF3D7_1027900	DHHC10	4	PF3D7_1324300		8
PF3D7_0406500		11	PF3D7_1027800	RPL3	6	PF3D7_1324200		10
PF3D7_0406200	Pfs16	7	PF3D7_1027600		4	PF3D7_1324000		0
PF3D7_0406100		10	PF3D7_1027300	nPrx	9	PF3D7_1323900		13
PF3D7_0406000		10	PF3D7_1027100	MPP10	7	PF3D7_1323800	VPS52	7
PF3D7_0405900	ASP	9	PF3D7_1027000		13	PF3D7_1323700	GAPM1	10
PF3D7_0405700		7	PF3D7_1026800	RPS2	11	PF3D7_1323600		4
PF3D7_0405400	PRPF8	11	PF3D7_1026600		3	PF3D7_1323400	RPL23	6
PF3D7_0405200	MAg-1	1	PF3D7_1026400	CDC20	0	PF3D7_1323300		1
PF3D7_0405000	DDX51	5	PF3D7_1026300		14	PF3D7_1323200		5
PF3D7_0404700	DPAP3	9	PF3D7_1026100		4	PF3D7_1323100		6
PF3D7_0404600		13	PF3D7_1026000		11	PF3D7_1322900		4
PF3D7_0404300		8	PF3D7_1025900		1	PF3D7_1322800		13
PF3D7_0404000		3	PF3D7_1025500		0	PF3D7_1322700		5
PF3D7_0403900	SET8	14	PF3D7_1025400		14	PF3D7_1322500	DHHC5	10
PF3D7_0403800		4	PF3D7_1025300		6	PF3D7_1322400		13
PF3D7_0403700	CLF1	8	PF3D7_1025100	GFPT	10	PF3D7_1322300		0
PF3D7_0403600		13	PF3D7_1025000		1	PF3D7_1322200		8
PF3D7_0403400		8	PF3D7_1024800		8	PF3D7_1322100	SET2	7
PF3D7_0403300		1	PF3D7_1024700		9	PF3D7_1322000		10
PF3D7_0403200		8	PF3D7_1024400		0	PF3D7_1321700	SF1	7
PF3D7_0402300	RH1	3	PF3D7_1024100		5	PF3D7_1321300		3
PF3D7_0402100		12	PF3D7_1024000		3	PF3D7_1321100		9
PF3D7_0402000		6	PF3D7_1023900	CHD1	8	PF3D7_1321000		4
PF3D7_0401600.1		6	PF3D7_1023800		3	PF3D7_1320800		4
PF3D7_0532300		12	PF3D7_1023600		0	PF3D7_1320700		1
PF3D7_0532100	ETRAMP5	6	PF3D7_1023500		0	PF3D7_1320600	RAB11a	3
PF3D7_0531100		13	PF3D7_1023300		5	PF3D7_1320200		5
PF3D7_0531000		8	PF3D7_1023100		0	PF3D7_1320100	ClpS	9
PF3D7_0530900		1	PF3D7_1023000		3	PF3D7_1320000	GP1	10
PF3D7_0530800		4	PF3D7_1022800	GcpE	8	PF3D7_1319800		7
PF3D7_0530700		14	PF3D7_1022600		13	PF3D7_1319700		8
PF3D7_0530300		1	PF3D7_1022400	SRSF4	6	PF3D7_1319600		13
PF3D7_0530200	PPT	2	PF3D7_1022000		11	PF3D7_1319400		8
PF3D7_0530000		6	PF3D7_1021900		2	PF3D7_1319300		11
PF3D7_0529900		1	PF3D7_1021800	SEA1	9	PF3D7_1319200		0
PF3D7_0529800		8	PF3D7_1021700		3	PF3D7_1319100		10
PF3D7_0529100	ATrx2	5	PF3D7_1021600		5	PF3D7_1319000		13
PF3D7_0529000		3	PF3D7_1021500	ROK1	5	PF3D7_1318800	SEC63	2

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0528900		5	PF3D7_1021200		14	PF3D7_1318700		5
PF3D7_0528800		7	PF3D7_1021000		9	PF3D7_1318500		14
PF3D7_0528600		5	PF3D7_1020900	ARF1	9	PF3D7_1318400	SMC2	8
PF3D7_0528500	CPalpha	13	PF3D7_1020800	DLAT	13	PF3D7_1318300		13
PF3D7_0528400	DHHC7	1	PF3D7_1020700		5	PF3D7_1318000		9
PF3D7_0528200	EIF3E	11	PF3D7_1020600		0	PF3D7_1317900		7
PF3D7_0528000	UMP1	5	PF3D7_1020400	RCM1	5	PF3D7_1317800	RPS19	6
PF3D7_0527600		5	PF3D7_1020300		5	PF3D7_1317400		4
PF3D7_0527500	HIP	11	PF3D7_1020200		5	PF3D7_1317300		6
PF3D7_0527200	USP14	2	PF3D7_1020000	RBM34	14	PF3D7_1317200	ApiAP2	7
PF3D7_0527000	MCM3	14	PF3D7_1019800		8	PF3D7_1317100	MCM4	8
PF3D7_0526700		10	PF3D7_1019700		1	PF3D7_1316900		0
PF3D7_0526600		2	PF3D7_1019600		1	PF3D7_1316800	SEC20	0
PF3D7_0526500		11	PF3D7_1019300		13	PF3D7_1316700		4
PF3D7_0526100		4	PF3D7_1019100		1	PF3D7_1316600	CCT	2
PF3D7_0526000		0	PF3D7_1019000		2	PF3D7_1316300		4
PF3D7_0525900	NEK2	5	PF3D7_1018900		4	PF3D7_1316100	IPK2	11
PF3D7_0525800	IMC1g	3	PF3D7_1018800		5	PF3D7_1316000		4
PF3D7_0525600		14	PF3D7_1018700		8	PF3D7_1315800	MYB1	0
PF3D7_0525500		7	PF3D7_1018600		14	PF3D7_1315400		11
PF3D7_0525400		5	PF3D7_1018400		4	PF3D7_1315300		1
PF3D7_0525300		4	PF3D7_1018300		11	PF3D7_1314700		7
PF3D7_0525200	SMC6	10	PF3D7_1018200	PPP8	1	PF3D7_1314200	TERT	8
PF3D7_0525100	ACS10	3	PF3D7_1018000		1	PF3D7_1313900		7
PF3D7_0525000		11	PF3D7_1017900	RPN5	10	PF3D7_1313800		8
PF3D7_0524900		7	PF3D7_1017800		5	PF3D7_1313600		1
PF3D7_0524800	UFD1	13	PF3D7_1017700		5	PF3D7_1313400		14
PF3D7_0524600		1	PF3D7_1017600		11	PF3D7_1313300		13
PF3D7_0524300		14	PF3D7_1017500		3	PF3D7_1313100		8
PF3D7_0524100		7	PF3D7_1017400	PMM	5	PF3D7_1312900	EIF4G	11
PF3D7_0524000	KAS $\beta$	11	PF3D7_1017100	RON12	9	PF3D7_1312800		1
PF3D7_0523900		1	PF3D7_1017000		0	PF3D7_1312700		4
PF3D7_0523700		5	PF3D7_1016900	ETRAMP10.3	7	PF3D7_1312600	BCKDHA	5
PF3D7_0523400		3	PF3D7_1016800		3	PF3D7_1312100		2
PF3D7_0523300		13	PF3D7_1016300	GBP	6	PF3D7_1311900	vapA	2
PF3D7_0523000	MDR1	4	PF3D7_1016000		5	PF3D7_1311800	M1AAP	6
PF3D7_0522900		13	PF3D7_1015900	ENO	6	PF3D7_1311600		1
PF3D7_0522600		9	PF3D7_1015800		9	PF3D7_1311500	RPT1	10
PF3D7_0522400		10	PF3D7_1015600	HSP60	11	PF3D7_1311400	AP1M1	2
PF3D7_0522200	TAF10	11	PF3D7_1015400		2	PF3D7_1311100		0
PF3D7_0522100		10	PF3D7_1014900		1	PF3D7_1310700		3
PF3D7_0522000		7	PF3D7_1014800		0	PF3D7_1310400		3
PF3D7_0521900		0	PF3D7_1014600	ADA2	2	PF3D7_1310100		7
PF3D7_0521400		11	PF3D7_1014300		14	PF3D7_1310000	OSCP	13
PF3D7_0521300		14	PF3D7_1014100		3	PF3D7_1309800		5
PF3D7_0521200		13	PF3D7_1013900		11	PF3D7_1309700	VPS18	7
PF3D7_0521000		7	PF3D7_1013800		9	PF3D7_1309500	GAR1	5
PF3D7_0520900	SAHH	6	PF3D7_1013600		2	PF3D7_1309400		5
PF3D7_0520800		6	PF3D7_1013500	PI-PLC	11	PF3D7_1309300	PRPF3	8
PF3D7_0520700		6	PF3D7_1013400		0	PF3D7_1309200	PPM6	1
PF3D7_0520300	LSM2	4	PF3D7_1013200		10	PF3D7_1309100		11
PF3D7_0520000		6	PF3D7_1013100	UTP13	4	PF3D7_1308900	DCP2	7
PF3D7_0519800		14	PF3D7_1013000		1	PF3D7_1308800	INT	9
PF3D7_0519700		11	PF3D7_1012700	NIF4	2	PF3D7_1308700		2
PF3D7_0519500	CCR4	7	PF3D7_1012400	HGPRT	6	PF3D7_1308500		5
PF3D7_0519400	RPS24	11	PF3D7_1012300	QCR7	1	PF3D7_1308400		2
PF3D7_0519200		5	PF3D7_1012200	RA	9	PF3D7_1308300	RPS27	5
PF3D7_0518800	PSOP13	4	PF3D7_1012100		1	PF3D7_1308200	cpsSII	6
PF3D7_0518700	PUF1	14	PF3D7_1012000		4	PF3D7_1308000		9
PF3D7_0518600	WDR26	14	PF3D7_1011900	HO	1	PF3D7_1307900		9
PF3D7_0518500	DDX23	8	PF3D7_1011800	PREBP	1	PF3D7_1307700		8
PF3D7_0518400	CYC3	0	PF3D7_1011700	RAD23	7	PF3D7_1307600		9
PF3D7_0518200	MDM2	5	PF3D7_1011500		5	PF3D7_1307400		0
PF3D7_0518100		5	PF3D7_1011300	ARV1	13	PF3D7_1307300	DBP6	13
PF3D7_0518000		13	PF3D7_1010900		13	PF3D7_1307200		4
PF3D7_0517900		1	PF3D7_1010700		13	PF3D7_1307100	UTP6	11
PF3D7_0517800		4	PF3D7_1010500		0	PF3D7_1307000	RRP40	6
PF3D7_0517700	EIF3B	11	PF3D7_1010400		0	PF3D7_1306800		7
PF3D7_0517600	CP $\beta$	5	PF3D7_1010300	SDH4	2	PF3D7_1306500		7
PF3D7_0517500		5	PF3D7_1010200		6	PF3D7_1306400	RPT4	9
PF3D7_0517400	FACT-L	8	PF3D7_1010100		14	PF3D7_1306000		8

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0517300	SR1	11	PF3D7_1010000		13	PF3D7_1305900		7
PF3D7_0517200		4	PF3D7_1009900		13	PF3D7_1305400		4
PF3D7_0517000		6	PF3D7_1009800		7	PF3D7_1305300		8
PF3D7_0516900	RPL2	6	PF3D7_1009700		1	PF3D7_1305200	ApiAP2	5
PF3D7_0516800	ApiAP2	10	PF3D7_1009600		4	PF3D7_1305100		1
PF3D7_0516700		10	PF3D7_1009500		13	PF3D7_1305000		10
PF3D7_0516600	MB2	2	PF3D7_1009300		1	PF3D7_1304600		7
PF3D7_0516400	PhLP3	4	PF3D7_1009100		4	PF3D7_1304500		7
PF3D7_0516100	ATPase1	7	PF3D7_1008800	NOP5	5	PF3D7_1304400		8
PF3D7_0516000		11	PF3D7_1008700		7	PF3D7_1304100	LigI	0
PF3D7_0515900	NIF2	11	PF3D7_1008500		10	PF3D7_1304000		8
PF3D7_0515700	GAP40	9	PF3D7_1008400	RPT2	2	PF3D7_1303800		8
PF3D7_0515600		13	PF3D7_1008200		10	PF3D7_1303700		0
PF3D7_0515500		13	PF3D7_1008100		11	PF3D7_1303500	NHE	9
PF3D7_0515400		5	PF3D7_1008000	HDA2	2	PF3D7_1303400	LisH	11
PF3D7_0515300	PI3K	2	PF3D7_1007900	EIF3D	11	PF3D7_1303300		7
PF3D7_0515200		5	PF3D7_1007800		13	PF3D7_1303000		5
PF3D7_0515100	ROM9	5	PF3D7_1007700	ApiAP2	8	PF3D7_1302800		6
PF3D7_0515000	CWC2	11	PF3D7_1007400		10	PF3D7_1302700		11
PF3D7_0514900		10	PF3D7_1007200		14	PF3D7_1302500		0
PF3D7_0514700		0	PF3D7_1006900		4	PF3D7_1302400		4
PF3D7_0514500		5	PF3D7_1006800	GBP2	11	PF3D7_1302300		6
PF3D7_0514300		10	PF3D7_1006700		13	PF3D7_1302100	G27/25	2
PF3D7_0514200		7	PF3D7_1006600	PhLP1	5	PF3D7_1302000	PTP6	12
PF3D7_0514100	UvrD	13	PF3D7_1006500		5	PF3D7_1301700	CBP2	12
PF3D7_0514000	TTL	5	PF3D7_1006400		13	PF3D7_1301600	EBA140	3
PF3D7_0513800	RAB1a	4	PF3D7_1006300		4	PF3D7_1478600	PTP3	12
PF3D7_0513700	PSOP12	5	PF3D7_1006200	ALBA3	11	PF3D7_1478000	GEXP17	0
PF3D7_0513600		7	PF3D7_1006000		5	PF3D7_1477700	Pfg14-748	0
PF3D7_0513300	PNP	11	PF3D7_1005900		1	PF3D7_1476700		4
PF3D7_0513200		8	PF3D7_1005600		13	PF3D7_1476600		14
PF3D7_0513100		4	PF3D7_1005500	UPF1	13	PF3D7_1476500		7
PF3D7_0513000		5	PF3D7_1005200		5	PF3D7_1476300		10
PF3D7_0512900	AKAL	2	PF3D7_1005100	UTP25	13	PF3D7_1476200		12
PF3D7_0512800		13	PF3D7_1004900		13	PF3D7_1475500	CCp1	5
PF3D7_0512600	RAB1b	5	PF3D7_1004600		10	PF3D7_1475400	CRMP4	7
PF3D7_0512500		7	PF3D7_1004500		2	PF3D7_1475200		8
PF3D7_0512100		4	PF3D7_1004400		8	PF3D7_1475100		1
PF3D7_0512000		5	PF3D7_1004300		1	PF3D7_1474700		14
PF3D7_0511800	INO1	10	PF3D7_1004200		9	PF3D7_1474600		11
PF3D7_0511600	ARNP	9	PF3D7_1004000		6	PF3D7_1474500	SF3A1	8
PF3D7_0511500		11	PF3D7_1003800		11	PF3D7_1474400		14
PF3D7_0511300		13	PF3D7_1003700		7	PF3D7_1474300		14
PF3D7_0511200	SCD	10	PF3D7_1003600	IMC1c	3	PF3D7_1474200		0
PF3D7_0511000	TCTP	4	PF3D7_1003400		13	PF3D7_1474000		7
PF3D7_0510800		13	PF3D7_1003000		10	PF3D7_1473900		14
PF3D7_0510700		4	PF3D7_1002800		13	PF3D7_1473700	NUP116	0
PF3D7_0510500	TopoI	8	PF3D7_1002700		1	PF3D7_1473500		7
PF3D7_0510200	CYP87	7	PF3D7_1002200	PArT	6	PF3D7_1473400		3
PF3D7_0510100		2	PF3D7_1002100	PTP5	12	PF3D7_1473200		11
PF3D7_0510000		14	PF3D7_1001700		9	PF3D7_1473100		7
PF3D7_0509800	PI4K	3	PF3D7_1001600		9	PF3D7_1472900		7
PF3D7_0509700		4	PF3D7_1001500	ETRAPM10	12	PF3D7_1472800		13
PF3D7_0509600	AsnRS	4	PF3D7_1001200	ACBP2	6	PF3D7_1472700		5
PF3D7_0509500		4	PF3D7_1149200		12	PF3D7_1472600	PDI-14	3
PF3D7_0509400	RNAPI	11	PF3D7_1149100.1		6	PF3D7_1472500		14
PF3D7_0509200	LRR2	14	PF3D7_1149000	Pf332	6	PF3D7_1472400		6
PF3D7_0509100	SMC4	8	PF3D7_1148700	GEXP12	12	PF3D7_1472200	HDA1	8
PF3D7_0508900		10	PF3D7_1148000		8	PF3D7_1472000	ISY1	14
PF3D7_0508700	PRP5	7	PF3D7_1147600		8	PF3D7_1471900		0
PF3D7_0508600		1	PF3D7_1147300		2	PF3D7_1471800		4
PF3D7_0508500	RCC1	9	PF3D7_1147200		13	PF3D7_1471700		5
PF3D7_0508400		0	PF3D7_1147100		13	PF3D7_1471600		5
PF3D7_0508100	SET9	9	PF3D7_1146800		13	PF3D7_1471500		5
PF3D7_0507800		2	PF3D7_1146700		0	PF3D7_1471400		10
PF3D7_0507700	NPL4	1	PF3D7_1146600	NFYB	3	PF3D7_1471300		6
PF3D7_0507600	CAF40	10	PF3D7_1146300		5	PF3D7_1471100	EXP2	6
PF3D7_0507500	SUB1	3	PF3D7_1146100		4	PF3D7_1470800		9
PF3D7_0507200	SUB3	9	PF3D7_1145800		7	PF3D7_1470700		7
PF3D7_0507100	RPL4	6	PF3D7_1145600		14	PF3D7_1470600		7

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0506900	ROM4	1	PF3D7_1145400	DYN1	1	PF3D7_1470500	PDEdelta	5
PF3D7_0506800	TCF25	9	PF3D7_1145200		1	PF3D7_1470300		13
PF3D7_0506700		10	PF3D7_1144800		13	PF3D7_1470200		13
PF3D7_0506500		14	PF3D7_1144600	TFG2	4	PF3D7_1470100		0
PF3D7_0506400		4	PF3D7_1144400		0	PF3D7_1470000	ATG12	6
PF3D7_0506300		14	PF3D7_1144200		0	PF3D7_1469900		14
PF3D7_0506100		4	PF3D7_1144100		13	PF3D7_1469800		13
PF3D7_0506000		6	PF3D7_1143700		14	PF3D7_1469700	MED6	1
PF3D7_0505700		10	PF3D7_1143600		7	PF3D7_1469600	ACC	7
PF3D7_0505600		5	PF3D7_1143500		14	PF3D7_1469300	PN01	13
PF3D7_0505500	MSH6	8	PF3D7_1143400		11	PF3D7_1469100		5
PF3D7_0505400		8	PF3D7_1143300	RPC40	5	PF3D7_1469000	IF1	10
PF3D7_0505100	TRS85	2	PF3D7_1143200		2	PF3D7_1468900		8
PF3D7_0505000		8	PF3D7_1143100	AP2-O	3	PF3D7_1468800	U2AF2	0
PF3D7_0504900		1	PF3D7_1142900		1	PF3D7_1468700	eIF4A	6
PF3D7_0504800		2	PF3D7_1142800		7	PF3D7_1468600		0
PF3D7_0504700	CEP120	7	PF3D7_1142500	RPL28	6	PF3D7_1468400	D13	3
PF3D7_0504500		4	PF3D7_1142400	CPO	6	PF3D7_1468200		9
PF3D7_0504400		11	PF3D7_1142300		10	PF3D7_1468100		2
PF3D7_0504200	DDX27	1	PF3D7_1142100		8	PF3D7_1468000		0
PF3D7_0504000	ATPase3	6	PF3D7_1141900	IMC1b	5	PF3D7_1467900		10
PF3D7_0503800	RPL31	6	PF3D7_1141800		1	PF3D7_1467600		7
PF3D7_0503600	MyoB	3	PF3D7_1141600	DPM1	6	PF3D7_1467400	RPL22	10
PF3D7_0503400	ADF1	9	PF3D7_1141500		0	PF3D7_1467000		5
PF3D7_0503300	SRSF12	11	PF3D7_1141400	PIGH	9	PF3D7_1466900		11
PF3D7_0503200		13	PF3D7_1141300		10	PF3D7_1466800		13
PF3D7_0503100	IspE	6	PF3D7_1141100		5	PF3D7_1466600		0
PF3D7_0503000		5	PF3D7_1140900		11	PF3D7_1466500		4
PF3D7_0502600		0	PF3D7_1140800		14	PF3D7_1466300	RPN2	2
PF3D7_0502400	MSP8	12	PF3D7_1140700		0	PF3D7_1466200		7
PF3D7_0502300		4	PF3D7_1140600		4	PF3D7_1466000		4
PF3D7_0502200	MPV17	10	PF3D7_1140500	MyoF	14	PF3D7_1465900		6
PF3D7_0502000	VPS11	2	PF3D7_1140200		8	PF3D7_1465800		0
PF3D7_0501800	CAF1	8	PF3D7_1139800		2	PF3D7_1465700		4
PF3D7_0501600	RAP2	9	PF3D7_1139700		3	PF3D7_1465600		0
PF3D7_0501500	RAP3	9	PF3D7_1139600		0	PF3D7_1465100	COG6	9
PF3D7_0501400	FIRA	12	PF3D7_1139300	ApiAP2	8	PF3D7_1464600	UIS2	3
PF3D7_0501300	SBP1	12	PF3D7_1139100		6	PF3D7_1464500		2
PF3D7_0501200	PIESP2	6	PF3D7_1138800		8	PF3D7_1464000		10
PF3D7_0501000		6	PF3D7_1138700		3	PF3D7_1463900		9
PF3D7_0500800	MESA	6	PF3D7_1138600		8	PF3D7_1463700	CYC1	13
PF3D7_0630900	HAS1	11	PF3D7_1138500	PPM2	10	PF3D7_1463000		5
PF3D7_0630800		0	PF3D7_1138400	GCalpha	2	PF3D7_1462800	GAPDH	6
PF3D7_0630600		10	PF3D7_1138300		4	PF3D7_1462700		10
PF3D7_0630400		5	PF3D7_1138000		10	PF3D7_1462500		0
PF3D7_0630300		8	PF3D7_1137700		0	PF3D7_1462400		7
PF3D7_0630200	PSOP6	4	PF3D7_1137600		0	PF3D7_1462300		11
PF3D7_0630100		3	PF3D7_1137500		5	PF3D7_1462200		9
PF3D7_0630000		4	PF3D7_1137300		11	PF3D7_1462100		2
PF3D7_0629800		8	PF3D7_1137200	AEP	1	PF3D7_1461900		6
PF3D7_0629700	SET1	2	PF3D7_1136900	SUB2	3	PF3D7_1461800		7
PF3D7_0629500		3	PF3D7_1136600		11	PF3D7_1461700		5
PF3D7_0629200		2	PF3D7_1136400	SRP72	11	PF3D7_1461600	SF3B2	11
PF3D7_0628900		7	PF3D7_1136300	TSN	2	PF3D7_1461400		5
PF3D7_0628700		11	PF3D7_1136200		9	PF3D7_1461300		11
PF3D7_0628600		10	PF3D7_1136000		8	PF3D7_1461100		14
PF3D7_0628300.1		3	PF3D7_1135900		9	PF3D7_1461000		14
PF3D7_0628200	PK4	8	PF3D7_1135600		0	PF3D7_1460800		14
PF3D7_0628100		1	PF3D7_1135500		0	PF3D7_1460700	RPL27	11
PF3D7_0627800	ACS	2	PF3D7_1134800		1	PF3D7_1460600	ISP3	3
PF3D7_0627700		11	PF3D7_1134700	RPA2	7	PF3D7_1460500		11
PF3D7_0627300	RNF5	10	PF3D7_1134600		7	PF3D7_1460400	UCHL3	13
PF3D7_0627100		0	PF3D7_1134500		2	PF3D7_1460300		6
PF3D7_0626800	PyrK	6	PF3D7_1134300		10	PF3D7_1460100	FCP	5
PF3D7_0626500		0	PF3D7_1134200		11	PF3D7_1459800		5
PF3D7_0626400		5	PF3D7_1134100	PDI-11	9	PF3D7_1459600		8
PF3D7_0626000		4	PF3D7_1134000	HSP70	6	PF3D7_1459200		2
PF3D7_0625400		3	PF3D7_1133900		2	PF3D7_1459000	DBP5	11
PF3D7_0625300		2	PF3D7_1133800		4	PF3D7_1458600		8
PF3D7_0625200		1	PF3D7_1133700		5	PF3D7_1458500	SAS4	8
PF3D7_0625100	SMS2	4	PF3D7_1133600		13	PF3D7_1458300		10

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0624900		0	PF3D7_1133400	AMA1	3	PF3D7_1458000	FP1	11
PF3D7_0624600	ISWI	8	PF3D7_1133300		9	PF3D7_1457900		2
PF3D7_0624400		8	PF3D7_1133200		2	PF3D7_1457800		0
PF3D7_0624300		4	PF3D7_1133100		11	PF3D7_1457700		11
PF3D7_0624200		5	PF3D7_1132900	GCVH	6	PF3D7_1457600		1
PF3D7_0624100		5	PF3D7_1132600	PRP38A	2	PF3D7_1457400		8
PF3D7_0623900		0	PF3D7_1132400		10	PF3D7_1457300		6
PF3D7_0623800	TKL4	12	PF3D7_1132300		14	PF3D7_1457200	TRX1	9
PF3D7_0623700	SUV3	13	PF3D7_1132200	TCP1	8	PF3D7_1457100		5
PF3D7_0623100	NAB2	11	PF3D7_1132000		10	PF3D7_1457000	SPP	1
PF3D7_0623000	CS	6	PF3D7_1131800		4	PF3D7_1456800	VP1	6
PF3D7_0622900	ApiAP2	11	PF3D7_1131600		0	PF3D7_1456700		7
PF3D7_0622800		8	PF3D7_1131500		4	PF3D7_1456500		8
PF3D7_0622300		7	PF3D7_1131400		0	PF3D7_1456400		0
PF3D7_0622200		11	PF3D7_1131200		14	PF3D7_1456000	ApiAP2	11
PF3D7_0622100		13	PF3D7_1131100	SR1	10	PF3D7_1455900		6
PF3D7_0621900	SRP68	2	PF3D7_1130900		0	PF3D7_1455800	CCp2	5
PF3D7_0621700		14	PF3D7_1130700	SMC1	7	PF3D7_1455700		11
PF3D7_0621500	RPP1	8	PF3D7_1130500	COG2	8	PF3D7_1455500		7
PF3D7_0621400	ALV7	5	PF3D7_1130400	RPT5	1	PF3D7_1455300		3
PF3D7_0621300		11	PF3D7_1130200	PfP0	6	PF3D7_1455100	PTP1	3
PF3D7_0621200	PDX1	11	PF3D7_1130100	RPL38	6	PF3D7_1455000	PPM3	7
PF3D7_0621100		1	PF3D7_1130000	PAGM	11	PF3D7_1454900		5
PF3D7_0621000		8	PF3D7_1129900	MFR5	5	PF3D7_1454400	APP	6
PF3D7_0620700		1	PF3D7_1129800		7	PF3D7_1454300	KIN	7
PF3D7_0620600		0	PF3D7_1129600		5	PF3D7_1454200		7
PF3D7_0620500		10	PF3D7_1129500		4	PF3D7_1454000		1
PF3D7_0620400	MSP10	3	PF3D7_1129300		9	PF3D7_1453700	P23	6
PF3D7_0619800		6	PF3D7_1129200	RPN7	2	PF3D7_1453600		11
PF3D7_0619700		1	PF3D7_1129100	PV1	11	PF3D7_1453500		4
PF3D7_0619500	ACS12	6	PF3D7_1129000	SpdSyn	6	PF3D7_1453400		10
PF3D7_0619400		2	PF3D7_1128900		1	PF3D7_1453200		9
PF3D7_0619300		0	PF3D7_1128600	NOT2	13	PF3D7_1453100		5
PF3D7_0619200		2	PF3D7_1128500		1	PF3D7_1453000		2
PF3D7_0618800		14	PF3D7_1128300	PFK11	5	PF3D7_1452900		5
PF3D7_0618700		4	PF3D7_1128200	MBF1	5	PF3D7_1452700		8
PF3D7_0618500	MDH	4	PF3D7_1128100		13	PF3D7_1452600		8
PF3D7_0618300		6	PF3D7_1127900		5	PF3D7_1452500	BET1	1
PF3D7_0618100		14	PF3D7_1127800		2	PF3D7_1452400		13
PF3D7_0618000		9	PF3D7_1127600		4	PF3D7_1452000	RON2	9
PF3D7_0617900	H3.3	10	PF3D7_1127500		13	PF3D7_1451800		4
PF3D7_0617800	H2A	10	PF3D7_1127100	dUTPase	6	PF3D7_1451600	LAP5	5
PF3D7_0617200		11	PF3D7_1126800		14	PF3D7_1451300	NSE2	1
PF3D7_0617100		8	PF3D7_1126700	ATG23	3	PF3D7_1451200		11
PF3D7_0616900		11	PF3D7_1126600		13	PF3D7_1451100	eEF2	6
PF3D7_0616600		4	PF3D7_1126500		7	PF3D7_1451000		7
PF3D7_0616500	TLP	5	PF3D7_1126400		5	PF3D7_1450700		4
PF3D7_0616400		11	PF3D7_1126300.1		10	PF3D7_1450500		10
PF3D7_0616300		13	PF3D7_1126200		6	PF3D7_1450400		11
PF3D7_0616200	NDC80	10	PF3D7_1126000	ThrRS	11	PF3D7_1450000		3
PF3D7_0615900		14	PF3D7_1125900		3	PF3D7_1449900		5
PF3D7_0615800		2	PF3D7_1125800		3	PF3D7_1449700	RRP6	11
PF3D7_0615600		0	PF3D7_1125700		3	PF3D7_1449500	ApiAP2	8
PF3D7_0615500	CRK5	8	PF3D7_1125400	TIM44	8	PF3D7_1449400	MUS81	6
PF3D7_0615400		8	PF3D7_1125300	mtRNAP	13	PF3D7_1449200		9
PF3D7_0615300	GWT1	13	PF3D7_1125200		13	PF3D7_1449100		0
PF3D7_0615200		4	PF3D7_1125100		14	PF3D7_1449000	GEST	13
PF3D7_0615000		4	PF3D7_1125000		5	PF3D7_1448900		4
PF3D7_0614900		1	PF3D7_1124900		1	PF3D7_1448600	VTI1	5
PF3D7_0614800		1	PF3D7_1124800		13	PF3D7_1448500		8
PF3D7_0614700		14	PF3D7_1124700	MGE1	5	PF3D7_1448400	HRD3	1
PF3D7_0614600		14	PF3D7_1124600	EK	2	PF3D7_1448300		2
PF3D7_0614500	RPL19	11	PF3D7_1124500	pdhA	5	PF3D7_1448200		0
PF3D7_0614400	CWF7	13	PF3D7_1124400	LSM1	4	PF3D7_1447800		7
PF3D7_0614300	MFR1	2	PF3D7_1124300		11	PF3D7_1447600		4
PF3D7_0614200	NAR1	10	PF3D7_1124200		0	PF3D7_1447400		5
PF3D7_0614100		1	PF3D7_1124100		10	PF3D7_1447200		5
PF3D7_0614000		3	PF3D7_1124000	ERO1	13	PF3D7_1447000		11
PF3D7_0613900	myoE	3	PF3D7_1123800	SMC5	2	PF3D7_1446900		5
PF3D7_0613800	ApiAP2	9	PF3D7_1123600		13	PF3D7_1446700		5
PF3D7_0613700		10	PF3D7_1123500		2	PF3D7_1446600	CEN2	4

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0613600		5	PF3D7_1123400		6	PF3D7_1446500		8
PF3D7_0613500		13	PF3D7_1123300		8	PF3D7_1446300		14
PF3D7_0613400		9	PF3D7_1123200	LRR11	8	PF3D7_1446200	LAP	6
PF3D7_0613300	ROP14	9	PF3D7_1123100	CDPK7	10	PF3D7_1446100		5
PF3D7_0612900		5	PF3D7_1123000		8	PF3D7_1445900	DDX5	7
PF3D7_0612800	P12p	13	PF3D7_1122900		0	PF3D7_1445700		13
PF3D7_0612700	P12	3	PF3D7_1122800	CDPK6	3	PF3D7_1445600		8
PF3D7_0612500		4	PF3D7_1122600		0	PF3D7_1445400	CLK1	13
PF3D7_0612200	LRR6	10	PF3D7_1122500		8	PF3D7_1445200	MAK5	11
PF3D7_0611900		5	PF3D7_1122400		7	PF3D7_1445100		6
PF3D7_0611800		13	PF3D7_1122300		5	PF3D7_1445000		10
PF3D7_0611600		5	PF3D7_1122200		14	PF3D7_1444900		4
PF3D7_0611200	ApiAP2	6	PF3D7_1122100	GPI16	13	PF3D7_1444800	FBPA	12
PF3D7_0611000		4	PF3D7_1121900		1	PF3D7_1444500	IK1	11
PF3D7_0610900	SPT5	14	PF3D7_1121800		5	PF3D7_1444300	LPAAT	4
PF3D7_0610800	TK	11	PF3D7_1121700	GCN20	6	PF3D7_1444200		0
PF3D7_0610400	H3	10	PF3D7_1121600	EXP1	6	PF3D7_1444100		0
PF3D7_0610200	RBM25	7	PF3D7_1121400		7	PF3D7_1443900	HSP90	6
PF3D7_0610100	SLU7	5	PF3D7_1121300	TKL2	9	PF3D7_1443600		8
PF3D7_0609900		8	PF3D7_1121100		10	PF3D7_1443500	PARN	4
PF3D7_0609700		1	PF3D7_1121000	DHHC3	5	PF3D7_1443400	WLP1	2
PF3D7_0609600		5	PF3D7_1120700		5	PF3D7_1443100		10
PF3D7_0609400	CLS	13	PF3D7_1120600		0	PF3D7_1443000	SRPK2	13
PF3D7_0609300		5	PF3D7_1120500		7	PF3D7_1442900	SEC7	2
PF3D7_0609100	ZIP1	4	PF3D7_1120100	PGM1	6	PF3D7_1442700		10
PF3D7_0609000		8	PF3D7_1120000		11	PF3D7_1442400		10
PF3D7_0608900		2	PF3D7_1119900		10	PF3D7_1442300	tRIP	6
PF3D7_0608800	OAT	6	PF3D7_1119500		10	PF3D7_1442200		13
PF3D7_0608700	CCT6	11	PF3D7_1119400		8	PF3D7_1442100	RPA3	8
PF3D7_0608600		7	PF3D7_1119300	U2AF1	6	PF3D7_1441800		5
PF3D7_0608300		13	PF3D7_1119200		5	PF3D7_1441700	ATP23	4
PF3D7_0608100		6	PF3D7_1119100		5	PF3D7_1441500		8
PF3D7_0608000	DPH7	4	PF3D7_1119000		10	PF3D7_1441300		4
PF3D7_0607900		0	PF3D7_1118900		5	PF3D7_1441200		6
PF3D7_0607700		1	PF3D7_1118800	ARC40	13	PF3D7_1441100		11
PF3D7_0607400		8	PF3D7_1118700	MLC-B	4	PF3D7_1440900		5
PF3D7_0607300	UROD	14	PF3D7_1118400		1	PF3D7_1440600		0
PF3D7_0607200		13	PF3D7_1118300		11	PF3D7_1440400		10
PF3D7_0607100		0	PF3D7_1118200		5	PF3D7_1440200	SPP	6
PF3D7_0607000		6	PF3D7_1117900		13	PF3D7_1440100		9
PF3D7_0606900	GLP2	5	PF3D7_1117800	MLH	8	PF3D7_1440000		8
PF3D7_0606800		9	PF3D7_1117700	RAN	6	PF3D7_1439900	TIM	6
PF3D7_0606700		2	PF3D7_1117500	TyrRSapi	6	PF3D7_1439800		2
PF3D7_0606600		8	PF3D7_1117400		1	PF3D7_1439600		5
PF3D7_0606400		4	PF3D7_1117300		10	PF3D7_1439300		8
PF3D7_0606100		14	PF3D7_1117200		0	PF3D7_1439200		0
PF3D7_0606000		8	PF3D7_1117100	UCH54	1	PF3D7_1439100		8
PF3D7_0605800	RAD50	2	PF3D7_1116900		5	PF3D7_1438900	Trx-Px1	11
PF3D7_0605600		0	PF3D7_1116800	HSP101	12	PF3D7_1438800		13
PF3D7_0605100		0	PF3D7_1116700	DPAP1	6	PF3D7_1438700		0
PF3D7_0604800		14	PF3D7_1116400	SEC12	5	PF3D7_1438500	CPSF3	6
PF3D7_0604700	GILP	6	PF3D7_1116100		3	PF3D7_1438400	MCA2	1
PF3D7_0604500		8	PF3D7_1116000	RON4	9	PF3D7_1438000		11
PF3D7_0604400		4	PF3D7_1115900	DHHC9	5	PF3D7_1437900	ERdj3	11
PF3D7_0604300		2	PF3D7_1115800		12	PF3D7_1437400	PANK2	6
PF3D7_0604100	SIP2	9	PF3D7_1115700	FP2A	6	PF3D7_1437300		9
PF3D7_0603900		0	PF3D7_1115600	CYP19B	9	PF3D7_1437200		0
PF3D7_0603800	CEP76	9	PF3D7_1115400	FP3	10	PF3D7_1437000		10
PF3D7_0603600		14	PF3D7_1115200	SET7	5	PF3D7_1436600	PKG	3
PF3D7_0603400	TEX1	8	PF3D7_1115100		4	PF3D7_1436300	PTEX150	12
PF3D7_0603100		13	PF3D7_1114900		1	PF3D7_1436200		9
PF3D7_0603000		7	PF3D7_1114700	CLK3	0	PF3D7_1436100		8
PF3D7_0602900		1	PF3D7_1113900	MAPK2	0	PF3D7_1435600		1
PF3D7_0602700		4	PF3D7_1113800		2	PF3D7_1435500		1
PF3D7_0602600		8	PF3D7_1113400	DSK2	3	PF3D7_1435300		10
PF3D7_0602400	EF-G	6	PF3D7_1113300	UGT	10	PF3D7_1435200		5
PF3D7_0602100		7	PF3D7_1113100	PRL	11	PF3D7_1435100		0
PF3D7_0602000		14	PF3D7_1113000		9	PF3D7_1434600	METAP2	11
PF3D7_0731800	GEXP08	14	PF3D7_1112900		0	PF3D7_1434500		8
PF3D7_0731600	ACS5	6	PF3D7_1112700		0	PF3D7_1434400		1
PF3D7_0731500	EBA175	3	PF3D7_1112600		0	PF3D7_1434300	HOP	6

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0731100	PTP2	12	PF3D7_1112400		5	PF3D7_1433900		8
PF3D7_0730900	PTP4	12	PF3D7_1112300		10	PF3D7_1433800		8
PF3D7_0730500		10	PF3D7_1112200	COQ4	5	PF3D7_1433700		8
PF3D7_0730400		3	PF3D7_1112100		9	PF3D7_1433500	TOP2	0
PF3D7_0730300	AP2-L	3	PF3D7_1112000		13	PF3D7_1433400		2
PF3D7_0729900		14	PF3D7_1111800		5	PF3D7_1433200		0
PF3D7_0729700		0	PF3D7_1111200		6	PF3D7_1433100		1
PF3D7_0729500		7	PF3D7_1110900		7	PF3D7_1433000	NOP10	11
PF3D7_0729300	NMD3	11	PF3D7_1110800		14	PF3D7_1432800		10
PF3D7_0729100		2	PF3D7_1110600		10	PF3D7_1432700		8
PF3D7_0728900		14	PF3D7_1110500	VPS35	1	PF3D7_1432600		8
PF3D7_0728800		0	PF3D7_1110400		11	PF3D7_1432500		14
PF3D7_0728700		5	PF3D7_1110300		5	PF3D7_1432400	LRR5	0
PF3D7_0728600		8	PF3D7_1110200	PRPF6	2	PF3D7_1432300		13
PF3D7_0728500		14	PF3D7_1110100		1	PF3D7_1432200		9
PF3D7_0728300		5	PF3D7_1109900	RPL36	11	PF3D7_1432000	SYN11	2
PF3D7_0728100		8	PF3D7_1109400	ENP1	13	PF3D7_1431700		6
PF3D7_0728000		11	PF3D7_1109100		5	PF3D7_1431500	MAPK1	5
PF3D7_0727900		2	PF3D7_1108700		8	PF3D7_1431400		9
PF3D7_0727800		11	PF3D7_1108600	ERC	9	PF3D7_1431300	LSG1	5
PF3D7_0727500		6	PF3D7_1108500		1	PF3D7_1431200		7
PF3D7_0727300	DNMT	10	PF3D7_1108400	CK2alpha	11	PF3D7_1431100		7
PF3D7_0727200	NFS	4	PF3D7_1108200	DTD	14	PF3D7_1430800		13
PF3D7_0727000	VPS53	1	PF3D7_1108100		14	PF3D7_1430600		10
PF3D7_0726700		5	PF3D7_1108000		7	PF3D7_1430400	ATG5	4
PF3D7_0726500		8	PF3D7_1107900	MSCS	13	PF3D7_1430300		8
PF3D7_0726400		8	PF3D7_1107800	ApiAP2	1	PF3D7_1430200		5
PF3D7_0726300	PMS1	0	PF3D7_1107700	PES	8	PF3D7_1430100	PTPA	4
PF3D7_0725400		3	PF3D7_1107600		0	PF3D7_1430000		5
PF3D7_0725100		2	PF3D7_1107300	PAIP1	8	PF3D7_1429900	WRN	5
PF3D7_0725000		1	PF3D7_1107200		10	PF3D7_1429800		2
PF3D7_0724900		5	PF3D7_1107100		1	PF3D7_1429700		4
PF3D7_0724800		3	PF3D7_1106900		4	PF3D7_1429600	GEXP03	4
PF3D7_0724700		8	PF3D7_1106800		14	PF3D7_1429400		11
PF3D7_0724600		0	PF3D7_1106700	DNA2	8	PF3D7_1429300		5
PF3D7_0724200	TAP42	6	PF3D7_1106500		0	PF3D7_1429200	ApiAP2	14
PF3D7_0724100		10	PF3D7_1106300		6	PF3D7_1429100		2
PF3D7_0724000		2	PF3D7_1106200		8	PF3D7_1428900		2
PF3D7_0723900		8	PF3D7_1106100		1	PF3D7_1428500		14
PF3D7_0723800		10	PF3D7_1105800		10	PF3D7_1428400	WDTC1	7
PF3D7_0723700		10	PF3D7_1105600	PTEX88	12	PF3D7_1428300		11
PF3D7_0723400		8	PF3D7_1105500	CEN4	5	PF3D7_1428200	MFS5	3
PF3D7_0723300		5	PF3D7_1105400		6	PF3D7_1428100		13
PF3D7_0723100		4	PF3D7_1105100	H2B	10	PF3D7_1427900		11
PF3D7_0722900		4	PF3D7_1105000	H4	9	PF3D7_1427300		5
PF3D7_0722800		5	PF3D7_1104900		3	PF3D7_1427200	Sel4	10
PF3D7_0722600	UTP7	5	PF3D7_1104600		0	PF3D7_1427100		4
PF3D7_0722500	CWC15	5	PF3D7_1104500		0	PF3D7_1427000		5
PF3D7_0722300		7	PF3D7_1104400		2	PF3D7_1426900	QCR6	1
PF3D7_0722200	RALP1	9	PF3D7_1104300		8	PF3D7_1426800		0
PF3D7_0722000		4	PF3D7_1104200	SNF2L	2	PF3D7_1426700	PEPC	13
PF3D7_0721600		6	PF3D7_1104100	SYN13	12	PF3D7_1426600		10
PF3D7_0721500		13	PF3D7_1104000		6	PF3D7_1426500	ABCG2	5
PF3D7_0721300	DBP7	14	PF3D7_1103800	NOT1	14	PF3D7_1426400		0
PF3D7_0721100		2	PF3D7_1103500		5	PF3D7_1426300		0
PF3D7_0721000		8	PF3D7_1103400	SufD	2	PF3D7_1426200	PRMT1	6
PF3D7_0720900		6	PF3D7_1103100	RPP1	6	PF3D7_1426000	RPL21	6
PF3D7_0720800		13	PF3D7_1102800	ETRAMP11.2	12	PF3D7_1425800		14
PF3D7_0720700		10	PF3D7_1102700	ETRAMP11.1	12	PF3D7_1425700		1
PF3D7_0720600		8	PF3D7_1102500	GEXP02	2	PF3D7_1425600		0
PF3D7_0720200		4	PF3D7_1102300		13	PF3D7_1425500		14
PF3D7_0720100		5	PF3D7_1253100		12	PF3D7_1425400		5
PF3D7_0719900		1	PF3D7_1252800		12	PF3D7_1425300		3
PF3D7_0719600		11	PF3D7_1252300		12	PF3D7_1424400		6
PF3D7_0719500		0	PF3D7_1252100	RON3	9	PF3D7_1424100		6
PF3D7_0719400		7	PF3D7_1251800		13	PF3D7_1423800	VPS3	11
PF3D7_0719300	ARP6	13	PF3D7_1251700	aTrpRS	6	PF3D7_1423700		8
PF3D7_0719200	NEK4	5	PF3D7_1251200		3	PF3D7_1423600		4
PF3D7_0719100		5	PF3D7_1250900		6	PF3D7_1423500		7
PF3D7_0718500		7	PF3D7_1250800		8	PF3D7_1423400		5
PF3D7_0718100	EST	0	PF3D7_1250600		14	PF3D7_1423300	PP7	3

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0718000		0	PF3D7_1250500		1	PF3D7_1423200	CYP52	11
PF3D7_0717900		5	PF3D7_1250400		5	PF3D7_1423100		1
PF3D7_0717800		4	PF3D7_1250200		3	PF3D7_1423000	NOG2	11
PF3D7_0717700		10	PF3D7_1250100	G377	5	PF3D7_1422900		5
PF3D7_0717600		1	PF3D7_1250000		5	PF3D7_1422800	ARP4a	8
PF3D7_0717500	CDPK4	0	PF3D7_1249900		11	PF3D7_1422700		7
PF3D7_0717400		6	PF3D7_1249800	THO2	8	PF3D7_1422600		8
PF3D7_0717300		3	PF3D7_1249400		1	PF3D7_1422500	HRD1	5
PF3D7_0717200		10	PF3D7_1249300	PPM4	11	PF3D7_1422400		14
PF3D7_0717100		2	PF3D7_1249100		4	PF3D7_1422300		5
PF3D7_0716900	DMT2	5	PF3D7_1248900	RPT6	2	PF3D7_1422200		4
PF3D7_0716800	EIF3I	13	PF3D7_1248700		2	PF3D7_1422100		0
PF3D7_0716700		13	PF3D7_1248600		8	PF3D7_1422000	COX14	5
PF3D7_0716600	SufS	13	PF3D7_1248500		11	PF3D7_1421800		5
PF3D7_0716400		1	PF3D7_1248400		5	PF3D7_1421300		0
PF3D7_0716300		2	PF3D7_1248200	RBM22	5	PF3D7_1421000		4
PF3D7_0716200		5	PF3D7_1248100		9	PF3D7_1420900		1
PF3D7_0716100	SDA1	7	PF3D7_1247800	DPAP2	13	PF3D7_1420700	P113	6
PF3D7_0716000		7	PF3D7_1247700		1	PF3D7_1420400	GlyRS	6
PF3D7_0715900	CDF	2	PF3D7_1247500		5	PF3D7_1420200		3
PF3D7_0715800	DMT1	1	PF3D7_1247400	FKBP35	0	PF3D7_1419700		11
PF3D7_0715400	PSOP20	4	PF3D7_1246900	PKB	5	PF3D7_1419600		4
PF3D7_0715200		1	PF3D7_1246700		6	PF3D7_1419500		14
PF3D7_0715100	MYCBP	14	PF3D7_1246600	BUD13	4	PF3D7_1419400		8
PF3D7_0714600		11	PF3D7_1246400	MTIP	3	PF3D7_1419100	SPB4	8
PF3D7_0714500		8	PF3D7_1246300		9	PF3D7_1419000		8
PF3D7_0714300	DHHC4	5	PF3D7_1246200	ACT1	3	PF3D7_1418900	DBP4	0
PF3D7_0714200		2	PF3D7_1245800		3	PF3D7_1417900		4
PF3D7_0714100		0	PF3D7_1245600		0	PF3D7_1417800	MCM2	8
PF3D7_0714000	H2B.Z	2	PF3D7_1245500		9	PF3D7_1417700		6
PF3D7_0713900		2	PF3D7_1245400		5	PF3D7_1417600		11
PF3D7_0713500		0	PF3D7_1245200		0	PF3D7_1417500	CBF5	6
PF3D7_0711500		7	PF3D7_1245100	KLP8	7	PF3D7_1417200		11
PF3D7_0711400	SAP18	8	PF3D7_1245000		13	PF3D7_1416800		5
PF3D7_0711200		0	PF3D7_1244900		8	PF3D7_1416700		4
PF3D7_0711000	Cdc48	6	PF3D7_1244700		4	PF3D7_1416600		7
PF3D7_0710800		4	PF3D7_1244600	ARFGAP	10	PF3D7_1416400		13
PF3D7_0710700		5	PF3D7_1244400		2	PF3D7_1416200	MCA3	10
PF3D7_0710600	RPL34	6	PF3D7_1244100		3	PF3D7_1416100	SEY1	9
PF3D7_0710500		4	PF3D7_1244000		2	PF3D7_1415600		8
PF3D7_0710200		8	PF3D7_1243900	DOC2	14	PF3D7_1415400		11
PF3D7_0710100		4	PF3D7_1243700		3	PF3D7_1415100		5
PF3D7_0710000		8	PF3D7_1243600		11	PF3D7_1415000	UDG	8
PF3D7_0709900		5	PF3D7_1243100		4	PF3D7_1414700		13
PF3D7_0709700		4	PF3D7_1243000	SYN16	4	PF3D7_1414600	Pgt1	13
PF3D7_0709400		2	PF3D7_1242800	rabGDI	10	PF3D7_1414500	ABCK2	1
PF3D7_0709300	CG2	8	PF3D7_1242700		11	PF3D7_1414300		6
PF3D7_0709100		1	PF3D7_1242500		14	PF3D7_1414200		7
PF3D7_0708900		5	PF3D7_1242400		0	PF3D7_1414000	RPN13	1
PF3D7_0708800	HSP110c	11	PF3D7_1242200	TGT	0	PF3D7_1413700		3
PF3D7_0708700		5	PF3D7_1242000		5	PF3D7_1413400		5
PF3D7_0708600	IMC1d	4	PF3D7_1241900		11	PF3D7_1413300		5
PF3D7_0708500		14	PF3D7_1241800	DBP9	11	PF3D7_1413200		14
PF3D7_0708400	HSP90	11	PF3D7_1241400		13	PF3D7_1413000		0
PF3D7_0708000		13	PF3D7_1241200		0	PF3D7_1412900.1		7
PF3D7_0707900		1	PF3D7_1240900	VAR	5	PF3D7_1412700		0
PF3D7_0707800		2	PF3D7_1239900	VPS16	7	PF3D7_1412600	DHS	0
PF3D7_0707700		8	PF3D7_1239800		10	PF3D7_1412400		8
PF3D7_0707500		12	PF3D7_1239700	FTSH1	2	PF3D7_1412300	NTF2	6
PF3D7_0707400		11	PF3D7_1239500	GyrB	7	PF3D7_1412200		0
PF3D7_0707300	RAMA	9	PF3D7_1239400		4	PF3D7_1412100	MCMBP	6
PF3D7_0707200		11	PF3D7_1239200	ApiAP2	2	PF3D7_1411500		0
PF3D7_0706700	MSH2-2	0	PF3D7_1239000		1	PF3D7_1411400	PREX	6
PF3D7_0706500		11	PF3D7_1238800	ACS11	1	PF3D7_1411200	ROM8	3
PF3D7_0706400	RPL37	6	PF3D7_1238700		7	PF3D7_1410900		13
PF3D7_0706100		9	PF3D7_1238500		8	PF3D7_1410700		2
PF3D7_0706000		11	PF3D7_1238300	CWC22	11	PF3D7_1410400	RAP1	9
PF3D7_0705800		4	PF3D7_1238200		5	PF3D7_1410300		8
PF3D7_0705700		4	PF3D7_1238000		5	PF3D7_1410100		13
PF3D7_0705500		10	PF3D7_1237900		3	PF3D7_1410000	EMC2	5
PF3D7_0705400	MCM7	8	PF3D7_1237700		10	PF3D7_1409900	CDS	6

Gene ID	Name	Cluster	Gene ID	Name	Cluster	Gene ID	Name	Cluster
PF3D7_0705300	ORC2	8	PF3D7_1237600	PWP1	7	PF3D7_1409600		9
PF3D7_0705100		1	PF3D7_1237500		11	PF3D7_1409500		14
PF3D7_0705000		8	PF3D7_1237200		7	PF3D7_1409400		5
PF3D7_0704900	RF2	4	PF3D7_1237100		5	PF3D7_1409300	DDI1	10
PF3D7_0704800		1	PF3D7_1236900		1	PF3D7_1409100		1
PF3D7_0704700	PPAT	2	PF3D7_1236800		4	PF3D7_1408800		0
PF3D7_0704600	UT	8	PF3D7_1236600		0	PF3D7_1408700		11
PF3D7_0704500		9	PF3D7_1236500		4	PF3D7_1408600		6
PF3D7_0704300		9	PF3D7_1236200		7	PF3D7_1408500		11
PF3D7_0704100		1	PF3D7_1236100	CARP	11	PF3D7_1408400		0
PF3D7_0704000		8	PF3D7_1236000		7	PF3D7_1408200	AP2-G2	7
PF3D7_0703900		14	PF3D7_1235900	XAB2	11	PF3D7_1408100	HAP	11
PF3D7_0703800		0	PF3D7_1235800		0	PF3D7_1407800	PM4	2
PF3D7_0703600		10	PF3D7_1235600	SHMT	9	PF3D7_1407500		5
PF3D7_0703500		8	PF3D7_1235500		13	PF3D7_1407400		5
PF3D7_0703200		10	PF3D7_1235400		7	PF3D7_1407300	PRP38B	8
PF3D7_0703000		13	PF3D7_1235300	NOT4	11	PF3D7_1407100	NOP1	6
PF3D7_0702400	SEMP1	12	PF3D7_1235200	VP2	9	PF3D7_1407000	CCp3	5
PF3D7_0831700	HSP70	6	PF3D7_1235100		4	PF3D7_1406600	ClpC	2
PF3D7_0831600	CLAG8	9	PF3D7_1235000	PIH1	0	PF3D7_1406500	WDR65	0
PF3D7_0831300	GEXP13	14	PF3D7_1234900		4	PF3D7_1406400	PPR	10
PF3D7_0831200		8	PF3D7_1234800	SF3B3	7	PF3D7_1406300	GDPD	6
PF3D7_0830800	SURF8.2	7	PF3D7_1234700		5	PF3D7_1406200		11
PF3D7_0830500	TryThrA	12	PF3D7_1234600	TOC75	9	PF3D7_1406100		7
PF3D7_0830400		11	PF3D7_1234500		14	PF3D7_1405900		13
PF3D7_0829600	ETRAPP8	5	PF3D7_1234300		0	PF3D7_1405800	BOP1	7
PF3D7_0829500		14	PF3D7_1234100		11	PF3D7_1405700		14
PF3D7_0829300	LSM8	13	PF3D7_1234000		4	PF3D7_1405500	CBWD1	0
PF3D7_0829000		2	PF3D7_1233900	SENP1	8	PF3D7_1405300		0
PF3D7_0828800	GAMA	3	PF3D7_1233800		1	PF3D7_1405100		14
PF3D7_0828700		14	PF3D7_1233600	AARP1	8	PF3D7_1405000		13
PF3D7_0828600	FT1	14	PF3D7_1233400		14	PF3D7_1404900		11
PF3D7_0828500		1	PF3D7_1233300		6	PF3D7_1404800		12
PF3D7_0828300		8	PF3D7_1233100		13	PF3D7_1404500	RRP5	11
PF3D7_0828200		5	PF3D7_1232600		13	PF3D7_1404200		4
PF3D7_0828100		4	PF3D7_1232500		3	PF3D7_1404000	RPB4	4
PF3D7_0828000	ROM3	4	PF3D7_1232400		5	PF3D7_1403800	MISFIT	14
PF3D7_0827900	PDI8	2	PF3D7_1232300		4	PF3D7_1403500		0
PF3D7_0827800	SET3	1	PF3D7_1232200	LPD1	4	PF3D7_1403400		0
PF3D7_0827600		9	PF3D7_1232100	CPN60	10	PF3D7_1403300		0
PF3D7_0827500		8	PF3D7_1231900		13	PF3D7_1403200		5
PF3D7_0827400		4	PF3D7_1231800		2	PF3D7_1403100	CND1	14
PF3D7_0827300		0	PF3D7_1231600	PRP2	8	PF3D7_1403000		14
PF3D7_0827200		4	PF3D7_1231400		3	PF3D7_1402800		2
PF3D7_0827100	IF2c	7	PF3D7_1231300		14	PF3D7_1402700	SR140	6
PF3D7_0827000	DBP10	6	PF3D7_1231200		3	PF3D7_1402500		6
PF3D7_0826900		0	PF3D7_1231000		13	PF3D7_1402200		5
PF3D7_0826700	RACK	6	PF3D7_1230900	RIO1	0	PF3D7_1401900		13
PF3D7_0826500	UBE4B	8	PF3D7_1230800		11	PF3D7_1401600		9
PF3D7_0826200		13	PF3D7_1230600		13	PF3D7_1401400	ETRAPP14	12
PF3D7_0826100		11	PF3D7_1230100	TLAP1	5			

